HABITS: PLASTICITY, LEARNING AND FREEDOM

EDITED BY: Javier Bernacer, Jose Angel Lombo and Jose Ignacio Murillo
PUBLISHED IN: Frontiers in Human Neuroscience

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88919-673-9
DOI 10.3389/978-2-88919-673-9

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: **researchtopics@frontiersin.org**

HABITS: PLASTICITY, LEARNING AND FREEDOM

Topic Editors:
Javier Bernacer, University of Navarra, Spain
Jose Angel Lombo, Pontifical University of the Holy Cross, Italy
Jose Ignacio Murillo, University of Navarra, Spain

In present times, certain fields of science are becoming aware of the necessity to go beyond a restrictive specialization, and establish an open dialogue with other disciplines. Such is the case of the approach that neuroscience and philosophy are performing in the last decade. However, this increasing interest in a multidisciplinary perspective should not be understood, in our opinion, as a new phenomenon, but rather as a return to a classical standpoint: a proper understanding of human features –organic, cognitive, volitional, motor or behavioral, for example– requires a context that includes the global dimension of the human being. We believe that grand neuroscientific conclusions about the mind should take into account what philosophical reflection has said about it; likewise, philosophers should consider the organic constitution of the brain to draw inferences about the mind. Thus, both neuroscience and philosophy would benefit from each other's achievements through a fruitful dialogue.

One of the main problems a multidisciplinary group encounters is terminology: the same term has a different scope in various fields, sometimes even contradictory. Such is the case of habits: from a neuroscientific perspective, a habit is a mere automation of an action. It is, therefore, linked to rigidity and limitation. However, from a classical philosophical account, a habit is an enabling capacity acquired through practice, which facilitates, improves and reinforces the performance of certain kind of actions. From neuroscience, habit acquisition restricts a subject's action to the learnt habit; from philosophy, habit acquisition allows the subject to set a distance from the simple motor performance to cognitively enrich the action. For example, playing piano is a technical habit; considering the neuroscientific account, a pianist would just play those sequences of keystrokes that had been repeatedly practiced in the past. However, according to the philosophical perspective, it would allow the pianist to improvise and, moreover, go beyond the movements of their hands to concentrate in other features of musical interpretation.

In other words, a holistic view of habits focuses on the subject's disposition when facing both known and novel situations.

We believe neuroscience could contribute to achieve a deeper understanding of the neural bases of habits, whose complexity could be deciphered by a philosophical reflection. Thus, we propose this Research Topic to increase our understanding on habits from a wide point of view. This

collection of new experimental research, empirical and theoretical reviews, general commentaries and opinion articles covers the following subjects: habit learning; implicit memory; computational and complex dynamical accounts of habit formation; practical, cognitive, perceptual and motor habits; early learning; intentionality; consciousness in habits performance; neurological and psychiatric disorders related to habits, such as obsessive-compulsive disorder, stereotypies or addiction; habits as enabling or limiting capacities for the agent.

Citation: Bernacer, J., Lombo, J. A., Murillo, J. I., eds. (2015). Habits: Plasticity, Learning and Freedom. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-673-9

Table of Contents

frontiers in Human Neuroscience

EDITORIAL
published: 27 August 2015
doi: 10.3389/fnhum.2015.00468

Editorial: Habits: plasticity, learning and freedom

Javier Bernacer[1]*, *Jose A. Lombo*[2] and *Jose I. Murillo*[1]

[1] Mind-Brain Group, Institute for Culture and Society, University of Navarra, Pamplona, Spain, [2] School of Philosophy, Pontifical University of the Holy Cross, Rome, Italy

Keywords: procedural learning, multidisciplinary, habitual, goal-directed, routine

"During much of our waking lives, we act according to our habits, from the time we rise and go through our morning routines until we fall asleep after evening routines. Taken in this way, habits have long attracted the interest of philosophers and psychologists, and they have been alternatively praised and cursed." This is a passage extracted from a highly cited article by Ann Graybiel, one of the most important researchers on habits in present times (Graybiel, 2008). Indeed, most people agree on considering habits one of the crucial aspects of human behavior. However, although laboratory experiments have contributed to advance our understanding on this issue, the difference between habitual and non-habitual behavior is not clear in real-life conditions. This distinction was established in animal research some decades ago (Dickinson, 1985), but the opposition between goal-directed actions and habits seems to fall short in the case of humans. Are habits definitely rigid, unconscious, automatic, and non-teleological? Motor routines, such as those learnt by a beginner piano player, are one of the main examples of habits in neuroscientific literature. If habits are assigned the four characteristics mentioned above, the relationship between well-learned motor routines and the ability of the experienced pianist to improvise will be rejected. How is it possible that motor routines lead to behavioral plasticity, such as improvisation? Habits allow human agents to release cognitive resources in the performance of well-known actions. This is essential for dual-tasking, as many experiments in neuroscience suggest. Furthermore, it is also useful to enhance the cognitive control of actions, as well as to direct habits to achieve a goal and consciously redress them when needed.

In this collection of articles we discuss the role of habits in human behavioral plasticity, learning, and freedom. To our knowledge, this is the first multidisciplinary approach on habits including contributions from the fields of neuroscience, philosophy, psychology, sociology, computation, history, education, psychiatry, neurology, linguistics, physics, and genetics. If we accept that habits are key elements of human behavior, as Graybiel states, they have to be analyzed from different perspectives. Thus, the opening contributions of this e-book are two multidisciplinary approaches from neuroscience and philosophy, in which we develop in depth our position regarding habits and behavioral plasticity (Bernacer and Murillo, 2014; Lombo and Giménez-Amaya, 2014). We then present a block of articles explaining the notion of habit through history, keeping in mind its influence on contemporary neuroscience. Within this block, Barandiaran and Di Paolo show the evolution of this concept from Aristotle to our days, and explain the development of two trends of thought: the organicist and the associationist (Barandiaran and Di Paolo, 2014). After that, our Research Topic contains a series of articles introducing several philosophers' interpretation of habits: Felix Ravaisson (Gaitan and Castresana, 2014), William James (Blanco, 2014), and Merleau-Ponty (Moya, 2014). Then, we present a set of articles with various views on habits: as configurators of personal identity (Wagner and Northoff, 2014), on their possible role on epigenetics (Novo, 2014), on the relationship between neuroplasticity and a characterization introduced in medieval philosophy (Larrivee and Gini, 2014), and from a dynamic systems perspective (Barrett, 2014). After that, we go deeper into the role of habits in behavioral plasticity from different points

OPEN ACCESS

Edited and reviewed by:
Hauke R. Heekeren, Freie Universität Berlin, Germany

***Correspondence:**
Javier Bernacer, jbernacer@unav.es

Received: *22 May 2015*
Accepted: *10 August 2015*
Published: *27 August 2015*

Citation:
Bernacer J, Lombo JA and Murillo JI (2015) Editorial: Habits: plasticity, learning and freedom. Front. Hum. Neurosci. 9:468. doi: 10.3389/fnhum.2015.00468

of view: a computational model of habits as consolidated patterns of motor behavior (Egbert and Barandiaran, 2014), an anthropological theoretical reflection on their role in the integral development of the person (Güell, 2014), and an explanation of their importance on the social basis of learning (Akrivou and Todorow di San Giorgio, 2014). Since the notion of "cognitive habit" may sound strange in neuroscience, we define it in one of our introductory contributions (Bernacer and Murillo, 2014). The following three articles of this e-book are theoretical and empirical examples of cognitive habits: Pagan-Cánovas and Valenzuela speculate about the possible role of habits in conceptual mapping under the umbrella of blending theory (Cánovas and Manzanares, 2014); also, Sanchez-Cañizares reflect on the role of consciousness in triggering cognitive habits (Sánchez-Cañizares, 2014); and finally, Rashidi-Najbar and collaborators present an empirical study on the importance of reading habits in different tasks (Rashidi-Ranjbar et al., 2014). After this approach on cognitive habits, we include several contributions on their role in human action, again from a multidisciplinary perspective: first, a reflection about the importance of habits as integrators of behavioral "systems" within the agent (Martinez-Valbuena and Bernacer, 2014); second, a computational approach on their importance in model averaging and optimal inference (FitzGerald et al., 2014); third, an opinion article about the relationship between procedural learning and behavioral plasticity (Crespo-Eguílaz et al., 2014); fourth, an experimental research on the link between model-based and model-free explanations of human behavior (Friedel et al., 2014); and finally, a theoretical contribution about the positive role of habits in the treatment of addiction (Güell and Nuñez, 2014). The final block of contributions is a set of four articles about the role of habits in learning: L'Ecuyer introduces her proposal about the importance of wonder in learning (L'Ecuyer, 2014), Balderas summarizes empirical research about the role of habits in learning, even in non-human animals (Balderas, 2014), Orón-Semper relates habits and self-control in adolescent maturation (Orón-Semper, 2014), and we close our Research Topic with an interesting empirical article about the link between episodic and prospective memory (Meier et al., 2014).

In all, we hope that the reader finds this Research Topic encouraging and stimulating for future theoretical and empirical research. We are aware that even the title of our e-book may be considered an oxymoron according to the general view of habits in neuroscience: this is the main reason to propose a multidisciplinary approach on the subject, and to investigate about the importance of habits in our daily lives.

Acknowledgments

We are grateful to all members of the Mind-Brain Group for the fruitful discussions during the preparation of this collection of articles. This research has been supported by Obra Social La Caixa and Institute for Culture and Society (ICS).

References

Akrivou, K., and Todorow di San Giorgio, L. (2014). A dialogic conception of Habitus: allowing human freedom and restoring the social basis of learning. *Front. Hum. Neurosci. Neurosci.* 8:432. doi: 10.3389/fnhum.2014.00432

Balderas, G. (2014). Habits as learning enhancers. *Front. Hum. Neurosci.* 8:918. doi: 10.3389/fnhum.2014.00918

Barandiaran, X. E., and Di Paolo, E. A. (2014). A genealogical map of the concept of habit. *Front. Hum. Neurosci.* 8:522. doi: 10.3389/fnhum.2014.00522

Barrett, N. F. (2014). A dynamic systems view of habits. *Front. Hum. Neurosci.* 8:682. doi: 10.3389/fnhum.2014.00682

Bernacer, J., and Murillo, J. I. (2014). The Aristotelian conception of habit and its contribution to human neuroscience. *Front. Hum. Neurosci.* 8:883. doi: 10.3389/fnhum.2014.00883

Blanco, C. A. (2014). The principal sources of William James' idea of habit. *Front. Hum. Neurosci.* 8:274. doi: 10.3389/fnhum.2014.00274

Cánovas, C. P., and Manzanares, J. V. (2014). Conceptual mappings and neural reuse. *Front. Hum. Neurosci.* 8:261. doi: 10.3389/fnhum.2014.00261

Crespo-Eguílaz, N., Magallón, S., and Narbona, J. (2014). Procedural skills and neurobehavioral freedom. *Front. Hum. Neurosci.* 8:449. doi: 10.3389/fnhum.2014.00449

Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Philos. Trans. R. Soc. B Biol. Sci.* 308, 67–78. doi: 10.1098/rstb.1985.0010

Egbert, M. D., and Barandiaran, X. E. (2014). Modeling habits as self-sustaining patterns of sensorimotor behavior. *Front. Hum. Neurosci.* 8:590. doi: 10.3389/fnhum.2014.00590

FitzGerald, T. H. B., Dolan, R. J., and Friston, K. J. (2014). Model averaging, optimal inference, and habit formation. *Front. Hum. Neurosci.* 8:457. doi: 10.3389/fnhum.2014.00457

Friedel, E., Koch, S., and Wendt, J. (2014). Devaluation and sequential decisions: linking goal-directed and model-based behavior. *Front. Hum. Neurosci.* 8:587. doi: 10.3389/fnhum.2014.00587

Gaitan, L. M., and Castresana, J. S. (2014). On habit and the mind-body problem. The view of Felix Ravaisson. *Front. Hum. Neurosci.* 8:684. doi: 10.3389/fnhum.2014.00684

Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. *Ann. Rev. Neurosci.* 31, 359–387. doi: 10.1146/annurev.neuro.29.051605.112851

Güell, F. (2014). Pre-dispositional constitution and plastic disposition: toward a more adequate descriptive framework for the notions of habits, learning and plasticity. *Front. Hum. Neurosci.* 8:341. doi: 10.3389/fnhum.2014.00341

Güell, F., and Nuñez, L. (2014). The liberating dimension of human habit in addiction context. *Front. Hum. Neurosci.* 8:64. doi: 10.3389/fnhum.2014.00664

L'Ecuyer, C. (2014). The wonder approach to learning. *Front. Hum. Neurosci.* 8:764. doi: 10.3389/fnhum.2014.00764

Larrivee, D., and Gini, A. (2014). Is the philosophical construct of "habitus operativus bonus" compatible with the modern neuroscience concept of human flourishing through neuroplasticity? A consideration of prudence as a multidimensional regulator of virtue. *Front. Hum. Neurosci.* 8:731. doi: 10.3389/fnhum.2014.00731

Lombo, J. A., and Giménez-Amaya, J. M. (2014). The unity and the stability of human behavior. An interdisciplinary approach to habits between philosophy and neuroscience. *Front. Hum. Neurosci.* 8:607. doi: 10.3389/fnhum.2014.00607

Martinez-Valbuena, I., and Bernacer, J. (2014). Behavioral duality in an integrated agent. *Front. Hum. Neurosci.* 8:614. doi: 10.3389/fnhum.2014.00614

Meier, B., Matter, S., Baumann, B., Walter, S., and Koenig, T. (2014). From episodic to habitual prospective memory: ERP-evidence for a linear transition. *Front. Hum. Neurosci.* 8:489. doi: 10.3389/fnhum.2014.00489

Moya, P. (2014). Habit and embodiment in merleau-ponty. *Front. Hum. Neurosci.* 8:542. doi: 10.3389/fnhum.2014.00542

Novo, J. (2014). Habit acquisition in the context of neuronal genomic and epigenomic mosaicism. *Front. Hum. Neurosci.* 8:255. doi: 10.3389/fnhum.2014.00255

Orón-Semper, J. V. (2014). Toward a new conception of habit and self-control in adolescent maturation. *Front. Hum. Neurosci.* 8:525. doi: 10.3389/fnhum.2014.00525

Rashidi-Ranjbar, N., Goudarzvand, M., Jahangiri, S., Brugger, P., and Loetscher, T. (2014). No horizontal numerical mapping in a culture with mixed-reading habits. *Front. Hum. Neurosci.* 8:72. doi: 10.3389/fnhum.2014.00072

Sánchez-Cañizares, J. (2014). The role of consciousness in triggering intellectual habits. *Front. Hum. Neurosci.* 8:312. doi: 10.3389/fnhum.2014.00312

Wagner, N.-F., and Northoff, G. (2014). Habits: bridging the gap between personhood and personal identity. *Front. Hum. Neurosci.* 8:330. doi: 10.3389/fnhum.2014.00330

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in
HUMAN NEUROSCIENCE

OPINION ARTICLE
published: 11 August 2014
doi: 10.3389/fnhum.2014.00607

The unity and the stability of human behavior. An interdisciplinary approach to habits between philosophy and neuroscience

José A. Lombo[1]* and José M. Giménez-Amaya[2]

[1] *School of Philosophy, Pontifical University of the Holy Cross, Rome, Italy*
[2] *Research group Science, Reason and Faith (CRYF), University of Navarre, Pamplona, Spain*
**Correspondence: lombo@pusc.it*

Edited by:
Jose Ignacio Murillo, University of Navarra, Spain

Reviewed by:
Walter Adriani, Istituto Superiore di Sanità, Italy

Keywords: habit, learning, memory, basal ganglia, reason, plasticity

INTRODUCTION

The study of learning and memory through routines is acquiring a growing interest in the present neuroscience. Many works focusing on "habit learning" highlight the relevance of such studies and much weight is given to it in the understanding of individual's behavior. The aim of this paper is to connect the concept of habit that arises from a neurobiological viewpoint and from a philosophical one. This will require a precise terminological distinction and connection between the two fields from an interdisciplinary approach.

One of the most recent studies of the use of the term "habit" in neuroscience is the review of Carol Seger y Brian Spiering, published in Frontiers in Systems Neuroscience in 2011. In this work, the authors make a historical evaluation of expressions "habit" and "habit learning" in these terms: ≪ "Habit" roughly corresponded to the resulting motor behavior [. . .], and habit learning to acquisition of these behaviors in an instrumental learning context.≫ "Habit" presents therefore these characteristics: inflexible, slow or incremental, unconscious, automatic and insensitive to reinforcer devaluation.

On the other hand, philosophical concept of habit (hexis, habitus) is not only—nor even mainly—related to a repetitive behavior, but to the control or "possession" of one's action (se habere ad). According to this, "habitus" is closer to the concept of "quality" or "skill" than to that of "stereotype" and appears as a vehicle of free will.

Apparently both approaches, neurobiological and philosophical, seem different and unconnected. Nevertheless, a possible bridge between them is the consideration of habit as "stable disposition for self-development." In order to illustrate this statement, we will follow several steps throughout the paper. First, we will expose how the concept of life implies a kind of self-activity and self-control, which requires either stability and change at different levels. Second, we will deal with the neurobiological understanding of stable behavior as seen in neurobiological processes such as learning and memory, "habit learning," etc. And thirdly, we will explain how the philosophical concept of "habit" corresponds to dispositive quality as control of one's action. Finally, we will try to integrate the two perspectives.

HABITS AS STABLE CONTROL OF ACTIONS

From a philosophical point of view, life consists in the activity of an individual over itself (self-activity, self-control). A living being is capable of actions whose outcome doesn't remain only outside, but inside the living being itself. This kind of activity implies a general scheme of "feedback" and corresponds with the Aristotelian concept of "praxis," as different to the concept of "poiesis." "Praxis" is an activity whose aim is the activity itself, and so its outcome remains in the individual (Metaphysics, IX, 6, 1048 b 18–35; Aristotle, 1924). Instead, "poiesis" is an activity that produces something different from the action itself and so it has an external outcome (Nicomachean Ethics, VI, 4, 1140a 1–6; Aristotle, 2011). Even though poiesis and praxis are different, they are not necessarily separable, but continuously interwoven in living beings endowed with knowledge. Life as a whole is praxis, but particular life activities include both poiesis and praxis (Aristotle, Politics, I, 2, 1254 a 7–8, Aristotle, 1990; Vicente Arregui and Choza, 1991).

In the interaction of poiesis and praxis, the living being not only maintains its own structure, but it progressively develops it. In general, this development consists in the extension or amplification of one's own physical structure. Nevertheless, in the case of living beings endowed with knowledge, development has also an intensive dimension, as they can acquire new capabilities through their interaction with other beings. This intensive development can be understood as "learning" or "accumulated experience." It results from single actions, but it differs from them as an acquired and stable capability. Aristotle called that capability "habit" (hexis, habitus) and understood it as making the subject of it able to perform new actions (Nicomachean Ethics, II, 4, 1106 b 36; Aristotle, 2011).

NEUROBIOLOGICAL UNDERSTANDING OF STABLE BEHAVIOR: LEARNING AND MEMORY, "HABIT LEARNING," etc.

On the part of empirical research, modern psychology has studied the concept of "habit" quite in detail. The context of it has been the study of learning and, more in general, of animal behavior: see,

for example, James (1890) and Watson (1919). This perspective has gained in depth thanks to the development of cognitive experimental psychology and the studies on learning and memory during XX century (Bernácer and Giménez-Amaya, 2013).

Since the second half of the 50's, neuroscientific studies showed the progressive implication of structures of the temporal lobe in memory and in learning (Scoville and Milner, 1957; Bernácer and Giménez-Amaya, 2013). Those studies defined a distinction between an explicit memory and an implicit one: in the former, cortical structures are mostly involved, mainly medial portions of temporal lobe; in the latter, some subcortical structures stand out, which belong to the basal ganglia.

In sum, there has been a progressive separation of two neurobiological processes related to memory. On the one hand, some mnesic processes reveal learning as related to processes of plasticity, which imply a high cortical activity (explicit memory). On the other hand, other processes evince learning as the stabilization of patterns of behavior—mainly motor—, in which some subcortical structures intervene, as the aforementioned basal ganglia (implicit memory).

The concept of "habit learning" was introduced in cognitive neuroscience through these premises. According to Seger and Spiering (2011): "The concept of habit learning has developed through the fruitful interaction of researchers in several intellectual domains, including animal learning, cognitive psychology, cognitive neuropsychology, and behavioral neuroscience." In large measure, the concept of "habit learning" has been related to subcortical structures of basal ganglia and, therefore, to processes of learning involved in implicit memory: see, for example, reviews of Seger and Spiering (2011) and Graybiel (2008).

Basal ganglia are structures strongly connected among themselves, with a fundamental role in the organization of complex circuits of cortical and subcortical feedback (Mengual et al., 1999; Obeso et al., 2002; Packard and Knowlton, 2002; Lanciego et al., 2012). Two traits make them especially relevant to study processes of learning and memory. First, their neural circuits of feedback are much wider and more complex than what was originally thought. In fact, basal ganglia are not only related to motor system in itself, but they are also important as nodal points in broad neural networks, which integrate motor behavior with emotional and motivational life, particularly frontostriatal circuits and limbic areas: see, for instance, reviews of Haber and Rauch (2010) and of Hwang (2013). Second, they are privileged structures of central nervous system for the understanding, at a molecular and synaptic level, of the strong interaction between neurotransmitters and neuromodulators involved in networks of implicit memory (Kreitzer and Malenka, 2008). This permits to establish complex patterns of cellular integration and of relations of nervous cells among them.

These remarks show the significance of basal ganglia in the study of "habit learning." This kind of learning has been described as "inflexible, slow or incremental, unconscious, automatic, and insensitive to reinforcer devaluation" (Seger and Spiering, 2011). Nevertheless, there is increasing evidence that, through their cortical and subcortical circuits, some degree of flexibility and control can be established (Smith and Graybiel, 2014). In fact, "habit learning" seems to be open to include instances of plasticity, learning and memory (Graybiel, 2008; Howe et al., 2011). As a result, several approaches to "habit learning" are increasingly seeing it as a balance between behavioral flexibility and fixity (Smith and Graybiel, 2014).

On one hand, some authors have regarded the idea of "habit learning" as the performance of an action, previously learned after many repetitions, in an unconscious manner, and whose execution is inflexible and independent to the outcome (Seger and Spiering, 2011; Bernácer and Giménez-Amaya, 2013). On the other hand, this perspective should be integrated with other view that recognizes sensitivity to the outcome and hence different levels of flexibility and feedback, allowing integrating changes onto behavioral processes or strategies. In this way, the whole system allows several levels of increase and development (Lombo and Giménez Amaya, 2013; Smith and Graybiel, 2014).

ADAPTATION AND CHANGE: STABILITY vs. RIGIDITY

From the mentioned approach, some opposition can be established between two ways of understanding "habit learning." On the one hand, it appears as a rigid and stereotyped behavior (Seger and Spiering, 2011). On the other hand, it can be understood in a more open and flexible way, what allows the incorporation of phenomena of variability within a general scheme of control (Graybiel, 2008; Smith and Graybiel, 2014).

Deep in this opposition, we discover that the second view does not exclude the first one, but rather it presupposes it. In fact, a habit is not a mere automatism or a repetitive behavior, but a stable disposition for action (practical skill). The difference between habits and automatisms or simple routines is that the former give control over actions, while the latter don't (Nicomachean Ethics, II, 1, 1103 a 14-b 25; Aristotle, 2011). As a consequence, the stability of habits differs from the rigidity of automatisms.

Consequently, rigidity and the stereotyped character of "habit learning" should be understood as "stability" of behavior, rather than as an irremovable configuration of it. This is therefore a richer concept, from a semantic standpoint, and points out to a stable basic structure on which living being's behavior is organized in a flexible manner. This flexibility allows adaptation to new stimuli and development of new abilities. On the other hand, excessive inflexibility makes adaptation impossible and may lead to behavioral disorders, like obsessive-compulsive personality disorder, for instance (De Reus and Emmelkamp, 2012).

This neurobiological view of "habit learning" and recent experimental contributions—especially those of professor Graybiel—are consistent with the philosophical concept of "habit" in human being. This one is essentially based on two aspects: (a) the stable character of an acquired quality; and (b) the capacity for new actions that arises from that quality (Millán-Puelles, 2002).

In first place, habit is related to "having," as the term indicates in its Latin original form ("habitus" comes from "habere," to have). According to Aristotle, a subject may have other realities or may

have itself as related to other realities (Nicomachean Ethics, II, 4, 1105 b 25–26; Aristotle, 2011). This "having himself" as related to something means actually "to be disposed in relation to something": Aristotle's Metaphysics, V, 20; 1022 b 10–12 (Aristotle, 1924). For this author, habit ("hexis") is not a simple reaction to the influence or activity of other subjects (he calls this influences "pathe," passions). It is rather to "dispose himself" from that influence, acquiring a stable capacity to accomplish something in a way that becomes usual. A habit can be described therefore as a usual way of behaving, so that Aristotle refers to it also as a "second nature": Aristotle's Categories, VIII 9 a 4 (Aristotle, 1930). Inasmuch as "habit" is not a simple reaction, but a stable disposition to action, it has been compared with cybernetic processes (Polo, 2002). This disposition, in fact, is stable and progressive, but not properly rigid.

CONCLUSIVE REMARKS

As we have seen, neurobiological concept of "habit" is reflected in the so called "habit learning." This implies two main aspects, that's to say, stability of behavior (that can be interpreted as "rigidity" or "stereotype") and its flexibilization in front of new stimuli (Seger and Spiering, 2011; Smith and Graybiel, 2014). This is clearly verified in superior mammals, but in the case of human being we find a special richness in his behavioral response. Neurobiological ground of that higher development can be found in the remarkable growth of his cortical and subcortical networks (basal ganglia, among other structures), and in his extraordinary cellular and high synaptic variety (see for example, Nijhuis et al., 2013).

We can discover, in sum, a connection between neurobiological and philosophical standpoints. On one hand, "habit learning" implies a stabilization of neurobiological information that subsequently allows its storage and re-utilization in front of new stimuli. On the other, philosophical description of "habit" presents it as feedback of human activity. This feedback allows not only to keep our activities, but also to use them again in front of new phenomena, making possible continuity and articulation of experience.

REFERENCES

Aristotle. (1924). *Metaphysics.* ed W. D. Ross. Oxford: Clarendon Press.

Aristotle. (1930). *Categories*. ed E. H. Edghill. Oxford: Clarendon Press.

Aristotle. (1990). *Politics.* ed H. Rackham. Cambridge; London: Harvard University Press.

Aristotle. (2011). *Nicomachean Ethics*. eds S. D. Collins and R. C. Bartlett. Chicago, IL: University of Chicago Press. doi: 10.7208/chicago/9780226026763.001.0001

Bernácer, J., and Giménez-Amaya, J. M. (2013). "On habit learning in neuroscience and free will," in *Is Science Compatible with Free Will? Exploring Free Will and Consciousness in the Light of Quantum Physics and Neuroscience*, eds A. Suarez and P. Adams (New York, NY: Springer), 177–193.

De Reus, R. J. M., and Emmelkamp, P. M. G. (2012). Obsessive–compulsive personality disorder: a review of current empirical findings. *Pers. Ment. Health* 6, 1–21. doi: 10.1002/pmh.144

Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. *Annu. Rev. Neurosci.* 31, 359–387. doi: 10.1146/annurev.neuro.29.051605.112851

Haber, S. N., and Rauch, S. L. (2010). Neurocircuitry: a window into the networks underlying neuropsychiatric disease. *Neuropsychopharmacology* 35, 1–3. doi: 10.1038/npp.2009.146

Howe, M. W., Atallah, H. E., McCool, A., Gibson, D. J., and Graybiel, A. M. (2011). Habit learning is associated with major shifts in frequencies of oscillatory activity and synchronized spike firing in striatum. *Proc. Natl. Acad. Sci. U.S.A.* 108,16801–16806. doi: 10.1073/pnas.1113158108

Hwang, E. J. (2013). The basal ganglia, the ideal machinery for the cost-benefit analysis of action plans. *Front. Neural Circuits* 7:121. doi: 10.3389/fncir.2013.00121

James, W. (1890). *Principles of Psychology*. New York, NY: Henry Holt. doi: 10.1037/11059-000

Kreitzer, A. C., and Malenka, R. C. (2008). Striatal plasticity and basal ganglia circuit function. *Neuron* 60, 543–554. doi: 10.1016/j.neuron.2008.11.005

Lanciego, J. L., Luquin, N., and Obeso, J. A. (2012). Functional neuroanatomy of the basal ganglia. *Cold Spring Harb. Perspect. Med.* 2:a009621 doi: 10.1101/cshperspect.a009621

Lombo, J. A., and Giménez Amaya, J. M. (2013). *La Unidad de la Persona. Aproximación Interdisciplinar Desde la Filosofía y la Neurociencia.* Pamplona: EUNSA.

Mengual, E., de las Heras, S., Erro, E., Lanciego, J. L., and Giménez-Amaya, J. M. (1999). Thalamic interaction between the input and the output systems of the basal ganglia. *J. Chem. Neuroanat*. 16, 187–200. doi: 10.1016/S0891-0618(99)00010-1

Millán-Puelles, A. (2002). *Léxico Filosófico*. Madrid: Rialp.

Nijhuis, E. H., van Cappellen van Walsum, A. M., and Norris, D. G. (2013). Topographic hub maps of the human structural neocortical network. *PLoS ONE* 8:e65511. doi: 10.1371/journal.pone.0065511

Obeso, J. A., Rodríguez-Oroz, M. C., Rodríguez, M., Arbizu, J., and Giménez-Amaya, J. M. (2002). The basal ganglia and disorders of movement: pathophysiological mechanisms. *News Physiol. Sci*. 17, 51–55.

Packard, M. G., and Knowlton, B. J. (2002). Learning and memory functions of the Basal Ganglia. *Annu. Rev. Neurosci.* 25, 563–593. doi: 10.1146/annurev.neuro.25.112701.142937

Polo, L. (2002). La cibernética como lógica de la vida. *Stud. Poliana* 4, 9–17.

Scoville, W. B., and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* 20, 11–21. doi: 10.1136/jnnp.20.1.11

Seger, C. A., and Spiering, B. J. (2011). A critical review of habit learning and the Basal Ganglia. *Front. Syst. Neurosci.* 5:66. doi: 10.3389/fnsys.2011.00066

Smith, K. S., and Graybiel, A. M. (2014). Investigating habits: strategies, technologies and models. *Front. Behav. Neurosci.* 8:39. doi: 10.3389/fnbeh.2014.00039

Vicente Arregui, J., and Choza, J. (1991). *Filosofía del Hombre. Una Antropología de la Intimidad.* Madrid: Rialp.

Watson, J. B. (1919). *Psychology from the Standpoint of a Behaviorist*. New York, NY: Norton. doi: 10.1037/10016-000

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 May 2014; accepted: 21 July 2014; published online: 11 August 2014.

Citation: Lombo JA and Giménez-Amaya JM (2014) The unity and the stability of human behavior. An interdisciplinary approach to habits between philosophy and neuroscience. Front. Hum. Neurosci. ***8****:607. doi: 10.3389/fnhum.2014.00607*

This article was submitted to the journal Frontiers in Human Neuroscience.

frontiers in
HUMAN NEUROSCIENCE

HYPOTHESIS AND THEORY ARTICLE
published: 03 November 2014
doi: 10.3389/fnhum.2014.00883

The Aristotelian conception of habit and its contribution to human neuroscience

Javier Bernacer* and Jose Ignacio Murillo

Mind-Brain Group, Institute for Culture and Society, University of Navarra, Pamplona, Navarra, Spain

Edited by:
Jose Angel Lombo, Pontifical University of the Holy Cross, Italy

Reviewed by:
Katie A. Jennings, University of Oxford, UK
Carol Seger, Colorado State University, USA

****Correspondence:***
Javier Bernacer, Mind-Brain Group, Institute for Culture and Society, University of Navarra, Edificio Biblioteca, Office #2490, Campus Universitario s/n, Pamplona, Navarra 31008, Spain
e-mail: jbernacer@unav.es

The notion of habit used in neuroscience is an inheritance from a particular theoretical origin, whose main source is William James. Thus, habits have been characterized as rigid, automatic, unconscious, and opposed to goal-directed actions. This analysis leaves unexplained several aspects of human behavior and cognition where habits are of great importance. We intend to demonstrate the utility that another philosophical conception of habit, the Aristotelian, may have for neuroscientific research. We first summarize the current notion of habit in neuroscience, its philosophical inspiration and the problems that arise from it, mostly centered on the sharp distinction between goal-directed actions and habitual behavior. We then introduce the Aristotelian view and we compare it with that of William James. For Aristotle, a habit is an acquired disposition to perform certain types of action. If this disposition involves an enhanced cognitive control of actions, it can be considered a "habit-as-learning". The current view of habit in neuroscience, which lacks cognitive control and we term "habit-as-routine", is also covered by the Aristotelian conception. He classifies habits into three categories: (1) theoretical, or the retention of learning understood as "knowing that *x* is so"; (2) behavioral, through which the agent achieves a rational control of emotion-permeated behavior ("knowing how to behave"); and (3) technical or learned skills ("knowing how to make or to do"). Finally, we propose new areas of research where this "novel" conception of habit could serve as a framework concept, from the cognitive enrichment of actions to the role of habits in pathological conditions. In all, this contribution may shed light on the understanding of habits as an important feature of human action. Habits, viewed as a cognitive enrichment of behavior, are a crucial resource for understanding human learning and behavioral plasticity.

Keywords: goal-directed actions, Aristotle, basal ganglia, cognitive control, prefrontal cortex, implicit memory, procedural learning

INTRODUCTION

In order to achieve a deep understanding of the main topics concerning the human mind, neuroscience must dialog with other sources of knowledge. In addition, from time to time, it is necessary to take a break from experimental work and ponder whether certain things taken for granted need to be revisited. Such is the case, in our opinion, with the concept of habit and habit learning. This revisiting has been profitably carried out in previous approaches to other topics, such as the self (Northoff, 2012).

In general terms, on the basis of experimental research in neuroscience, a habit is defined as a motor or cognitive routine that, once it is triggered, completes itself without conscious supervision. Furthermore, it has always been characterized via terms such as "unconscious", "rigid", "automatic" and, more importantly, "non-teleological": that is, as the opposite of goal-directed. However, the original and most elegant description of habits, which goes back to Aristotle, defines them as acquired dispositions that improve the agent's performance, making him/her more successful in the quest to achieve a goal. The neuropsychological distinction between goal-directed actions and habits (Dickinson, 1985) is, therefore, hardly compatible with this perspective. This distinction is based on two key phenomena: some behavior is habitual if it is performed even after (1) outcome devaluation or (2) a degradation of the action-outcome contingency. In other words, "habitual" behavior according to neuroscience is defined by the absence of self-proposed goals and a lack of cognitive control. These two elements, however, are crucial in the Aristotelian conception of habit.

This article will review, very briefly, the mainstream view of habit in neuroscience, its philosophical inspiration, as well as the challenges that recent research projects are encountering due to their reliance on this definition. We propose a multidisciplinary revised version of the notion of habit based on Aristotle's work, and we explain to what extent it may help neuroscientific research. Finally, we suggest certain novel approaches to experimental research on habits, in order to attain a deeper understanding of the human mind.

In this article, our main purpose is not just to expose a terminological confusion that exists between neuroscience and philosophy. In fact, the common view of habit in neuroscience

derives from a more specific view, which has its own history (Barandiaran and Di Paolo, 2014; Blanco, 2014). In our opinion, the notion of habit drawn from classical philosophy allows a better understanding of learning, including the role of routines and automatisms, in human behavior. We also believe that a richer view of habits in neuroscience may provide an improved interpretation of such "dichotomies" as conscious-unconscious, automatic-controlled, or teleological-ateleological, and may ultimately help to demonstrate that, in the case of human beings, these are not black and white processes employing binary variables, but arise from the complex interplay that configures human action.

CURRENT VIEW OF HABITS IN NEUROSCIENCE AND ITS THEORETICAL INSPIRATION

An extensive review of the notion of habit in neuroscience is beyond the scope of this article, and we refer the reader to the works and reviews cited below for further reading. However, we summarize in a few paragraphs the conceptual background where habits reside in neuroscience, mainly based on the works by Anthony Dickinson and Ann Graybiel. The explicit investigation of habits in neuroscience is quite recent. This is a remarkable issue, if we accept that "we act according to our habits, from the time we rise and go through our morning routines until we fall asleep after evening routines" (Graybiel, 2008). All throughout the twentieth century, research on habits has centered on animal research, specifically on how behavioral patterns, i.e., motor routines, are developed and executed in non-human animals (see, for a historical review, the article by Seger and Spiering, 2011). One of the most important topics when studying habits in the field of neuroscience has been the relationship between actions, habits and goals. In that sense, the work by Dickinson (1985) was the seminal contribution. In his work, habits are overtly opposed to teleological actions, and identified with stimulus-response pairings. The main difference between these two processes is that, whereas actions are outcome-oriented and thus sensitive to reward devaluation or extinction, habits are just guided by the stimulus itself, and not by the outcome it leads to (Adams and Dickinson, 1981). Thus, a behavior is considered a habit when the animal insists on its performance in spite of outcome devaluation, or of the degradation of the contingency between the action and the outcome (Balleine and Dickinson, 1998; Yin and Knowlton, 2006). This is the mainstream view of habits in various subdisciplines within neuroscience, such as experimental psychology (Dickinson et al., 1998), psychiatry (Gillan et al., 2011), neuroanatomy (Yin and Knowlton, 2006) and neurocomputation (Balleine et al., 2008). Graybiel (2008) wraps this view up stating that habits are largely learned after extensive experience, remaining fixed and performed automatically, and they involve a structured action sequence triggered by a stimulus. Graybiel's view on habits is not so clearly focused on the opposition between goal-directed actions and habits, although the defining characteristics of the latter, as proposed by her, are oriented towards those characteristics as defined by Dickinson. Goals are explicitly present during action evaluation and selection, but they increasingly blur the more an action is repeated. The main examples of habits Graybiel proposes are fixed action patterns, i.e., complex repetitive behaviors in animals, and repetitive behaviors and thoughts in human pathological conditions, such as Tourette syndrome, obsessive compulsive disorders, and stereotypies in Huntington's and Parkinson's diseases, as well as in addictive disorders. Therefore, a habit completely disengaged from a goal becomes either a stimulus-response pair in animals, or a pathological trait in humans. Graybiel also thinks that habits play an important role in societal terms, when they are shaped as mannerisms and rituals. However, the link between the anatomical and physiological bases of habits and their social expression is not clear at all, mostly because the majority of the experiments are carried out in laboratory animals. At a theoretical level, Graybiel describes an intuitive classification of habits as "neutral", "good" or "bad". Good habits would be those we strive to incorporate in our behavior, whereas bad habits are those that powerfully take control of our behavior. This categorization seems to leave a door open to include goals as drivers of habits: we interpret Graybiel's "good" habits as those rationally directed to a goal, and "bad" habits as behavioral dispositions to perform rigidly defined actions uncontrolled by cognitive processes.

This is, very broadly, the current view of habits in neuroscience. In our opinion, it is of great interest to analyze the theoretical foundations of this conception, and to consider other alternatives that could enrich the study of habits in human neuroscience. As Seger and Spiering (2011) state in their recent review, the notion of habit in neuroscience is inspired by the view of the psychologist and philosopher William James (1890). James is also credited by Dickinson and Graybiel in their works. A succinct but clear explanation of the influences received by James in the formulation of his conception of habit has been recently published (Blanco, 2014), and we outline it here. According to James, habits can be innate or learned. In both cases they are based on plasticity, understood as a feature of inert matter, an idea taken from the French psychologist Léon Dumont: habit is just an analogy of the natural laws that affect the inanimate universe, but applied to living beings. Another important influence on William James's idea of habit is that of William Benjamin Carpenter, an expert on comparative neurology whose conception of the unconscious inspired James. However, the main philosophical branch that influenced James was associationism, as understood by Alexander Bain and John Stuart Mill. This is the theoretical background that has had the greatest impact on the study of habit by neuroscience. The main idea is as follows: habits are based on the plasticity of matter, and they subserve adaptive purposes. Moreover, a habit can be chunked into smaller pieces that are automatically assembled: this is the main feature of associationism, and the start point of the Pavlovian stimulus-response pairing.

Another recent publication gives an extremely interesting genealogy of the concept of habit over the course of the history of philosophy and neuroscience (Barandiaran and Di Paolo, 2014). These authors state that "neuroscientific research on habit remains rooted within a narrow theoretical tradition". Interestingly for the purpose of our manuscript,

Barandiaran and Di Paolo (2014) acknowledge that the first description of habits was developed by Aristotle, and further interpretations of his findings have given rise to two opposed branches: organicism and associationism. According to these authors, the latter is the only theoretical influence in the study of habits in neuroscience. It is based on the idea that mental states are formed by the association of simpler units. Furthermore, the probability of one unit's occurrence automatically following another's increases when they have been contiguously presented in the past. The organicist view, although it has its origin in Aristotle, has been developed on the basis of the conception of *conatus* introduced by Spinoza: *conatus* is the essence of any "finite mode" (let's say, as an example, any natural being), and it refers to the struggle to keep being oneself. Habits, then, are intended to preserve homeostasis of the organism, a view that differs from the original Aristotelian view, since according to the Greek philosopher good habits imply an increasing improvement of the agent. Going back to the organicist view, a habit includes the organism as a whole and its environment. The main difference between the associationist and the organicist interpretations of habits is, therefore, that the former views habits as an assembly of small mechanisms, whereas the latter considers them to be a resource of the organism—an embodied mind in a complex environment—that works to maintain homeostasis. Barandiaran and Di Paolo place William James in the associationist branch of the theoretical conceptualization of habit. The associationist heritage of James is clear in Chapter 4 ("Habit") of "The Principles of Psychology". Habits are based on the plasticity of brain matter and are characterized by the sequential functioning of different brain regions: "a simple habit, as every other nervous event (. . .) is, mechanically, nothing but a reflex discharge. (. . .) The most complex habits, (. . .) are, from the same point of view, nothing but concatenated discharges in the nerve-centers".

At this point, we believe it may be useful to move back and forth between William James and the current notion of habit in neuroscience, in order to understand their similarities and differences. One of the results of James's research is that "habit diminishes the conscious attention with which our acts are performed". When we are learning a skill, "we interrupt ourselves at every step by unnecessary movements and false notes. When we are proficient, on the contrary, the results not only follow with the very minimum of muscular action requisite to bring them forth, they also follow from a single instantaneous 'cue'". This is the conceptual origin of the physiological chunking proposed by Graybiel and others (Graybiel, 1998; Barnes et al., 2005; Seger, 2008): when an animal is learning a motor sequence *de novo*, there is a continuous activation of the projection neurons located in the sensorimotor striatum; however, when the sequence is well-learned, these cells are significantly activated just in certain landmarks of the task, such as at the beginning and the end of the sequence. The decline in conscious attention suggested by James is also supported by current neuroscience, since the consolidated chunked activity in the motor aspects of the striatum during the performance of a well-learned motor sequence correlates with a decreased activity in the cognitive part of this brain area (Smith and Graybiel, 2014; Thorn and Graybiel, 2014).

We have outlined here the influence of William James on the current notion of habit in neuroscience, although much more could be said about this topic. The main conclusions of this initial part of our research are as follows: (1) most neuroscientists working on habits overtly credit the influence of James in their research; (2) James's proposal is based on associationism, that is, small units that mechanically follow each other; (3) this association is the theoretical inspiration for the physiological chunking proposed as the neural bases of habits; and (4) having this theoretical background, experimental neuroscience has set the following condition for an action to be considered a habit: its performance must remain unchanged in the face of outcome devaluation and degradation of action-outcome contingency.

LIMITATIONS

A notion of habit based solely on William James's thought entails obvious benefits: neuroscientific research has achieved major advances in the study of the neurobiological foundations of motor routines, the relation of consciousness with habits, the mechanisms of instrumental learning in animals and the implication of these phenomena in human disorders, for example. However, this view is also limited to some extent, and we suggest overcoming these limitations with a different theoretical interpretation of habits.

The main shortcoming is that the opposition between goal-directed actions and habits—founded on the associationist view and developed on the basis of excellent animal research (Dickinson, 1985)—works experimentally, but it is far from explaining the complexity of human habits. This opposition has the strong point of being impeccable from an experimental point of view: a goal-directed action is driven by the outcome it leads to, whereas a habit is carried out even in the case of outcome devaluation or degradation of the action-outcome contingency. Therefore, experimentally and by definition, there cannot be goal-directed habits. However, this is not what we observe in human behavior, where many habits, even the simplest ones, such as tying one's shoelaces, are goal-directed. As we will explain later, the fact of being or not being goal-directed is not necessarily the critical issue for distinguishing non-habitual from habitual behavior. It is interesting to note that the identification of habits as ateleological behavior is an interpretation of James's work, although he himself does not put it in those terms. In fact, the first conclusion of his analysis is that "habits simplify the movements required to achieve a given result". Thus, habits can be oriented towards a goal. When giving examples of human habits, he tended to mention musical performance, although the current notion in neuroscience is based more intensely on his "negative" examples, generally termed "slips-of-action": once a behavioral sequence is initiated it can continue even beyond the intention that has elicited it: "Who is there that has never wound up his watch on taking off his waistcoat in the daytime, or taken his latchkey out on arriving at the door-step of a friend?". Slips-of-action have been studied in the context of obsessive-compulsive disorder, where patients' performance seems to be ateleological and, as some authors see it, "habitual". Undoubtedly, acquiring a habit implies some determination; however, it is also worth noting that James leaves a door open to the conscious control of habits, since

they "immediately call our attention if they go wrong". Although we will not discuss the issue here for the sake of brevity, recent neuroscience research works accept that goals and habits are not strangers to each other: they can be intertwined in various ways, and not just during habit acquisition (Wood and Neal, 2007; Dezfouli and Balleine, 2013; Duncan, 2013).

Seger and Spiering uncover more limitations of this narrow view of habits, although they do not question the theoretical background that underlies them. When referring to habits and habit learning, the common interpretation of these phenomena employs several dichotomies to clearly distinguish them from goal-directed actions: rigid/flexible, slow/fast, unconscious/conscious, automatic/controlled, insensitive/sensitive to outcome revaluation (Seger and Spiering, 2011). At this point, we want to stress that these authors challenge a restrictive view of these "defining" characteristics of habits: (1) action categorization in the basal ganglia makes it possible to deal with new stimuli as if they were well-learned, allowing for some flexibility; (2) it is not clear how many trials are necessary for an "action" to become a "habit", and for reaching a behavioral asymptote; (3) the various aspects of the basal ganglia (associative, sensorimotor and limbic) seem to be involved in conscious and unconscious learning, the distinction between which is far from being sharp (Horga and Maia, 2012); (4) automaticity has usually been assessed by dual-task performance, which is not actually an exclusive indicator of automaticity; and (5) outcome revaluation is a straight-forward method to be used in laboratory animals, but not quite so in humans.

In our opinion, the main problem with applying these categories to human behavior comes from the direct extrapolation of animal experiments to human research. Human cognitive resources are clearly different from those in animals. An inflexible comparison between the results of animal and human research would only shed light on the lowest levels of human cognition. If we focus our research strictly on those habits that animals are able to perform, or on those that fulfill the current theoretical model, we will constrain research on human neuroscience to investigating very simple habits. This interspecies correspondence could be also a consequence of the associationist heritage of the concept of habit, if the researcher assumes that the smaller units that constitute habits are the same in humans and non-human animals. Therefore the main limitation is, in our opinion, that habits are held as being apart from cognition, and this is why they are considered ateleological, rigid, unconscious, automatic and insensitive to outcomes. We next intend to demonstrate that the first conception of habit, found in classical Greek philosophy, incorporates cognitive control as a crucial element in its acquisition and performance, and that this theoretical framework may help in overcoming the limitations posed by the associationist view of habits in neuroscience.

THE ORIGINAL DEFINITION OF HABIT

As we have seen, the dominant vision of habits in neuroscience conceives of them as a routine, very similar to the releasing mechanism that ethologists employ to analyze instinct (Tinbergen, 1951). The main difference between the two is that habits are not innate but acquired. After acquisition, they are considered to behave in a similar way as instincts: inflexibly, automatically and unconsciously.

However, this is not the first characterization of habit, historically speaking. The pioneering definition and analysis of habits were carried out by Aristotle, whose view has the great advantage of not being conditioned by the sharp distinction between conscious and unconscious processes, a dichotomy which is frequent in modern and contemporary thought. He explains his conception of habit in his book *Nicomachean Ethics* (Aristotle, 2002). Our analysis is based on the original version in ancient Greek, although we will cite versions translated into English for clarity. We have also freely translated some terms to show their similarity with concepts currently used in neuroscience, as we explain below.

According to the Aristotelian view, acquired habits presuppose behavioral plasticity, so that the agent can acquire new patterns of behavior in order to achieve a desired adaptation and so be more successful in the pursuit of his/her goals. Aristotle characterizes habits as dispositions, that is, particular arrangements of human capacities. The cornerstone that underlies the Aristotelian theory of action is the following: when an agent does or makes something, there is an effect not only on the receiver of the action or the product made, but also—and more importantly—on the agent. This is mainly explained in Book IX, chapter 8, 1050b 23–38 of *Metaphysics* (Aristotle, 2007) and in Book 3, chapter 7, 431a 5–9 of *On the soul* (Aristotle, 1986). Since human actions are driven and controlled by cognition, each new action leaves a footprint in the agent as a kind of learning: a disposition to face further similar situations in a certain way, which includes the interpretation of that situation and the possible ways of dealing with it. In some types of learning, this disposition also includes affective control and corporal skills. Please note that although Aristotle is highly subtle in his analysis, his conclusions are plain: one acquires a new ability by doing or making things related to that ability. As he states in Book 2 of the *Nicomachean Ethics*: "for the things we have to learn before we can do them, we learn by doing them, e.g., men become builders by building and lyre players by playing the lyre; so too we become just by doing just acts, temperate by doing temperate acts, brave by doing brave acts" (Aristotle, 2002). We add: through our actions, we acquire the disposition or habit of being builders, mathematicians, piano players or temperate. Since these habits are gained through practice, this process is goal directed.

In these first paragraphs we have outlined the Aristotelian view on habits, their place in his philosophical system and their characterization as learned dispositions. Next, we will show the ability of this conception to account for "good" and "bad" habits. As we mentioned above, this categorization has been suggested by Graybiel, who considers good habits as being those which we try to incorporate in our behavior, and bad habits as those that take control of our behavior (Graybiel, 2008). If we consider a habit to be a mere motor routine (or a behavior that remains unchanged after outcome devaluation or degradation of action-outcome contingency), it is hard to categorize it as good or bad in itself, because this usually depends on the context in which it is triggered. In our opinion, there is a key factor involved in considering habits as good or bad, appropriate or inappropriate: cognitive

control. Through it, the agent can direct his or her behavior more adequately to the goal. If the acquisition of a habit implies a better cognitive control of the actions related to that habit, it can be considered as "good". Otherwise, if it involves rigidity and blurs the goal, it is a "bad" habit. Since "good" and "bad" (or the Aristotelian terms "virtue" and "vice") may sound odd to the neuroscientific community, we will term them "habits-as-learning" and "habits-as-routines", thus highlighting the behavioral plasticity or the rigidity they lead to. In Book V of Aristotle's *Metaphysics*, he states that "'habit' means a disposition according to which that which is disposed is either well or ill disposed, and either in itself or with reference to something else" (Aristotle, 2007). This, in our opinion, links habits to cognitive control and goals. Please note that the usual view of habit in neuroscience, inherited from associationism, corresponds to that subtype of habit that we have termed habits-as-routine. Therefore, habits-as-routines could be considered a cognitively-impoverished type of habits-as-learning. This is not surprising if we consider that such view has been elaborated on the basis of animal experiments, whose cognitive abilities are very limited in comparison with adult humans. We will elaborate on this point in the next section of our article, in order to demonstrate how the Aristotelian view gives an account of habits as motor routines, addictions and slips-of-action.

The main reason why Aristotle analyzed the nature of habits was to focus on ethics. From his point of view, ethics imply a broad study of human behavior (please note that "ethics" derives from "ethos", which means conduct or behavior). However, the path he followed included a classification of acquired dispositions that can be of great interest for neuroscience. He distinguished three kinds of acquired habits, originally termed dianoethical, ethical and technical (Aristotle, 2002). In order to assist in a better understanding of the three types, we will use an updated terminology: theoretical, behavioral and technical. First, theoretical habits consist in the retention of learning. This is different from memory, the plain retention of former experiences. Theoretical habits are not acquired through mere experience and repetition, but require comprehension. A good example is the understanding of a mathematical discipline, like geometry, and the capacity to understand its internal coherence. A human being does not become a mathematician through simple repetition of operational routines; instead, he or she must understand mathematical concepts and theorems along with the deductions that prove them. Therefore, while comprehension is a key element of this type of habit, this is not the case for repetition: it depends on the quality of the action whether or not repetition improves the ability to understand the internal coherence of the discipline. Once acquired, a theoretical habit allows the agent to understand new concepts and propositions, and even to improve that particular discipline. This kind of habit is some sort of "know that". In spite of the theoretical nature of these habits, they have a major influence on praxis because they allow cognitive abilities to develop. In neuroscience, this type of habit is usually studied as explicit memory (Schacter, 1987; Gabrieli et al., 1998), the "aha effect" (Luo et al., 2004)—the positive emotional response after understanding a concept or solving a problem—or the learning of a cognitive skill (Ashkenazi et al., 2013), for example.

The two remaining types of habits, however, improve behavior as well as the cognitive abilities that make it possible, rather than the theoretical abilities of the agent. In any case, Aristotle understands them as cognitive capacities as well, instead of mere routines. The second type is the behavioral habit, which depends on and is oriented towards *phronesis*: the habit of choosing and carrying out the best option for the agent in every situation. As Aristotle writes in the *Nicomachean Ethics*, "just as to practice medicine and healing consists not in applying or not applying the knife, in using or not using medicines, but in doing so in a certain way". *Phronesis* is a Greek term usually translated as prudence, which is the perfection of practical reason. By exercising *phronesis*, the agent achieves a rational control of desires (*epithymia*: temperance) and impulses (*thymós*: fortitude). In turn, desires are another kind of behavioral habits connected to emotions. Hence, the key point here is that emotions can be rationally governed, and behavioral habits are the improvement of such control through qualified practice. *Phronesis*, or practical wisdom, also affects decision making by way of this adaptation of emotional responses to rationally proposed goals. Therefore, behavioral habits can be defined as knowing how to act; they are the basis of ethics and are studied in neuroscience under the umbrella of decision making (Caspers et al., 2011), moral judgments (Moll and de Oliveira-Souza, 2007) and in the context of the interplay between cognition and emotion (Pessoa, 2013).

The third type is technical habits, which include those learned skills of doing or making things qua directed to an external goal. They usually entail embodied skills, as in the case of playing a musical instrument, painting or competitive running. Motor routines, understood as habits by the associationist view and by neuroscience, would be included under the umbrella of technical habits, since in general technical habits involve the acquisition of psychomotor skills that, of course, are improved through practice. However, this third Aristotelian class of habits are not *just* habits-as-routines, since technical habits are also rationally controlled and, ultimately, goal-directed: knowing how to play the piano involves mastering certain motor skills, but also—and more importantly—putting them into practice in the right way and at the right moment. As Averroes—a philosopher of the 12th century and an expert on Aristotle—wrote, "habit is that whereby we act when we will". This third kind of disposition consists therefore in knowing how to make or how to do. In neuroscience today, these habits are mainly analyzed as procedural learning (Censor et al., 2014; Pinho et al., 2014).

CONTRIBUTIONS OF THIS DEFINITION OF HABIT TO NEUROSCIENCE

This section intends to show why the Aristotelian theory of habits should be of interest for neuroscience. Before proceeding, we would like to clarify the main conclusions drawn from our analysis of the Aristotelian conception of habit presented in the previous chapter: (1) an acquired habit is an acquired disposition to perform certain types of actions; (2) this disposition, usually acquired by means of repetition of one or more actions, makes the execution of these actions prompter, more spontaneous and autonomous from continuous conscious supervision, all of which generally leads to a better performance; and (3) if the habit

increases cognitive control of the actions, it can be termed a habit-as-learning; if on the contrary it increases their rigidity, it is a habit-as-routine.

As we mentioned above, the associationist view based on William James's thought and introduced in experimental psychology has been extremely successful for understanding habits-as-routines in animals; however, it is quite limited when applied to human behavior, such as satisfactorily explaining good and bad habits, resolving the opposition between goal-directed actions and habits, overcoming the sharp distinction between conscious and unconscious processes or, more importantly, clarifying the role of cognitive control in human habits.

First, we would like to focus on the most important consequence of Aristotle's research on this topic: habits contribute to the cognitive enrichment of actions. As in the case of the notion of habits-as-routines, all these capacities are acquired through a variable amount of practice: we become scientists through correct intellectual activity (Aristotle, 2002), and we improve the performance of a sequential finger motor routine through repetition (Nissen and Bullemer, 1987). However, habits-as-learning are not just the acquisition of a way of acting, but rather involve a cognitive capacity connected to the habit that can be flexibly utilized in different situations. As in the case of habits-as-routines, this capacity eliminates the need for fully conscious control of the basic components of the action in order to make possible the agent's orientation to further and higher goals. For example, the pianist who can easily read the notes from the score (a mostly theoretical habit), and whose fingers appropriately respond to this reading (a technical habit) is able to exploit the expressive possibilities of the instrument. In summary, this feature of habits-as-learning is very important in order that this kind of habit may be read as a cognitive enrichment of behavior rather than as the acquisition of a routine. In the case of behavioral and technical habits they imply the availability of motor skills for complex activities, as well as the modulation of tendencies and desires to respond positively to conscious and rational goals. Therefore, they involve the acquisition of habits-as-routines, but their critical characteristics go beyond their motor aspects.

Second, another important difference between habits-as-routines and habits-as-learning is their differing relation to consciousness. For the former, habit performance is fully unconscious. In the latter, habits reduce or eliminate consciousness of basic elements of the action in order to concentrate on higher goals, while preserving at all times the possibility of recovering them for conscious attention. Although they seem unconscious and routinely performed, they are at the disposition of consciousness. They are not, in any case, rigid sequences. The possibility of developing habits-as-learning lies precisely in the feasibility of chunking those movements, actions and sequences, in order to organize them in other ways to perform different actions. Thus, pianists can learn how to play piano by repeating several motor routines, but they are not restricted to playing the routines they practice: their ability goes beyond those fixed movements to include improvisation. Therefore, the definition of habits-as-learning does not depend on the dichotomy of consciousness vs. unconsciousness. This is particularly important when this opposition is at stake in certain authors (Horga and Maia, 2012; Cleeremans, 2014).

The third contribution is two-fold: the Aristotelian view on habits allows us to understand the classification into good and bad habits, as well as to explain habits-as-routines (the notion of habit currently used in neuroscience) as a subtype of habits-as-learning. As we have outlined above, "good habits" can be defined as those that improve cognitive control, whereas "bad habits" are rigid behaviors nearly impossible to be cognitively regulated by the agent. This is intimately related to the interplay between habits and goals: since "good" habits involve an enhanced cognitive control, they lead us to a rationally proposed goal. In turn, this goal is enriched by the habit, as we explained in the case of the experienced pianist, who can concentrate on a better interpretation of the musical piece. On the other hand, the rigidity of "bad" habits leads us towards unwanted (non-rational) goals or away from rationally selected aims. Thus, addictions (Everitt and Robbins, 2005), compulsions (Gillan et al., 2011) and the susceptibility to slips-of-action (Norman, 1981) may be considered "bad" habits (or habits-as-routines) that could be due to a cognitive impoverishment of learned skills among other reasons. The acquisition of theoretical, behavioral and technical habits requires repetition; however, a high amount of it is not strictly necessary in the case of theoretical habits, since they can be acquired even by a single comprehension. In any case, it is important to emphasize that this repetition has to be qualified, rather than plain: the budding pianists have to have self-discipline in order to acquire habits that will help them to become virtuous. If they get used to performing the wrong movements, their ability will deteriorate. What is the critical feature that distinguishes a virtuous pianist from a regular piano player? It is, in our opinion, behavioral plasticity. If a student has acquired a set of cognitive-driven routines such that he or she can use them when they want to, it will result in a flexible performance. On the contrary, when routines have been learned through non-cognitive repetition, the final performance will be reduced to that set of routines.

This is also the case for behavioral habits: repetition of wrong behaviors causes the acquisition of habits with a poor or non-existent cognitive content. These are more similar to those routines that trigger "irrational" behaviors, such as addictions, compulsions and slips-of-actions. We would also include here unconscious biases that lead the agent to make a decision without considering all the relevant information (Kahneman, 2011). This would be a sort of "intellectual slip-of-action" that leads the agent to inadequately constrain the environment to be considered when making a decision. Rigidity is a consequence of the acquisition of habits that do not imply a cognitive enrichment of the action. Moreover, it is possible that some acquired skills may fall into rigidity and automation as a consequence of the decaying of higher cognitive functions, which by definition are in charge of controlling, reorganizing and reassessing acquired patterns. In fact, inappropriate habits imply a "negative" learning style that causes rigidity, as in the case of addiction or those technical habits-as-routines that cause difficulties with taking advantage of our possibilities—as in the case of the regular, but not virtuous, pianist.

Finally, the Aristotelian view on habits may provide new insights on the emotional response of the agent after habit acquisition. This is related, as we explain below, to the wanting/liking unbalance in drug addiction (Robinson and Berridge, 1993). Since cognitively controlled habits help the agent achieve rationally proposed goals, they tend to increase the enjoyment of the agent when performing such actions. However, the rigidity of habits-as-routines and the consequent blurring of goals diminish this enjoyment. Interestingly, this is in line with the current experimental approach to habits-as-routines in neuroscience: if the animal performs a goal-directed action, it has the pleasure of obtaining the reward; however, if its behavior has become a habit-as-routine, the "pleasure" is transferred to the response itself or to the cue that anticipates its performance. This results in an increased craving and a decreased pleasure after the outcome (Volkow et al., 2010).

NEW PERSPECTIVES FOR AN INTERDISCIPLINARY RESEARCH

How can the Aristotelian notion of habit be of use in future research in neuroscience? In this last section, we would like to point to possible new directions for research on habits in human neuroscience. Whereas the successful experimental approach employing the associationist view of habits focuses on outcome devaluation and the degradation of the action-outcome contingency (Adams and Dickinson, 1981; Adams, 1982; Dickinson, 1985), we propose new criteria to be considered when researching human habits as a whole (both habits-as-learning and habits-as-routines). First, a habit will have been incorporated when its related actions are performed more spontaneously, that is, with greater promptitude. This could be quantified by a decrease in the reaction time of the deliberation prior to the action. Second, habit acquisition would imply a more accurate performance of the action, especially in the case of technical habits, measured by a decreased number of errors. Third, a categorization as habit-as-learning or habit-as-routine could be done by assessing cognitive control; behaviorally, this could be tested by error monitoring and adequately switching to a different task; neuroanatomically, by the recruitment of prefrontal regions and cognitive aspects of the basal ganglia. Finally, considering the relationship between the cognitive control of habits-as-learning and the enjoyment of their performance, a further indicator of their acquisition would be the recruitment of the reward system both before and after performance, whereas habits-as-routines would mainly involve the neuroanatomical "wanting" (incentive salience) system.

After these general experimental considerations, we focus on other topics within neuroscience where the dichotomy between habits-as-learning and habits-as-routines could be of great use.

HABITUAL DECISION MAKING

In a recent publication, we highlighted the difficulty of defining conscious (vs. unconscious) processes in "habitual decision making" (Bernacer et al., 2014). We are aware that this concept may sound provocative in neuroscience, since habits are related with unconscious phenomena, and decision making is mainly considered conscious, at least according to some accounts (Newell and Shanks, 2014). However, the nature of a decision should be considered with reference to its final goal. Driving is a technical habit that entails a high number of decisions, most of which are unconsciously made and performed: changing gear, putting the clutch in, switching on the indicator when turning, etc. However, driving is a conscious process overall: we decide to start the process, we consciously set the goal, and our driving is continuously available to conscious supervision. This framework is similar to the hierarchical model by Dezfouli and Balleine (2013), according to which habits are at the service of goal-directed behaviors. If we keep maintaining the extreme dichotomy between goal-directed actions and habits, we will be ruling most human activities out of the reach of neuroscience.

Furthermore, neuroscience can study the interplay between habits and decision making from another perspective. All three types of habits considered by Aristotle (theoretical, behavioral and technical) can be viewed as dispositions to configure one's acting and, therefore, decision making. In recent years there have been a plethora of studies to determine the neural bases of human decision making (see, for example, the editor's introduction to a special issue on this topic (Doya and Shadlen, 2012)). In short, it seems to be clear that the main players are the ventromedial prefrontal cortex (Levy and Glimcher, 2012), striatum and substantia nigra (Balleine et al., 2007). In addition, more dorsal aspects of the prefrontal cortex supervise the whole process (Manes et al., 2002), and other cortical regions are especially active in highly uncertain decisions (Hsu et al., 2005; Goñi et al., 2011). In a very simplistic—albeit accurate—way, humans decide to perform the action that carries the highest subjective value. This value depends on personal preferences, which in turn rely on the history of actions, decisions, skills and dispositions that the agent has carried out or acquired during his or her life. Thus, in many cases, decision making depends on habits. For example, it is well known that temporal discounting depends on personal preferences: people may tend to be either impulsive or else patient, and temporal discounting has been reported to correlate with the BOLD signal in the ventromedial prefrontal cortex (Kable and Glimcher, 2007). But, how do we initially become impulsive or patient? Is it encoded in our genes or in our neurotransmitters? Can our actions change this feature of our personality, as well as its in-brain correlate? In our opinion, the role of habits in decision making is a key topic for future research in cognitive neuroscience.

RESEARCHING HABITS-AS-LEARNING IN NEUROSCIENCE

In the classical Aristotelian view, when the agent acquires a (good) habit, he or she performs the action: (1) more easily; (2) more efficiently; and (3) with higher enjoyment. This can be exemplified with the healthy habit of running: at the beginning, the jogger has to struggle to find the perfect time to go out, he or she can only run a short distance, and finds it definitely painful. However, as days go by, all three nuisances become increasingly tolerable.

Habits contribute to improving action performance because they release consciousness from having to focus on immediate goals, and allow all cognitive resources to focus instead on higher goals. This is the key idea for understanding how habits induce behavioral plasticity. A good pianist is able to improvise and concentrate on the artistic eloquence of the piece, because

his or her acquired habit allows the player to go beyond the mere movements of his or her hands. This has been partially studied in neuroscience under the umbrella of "dual-tasking", one of the measures of habits-as-routines. When a motor task is being learned, it requires the agent to expend a high amount of energy, since many executive brain areas are active; however, after practicing, brain activation is more restricted and energy consumption is thus lower (Poldrack et al., 2005). At the beginning, different aspects of the prefrontal cortex as well as their striatal targets—mainly the caudate nucleus—are in charge of the process; however, when the task is mastered, the activation of these areas is decreased and the putamen, globus pallidus and supplementary motor area of the cortex have a higher BOLD signal. This allows the prefrontal cortex and caudate nucleus to engage in a novel task when performing the well learned sequence.

This neuroanatomical framework is useful to understand those aspects of habits related with the automation of behavior. Automation is a condition for developing most habits-as-learning, since it releases executive areas from a continuous supervision of certain tasks. Thus, automations allow a cognitive enrichment of actions. It would be interesting to research, from the point of view of neuroscience, how this interplay between executive and "habit related" areas is carried out not only in motor routines, but also in more cognitive habits. For example, solving a Sudoku puzzle for the first time may seem overwhelming. With practice, the player discovers that as a result of performing certain intellectual routines the puzzle is easier to tackle. Furthermore, once these routines are acquired, it is easier for the player to monitor for errors and deal with new challenges within the puzzle. An area of possible future research is opened here, since error monitoring and problem solving will find their neural correlate in the prefrontal and anterior cingulate cortex; will, however, the intellectual routines be coded in the posterior putamen and premotor areas of the cortex?

Another interesting subject to investigate in the future is increased enjoyment in habit performance. Since this could be an extremely broad topic, we will only suggest its outlines here. For a start, it will be necessary to have an adequate characterization of pleasure and enjoyment. Human neuroscience assumes the reward circuit is an analog to that of non-human animals: unquestionably, regions such as the substantia nigra and the ventral striatum are active when an animal—rat, monkey or human—receives a primary reward (Schultz, 1997). It also happens when humans are granted a secondary reward, such as money (Delgado, 2007). However, human beings are also able to interpret as rewards things that are far from being pleasurable, including physically painful experiences. Thus, it may be appropriate to search for the brain correlates of these phenomena.

ADDICTIONS, COMPULSIONS, STEREOTYPIES AND SLIPS-OF-ACTION: BAD HABITS

Our article suggests a concept of habit that broadens the current view in neuroscience; for that reason, it has to be compatible with that very view. In this section we will briefly clarify how habits-as-learning can decay in humans to being habits-as-routines, leading thus to a behavioral rigidity of the agent, instead of flexibility. As we mentioned above, a habit turns into a mere automatic routine when its associated cognitive control declines. The role of cognition in habits is found first in goal setting. The initiation of a set of motor routines is meaningful if they serve a goal; otherwise, they can evolve into a compulsion or stereotypy. In his interesting review, Duncan (2013) cites the work of the Italian psychiatrist Bianchi (1922), who ablated different parts of the frontal lobe in monkeys to investigate changes in their behavior. Duncan highlights the following section of Bianchi's work: "The monkey which used to jump on to the window-ledge, to call out to his companions, after the operation jumps on to the ledge again, but does not call out. The sight of the window determines the reflex of the jump, but the purpose is now lacking, for it is no longer represented in the focal point of consciousness... Evidently there are lacking all those other images that are necessary for the determination of a series of movements coordinated towards one end". Therefore, the monkey can perform goal-directed motor habits while its frontal lobe is intact; when its cognitive function is damaged, the motor habits disengage from their goal and become a plain meaningless routine.

In any case, we believe that animal research on habits should be given due weight when transferred to humans, since the latter are able to acquire and perform habits that the former are not. The reason for this has been just outlined: cognitive control. The brain area involved in this phenomenon is the prefrontal cortex, which finds its greatest evolution in humans (Miller, 2000). Even though most of our knowledge on human neuroscience comes from animal studies, these are informative in the case of habits-as-routines, rather than habits-as-learning, because of the role of the cognitive enrichment of actions.

Finally, the lack of cognitive control in habits could be also a crucial feature in the case of compulsions in obsessive-compulsive disorder: washing one's hands triggers a set of motor routines towards the goal of personal hygiene. However, when someone washes them repeatedly without a purpose, it becomes a compulsion. Slips-of-action may be understood as a temporary disengagement between a habit and its goal. They happen when the agent starts a goal-directed set of routines—for example, driving to a friend's party—and ends up reaching an unwanted goal—arriving to the office instead of the party because the initial part of the driving routine is the same. In fact, these action errors have been related to obsessive-compulsive disorder (Gillan et al., 2011). Our proposal here, very briefly, is that routines could be incorporated in the agent—that is, coded into his/her brain, but would remain inhibited most of the time. Only when that routine needs to be executed do higher executive areas allow its performance through disinhibition. Again, in situations when cognitive control is compromised, the routine may be executed unwantedly (Mendez et al., 1997).

In sum, we believe that future research on stereotypies and compulsions should maintain its connection to the study of habits, but should focus on the lack of cognitive control rather than on the neural bases of the established motor routine, and bearing in mind that encouraging the patient to acquire cognitive-driven habits may help overcome rigid routines (Güell and Nuñez, 2014).

CONCLUSION

Our interdisciplinary research seeks to provide a more adequate framework for the study of habits in neuroscience. We have demonstrated that this perspective is compatible with past and current experimental research, and it expands its scope when applied to humans. From a holistic view of human behavior, habits are very important aspects for behavioral plasticity and learning, since they release cognitive areas to focus on higher goals. Further, even though repetition and routines are important for habit acquisition, they can also be considered a crucial element for human behavioral freedom (Bernacer and Gimenez-Amaya, 2013), inasmuch as they increase the repertoire of actions and allow a better cognitive control of behavior.

Revisiting the ideas of classical philosophers may be very useful for neuroscience, since the cognitive and psychological substrate of neuroscience is composed, at least in a high proportion, of the ideas developed by philosophers across the centuries. We believe interdisciplinary research may help achieve a better understanding of human behavior, and provide the neuroscience community with an adequate theoretical background for undertaking new experimental approaches and tackling the challenges resulting from them.

ACKNOWLEDGMENTS

We are grateful to all members of the Mind-Brain Group for the fruitful discussions during the preparation of this manuscript. This research has been supported by Obra Social La Caixa.

REFERENCES

Adams, C. (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Q. J. Exp. Psychol. Sect. B Comp. Physiol. Psychol.* 34, 77–98. doi: 10.1080/14640748208400878

Adams, C., and Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *Q. J. Exp. Psychol. Sect. B Comp. Physiol. Psychol.* 33, 109–121. doi: 10.1080/14640748108400816

Aristotle. (1986). *On the Soul.* New York: Penguin Books.

Aristotle. (2002). *Nicomachean Ethics.* New York: Oxford University Press.

Aristotle. (2007). *Metaphysics.* Mineola, NY: Dover.

Ashkenazi, S., Black, J. M., Abrams, D. A., Hoeft, F., and Menon, V. (2013). Neurobiological underpinnings of math and reading learning disabilities. *J. Learn. Disabil.* 46, 549–569. doi: 10.1177/0022219413483174

Balleine, B. W., Daw, N., and O'Doherty, J. (2008). "Multiple forms of value learning and the function of dopamine," in *Neuroeconomics: Decision Making and the Brain*, eds P. W. Glimcher, C. F. Camerer, R. A. Poldrack and E. Fehr (New York: Academic Press), 367–387.

Balleine, B. W., Delgado, M. R., and Hikosaka, O. (2007). The role of the dorsal striatum in reward and decision-making. *J. Neurosci.* 27, 8161–8165. doi: 10.1523/jneurosci.1554-07.2007

Balleine, B. W., and Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37, 407–419. doi: 10.1016/s0028-3908(98)00033-1

Barandiaran, X. E., and Di Paolo, E. A. (2014). A genealogical map of the concept of habit. *Front. Hum. Neurosci.* 8:522. doi: 10.3389/fnhum.2014.00522

Barnes, T. D., Kubota, Y., Hu, D., Jin, D. Z., and Graybiel, A. M. (2005). Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature* 437, 1158–1161. doi: 10.1038/nature04053

Bernacer, J., Balderas, G., Martinez-Valbuena, I., Pastor, M. A., and Murillo, J. I. (2014). The problem of consciousness in habitual decision making. *Behav. Brain Sci.* 37, 21–22. doi: 10.1017/s0140525x13000642

Bernacer, J., and Gimenez-Amaya, J. (2013). "On habit learning in neuroscience and free will," in *Is Science Compatible with Free Will?*, eds A. Suarez and P. Adams (New York: Springer), 177–193.

Bianchi, L. (1922). *The Mechanism of the Brain and the Function of the Frontal Lobe.* Edinburgh: Livingston.

Blanco, C. A. (2014). The principal sources of William James' idea of habit. *Front. Hum. Neurosci.* 8: 274. doi: 10.3389/fnhum.2014.00274

Caspers, S., Heim, S., Lucas, M. G., Stephan, E., Fischer, L., Amunts, K., et al. (2011). Moral concepts set decision strategies to abstract values. *PLoS One* 6:e18451. doi: 10.1371/journal.pone.0018451

Censor, N., Dayan, E., and Cohen, L. G. (2014). Cortico-subcortical neuronal circuitry associated with reconsolidation of human procedural memories. *Cortex* 58, 281–288. doi: 10.1016/j.cortex.2013.05.013

Cleeremans, A. (2014). Connecting conscious and unconscious processing. *Cogn. Sci.* 38, 1286–1315. doi: 10.1111/cogs.12149

Delgado, M. R. (2007). Reward-related responses in the human striatum. *Ann. N Y Acad. Sci.* 1104, 70–88. doi: 10.1196/annals.1390.002

Dezfouli, A., and Balleine, B. W. (2013). Actions, action sequences and habits: evidence that goal-directed and habitual action control are hierarchically organized. *PLoS Comput. Biol.* 9:e1003364. doi: 10.1371/journal.pcbi.1003364

Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Philos. Trans. R. Soc. B Biol. Sci.* 308, 67–78. doi: 10.1098/rstb.1985.0010

Dickinson, A., Squire, S., Varga, Z., and Smith, J. (1998). Omission learning after instrumental pretraining. *Q. J. Exp. Psychol.* 51B, 271–286.

Doya, K., and Shadlen, M. N. (2012). Decision making. *Curr. Opin. Neurobiol.* 22, 911–913. doi: 10.1016/j.conb.2012.10.003

Duncan, J. (2013). The structure of cognition: attentional episodes in mind and brain. *Neuron* 80, 35–50. doi: 10.1016/j.neuron.2013.09.015

Everitt, B. J., and Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat. Neurosci.* 8, 1481–1489. doi: 10.1038/nn1579

Gabrieli, J., Poldrack, R., and Desmond, J. (1998). The role of left prefrontal cortex in language and memory. *Proc. Natl. Acad. Sci. U S A* 95, 906–913. doi: 10.1073/pnas.95.3.906

Gillan, C. M., Papmeyer, M., Morein-Zamir, S., Sahakian, B. J., Fineberg, N. A., Robbins, T. W., et al. (2011). Disruption in the balance between goal-directed behavior and habit learning in obsessive-compulsive disorder. *Am. J. Psychiatry* 168, 718–726. doi: 10.1176/appi.ajp.2011.10071062

Goñi, J., Aznárez-Sanado, M., Arrondo, G., Fernández-Seara, M., Loayza, F. R., Heukamp, F. H., et al. (2011). The neural substrate and functional integration of uncertainty in decision making: an information theory approach. *PLoS One* 6:e17408. doi: 10.1371/journal.pone.0017408

Graybiel, A. M. (1998). The basal ganglia and chunking of action repertoires. *Neurobiol. Learn. Mem.* 70, 119–136. doi: 10.1006/nlme.1998.3843

Graybiel, A. M. (2008). Habits, rituals and the evaluative brain. *Annu. Rev. Neurosci.* 31, 359–387. doi: 10.1146/annurev.neuro.29.051605.112851

Güell, F., and Nuñez, L. (2014). The liberating dimension of human habit in addiction context. *Front. Hum. Neurosci.* 8:664. doi: 10.3389/fnhum.2014.00664

Horga, G., and Maia, T. V. (2012). Conscious and unconscious processes in cognitive control: a theoretical perspective and a novel empirical approach. *Front. Hum. Neurosci.* 6:199. doi: 10.3389/fnhum.2012.00199

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., and Camerer, C. F. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310, 1680–1683. doi: 10.1126/science.1115327

James, W. (1890). *Principles of Psychology.* New York: Henry Holt.

Kable, J. W., and Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nat. Neurosci.* 10, 1625–1633. doi: 10.1038/nn2007

Kahneman, D. (2011). *Thinking, Fast and Slow.* New York: Farrar, Straus and Giroux.

Levy, D. J., and Glimcher, P. W. (2012). The root of all value: a neural common currency for choice. *Curr. Opin. Neurobiol.* 22, 1027–1038. doi: 10.1016/j.conb.2012.06.001

Luo, J., Niki, K., and Phillips, S. (2004). Neural correlates of the "Aha! reaction". *Neuroreport* 15, 2013–2017. doi: 10.1097/00001756-200409150-00004

Manes, F., Sahakian, B., Clark, L., Rogers, R., Antoun, N., Aitken, M., et al. (2002). Decision-making processes following damage to the prefrontal cortex. *Brain* 125, 624–639. doi: 10.1093/brain/awf049

Mendez, M., Perryman, K., Miller, B., Swartz, J., and Cummings, J. L. (1997). Compulsive behaviors as presenting symptoms of frontotemporal dementia. *J. Geriatr. Psychiatry Neurol.* 10, 154–157. doi: 10.1177/089198879701000405

Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nat. Rev. Neurosci.* 1, 59–65. doi: 10.1038/35036228

Moll, J., and de Oliveira-Souza, R. (2007). Moral judgments, emotions and the utilitarian brain. *Trends Cogn. Sci.* 11, 319–321. doi: 10.1016/j.tics.2007.06.001

Newell, B. R., and Shanks, D. R. (2014). Unconscious influences on decision making: a critical review. *Behav. Brain Sci.* 37, 1–19. doi: 10.1017/s0140525x12003214

Nissen, M., and Bullemer, P. (1987). Attentional requirements of learning: evidence from performance measures. *Cogn. Psychol.* 19, 1–32. doi: 10.1016/0010-0285(87)90002-8

Norman, D. (1981). Categorization of action slips. *Psychol. Rev.* 88, 1–15. doi: 10.1037//0033-295x.88.1.1

Northoff, G. (2012). Immanuel Kant's mind and the brain's resting state. *Trends Cogn. Sci.* 16, 356–359. doi: 10.1016/j.tics.2012.06.001

Pessoa, L. (2013). *The Cognitive-Emotional Brain.* Cambridge, MA: MIT Press.

Pinho, A. L., de Manzano, Ö., Fransson, P., Eriksson, H., and Ullén, F. (2014). Connecting to create: expertise in musical improvisation is associated with increased functional connectivity between premotor and prefrontal areas. *J. Neurosci.* 34, 6156–6163. doi: 10.1523/jneurosci.4769-13.2014

Poldrack, R. A., Sabb, F. W., Foerde, K., Tom, S. M., Asarnow, R. F., Bookheimer, S. Y., et al. (2005). The neural correlates of motor skill automaticity. *J. Neurosci.* 25, 5356–5364. doi: 10.1523/jneurosci.3880-04.2005

Robinson, T. E., and Berridge, K. C. (1993). The neural basis of drug craving: an incentive-sensitization theory of addiction. *Brain Res. Rev.* 18, 247–291. doi: 10.1016/0165-0173(93)90013-p

Schacter, D. L. (1987). Implicit memory: history and current status. *J. Exp. Psychol.* 13, 501–517. doi: 10.1037/0278-7393.13.3.501

Schultz, W. (1997). Dopamine neurons and their role in reward mechanisms. *Curr. Opin. Neurobiol.* 7, 191–197. doi: 10.1016/s0959-4388(97)80007-4

Seger, C. A. (2008). How do the basal ganglia contribute to categorization? Their roles in generalization, response selection and learning via feedback. *Neurosci. Biobehav. Rev.* 32, 265–278. doi: 10.1016/j.neubiorev.2007.07.010

Seger, C. A., and Spiering, B. J. (2011). A critical review of habit learning and the basal ganglia. *Front. Syst. Neurosci.* 5:66. doi: 10.3389/fnsys.2011.00066

Smith, K. S., and Graybiel, A. M. (2014). Investigating habits: strategies, technologies and models. *Front. Behav. Neurosci.* 8:39. doi: 10.3389/fnbeh.2014.00039

Thorn, C. A., and Graybiel, A. M. (2014). Differential entrainment and learning-related dynamics of spike and local field potential activity in the sensorimotor and associative striatum. *J. Neurosci.* 34, 2845–2859. doi: 10.1523/jneurosci.1782-13.2014

Tinbergen, N. (1951). The study of instinct. Available online at: http://psycnet.apa.org/psycinfo/2004-16480-000. Accessed on April 3, 2014.

Volkow, N. D., Wang, G.-J., Fowler, J. S., Tomasi, D., Telang, F., and Baler, R. (2010). Addiction: decreased reward sensitivity and increased expectation sensitivity conspire to overwhelm the brain's control circuit. *Bioessays* 32, 748–755. doi: 10.1002/bies.201000042

Wood, W., and Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychol. Rev.* 114, 843–863. doi: 10.1037/0033-295x.114.4.843

Yin, H. H., and Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nat. Rev. Neurosci.* 7, 464–476. doi: 10.1038/nrn1919

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 June 2014; accepted: 13 October 2014; published online: 03 November 2014.

Citation: Bernacer J and Murillo JI (2014) The Aristotelian conception of habit and its contribution to human neuroscience. Front. Hum. Neurosci. ***8****:883. doi: 10.3389/fnhum.2014.00883*

This article was submitted to the journal Frontiers in Human Neuroscience.

frontiers in
HUMAN NEUROSCIENCE

MINI REVIEW ARTICLE
published: 21 July 2014
doi: 10.3389/fnhum.2014.00522

A genealogical map of the concept of *habit*

Xabier E. Barandiaran[1,2]* and Ezequiel A. Di Paolo[2,3,4]

[1] Department of Philosophy, University School of Social Work, UPV/EHU University of the Basque Country, Vitoria-Gasteiz, Spain
[2] Department of Logic and Philosophy of Science, IAS-Research Center for Life, Mind, and Society, UPV/EHU University of the Basque Country, Donostia - San Sebastián, Spain
[3] Ikerbasque, Basque Foundation for Science, Bilbao, Spain
[4] Department of Informatics, Centre for Computational Neuroscience and Robotics, University of Sussex, Brighton, UK

Edited by:
Jose Ignacio Murillo, University of Navarra, Spain

Reviewed by:
Nathaniel Frost Barrett, Institute for Culture and Society, Spain
Clare Carlisle, King's College London, UK

***Correspondence:**
Xabier E. Barandiaran, Escuela Universitaria de Trabajo Social, UPV/EHU University of the Basque Country, Dpto. de Filosofía, C/ Los Apraiz, 2. 01006 – Vitoria-Gasteiz, Araba, Spain
e-mail: xabier.academic@barandiaran.net

The notion of information processing has dominated the study of the mind for over six decades. However, before the advent of cognitivism, one of the most prominent theoretical ideas was that of *Habit*. This is a concept with a rich and complex history, which is again starting to awaken interest, following recent embodied, enactive critiques of computationalist frameworks. We offer here a very brief history of the concept of habit in the form of a genealogical network-map. This serves to provide an overview of the richness of this notion and as a guide for further re-appraisal. We identify 77 thinkers and their influences, and group them into seven schools of thought. Two major trends can be distinguished. One is the associationist trend, starting with the work of Locke and Hume, developed by Hartley, Bain, and Mill to be later absorbed into behaviorism through pioneering animal psychologists (Morgan and Thorndike). This tradition conceived of habits atomistically and as automatisms (a conception later debunked by cognitivism). Another historical trend we have called organicism inherits the legacy of Aristotle and develops along German idealism, French spiritualism, pragmatism, and phenomenology. It feeds into the work of continental psychologists in the early 20th century, influencing important figures such as Merleau-Ponty, Piaget, and Gibson. But it has not yet been taken up by mainstream cognitive neuroscience and psychology. Habits, in this tradition, are seen as ecological, self-organizing structures that relate to a web of predispositions and plastic dependencies both in the agent and in the environment. In addition, they are not conceptualized in opposition to rational, volitional processes, but as transversing a continuum from reflective to embodied intentionality. These are properties that make habit a particularly attractive idea for embodied, enactive perspectives, which can now re-evaluate it in light of dynamical systems theory and complexity research.

Keywords: habit, associationism, organicism, history of psychology, history of philosophy

INTRODUCTION

For over 60 years the most basic theoretical concept in psychology, neuroscience, and cognitive science has been the processing of information and the associated notion of "mental representation." Neuroscientists search for modules and regions that process, store, retrieve or integrate information that is encoded or represented in the brain. But this hasn't always been the case. Before the advent of cognitivism in the 1950s one of the most prominent concepts for the study of mind was that of *Habit*. Despite constituting only very coarse evidence, the sub-plot in **Figure 1** (top-left) shows trends in the use of the words "habit" and "representation" since 1850. It is noteworthy that for most of the second half of the 20th century mentions of "habit" decrease and those of "representation" increase in a sustained manner. The anti-correlation is maintained with the reversal of these tendencies at the start of the 21st century, roughly indicating that habit is again becoming a notion of interest. This is no coincidence. Current embodied dissatisfactions with the information-processing framework (Varela et al., 1991; Kelso, 1995; Van Gelder, 1998; Thompson, 2007; Chemero, 2009; Di Paolo et al., 2010; Hutto and Myin, 2013) call for a reappraisal of this notion. The task, one quickly finds, is huge. The richness and polysemy of the notion of habit and its transformations since ancient Greece to the present day, all militate against the naïve idea of producing an off-the-shelf alternative theoretical primitive for psychology and neuroscience.

In this mini-review we offer a brief genealogy of the concept of habit in the form of a network-map. We place those thinkers who have worked on this concept in a historical relation. Our objective is to outline the genealogy of the notion of habit and identify major trends and schools of thought that have had an impact on current neuroscientific conceptions of habit and those that have not but still deserve attention.

As in the case with real maps, there is potentially no end to the amount of detail that may be included. The more detailed the map, the better the chances for efficient local navigation, but often at the price of losing the big picture. We have chosen to draw only the big picture (**Figure 1**). For this reason, links represent a general notion of "influence" between two thinkers, without going into relevant details such as, e.g., whether the

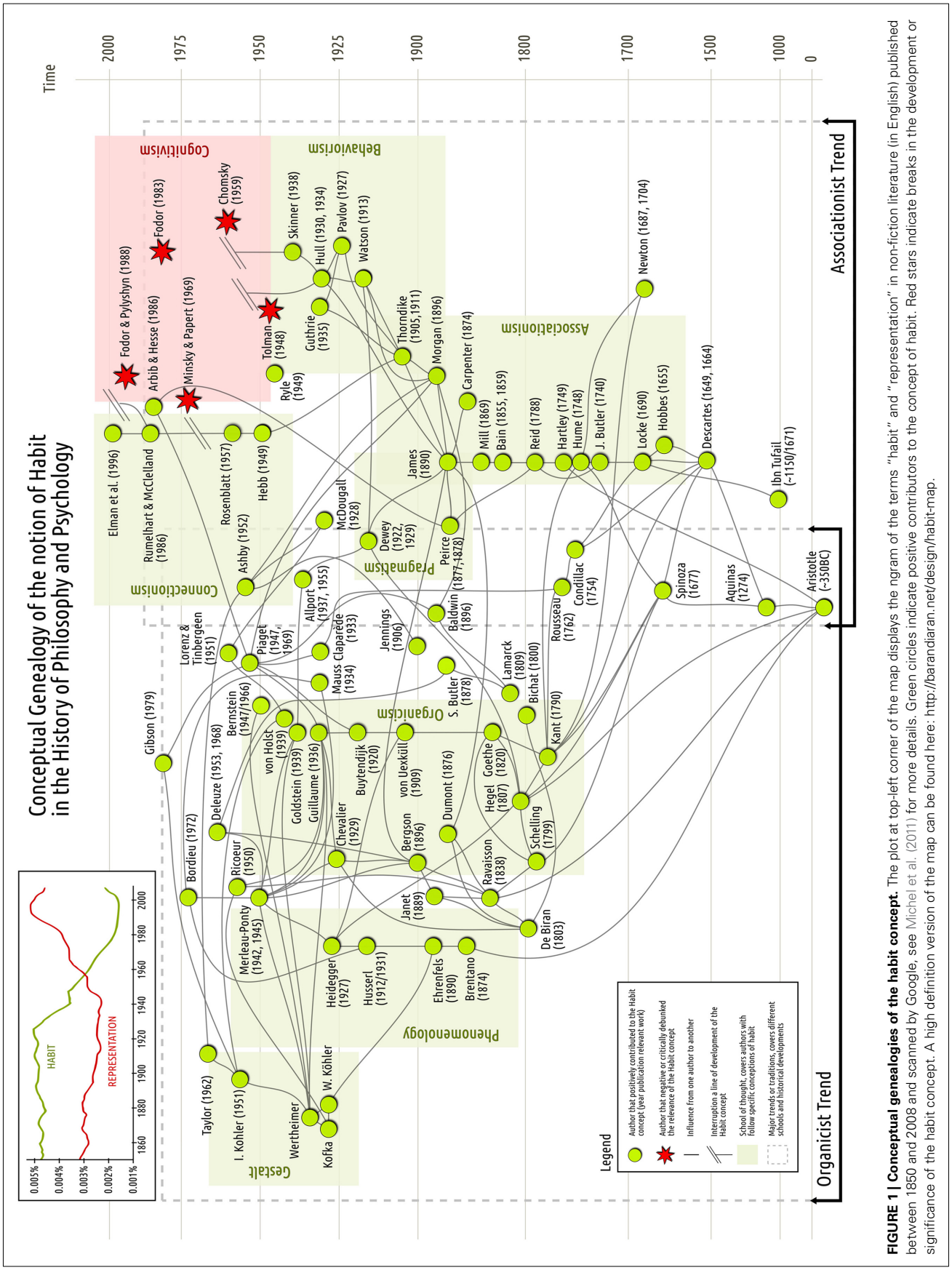

FIGURE 1 | Conceptual genealogies of the habit concept. The plot at top-left corner of the map displays the ngram of the terms "habit" and "representation" in non-fiction literature (in English) published between 1850 and 2008 and scanned by Google, see Michel et al. (2011) for more details. Green circles indicate positive contributors to the concept of habit. Red stars indicate breaks in the development or significance of the habit concept. A high definition version of the map can be found here: http://barandiaran.net/design/habit-map.

influence has been positive or critical, whether it is manifested as an explicit conceptual debt or as more subtle forms of inspiration, or indeed whether the same thinker's notion of habit has evolved significantly at different stages and under different influences.

It is likely that no two links in our map depict the exact same kind of influence. But a link describes at least an acknowledged or clearly recognized impact, which in most cases will be manifested as a direct reference to the influencing thinker in the works listed on **Table 1**. As a general rule transitive influences have not been drawn on the map and antagonistic links are also left out unless the critique of a previous conception of habit leads to a richer conception that integrates the view of the criticized author.

We have taken the general rule that all authors presented on the map should have discussed habits explicitly. But there are a few exceptions to this rule. For instance, Kant did not elaborate a strong positive contribution to the notion of habit—in fact, he is accountable for the ensuing divide between habit and reason in ethics—yet, his insights into the nature of teleology and self-organization strongly influenced the notion of habit, plasticity and holistic interdependence in various thinkers. Others do not make direct use of the term habit, but use parallel notions that were later (or previously) conceptualized as habits (such as von Uexküll's "functional cycles" or Pavlov's "reflexes"). The map still leaves out a considerable amount of literature on habit or habit-related research, e.g., work in economics, anthropology, psychoanalysis and research on habituation and addiction.

The timeline reaches up to the 1980s with some additional references to later work in the cognitivist and connectionist traditions added for completeness (Elman, Rumelhart, and McClelland, Arbib, Fodor, etc.). It is worth noting that a few authors appear (almost) without connections (von Holst, Bernstein, Ryle), yet their contributions are nowadays considered important. Gestalt psychologists, who together exert a notable influence on the habit concept without addressing it directly in their work, appear without a reference in **Table 1**.

The reader might still be left with a fundamental question regarding the key contribution of this map: What is the value of this genealogy for contemporary neuroscience? Whereas much work in human neuroscience appears informed by a rich philosophical, psychological and theoretical tradition (e.g., the neuroscience of perception, emotion or consciousness, cognitive or large-scale neuroscience), we believe that neuroscientific research on habit remains rooted within a narrow theoretical tradition. For instance, in an otherwise excellent review of recent work, Graybiel (2008) makes only a sparse reference to William James. Similarly, Wood and Neal (2007) only mention Thorndike and Skinner as conceptual precursors. This is understandable, as the history of habit is indeed complex and relatively unexplored. Our inherited conception appears historically distorted—only a few recent studies examine the genealogy of the concept (see Pollard, 2008; Carlisle and Sinclair, 2011; Carlisle, 2014). The map we present is an attempt to fill in this gap, providing a birds-eye view that can be used to navigate the history of the concept.

TRACING THE GENEALOGIES OF HABIT

Let us attempt a broad reading of the map. We identify two major historical trends, associationism and organicism, taking

Table 1 | List of authors and their most significant work related to habits. The year corresponds to the original publication and the title to the English translation (if available).

Year	Author	Work
−350	Aristotle	Nichomachean ethics, Metaphysics, De anima, De memoria and, Categories
∼1150/1671	Ibn Tufail	Philosophus autodidactus [Risala Hayy ibn Yaqzan fi asrar al-hikmat al-mashriqiyya]
1274	T. Aquinas	Summa theologica (Treatise on habit QQ49-54)
1649, 1664	R. Descartes	The passions of the soul and Treatise of man
1655	T. Hobbes	De corpore
1677	B. Spinoza	Ethics
1687, 1704	I. Newton	Philosophiae naturalis Principia mathematica and Opticks
1690	J. Locke	An essay concerning human understanding
1739, 1748	D. Hume	A treatise of human nature and Enquiry concerning human understanding
1740	J. Butler	The analogy of religion, natural and revealed, to the constitution and course of nature
1749	D. Hartley	Observations on man, his frame, his duty, and his expectations
1754	E. B. de Condillac	Treatise on the sensations
1762	J-J Rousseau	Émile or On education
1788	T. Reid	Essays on the active powers of the human mind
1790	E. Kant	Critique of judgment
1799	F. W. J. Schelling	First outline for a system of a philosophy of nature
1800	X. Bichat	Recherches physiologiques sur la vie et la mort
1803	M. de Biran	Influence de l'habitude sur la faculté de penser
1809	J. B. P. Lamarck	Zoological philosophy
1820	J. W. Goethe	Outline for a general introduction comparative anatomy, Commencing osteology
1830	G. W. F. Hegel	The philosophy of mind (Part 3 of the Encyclopaedia of the philosophical sciences)
1838	F. Ravaisson	Of habit
1855, 1859	A. Bain	Senses and the intellect, The emotions and the will
1869	J. Mill	Analysis of the phenomena of the human mind
1874	W. B. Carpenter	Principles of mental physiology
1874	F. C. Brentano	Psychology from an empirical standpoint
1876	L. Dumont	De l'habitude (Rev. Phil de la France et de l'Etranger)

(Continued)

Table 1 | Continued

Year	Author	Work
1877, 1878	C. S. Peirce	The fixation of belief, How to make ideas clear (see also Collected Papers)
1878	S. Butler	Life and habit
1889	P. Janet	L'Automatisme psychologique
1890	C. von Ehrenfels	Über Gestaltqualitäten
1890	W. James	Principles of psychology (Ch. 4 Habit)
1896	J. M. Baldwin	Mental development in the child and the race: Methods and processes
1896	C. L. Morgan	Habit and instinct
1896	H. Bergson	Matter and memory
1905, 1911	E. Thorndike	Elements of psychology, Animal intelligence: Experimental studies
1906	H. S. Jennings	Behavior of the lower organisms
1909	J. von Uexküll	Umwelt und Innenwelt der Tiere
1912	E. Husserl	Ideas: General introduction to pure phenomenology (Part II)
1913	J. B. Watson	Psychology as the behaviorist views it
1920	F. J. J. Buytendijk	Psychologie der dieren
1922, 1929	J. Dewey	Human nature and conduct, Experience and Nature
1927	M. Heidegger	Being and time
1927	I. P. Pavlov	Conditioned reflexes
1928	W. McDougall	Body and mind; A history and a defence of animism
1929	J. Chevallier	L'habitude: essai de métaphysique scientifique
1930, 1934	C. L. Hull	Knowledge and purpose as habit mechanisms, The concept of the habit-family hierarchy and maze learning
1933	E. Claparède	La Genèse de l'hypothèse: étude expérimentale
1934	M. Mauss	Techniques of the body
1934	K. Goldstein	The organism
1935	E. von Holst	Relative coordination as a phenomenon and as a method of analysis of central nervous system function
1935	E. R. Guthrie	The psychology of learning
1936	P. Guillaume	La formation des habitudes
1937, 1955	G. Allport	The functional autonomy of motives, Becoming
1938	B. F. Skinner	The behavior of organisms: An experimental analysis
1942, 1945	M. Merleau-Ponty	The structure of behavior, Phenomenology of perception
1947, 1969	J. Piaget	The psychology of intelligence, Biology and knowledge
1947/1967	N. Bernstein	The co-ordination and regulation of movements see also Dexterity and its development (1996)
1948	E. C. Tolman	Cognitive maps in rats and men
1949	G. Ryle	The concept of mind
1949	D. Hebb	Organization of behavior
1950	P. Ricoeur	Freedom and nature: The voluntary and the involuntary
1951	K. Lorenz & N. Timbergeen	The study of instinct
1951	I. Kohler	The formation and transformation of the perceptual world (1964)
1952	W. R. Ashby	Design for a brain
1953, 1968	G. Deleuze	Difference and repetition, Empiricism and subjectivity
1957	F. Rosenblatt	The perceptron: A probabilistic model for information storage and organization in the brain
1959	N. Chomsky	A review of B. F. Skinner's Verbal behavior
1962	J. G. Taylor	The behavioral basis of perception
1969	M. L. Minsky & S. Papert	Perceptrons: An introduction to computational geometry
1972	P. Bourdieu	Outline of a theory of practice (1977)
1979	J. J. Gibson	The ecological approach to visual perception
1983	J. Fodor	The modularity of mind: an essay on faculty psychology
1986	M. A. Arbib and M. B. Hesse	The construction of reality
1986	D. E. Rumelhart, J. L. McClelland and PDP Group	Parallel distributed processing, Vol. 1: Foundations
1988	J. Fodor and Z. W. Pylyshyn	Connectionism and cognitive architecture: A critical analysis
1996	J. L. Elman et al.	Rethinking innateness: A connectionist perspective on development

their names from the most salient school of thought in each trend. But we shall first start from the Greek and Aristotelian polysemic conception of habit.

The Latin term *habitus*, from which the English *habit* comes, can be traced back to two Greek words: *ethos* (ἔθος), and *hexis* (ἕξις). The etymology of *ethos*, from which the English term *ethics* derives, is particularly revealing because it contains a profound duality. It means both "an accustomed place" in which human and animals live or in-habit (a "habitat") and "a disposition or character" denoting the personality that develops along a person's lifetime. According to Aristotle, the term *hexis* (having or being in possession of something) is a relational and active category: "a kind of activity of the haver and of what he has—something like an action or movement" [Met. 5.1022b][1], it is also a normative dispositional category "'Having' or 'habit' means a disposition according to which that which is disposed is either well or ill disposed" [Met. 5.1022b]. The ethical implications of this conception of habit extend to a self-modifying practice,

[1]References given between square brackets correspond to works used in constructing the map. They are listed in **Table 1**.

exercised so as to attain a virtuous character wherein spontaneity, joy, and norms converge.

We can interpret the Aristotelian conception of habit as an arrangement of behavioral mediations between subject and object (or between a subject and herself—in the future or past) that is well or ill-disposed in relation to essence or form and the "immediate substrate in which it is naturally produced." Habit arises from custom or repetition in a manner that constitutes a sort of second nature for the subject. In this sense, Aristotle can be said to be one of the early precursors of the organicist trend in the conception of habit. But he is also credited for inspiring the central claim of associationism (Buckingham and Finger, 1997).

THE ASSOCIATIONIST TREND

Associationism can be summarized as the view that mental phenomena are formed by combination or association of simple elements. This association follows the principle that the occurrence of event B given event A will be favored if B has repeatedly followed A in the past (often, the strength of A or B, their similarity, space-time contiguity, etc. are taken as strengthening this association). A and B are generally considered as mental states or ideas arising from sensations (often interpreted in terms of nervous activation).

The work of Ibn Tufail (12th century), translated into Latin as *Philosophus Autodidactus* [1671], tells the story of a child that reconstructs a full philosophical and theological system without the help of a social or cultural environment. It influenced the first associationists, particularly John Locke whose notion of *tabula rasa* was almost directly taken from Ibn Tufail (Russell, 1994). Locke's empiricist principle—that sense data had to fill in a blank slate—provided the basis of what was to come although he didn't provide a detailed account of associationism[2].

It was David Hume [1748] who proposed the notion of "habit," "custom," or "association" as the fundamental mechanism for the development of psychological and epistemological complexes. Atomized ideas are the direct result of sensations, while the law of habit becomes the general principle of mental organization by linking these ideas. Newton's influence on this conception of habit is apparent. Although the principles established by Hume are not fundamentally modified by Hartley's work, the latter was capable of extending them to many psychological phenomena (from memory to language, psychological development and emotions). Perhaps one of the most salient contributions was Hartley's account of habits as arising from "corporeal matter," completing Hume's philosophical approach with an influential neuro-physiological theory of associations based on the operations of the brain and the spinal cord, in accordance with the "doctrine of vibrations" previously suggested by Newton (Glassman and Buckingham, 2007). Further contributors to the associationist school (Bain, Mill, Carpenter, etc.) conserved most of the principles and theoretical assumptions of Hume and Hartley until a scientific formulation of some of these principles by behaviorist precursors came from the scientific study of animal behavior (Morgan, Thorndike and Pavlov).

The subsequent development of the notion of habit was subordinated to the available methods of measurement and intervention, which aimed at the "prediction and control of behavior" [Watson, 1913: 158]. The contribution of behaviorism to this trend can be summarized in two main aspects that result from the epistemological constraints of logical positivism (Smith, 1986) on the notion of habit: (a) the progressive externalization of the units of association in terms of stimulus and response (removing any reference to intermediate neurological or psychological processes) and (b) the mathematical treatment of the relationship between external operators and observables (stimulus, response, reinforcers) in terms of conditional probabilities. Skinner even rejected learning theories (Skinner, 1950) and purified the available terminology dropping the notion of habit altogether in favor of "rate of conditioned response."

At this point, together with the advent of computational and information theory, the ground was prepared for the now much impoverished notion of habit to disappear altogether from the set of theoretical primitives in psychology and neuroscience. Through experimental [Tolman, 1948] and theoretical [Chomsky, 1959; Fodor, 1983] arguments against behaviorism, habit was soon replaced by "mental representation" and the notion of "association" was substituted by that of "computation." Some of the associationist (and also organicist) principles were revived in neuroscience [Hebb, 1949] and, particularly, in connectionism [Rosenblatt, 1957; Rumerhart and McClelland, 1987] only to be fiercely attacked again by cognitivists [Minsky and Papert, 1969; Fodor and Pylyshyn, 1988][3]. The result of this development is the current convergence of machine learning and reinforcement learning with neuroscience (see Sutton and Barto, 1998; Daw et al., 2005; Dezfouli et al., 2012) where habits have been subsumed under networks of conditional probabilities of expected rewards associated with a set of available actions under specific conditions, or simply reduced to stimulus-triggered responses reinforced only by repetition (Dickinson, 1985). Associationist principles still exert an influence in neuroscience under the form of Hebbian learning and activity-dependent plasticity (Abbott and Nelson, 2000).

THE ORGANICIST TREND

Somewhat parallel to the development of the associationist trend we encounter an organicist tradition (left of **Figure 1**). Habits in this tradition are examined along what we would call today more ecological, self-organizing lines. Habits are both cause and effect of their own enactment and therefore constitute their own principle of individuation (Toscano, 2006), as opposed to being the passive result of the recurrence of an otherwise pre-established set of entities (ideas, stimulus, rewards, etc.). For organicism, habits are also related to a plastic equilibrium that involves the totality of the organism, including other habits, the body and the habitat they co-determine.

Spinoza's notion of *conatus*, as the striving for perseverance that defines the essence of organisms, prefigured the internalist and naturalistic conception of individuality and teleology

[2]Except for chapter 33 in his Essay introduced only in the 4th edition and dealing mostly with the origin of confusion and mistaken ideas.

[3]For a detailed account of these developments and intellectual battles, see Margaret Boden's monumental history of cognitive science (Boden, 2006).

that characterizes organicism. Kant [1790] provided a regulative notion of teleology in terms of the intertwinement of means and ends in the self-organized nature of organic life, thereby insinuating a way out of the tight mechanistic framework established by Descartes and Newton. Hegel [1830], in deep dialog with the Aristotelian tradition, emphasized the plasticity of habit as the mediating term in the resolution of the mind's contradictory tendencies toward world-independence and self-determination on the one hand, and over-stimulation and world-determination, on the other. By becoming second nature, habit prevents the mind from falling into either extreme that would lead to insanity. Goethe (though not directly addressing the notion of habit) deeply influenced subsequent conceptions of organic life by coining the term "morphology" and proposing the *law of compensation* to refer to the plastic change of natural forms in accordance with inner forces that respect the balance of the totality [1820]. Ravaisson's *De l'habitude* [1838] constitutes a cornerstone within this trend. Ravaisson puts habits at the center of metaphysics, extending from vegetative life to deliberative thought, defining habits as dynamical processes that transverse a continuum between reflective/self-aware and pre-reflective/embodied forms of intentionality (Sinclair, 2011).

Further development of the habit notion within the organicist school made it possible to expand on the dialectics between the inner tendencies of organic individuality and its co-development with the environment. von Uexküll [1909] used the term *Umwelt* to designate the *habitat* of the organism, that is, the carving of a world (from an undifferentiated environment) through functional sensorimotor cycles. His work was part of an organicist revival in Central Europe during the first half of the 20th century (Grene, 1965; Harrington, 1996), with notable exponents like the phenomenologically-informed psychologist/ethologist F. J. J. Buytendijk and neurologist Kurt Goldstein, whose studies of abstract vs. concrete behaviors in patients with brain lesions led him to holistic notions of the organism as seeking the equilibrium of preferred behaviors.

Somewhat intertwined within the organicist school, phenomenology and Gestalt psychology enriched this tradition in various ways. Husserl, for instance, acknowledged that habit is "intimately involved in the constitution of *meaningfulness*" at all levels, from perception to society (Moran, 2011). Merleau-Ponty [1945] drew inspiration from Paul Guillaume's Gestalt approach and Goldstein's experiments to develop a notion of habits as incorporated styles of being-in-the-world, thus revealing their inherent corporeal intentionality in contrast to notions of habits as blind automatisms. Gestalt psychology provided a systematic and experimental basis for holistic phenomena in perception, displacing atomistic metaphors in psychology in favor of fields and systems theory. Of particular significance are the experiments with vision distorting goggles by Ivo Kohler [1962] who emphasized the importance of action for perception, combining an active notion of habit with Gestalt principles.

Overall, the exponents of the organicist tradition in the 20th century pronounced themselves explicitly against atomistic tendencies, such as the localization of brain function and theories of reflex conditioning. The trend also influenced pragmatist thinkers such as James [1890], and particularly Dewey [1922], who also saw habits as communicating wholes affecting each other and as the substrate of self-transforming human nature. He resisted the reductionist implications of the reflex-arc concept by highlighting the active role of the organism in the selection of stimuli.

Organicism, whose ramifications appear less unified and cumulative than the associationist line, has influenced a variety of positions ranging from the integrative work of Piaget (his treatment of habit marks the starting point for a dynamic conception of cognitive development) to ecological psychology [Gibson, 1979], and the sociological conception of *habitus* as structured and structuring practices [Mauss, 1934; Bourdieu, 1972].

Current sensitivity to the organicist trend is manifest in large-scale neuroscience (Edelman and Tononi, 2001; Freeman, 2001; Llinas, 2001), constructivist developmental neuroscience (Quartz and Sejnowski, 1997; Johnson, 2001), embodied-enactive cognitive science (Varela et al., 1991; Thompson, 2007; Di Paolo et al., 2010), robotics (Di Paolo, 2003; Egbert and Barandiaran, under review), sensorimotor approaches to cognition (O'Regan and Noë, 2001; Noë, 2004) and cognitive neuroscience (Engel et al., 2013). In most cases the concept of habit forged by the organicist tradition has been modified to avoid the critiques against behaviorism, and its legacy appears to be masked under related notions such as skill, sensorimotor organization, neuroplasticity, etc.

CONCLUSIONS

We have provided a map and a very broad survey of the various ways in which the concept of habit has evolved from ancient Greece to the late 1980s, identifying two major traditions. The associationist trend conceives of habits atomistically as units that result from the association of ideas or between stimulus and response. The organicist trend, in contrast, sees habits as dynamically configured stable patterns, strengthened and individualized by their enactment. Associationism provides a statistical or combinatorial relationship between the components of a habit (based on time lapses between events, their similarity, etc.). Organicism, in contrast, proposes a more holistic view, wherein embodied relational constraints and plastic interdependencies determine the formation and maintenance of habits. Finally, the associationist trend keeps habit within the realm of reactive sub-personal automatisms (in opposition to the intentional, rational, and personal levels of cognitive processing). For organicism, in contrast, habits transition between nature and will, forming an integral part of individual embodied intentionality; they are the systemic conditions of the possibility of experience—their significance becomes clearly manifested when habits are disturbed yet they remain continuously present, configuring the identity and world of the cognitive subject.

Unlike many notions in organicism, associationist ideas were ready-made for translation into scientific hypotheses during the 20th century, even if it was ironically the subsequent development of such formalisms that fueled the cognitivist rejection of the notion of habit. While neuroscience has been partially influenced by this rejection, related ideas have survived, particularly in theories of neuroplasticity and Hebbian learning. These habit-like notions are generally associationist in character, but they have also given rise to theories of neural assemblies and neural

self-organization (Varela, 1995; Freeman, 2001), which are more organicist-friendly. Similarly, in other areas of cognitive science, dynamical systems formalisms, modeling and experimental techniques now provide the necessary tools for investigating more organicist conceptions of habit.

ACKNOWLEDGMENTS

This work is funded by the eSMCs: Extending Sensorimotor Contingencies to Cognition project, FP7-ICT-2009-6 no: 270212. XEB hold a Postdoc with the FECYT foundation (funded by Programa Nacional de Movilidad de Recursos Humanos del MEC-MICINN, Plan I-D+I 2008-2011, Spain) during the development of this work and acknowledges IAS-Research group funding IT590-13 from the Basque Government.

REFERENCES[4]

Abbott, L. F., and Nelson, S. B. (2000). Synaptic plasticity: taming the beast. *Nat. Neurosci.* 3, 1178–1183. doi: 10.1038/81453

Boden, M. (2006). *Mind as Machine: A History of Cognitive Science,* Vol. 2. Oxford: Oxford University Press.

Buckingham, H. W., and Finger, S. (1997). David Hartley's psychobiological associationism and the legacy of Aristotle. *J. Hist. Neurosci.* 6, 21–37. doi: 10.1080/09647049709525683

Carlisle, C. (2014). *On Habit*. London: Routledge.

Carlisle, C., and Sinclair, M. (eds.). (2011). Habit. *J. Br. Soc. Phenomenol.* 42:1.

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, MA: MIT Press.

Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711. doi: 10.1038/nn1560

Dezfouli, A., Balleine, B. W., Dezfouli, A., and Balleine, B. W. (2012). Habits, action sequences and reinforcement learning. *Eur. J. Neurosci.* 35, 1036–1051. doi: 10.1111/j.1460-9568.2012.08050.x

Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 308, 67–78. doi: 10.1098/rstb.1985.0010

Di Paolo, E. A. (2003). "Organismically-inspired robotics: homeostatic adaptation and teleology beyond the closed sensorimotor loop," in *Dynamical Systems Approaches to Embodiment and Sociality* eds K. Murase and Asakura (Adelaide: Advanced Knowledge International), 19–42.

Di Paolo, E. A., Rohde, M., and De Jaegher, H. (2010). "Horizons for the enactive mind: values, social interaction, and play," in *Enaction. Toward a New Paradigm for Cognitive Science,* eds J. Stewart, O. Gapenne and E. A. Di Paolo (Cambridge, MA: MIT Press), 33–87.

Edelman, G., and Tononi, G. (2001). *A Universe of Consciousness How Matter Becomes Imagination*. New York, NY: Basic Books.

Engel, A. K., Maye, A., Kurthen, M., and König, P. (2013). Where's the action? The pragmatic turn in cognitive science. *Trends Cogn. Sci.* 17, 202–209. doi: 10.1016/j.tics.2013.03.006

Freeman, W. J. (2001). *How Brains Make Up their Minds.* 1st Edn. New York, NY: Columbia University Press.

Glassman, R. B., and Buckingham, H. W. (2007). "David Hartley's neural vibrations and psychological associations," in *Brain, Mind and Medicine: Essays in Eighteenth-Century Neuroscience*, eds H. Whitaker, C. U. M. Smith, and S. Finger (New York, NY: Springer), 177–190.

Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. *Ann. Rev. Neurosci.* 31, 359–387. doi: 10.1146/annurev.neuro.29.051605.112851

Grene, M. (1965). *Approaches to a Philosophical Biology*, New York, NY: Basic Books.

Harrington, A. (1996). *Reenchanted Science: Holism in German Culture from Wilhem II to Hitler*. Princeton, NJ: Princeton University Press.

Hutto, D. D., and Myin, E. (2013). *Radicalizing Enactivism: Basic Minds without Content*. Cambridge, MA: MIT Press.

Johnson, M. H. (2001). Functional brain development in humans. *Nat. Rev. Neurosci.* 2, 475–483. doi: 10.1038/35081509

Kelso, J. A. S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Cambridge, MA: MIT Press.

Llinas, R. R. (2001). *I of the Vortex: From Neurons to Self*. Cambridge, MA: MIT Press.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., et al. (2011). Quantitative analysis of culture using millions of digitized books. *Science* 331, 176–182. doi: 10.1126/science.1199644

Moran, D. (2011). Edmund Husserl's phenomenology of habituality and habitus. *J. Br. Soc. Phenomenol.* 42, 53–77.

Noë, A. (2004). *Action in Perception*. Cambridge, MA: MIT Press.

O'Regan, J. K., and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* 24, 939–1031. doi: 10.1017/S0140525X01000115

Pollard, B. (2008). *Habits in Action. A Corrective to the Neglect of Habits in Contemporary Philosophy of Action,* Saarbrücken: VDM Verlag Dr. Mueller.

Quartz, S. R., and Sejnowski, T. J. (1997). The neural basis of cognitive development: a constructivist manifesto. *Behav. Brain Sci.* 20, 537–556. doi: 10.1017/S0140525X97001581

Russell, G. A. (1994). "The impact of the philosophus autodidactus: Pocockes, John Locke, and the society of friends," in *The 'Arabick' Interest of the Natural Philosophers in Seventeenth-Century England*, ed G. A. Russell (Leiden: E. J. Brill), 224–265.

Sinclair, M. (2011). Ravaisson and the force of habit. *J. Hist. Philos.* 49, 65–85.

Skinner, B. F. (1950). Are theories of learning necessary? *Psychol. Rev.* 57, 193–216. doi: 10.1037/h0054367

Smith, L. D. (1986). *Behaviorism and Logical Positivism: A Reassessment of the Alliance*. Stanford, CA: Stanford University Press.

Sutton, R. S., and Barto, A. G. (1998). *Reinforcement Learning*. Cambridge, MA: MIT Press.

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge, MA: Harvard University Press.

Toscano, A. (2006). *The Theatre of Production: Philosophy and Individuation between Kant and Deleuze*. Basingstoke: Palgrave Macmillan. doi: 10.1057/9780230514195

Van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behav. Brain Sci.* 21, 615–628. doi: 10.1017/S0140525X98001733

Varela, F. J. (1995). Resonant cell assemblies: a new approach to cognitive functions and neuronal synchrony. *Biol. Res.* 28, 81–95.

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

Wood, W., and Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychol. Rev.* 114, 843–863. doi: 10.1037/0033-295X.114.4.843

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 March 2014; accepted: 27 June 2014; published online: 21 July 2014.
Citation: Barandiaran XE and Di Paolo EA (2014) A genealogical map of the concept of habit. Front. Hum. Neurosci. ***8****:522. doi: 10.3389/fnhum.2014.00522*
This article was submitted to the journal Frontiers in Human Neuroscience.

[4]References to the work of authors appearing in the map are marked with square brackets in the text (e.g. Ravaisson [1838]) and are listed in **Table 1**.

frontiers in
HUMAN NEUROSCIENCE

OPINION ARTICLE
published: 09 September 2014
doi: 10.3389/fnhum.2014.00684

On habit and the mind-body problem. The view of Felix Ravaisson

Leandro M. Gaitán[1]* and Javier S. Castresana[2]

[1] *Unit of Medical Education and Bioethics, University of Navarra School of Medicine, Pamplona, Spain*
[2] *Department of Biochemistry and Genetics, University of Navarra School of Sciences, Pamplona, Spain*
**Correspondence: lemgaitan@yahoo.com.ar*

Edited by:
Jose Angel Lombo, Pontifical University of the Holy Cross, Italy

Reviewed by:
Simon Boag, Macquarie University, Australia
Gerald Wiest, Medical University Vienna, Austria

Keywords: habit, mind-body problem, metaphysics, anthropology, consciousness

In his book *De l'habitude* (*Of Habit*), 1838, the philosopher and archeologist Felix Ravaisson deals with the study of habit by using a broad spectrum of sources ranging from Aristotle to Butler, Leibniz, Hume, Main de Biran and Schelling, among others. The combination of these authors together with the originality of Ravaisson's own results produces a work which, though brief, has inspired some of the most important contemporary philosophers[1]. Moreover, this book has recently (2008) been translated into English for the first time, which has favored its rediscovery and redefinition in the context of current debates. Indeed, Ravaisson seems to have found in the study of habit a key point for the solution of some of the fundamental problems of philosophy such as the relationship between mind and body, nature and freedom, and nature and culture (Carlisle, 2013). Moreover, his approach is distinguished from the dominant method that from Descartes onwards has been focused on the study of consciousness rather than precisely on habit.

Ravaisson's approach currently remains as challenging as in his own time. This is because the anthropological conception of cognitive science is based on a clearly defined and tacitly assumed axiom: that human beings are essentially thinking beings, demonstrating that cartesianism is as valid today as in the days of Ravaisson. And for that reason, the traditional position of neuroscience tends to ignore the importance of habit (Noë, 2009). In this article we will examine how the study of the nature of habit applies to the mind-body problem and discuss the ontological status of habit, as well as the habit-consciousness relationship.

In his essay *La vie et l' oeuvre de Ravaisson* (1938/2009), the philosopher Henri Bergson says that the work *Of Habit*, despite having a modest title, is a treatise on the philosophy of nature, as it offers answers to key questions such as "What is nature? How to represent its inside? What does it hide under the regular succession of causes and effects? Does it cover something or is it reduced, in short, to a whole array of completely superficial movements that mechanically engage one another?" (pp. 266–267). This last question is key regarding the issue addressed here. Ravaisson, naturally reluctant to the great metaphysical constructs, will find the right tool to answer this question in such a daily occurrence as the habit. According to Bergson's interpretations, the inner experience shows that the habit is an activity that has passed, by insensible degrees, from consciousness to unconsciousness, and from voluntary to involuntary action. This seems to suggest that nature is a kind of obscured consciousness, or sleepy will[2].

Now, why does the habit play such a crucial role in the ravaissonian conception of the mind-body or mind-matter problem? One could start by saying that the habit "is [. . .] a disposition relative to change, which is engendered in a being by the continuity or the repetition of this very same change" (Ravaisson, 1838/2008; p. 25) and is "a general, permanent way of being" (Ravaisson, 1838/2008; p. 25). The habit, according to Ravaisson, is not possible at the inorganic level (physical, chemical, and mechanical), but it is organically possible. This is because the physical bodies are subject to external influences, i.e., to the general laws of matter, while living things have a nature that remains constant in the midst of change. For this reason, there is individuality only where there is life.

The author defines the realm of the inorganic as the "empire of Destiny," and the organic realm as the "empire of Nature" (Ravaisson, 1838/2008; p. 31). So, the habit occurs, ontologically speaking, in nature, in the living world. And the law of habit is the development of a spontaneity that runs through the dichotomy between the "mechanical Fatality" and the "reflective Freedom," as it is not identified with either of those. The habit is an "inclination that follows from the will" (Ravaisson, 1838/2008; p. 55), i.e., an idea—result of reflection and willingness—that gradually transforms in being, in "substantial idea" (Ravaisson, 1838/2008; p. 55) or in "thought in action" (Ravaisson, 1838/2008; p. 59). In other words, an idea that gradually naturalizes, an action that, as a result of repetition, imperceptibly moves from the understanding and the will, to nature. So Nature is the limit of habit: "In descending gradually from the clearest regions of consciousness, habit carries with it light from those regions into the depths and dark night of nature. Habit

[1] Maurice Merleau-Ponty and Paul Ricoeur in phenomenology, Henri Bergson and Gilles Deleuze in vitalism, and William James and John Dewey in american pragmatism.

[2] Bergson finds in Ravaisson the bases of his theory of *élan vital*, and of nature as obscured consciousness.

is an acquired nature, a second *nature* that has its ultimate ground in primitive nature, but which alone explains the latter to the understanding" (Ravaisson, 1838/2008; p. 59). The purely biological sphere is a sort of lower limit, while the sphere of reason and the will is the upper limit. Therefore, the habit flows from the upper limit to the lower limit, revealing a continuity underlying along the whole spectrum[3].

Certainly, it is in connecting those limits that habit plays a more prominent role and in which is revealed as a key to search for answers to the mind-body problem. As mentioned above, the habit is an action that harmoniously unites the area of freedom, intentionality, reflection and will with our most primitive nature[4] and includes therefore two vectors: an open *temporality* in which the future is not contained in the present, but where the present places certain regularities or patterns that anticipate what the future may include; and a *living being* whose activities may be modified by the incorporation of stereotyped behaviors (Grosz, 2013). Considering both vectors, the habit can be conceived as a complex phenomenon that is part, concomitantly, of our consciousness, and of our natural tendencies or impulses. One could argue that the habit is, then, a kind of instinct, or learned impulse that becomes standard of behavior.

But Ravaisson is cautious in speaking of habit and instinct. These functions are not identifiable because there is a difference of degree among them. Instinct is thoughtless, necessary, and perfectly spontaneous; devoid of any will and consciousness. The habit, however, has its starting point in consciousness and never completely ignores it (Malabou, 2008). However, the difference between habit and instinct can be reduced *ad infinitum* as the habit is strengthened by repeated and prolonged exercise. As pointed out by one of its most important scholars, in habit "the facility in an action gained through its repetition can become a pre-reflective desire, tendency or inclination to carry out the act [. . .] but this inclination, in turn, can develop into the almost completely involuntary phenomena that we know as tics" (Sinclair, 2011a,b). This ravaissonian idea is deeply original and important, because it highlights an aspect that is not present in other authors (including neuroscientists and contemporary philosophers). This aspect refers to the existence of an imperceptible gradualness in the process of acquiring the habit, and therefore, to the existence of habits with different degrees of strengthening or consolidation. For example, novice driver has certain visual-motor skills that undoubtedly constitute a habit. But the level of strengthening of that habit is not comparable to the case of a rally driver.

In the novice driver, the habit is not yet sufficiently near to the lower limit. In between the extremes—the beginner level and expert level—, there are countless intermediate levels. In the beginner, reasoning and free will still have a huge role, while the habit of driving is almost instinctive in the expert driver. According to the ravaissonian thesis, there seems to be an inverse relationship between consciousness and habit: more consciousness, less habit; more habit less consciousness. However, it should be stressed that, according to the author, at no time is consciousness completely eliminated. Recently, neuroscience has verified Ravaisson's assertions: experts with very entrenched habits significantly drop their brain activation level; that is, the more established you have a habit, the brain must work less. Which implies a significant reduction in muscle activity, gain in precision and elegance, and energy savings (Noë, 2009).

It is also possible to correlate the philosophical concept of habit and brain plasticity. The presence of habits in the organic world reveals the existence of a limit for change. Without habits, lifetime would be subject to the circumstances and completely adrift. Conversely, if habits would prevent any possibility of change, life would be reduced to a mere mechanism. The concept of plasticity, understood in the terms applied to the brain, i.e., its own ability to change itself, summarizes the two conditions of habit: (a) the condition of resistance to change; and (b) the condition for flexibility and variation. In other words, the habit is a form of resistance to change gradually acquired, that shows at the same time, the ability of living beings to change. On this, Carlisle states: ". . .while contemporary accounts of the brain's plasticity help us to understand the processes of habit formation, philosophical reflection on habit helps us to understand the significance of plasticity"[5] (Carlisle, 2014; p. 22). Therefore, Ravaisson's ideas about the habit and the theory of neural plasticity can be mutually reinforcing.

On the other hand, the process of acquiring the habit modifies both the mind and body, "there is, therefore, a single force, a single intelligence that is, in the life of man, the principle of all this functions and forms" (Ravaisson, 1838/2008; p. 65). According to the latter reference, the mind-body relationship would not be explained as the articulation of two substances, even two properties. Mind and body form the ends (upper and lower limits) of a *continuum*[6] in which the habit gradually down-flows. Certainly Ravaisson admits it is not possible to apodictically prove the absolute continuity between the two limits, and therefore, the existence of one and the same principle. The *continuum*, the underlying dark unit, harmonizing principle postulated by the philosopher of Namur, is only a possibility and an assumption that cannot be verified in nature. However, this presumption is inferred from the progression of habit because ". . . it draws its proof from it, by the most powerful of analogies" (Ravaisson, 1838/2008; p. 65). Ravaisson's argument compels us to think the habit from outside the predominant dualistic paradigm of modernity, and offers a phenomenological and metaphysically superior explanation.

Then, the habit is not a mere accident in the world of life, but the key to their organization and their subsistence, being a structural component in it, regardless of their level of complexity or stage of development. On the other hand, considered from the social point of view, "When we contract habits from others by sharing spaces, practices, routines and rhythms, and a language, communication and interaction become easier and less effortful, and communal life becomes

[3] Ravaisson refers primarily to motor habits.

[4] The ravaissonian thesis unifying the ideal and the reality by habit, reflects the influence of Schelling.

[5] For a neuroscientific approach to processes of habit formation, see Graybiel (2008).

[6] Ravaisson inherited the notion of *continuum* from Leibniz.

more harmonious" (Carlisle, 2013). The habit, whatever the angle from which it is considered, is a unifying element that reveals the existence of continuities in the human being individually or collectively understood. Where there is habit, there is order and connection. Considering all this, it is not absurd to claim that the habit is the clearest expression of the *continuum*. This seems to be the ontological value of the habit in Ravaisson's work.

But this is not all. It should be added that the habit does not mechanize or reduce consciousness to unconsciousness or to mere automatism, but "it brings about a new kind of consciousness, one not aware of itself but prone to act, that is activated by the possibility of its acting, that knows but cannot know that it knows" (Grosz, 2013). The ravaissonian conception of consciousness differs substantially from the cartesian conception that identifies it with reflective thought, will and therefore with knowledge. The ravaissonian consciousness has degrees (like the habit); in fact, in some of them it does not know but it acts, and acting produces effects (actions and feelings); that consciousness is always near instinct, and in its daily application through habit it opens to the possibility of creation, transformation, and learning. Thus seen, the habit, far from being mere mechanical automation, is possibility of innovation through the acquisition of new traits and skills, and openness to the future.

The author shows the habit manifests the inhabitation of freedom and intelligence in the body (Carlisle, 2013). Indeed, the process of acquiring the habit involves a shift from free reflection to the primitive nature in order to obtain that second nature (to which we refer above), but this in turn serves as a platform for further actions of free reflection. Put succinctly, the habit is the condition of possibility of conscious actions. For example, if a musician composes a song, the realization of this purpose involves the previous acquisition of physical and intellectual habits such as management of musical instruments and of singing techniques, mastery of musical notation and music theory, etc. The most original manifestations of intelligence and freedom are the result of habit.

So, the habit operates in two directions fulfilling a sort of recursive function within the mind-body *continuum*: downwards, ranging from consciousness to nature (in the process of acquisition); and upwards, ranging from nature to consciousness (once it has taken hold). This double movement attributed by Ravaisson to habit, shows an original anthropological conception, refractory of any dualism or reductionism. Indeed, according to the philosopher, humanity is not confined to the *res cogitans*, or mere *brainhood* (as postulated in the mainstream of current neuroscience). The human being is an embodied subjectivity, is a *self*, the most genuine form of unity.

Thus, the study on habit done by Ravaisson offers a phenomenologic-metaphysical answer to the so-called *hard problem* of philosophy of mind; an answer long forgotten and hard to locate in the complex map of current theories, which can still provide interesting clues not only to philosophy but also to the current neuroscience of habit.

REFERENCES

Bergson, H. (1938/2009). "La vie et l'ouvre de Ravaisson," in *La Pensée et le Mouvant: Essais et Conférences, 16th Édn.* (Paris: Quadrige, P.U.F.), 253–291.

Carlisle, C. (2013). The question of habit in theology and philosophy: from hexis to plasticity. *Body Soc.* 19, 30–57. doi: 10.1177/1357034X12474475

Carlisle, C. (2014). *On Habit*. New York, NY: Routledge.

Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. *Annu. Rev. Neurosci.* 31, 359–387. doi: 10.1146/annurev.neuro.29.051605.112851

Grosz, E. (2013). Habit today: Ravaisson, Bergson, Deleuze and us. *Body Soc.* 19, 217–239. doi: 10.1177/1357034X12472544

Malabou, C. (2008). "Addiction and grace: preface to Félix Ravaisson's of habit," in *On Habit*, eds F. Ravaisson, C. Carlisle, and M. Sinclair (London: Continuum International Publishing Group), vii–xx.

Noë, A. (2009). *Out of Our Heads. Why You are Not Your Brain, and Other Lessons from the Biology of Consciousness.* New York, NY: Hill and Wang.

Ravaisson, F. (1838/2008). *Of Habit*. London: Continuum International Publishing Group.

Sinclair, M. (2011a). Is habit 'The Fossilised Residue of a Spiritual Activity? Ravaisson, Bergson, Merleau-Ponty. *J. Br. Soc. Phenomenol.* 42, 33–52.

Sinclair, M. (2011b). Ravaisson and the force of habit. *J. Hist. Philos.* 49, 65–85. doi: 10.1353/hph.2011.0013

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 14 April 2014; accepted: 14 August 2014; published online: 09 September 2014.

Citation: Gaitán LM and Castresana JS (2014) On habit and the mind-body problem. The view of Felix Ravaisson. Front. Hum. Neurosci. ***8****:684. doi: 10.3389/fnhum.2014.00684*

This article was submitted to the journal Frontiers in Human Neuroscience.

frontiers in
HUMAN NEUROSCIENCE

GENERAL COMMENTARY
published: 08 May 2014
doi: 10.3389/fnhum.2014.00274

The principal sources of William James' idea of habit

Carlos A. Blanco *

Instituto de Cultura y Sociedad, Universidad de Navarra, Pamplona, Navarra, Spain
**Correspondence: carlos.s.blanco@gmail.com*

Edited by:
Jose Angel Lombo, Pontifical University of the Holy Cross, Italy

Reviewed by:
Rafael González-Redondo, Horacio Oduber Hospitaal, Aruba

Keywords: James, habit, plasticity, empiricism

A commentary on

The Principles of Psychology
by James, W. (1890). New York, NY: Holt.

James consecrated the fourth chapter of his *Principles of Psychology* to the explanation of the idea of habit, for "when we look at living creatures from an outward point of view, one of the first things that strike us is that they are bundles of habits. In wild animals, the usual round of daily behavior seems a necessity implanted at birth; in animals domesticated, and especially in man, it seems, to a great extent, to be the result of education. The habits to which there is an innate tendency are called instincts; some of those due to education would by most persons be called acts of reason" (James, 1890).

The first relevant idea exposed by William James concerns the importance of plasticity in the development of all organic forms. Habit, enabled by this universally manifested–though in growing degrees-plasticity, is the biological correlate of the idea of natural law in the inanimate universe. In his own words, "the laws of Nature are nothing but the immutable habits which the different elementary sorts of matter follow in their actions and reactions upon each other." Habit as the organic transposition of a natural law constitutes one of the guiding principles of James approach to this category. Its sources can be found in several authors. One of them is Léon Dumont (1837–1877), a French psychologist whose essay *De l'Habitude* (Dumont, 1876) is quoted by James in the above mentioned chapter. In this text, Dumont, following August Comte (1798–1857), had written that the idea of habit expresses, better than anyone else, the notion of a gradual acquisition of new faculties. According to him, the evolutionary perspective (recently discovered at his time) finds a good ally in the idea of habit, for it contains the progressive perfectibility of all beings, including man. In his studies of habit, sensibility and evolution, Dumont understood habit in analogy with the laws of inanimate nature.

A second major source of influence on James is the work of William Benjamin Carpenter (1813–1885), an English physician and physiologist who had done extensive work on comparative neurology. He spoke in terms of "adaptive unconscious" (Carpenter, 1874), in which there are resonances of Hermann von Helmholtz's (1821–1894) conception of thought and perception as drawing unconscious hypotheses and inferring probabilistic accounts about the surrounding environment (Helmholtz, 1867). According to this theory, thought and perception would operate, to a large extent, without awareness, and we would remain unconscious about a substantial body of mental phenomena which we consider rooted in the deepest powers of consciousness. As in the case of Dumont, in Carpenter there is a clear influence of Darwin's theory of evolution.

James conceived of a habit as the fruit of the exceptional plasticity of organic life, whose versatility would have played a significant role in favoring the adaptation to different environment, needs, and challenges. But beyond the biological and evolutionary basis of habits, James wanted to unfold the formation of this kind of automatized behavior. To answer this question, he found inspiration in the work of English utilitarian philosophers like Alexander Bain (1818–1903) and John Stuart Mill (1806–1873). Bain, a Scottish psychologist and a leading figure of empiricism, had like Mill (whom he revered) endorsed an associationist approach to the acquisition of new behaviors (Bain, 1868).

James went a step further and delineated a refined view of habits in which the ideas of plasticity, automatization, and association were carefully bounded. For him, a habit corresponded to a general form of discharge that helped concentrate energies on unpredicted challenges. As he wrote, following Carpenter's idea that our nervous system grows to the modes in which it has been exercised, "habit simplifies the movements required to achieve a given result, makes them more accurate and diminishes fatigue" (James, 1890). In James' view, this is perhaps the most remarkable feature of a habit: it diminishes the conscious attention with which our acts are performed. The precedents to this idea can be found in the work of the French spiritualist philosopher François-Pierre Maine de Biran (1766–1824). According to James, the ability to act without the concourse of will has clear advantages. In a habitual action, mere sensation suffices for eliciting muscular movements, so that the upper regions of the brain and mind are set "comparatively free," unless they go wrong and they immediately call our attention (James, 1890). This liberation shows extremely beneficial for displaying a larger array of actions.

ACKNOWLEDGMENTS

Supported by Obra Social "La Caixa."

REFERENCES

Bain, A. (1868). *Mental and Moral Science: A Compendium of Psychology and Ethics*. New York, NY: Longmans.

Carpenter, W. B. (1874). *Principles of Mental Physiology*. New York, NY: Appleton.

Dumont, L. (1876). "De l'habitude," in *Revue Philosophique de la France et de l'Étranger* (Paris: PUF), 321–366.

Helmholtz, H. (1867). *Handbuch Der Physiologischen Optik*. Leipzig: L. Voss.

James, W. (1890). *The Principles of Psychology*. New York, NY: Holt.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 27 March 2014; accepted: 13 April 2014; published online: 08 May 2014.

Citation: Blanco CA (2014) The principal sources of William James' idea of habit. Front. Hum. Neurosci. ***8****:274. doi: 10.3389/fnhum.2014.00274*

This article was submitted to the journal Frontiers in Human Neuroscience.

frontiers in
HUMAN NEUROSCIENCE

OPINION ARTICLE
published: 25 July 2014
doi: 10.3389/fnhum.2014.00542

Habit and embodiment in Merleau-Ponty

Patricia Moya*

Philosophy Department, Universidad de los Andes, Santiago, Chile
**Correspondence: pmoya@uandes.cl*

Edited by:
Javier Bernacer, University of Navarra, Spain

Reviewed by:
Javier Bernacer, University of Navarra, Spain
Leandro Martín Gaitán, University of Navarra, Spain

Keywords: habit, Merleau-Ponty, embodiment, pre-reflective knowledge, Gallagher, Zahavi

INTRODUCTION

Merleau-Ponty (French phenomenological philosopher, born in 1908 and deceased in 1961) refers to habit in various passages of his *Phenomenology of Perception* as a relevant issue in his philosophical and phenomenological position. Through his exploration of this issue he explains both the pre-reflexive character that our original linkage with the world has, as well as the kind of "understanding" that our body develops with regard to the world. These two characteristics of human existence bear a close relation with the vision of an embodied mind sustained by Gallagher and Zahavi in their work *The Phenomenological Mind: An Introduction to Philosophy of Mind and Cognitive Science*. Merleau-Ponty uses concepts like those of the *lived* or *own* body and of *lived* space in order to emphasize, from a first-person perspective, the co-penetration that exists between subject and world.

Gallagher and Zahavi have regained the experience of phenomenology, especially that of Merleau-Ponty and Sartre, to contribute to the development of the cognitive sciences. Via the phenomenological approach to the reality of habit, a new understanding of the body becomes possible for us, such that it becomes characterized "as subject, as experiencer, as agent," and at the same time we can understand "the way the body structures our experience" (Gallagher and Zahavi, 2008). Additionally, the idea of a pre-reflexive understanding is conceived of by these authors as a way for refuting those introspective or reflexive explanations that derive from the Cartesian tradition and which are promoted by certain contemporary authors (see, for instance, Dennett, 1991; Price and Aydede, 2005).

In this article I propose to explain the role that habit plays in the phenomenology of Merleau-Ponty and the use that Gallagher and Zahavi make of his theory in their work on cognitive science. The goal of these authors in the work mentioned above goes beyond that of an analysis of habit: they want to demonstrate that "phenomenology addresses issues and provides analyses that are crucial for an understanding of the true complexity of consciousness and cognition," and thereby reverse the contemporary situation where this perspective is frequently absent from current debates (Gallagher and Zahavi, 2008). For this reason, the neuroscientific community could know a more unified perspective of human behavior. The habit explanation given by Merleau-Ponty shows a kind of body knowledge that cannot be exclusively understood by neurological processes.

This paper could provide the neuroscientific community with a more unified perspective of human behavior. The explanation given by Merleau-Ponty of the habit shows a kind of corporeal *knowledge* which cannot be only clarified by neurological processes.

EMBODIED CONSCIOUSNESS

According to Merleau-Ponty, there is no hard separation between bodily conduct and intelligent conduct; rather, there is a unity of behavior that expresses the intentionality and hence the meaning of this conduct. In habits, the body adapts to the intended meaning, thus giving itself a form of embodied consciousness. Indeed, for our author, corporeal existence constitutes a third category that unifies and transcends the physiological and psychological (cf. Merleau-Ponty, 2012; see also Merleau-Ponty, 1964).

For this reason, Gallagher and Zahavi hold that the philosophy of Merleau-Ponty incorporates the body as "a constitutive or transcendental principle, precisely because it is involved in the very possibility of experience" (Gallagher and Zahavi, 2008). From the perspective of cognitive science, they propose that "the notion of an embodied mind or a minded body, is meant to replace the ordinary notions of mind and body, both of which are derivations and abstractions" (Gallagher and Zahavi, 2008). They note that, by way of confirming the priority of the body, the biological fact of the vertical position of the human body has consequences in the perception and action of the person (cf. Gallagher and Zahavi, 2008)[1].

HABIT AND UNDERSTANDING OF THE WORLD

Merleau-Ponty explains that the *lived* human body relates to a space that is also *lived*, i.e., that is already incorporated into the world understood as the horizon of its coming to be. According to this view, habit presupposes a form of "understanding" that the body has of the world in which it carries out its operations. An *operant intentionality* (*fungierende Intentionalität*) is established with the world, using the terminology of Husserl (see Merleau-Ponty, 2012). That is, the corporeal subject is inserted into a world that provokes certain questions or problems that must be resolved. Therefore, one can speak of a motivation on the part of the world,

[1]Cf. also the works that these authors cite by Straus (1966); Lakoff and Johnson (1980); Lakoff and Núñez (2001).

although not of a necessity, because the response is not mechanical or determined[2]. Between the movement of the body and the world, no form of representation is established, but rather the body "adapts" to the invitation of the world (cf. Merleau-Ponty, 2012). On the basis of this idea of Merleau-Ponty, Gallagher and Zahavi add: "The environment calls forth a specific body-style so that the body works with the environment and is included in it. The posture that the body adopts in a situation is its way of responding to the environment" (Gallagher and Zahavi, 2008). These affirmations are supported by studies that show that the nervous system does not process any information that does not proceed from corporeality (cf. Zajac, 1993; Chiel and Beer, 1997).

Habit bears a direct relation to this form of *dialog* between environment and subject. Its role is to establish in time those behaviors or forms of conduct that are appropriate for responding to the invitations of the environment. Merleau-Ponty, in establishing the etymological root of the term "habit," notes that the word *have* states a relation with what has been acquired by the subject as a possession, which in the case of the body is conserved as a dynamic corporeal scheme (Merleau-Ponty, 2012). Thanks to habit, the person establishes appropriate relations with the world that surrounds him or her without needing any prior reasoning, but rather in a spontaneous or immediate way (cf. Merleau-Ponty, 2012). Gallagher and Zahavi also refer to this form of pre-reflexive understanding, relating it to proprioception, i.e., those sensations by which we know where and how our body is, and that are in our consciousness in a tacit manner (cf. Gallagher and Zahavi, 2008; see also Legrand, 2006)[3]. This perspective allows them to distance themselves from representationalist interpretations—for instance, those of Damasio (1999) and Crick (1995), among others—that do not recognize that perception is meaningful in itself (cf. Gallagher and Zahavi, 2008)

We can speak of an engagement of body and world, in which a relation is created that serves as the basis or ground for the rest of the actions of the subject, and which permits him or her to be especially "at home," comfortable, able to move in an oriented way in a given space (cf. Talero, 2005; Merleau-Ponty, 2012). Just as Gallagher and Zahavi note, this connection with the world does not only mean knowing the physical environment in which the body is situated, "but to be in rapport with circumstances that are bodily meaningful" (Gallagher and Zahavi, 2008).

HABITUAL AND ACTUAL BODY

According to Merleau-Ponty, the situated character of the person explains that there is, at the same time, a "general" existence as well as an existence that is linked with the effectiveness of action, and which we can call "personal." Being anchored in the world makes the person renounce a part of his or her protagonism because he or she already possesses a series of habitualities. In this counterpoint between the general and the protagonistic, there occurs "this back-and-forth of existence that sometimes allows itself to exist as a body and sometimes carries itself into personals acts" (Merleau-Ponty, 2012). Merleau-Ponty distinguishes the habitual body—that of general and pre-reflexive existence—from the actual—that of personal and reflexive existence—understanding that both always co-penetrate each other. He explains that in the behaviors of mentally ill or brain damaged persons the nexus between the habitual and the actual body are broken (cf. Merleau-Ponty, 2012). In these cases, the person can reproduce certain habitual movements, but not those that require an actual understanding of the situation. For instance, a person can perform movements like touching his or her nose with a hand, but cannot respond to an order to touch the nose with a ruler. In contrast, in the non-pathological subject there is no rupture between either form of movement, since he or she is able to grasp this analogous form of movement toward the nose that the sick person cannot achieve (cf. Merleau-Ponty, 2012). The healthy person is able to come and go from the habitual to the actual. He or she is able to readjust the habitual to the actual. The world appears to the healthy subject as unfinished, offering him or her a set of possibilities such that experience "is shaped by the insistence of the world as much as it is by my embodied and enactive interests" (Gallagher and Zahavi, 2008).

THE PRIMACY OF PRACTICAL ACTION AND THE GRASPING OF MEANING

In the linkage of the subject with the world, effective, practical action has primacy. In the words of our philosopher, there is always "another self that has already sided with the world, that is already open to certain of its aspects and synchronized with them" (Merleau-Ponty, 2012; see also Talero, 2005). Merleau-Ponty frequently expresses the close relation between body and world with the term "inhabit," as referring to that which is known by the body and which translates into a knowledge of what to do with an object without any reflexion coming in between (cf. Merleau-Ponty, 2012)[4]. Gallagher and Zahavi corroborate these affirmations with research that relates perception and kinesthesia, as well as with the "enactive theory of perception" (see Varela et al., 1991). In their studies, they show that perception is not a passive reception of information, but instead implies activity, specifically, the movement of our body[5].

Merleau-Ponty explains that habitual behavior arises on the basis of a set of situations and responses that, despite not being identical, constitute a community of meaning (cf. Merleau-Ponty, 2012). This is possible because the body "understands" the situation in the face of which it must act. For example, in the case of motor habits, such as dancing, the body "traps" and "understands" movement. This is explained by the fact that the subject integrates certain elements of general motility that permit him or her to grasp what is essential to the dance in question and perform it with an ease that is expressed in the mastery of the body

[2]Cf. Merleau-Ponty (2012). In chap. IV of the Introduction, entitled "The Phenomenal Field," he explains the vital communication with the world that we are given via sensation and perception.

[3]Gallagher and Zahavi show that Sartre also shares with Merleau-Ponty the idea of being one's own body, rather than possessing it; cf. Sartre (1956) and Merleau-Ponty (2012). In this work he affirms: "But I am not in front of my body, I am in my body, or rather I am my body."

[4]For a more detailed analysis, see Kelly (2007).

[5]These ideas, which were already present in Husserl's thought (1970), are taken up by authors such as Noë (2004); Gibbs (2006).

over the movements (cf. Merleau-Ponty, 2012). The ability acquired "will lead to performance without explicit monitoring of bodily movement; the skill becomes fully embodied and embedded within the proper context" (Gallagher and Zahavi, 2008). This corporealization of habit agrees fully with the idea of Merleau-Ponty that the body is a correlate of the world: "Habit expresses the power we have of dilating our being in the world, or of altering our existence through incorporating new instruments" (Merleau-Ponty, 2012). Gallagher and Zahavi take from Merleau-Ponty this non-automatic understanding of habitual acts that, despite not requiring an express intentionality, nonetheless form part of the operative intentionality that was mentioned at the beginning of this article (cf. Gallagher and Zahavi, 2008). Citing Leder, they state: "A skill is finally and fully learned when something that once was extrinsic, grasped only through explicit rules or examples, now comes to pervade my own corporeality. My arms know to swim, my mouth can at last speak the language" (Leder, 1990).

Gallagher and Zahavi are able, over the course of their book, to demonstrate the error of that naturalism that defends objective natural science as the only legitimate manner of understanding the mind (cf. Gallagher and Zahavi, 2008; one example, among others, of this posture is found in Sellars, 1963 and in Dennett, 1991).[6] In contrast, they hold that there is a reciprocal influence between science and phenomenology, just as Varela et al. (1991) understood it via his neurophenomenology based on aspects of the phenomenology of perception of Merleau-Ponty (cf. Gallagher and Zahavi, 2008; see also Gallagher, 1997).

[6]This concept deserves a treatment that I cannot give it in this article, especially after the appearance in 1999 of the book Naturalizing Phenomenology.

REFERENCES

Chiel, H. J., and Beer, R. D. (1997). The brain has a body: adaptive behaviors emerge from interactions of nervous system, body and environment. *Trends Neurosci.* 20, 553–557. doi: 10.1016/S0166-2236(97)01149-1

Crick, F. (1995). *The Astonishing Hypothesis*. London: Touchstone.

Damasio, A. R. (1999). *The Feeling of What Happens*. San Diego, CA: Harcourt.

Dennett, D. C. (1991). *Consciousness Explained*. Boston, MA: Little, Brown and Co.

Gallagher, S. (1997). Mutual enlightenment: recent phenomenology in cognitive science. *J. Conscious. Stud.* 4, 195–214.

Gallagher, S., and Zahavi, D. (2008). *The Phenomenological Mind: an Introduction to Philosophy of Mind and Cognitive Science.* New York, NY: Routledge

Gibbs, R. W. (2006). *Embodiment and Cognitive Science*. Cambridge: Cambridge University Press.

Husserl, H. (1970). *The Crisis of European Sciences and Transcendental Phenomenology. An Introduction to Phenomenology.* Transl. ed E. D. Carr. Evanston, IL: Northwestern University Press.

Kelly, S. (2007). "Seeing things in Merleau-Ponty," in *The Cambridge Companion to Merleau-Ponty*, eds T. Carman and M. B. N. Hansen (Cambridge: Cambridge University Press), 74–110.

Lakoff, G., and Johnson, M. (1980). *Metaphors We Live By*. Chicago, IL: University of Chicago Press.

Lakoff, G., and Núñez, R. E. (2001). *Where Mathematics Comes from: How the Embodied Mind Brings Mathematics into Being*. New York, NY: Basic Books.

Leder, D. (1990). *The Absent Body*. Chicago, IL: University of Chicago Press.

Legrand, D. (2006). The bodily Self. The sensorimotor roots of pre-reflexive self-consciousness. *Phenomenol. Cogn. Sci.* 5, 89–118. doi: 10.1007/s11097-005-9015-6

Merleau-Ponty, M. (1964). *Signs*. Transl. ed R. C. McCleary. Evanston, IL: Northwestern University Press.

Merleau-Ponty, M. (2012). *The Phenomenology of Perception*. Transl. ed D. A. Landes. London; New York: Routledge.

Noë, A. (2004). *Action in Perception*. Cambridge, MA: MIT Press.

Price, D. D., and Aydede, M. (2005). "The experimental use of introspection in the scientific study of pain and its integration with third-person methodologies: the experiential-phenomenology approach," in *Pain: New Essays on its Nature and the Methodology of its Study,* ed M. Aydede (Cambridge MA: MIT Press), 243–273.

Sartre, J. P. (1956). *Being and Nothingness*. Transl. ed H. E. Barnes. New York, NY: Philosophical Library.

Sellars, W. (1963). *Science, Perception and Reality*. London: Routledge and Kegan Paul.

Straus, E. (1966). *Philosophical Psychology*. New York, NY: Basic Books.

Talero, M. (2005). Perception, normativity and selfhood in Merleau-Ponty: the spatial 'level' and existential space. *Southern J. Philos. XLIII*, 443–461. doi: 10.1111/j.2041-6962.2005.tb01962.x

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

Zajac, F. E. (1993), Muscle coordination of movement: a perspective. *J. Biomech.* 26(Suppl. 1), 109–124. doi: 10.1016/0021-9290(93)90083-Q

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 27 March 2014; accepted: 02 July 2014; published online: 25 July 2014.

*Citation: Moya P (2014) Habit and embodiment in Merleau-Ponty. Front. Hum. Neurosci. **8**:542. doi: 10.3389/fnhum.2014.00542*

This article was submitted to the journal Frontiers in Human Neuroscience.

frontiers in
HUMAN NEUROSCIENCE

HYPOTHESIS AND THEORY ARTICLE
published: 21 May 2014
doi: 10.3389/fnhum.2014.00330

Habits: bridging the gap between personhood and personal identity

Nils-Frederic Wagner*[1,2] *and* ***Georg Northoff*[1]**

[1] *Royal Ottawa Health Care Group, Mind, Brain Imaging and Neuroethics, Institute of Mental Health Research, University of Ottawa, Ottawa, ON, Canada*
[2] *Taipei Medical University-Shuang Ho Hospital, Brain and Consciousness Research Center, New Taipei City, Taiwan*

Edited by:
Javier Bernacer, University of Navarra, Spain

Reviewed by:
Francisco Guell, University of Navarra, Spain
Marya Schechtman, University of Illinois at Chicago, USA

****Correspondence:***
Nils-Frederic Wagner, Royal Ottawa Health Care Group, Mind Brain Imaging and Neuroethics, Institute of Mental Health Research, University of Ottawa, 1145 Carling Avenue, Ottawa, ON K1Z 7K4, Canada
e-mail: nils-frederic.wagner@web.de

In philosophy, the criteria for personhood (PH) at a specific point in time (synchronic), and the necessary and sufficient conditions of personal identity (PI) over time (diachronic) are traditionally separated. Hence, the transition between both timescales of a person's life remains largely unclear. Personal habits reflect a decision-making (DM) process that binds together synchronic and diachronic timescales. Despite the fact that the actualization of habits takes place synchronically, they presuppose, for the possibility of their generation, time in a diachronic sense. The acquisition of habits therefore rests upon PI over time; that is, the temporal extension of personal decisions is the necessary condition for the possible development of habits. Conceptually, habits can thus be seen as a bridge between synchronic and diachronic timescales of a person's life. In order to investigate the empirical mediation of this temporal linkage, we draw upon the neuronal mechanisms underlying DM; in particular on the distinction between internally and externally guided DM. Externally guided DM relies on external criteria at a specific point in time (synchronic); on a neural level, this has been associated with lateral frontal and parietal brain regions. In contrast, internally guided DM is based on the person's own preferences that involve a more longitudinal and thus diachronic timescale, which has been associated with the brain's intrinsic activity. Habits can be considered to reflect a balance between internally and externally guided DM, which implicates a particular temporal balance between diachronic and synchronic elements, thus linking two different timescales. Based on such evidence, we suggest a habit-based neurophilosophical approach of PH and PI by focusing on the empirically-based linkage between the synchronic and diachronic elements of habits. By doing so, we propose to link together what philosophically has been described and analyzed separately as PH and PI.

Keywords: habits, personhood, personal identity, decision-making, default-mode network, resting state, fMRI

INTRODUCTION

What is a person? More precisely, which conditions are necessary for an entity to be a person at a discrete point in time; or, which features define an entity synchronically as a person? It is important to shed light on the constitutive features of personhood in order to be able to determine how persons persist, since entities of different kinds persists in different ways. Once the constitutive features of personhood have been settled, one can ask what it takes for the *same* person to exist at different times. Since John Locke added a chapter on identity and diversity to the second edition of his "Essay Concerning Human Understanding" (Locke, 1694/1975), these questions have been intensely discussed in philosophy, as well as in related disciplines.

In the philosophical discussion, traditionally, there has been a separation between the criteria of personhood and the necessary and sufficient conditions of personal identity. That is, the *synchronic* and the *diachronic* dimension of a person's life have mostly been discussed and analyzed separately. The traditional view in philosophy of mind is that the constitutive conditions of personhood at a specific point in time and the criteria for persons to persist through time are neither identical nor coextensive. What makes someone a person at time t_1 does not account for what makes this person persist; however, quite frankly, these two dimensions of a person's life are closely related. Only if we know the conditions of personhood, can we give a compelling account of personal identity over time. Similarly, only if we have an idea of how persons persist, can we coherently analyze their synchronic dimension. This is so, as we will elaborate throughout this paper in more detail, because at least one constitutive feature of personhood—namely self-reflectiveness, particularly in its role of planning agency—involves a temporal dimension. Disregarding the temporal transition from personhood to personal identity leaves not only a gap in an encompassing theory of what constitutes a person's life as a whole, but also limits the explanatory scope of each dimension on its own. It is for this reason that theories of personal identity must at least implicitly presuppose a view of personhood; and accounts of personhood must at least implicitly consider how personal identity is constituted. Our attempt is to offer some empirically informed suggestions of how this implicit linkage between personhood and personal identity can be elucidated. We believe that personal habits serve an explanatory purpose in how these different temporal dimensions of a person's life are linked. Yet, our hypothesis does not come out of the blue. In the philosophy of action there have

been some attempts to address this issue. Particularly, Frankfurt (1982, 1988), Korsgaard (1996, 2009), and Bratman (2000) offer conceptual resources of how human agency involves reflection and planning, which implies both the synchronic and diachronic dimension of a person's life. In the discussion section, we draw on some of Bratman's conceptual work and approximate how our hypothesis is in line with his account, and further, how it can be fruitfully complemented with the empirical evidence we discuss.

To start with, we will give a brief overview of the paradigmatic approaches in philosophy of both the synchronic question of personhood and the diachronic question of personal identity. For that purpose, quite a bit of conceptual ground-clearing will be necessary. We will reconstruct the criteria for personhood and personal identity that have been claimed to be most plausible. This discussion will suggest that the separate analysis of personhood and personal identity leaves an unnecessary gap between the synchronic and the diachronic dimension of a person's life. Subsequently, in order to make an attempt to bridge this gap, we will shed light on the conceptual role that personal habits play in the linkage between personhood and personal identity. In light of this conceptual analysis, we further investigate how the temporal linkage between synchronic and diachronic aspects of a person's life is mediated empirically. Finally, we will outline an account that shows how this empirical mediation can bridge the gap between personhood and personal identity. In so doing, we will analyze the synchronic dimension of personhood and the diachronic dimension of personal identity in the realm of decision-making, which will show how habits can be considered to reflect a balance between internally and externally guided decision-making. More specifically, we will show how decision-making in form of habitual behavior already implicates a particular balance between the diachronic and synchronic aspects of a person's life, thereby linking together these two different temporal dimensions.

PERSONHOOD AND ITS SYNCHRONIC CHARACTERIZATION

What do persons have that non-persons don't have? The philosophical goal has largely been to identify a set of mental features possessed by all and only persons. These features, both traditionally and in recent philosophical discussions, are determined first and foremost by higher-order cognitive functions. It is fairly agreed upon the view that a person is someone who acts from reasons. This conception of personhood has a long tradition, reaching back to John Locke who famously regarded the concept of a person as a "forensic term." Locke says, a person is "a thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places" (Locke, 1694/1975, p. 335). Locke established this rationality-based understanding of personhood as a foundation for his account of personal identity over time. This view has a great number of modern day successors, sometimes referred to as "Neo-Lockeans" (Shoemaker, 1970, 1984, 1997, 1999; Parfit, 1971, 1984, 2007; Perry, 1972; Lewis, 1976; Nozick, 1981; Nagel, 1986; Noonan, 2003).

With regard to the moral consideration of human life, Immanuel Kant makes similar remarks when he states that "every rational being exists as an end in himself and not merely as a means to be arbitrarily used by this or that will . . . rational beings are called persons inasmuch as their nature already marks them out as ends in themselves" (Kant, 1785/2012, p. 428). In the "Lectures on Anthropology," Kant once again emphasizes that moral considerations are closely related to rationality, he states: "The fact that the human being can have the representation "I" raises him infinitely above all the other beings on earth. By this he is a person. . . . [T]hat is, a being altogether different in rank and dignity from things, such as irrational animals, with which one may deal and dispose at one's discretion" (Kant, 1798/2012, p. 127). Rationality, in Kant's eyes, is the foundation for human dignity which distinguishes us from animals and holds us responsible for our actions. In the contemporary debate, Christine Korsgaard puts this point forward, combining elements of Kant, Plato and Aristotle (Korsgaard, 2009). Peter Singer is another prominent advocate of a rationality-based view of personhood. Singer sees the special moral value in a person's life preserved in four features: (1) Being rational and self-consciously aware of oneself as an extended body existing over an extended period of time. (2) Having desires and making plans. (3) Containing a necessary condition for the right to life that one desires to continue living. (4) Being autonomous (cf. Singer, 1979, pp. 78–84).

In what follows, we focus on the prevailing claim that rationality is the conceptual starting point for personhood. This has been fleshed out paradigmatically by Daniel Dennett, who aims to define necessary conditions of personhood that are fundamentally based on our cognitive abilities. In his seminal paper, Dennett claims that

> "being rational is being intentional is being the object of a certain stance. These three together are necessary but not sufficient conditions for exhibiting the form of reciprocity that is in turn a necessary but not sufficient condition of having the capacity for verbal communication, which is the necessary condition for having a special sort of consciousness, which is . . . , a necessary condition of moral personhood (Dennett, 1976, p. 179)."

Rationality is established as the necessary condition to acquire the additional features that together make up personhood. Therefore, all other features of personhood in Dennett's account can be seen as derivative to rationality. Dennett explicitly calls rationality "the *first* and most obvious theme" (Dennett, 1976, p. 177) of personhood. Subsequently, Dennett gives six defining conditions of personhood—he calls them *themes*. They can be summed up in the particular order of their appearance as listed in **Table 1**.

Dennett aims to account for the rationality-based conditions that need to be fulfilled in order to ensure that an entity at a given point in time qualifies as a person. This account is synchronic

Table 1 | Synchronic criteria of personhood.

1. Rationality
2. Conscious mental states and intentionality
3. Being the subject of a special stance or attitude of regard by other persons
4. Being able to give that regard back to others (reciprocity)
5. Capacity for verbal communication
6. Self-consciousness

because it is not concerned with the criteria that are necessary and sufficient for a person to persist through time. To illustrate the claim that Dennett's account is synchronic rather than diachronic, consider the following example. An entity X at time t_1 is a person by virtue of him meeting the criteria stated in **Table 1**. At time t_2 X continues to be a person because he still meets the criteria in **Table 1**. However, X at time t_2 might have lost all the memories, intentions, preferences, desires and so forth that he possessed at time t_1, and is therefore no longer the *same* person; nevertheless X is still *a* person. In other words, the criteria for someone to be a person at a given time and the criteria for a given person to persist through time are different.

Dennett's aim is to show how the features stated in **Table 1** are necessary conditions of personhood, dependent on each other. Rationality is seen as the starting point for the ascription of conscious mental states to other persons and intentionality. By claiming that persons are attributed to having states of consciousness, Dennett includes that persons have "Intentional predicates" (Dennett, 1976, p. 177). That is to say, in order to think or act intentionally, a person has to decide to treat the entity whose behavior is to be predicted as a rational agent. Subsequently, the person tries to figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then the person figures out what desires it ought to have, and finally the person predicts that this rational agent will act to further its goals in the light of its beliefs. By means of this kind of practical reasoning, the person is able to predict what the rational agent will do (cf. Dennett, 1996, p. 17).

One can easily imagine other intentional systems besides human persons. Dennett gives examples of dogs and chess playing computers. According to Dennett, intentionality is not a sufficient but surely a necessary condition of personhood: "Nothing to which we could not successfully adopt the Intentional stance, with its presupposition of rationality, could count as a person" (Dennett, 1976, p. 180). When Dennett further claims that "whether something counts as a person depends in some way on an attitude taken toward it" (Dennett, 1976, p. 177), he implicitly concedes that personhood is not entirely an intrinsic feature, but to some extend a matter of social ascription. The same holds true for reciprocity, by which Dennett emphasizes that the ascription of personhood is not something that is merely given, but also something that has to be returned. Therefore, reciprocity is the capacity to exhibit higher-order intentions and thus depends on the first three, but not on the fifth and sixth condition (cf. Dennett, 1976, p. 185). To establish verbal communication as a necessary condition of personhood is rather narrow. On these grounds this requirement has been criticized by a great deal of other philosophers. In Dennett's account, verbal communication serves the goal to further link personhood to morality and, by doing so, to exclude non human animals from full personhood. However, this also comes at the cost of excluding, among others, infants. Self-consciousness is another feature that Dennett believes only to be present in humans, and, since it is seen as a precondition for morality, it defines persons as the only beings capable of morality. Self-consciousness depends in Dennett's account on the previous established conditions and, rather surprisingly, not vice versa. In order to substantiate this claim, Dennett adverts to moral responsibility. To be held responsible for an action, Dennett says, a person must have been *aware* of that action: "Because only if I was aware of the action can I say what I was about, and participate from a privileged position in the question-and-answer game of giving reasons for my actions" (Dennett, 1976, p. 191). Once again, the emphasis lies on the rational capacity of acting from reasons and on constituting ourselves by choosing the actions in awareness of our responsibility for them. "The capacities for verbal communication and for awareness of one's actions are thus essential in one who is going to be amenable to argument or persuasion, and such persuasion, such reciprocal adjustment of interest achieved by mutual exploitation of rationality, is a feature of the optimal mode of personal interaction" (Dennett, 1976, p. 191). With reference to Harry Frankfurt's concept of "second-order volitions" (Frankfurt, 1971), i.e., the unique ability of persons to develop volitions *about* other volitions, Dennett points out that reflective self-evaluation is yet another person constitutive feature that is directly dependent (and therefore subsumed under) self-consciousness. Due to our ability of being able to self-reflectively questioning our own beliefs and desires, and eventually agree or refuse them, we move beyond a level of mere informing ourselves about our beliefs and desires toward a deliberative level of an "Anscombian reason-asker and persuader" (Dennett, 1976, p. 193).

Dennett admits that, although all the conditions he has established as being necessary for personhood, one cannot simply assume that their sum is sufficient. This is so, because personhood is an inescapably normative concept and to that extent, when it is applied to categorize entities ontologically, it is a regulative idea (or a heuristic device) rather than an actual achievable goal. However, the reasons Dennett gives for what makes it even in principle very difficult (if not impossible) to find sufficient conditions for personhood are somewhat peculiar. Dennett claims: "There is no objectively satisfiable sufficient condition for any entity's *really* having beliefs, and as we uncover apparent irrationality under an Intentional interpretation of an entity, our grounds for ascribing any beliefs at all wanes, especially when we have (what we always *can* have in principle) a non-Intentional, mechanistic account of the entity" (Dennett, 1976, p. 193 f.). Peculiar about this claim is how fundamental the connection of rationality and the ascription of beliefs are linked in Dennett's account. One could ask why an irrational action, even an action that is averse to a person's apparent beliefs, should make it altogether impossible to still ascribe this belief to the person. Having a belief does not to necessarily entail that a person always acts in accordance with this very belief, unless one assumes that persons are *ipso facto* and above all, rational beings. It seems this is exactly what Dennett intends to claim when he asserts that rationality is the necessary condition for personhood.

Even though philosophers differ in the details concerning the necessary conditions of personhood, rationality is in almost every account fundamental. For the purpose of this paper, we go with this standard view. Albeit, there are alternative approaches in philosophy to what constitutes personhood. Marya Schechtman convincingly argues for a view which is less demanding in terms of cognitive abilities, but rather focuses on the social constitution

of personhood as its most salient feature (Schechtman, 2010, 2014).

Having reconstructed the paradigmatic philosophical view of what constitutes persons synchronically, we now turn to ask how persons persist through time.

PERSONAL IDENTITY AND ITS DIACHRONIC CHARACTERIZATION

If you point to a child on an old photograph of your class, say 20 years ago, and proclaim: "This is me!"—an obvious question pops up: In which way are you related to the child on the photograph that makes it true that you today and the child on the photograph are identical, or the same person over time? This is a question of diachronic personal identity. In order to answer these kinds of questions, we must know the *criterion* of personal identity over time; i.e., the relation between a person at one point in time and a person at another point in time which makes them one and the same person.

When philosophers debate personal identity, they are mostly concerned with *numerical* identity, whereby they mean that, despite of qualitative changes, a person still remains numerically identical, and thus persists through time. For example, a person X radically changed in her personality traits, as well as in her appearance due to a religious conversion. These changes, however, do not make X cease to exist altogether, they rather alter her *qualitative* identity. In questions about numerical identity, we look at two names or descriptions, and ask whether these refer to one and the same person at different times, or rather to different persons. Philosophers focus on numerical identity, since in the concern about our own futures it is this kind of identity that we care about. However much X will change, X shall still be alive, if there will be someone living who will be numerical identical to X. For this reason, some philosophers prefer to use the term *survival* in order to ensure that numerical and not qualitative identity is at issue.

Some concerns have been raised about this understanding of personal identity. Ludwig Wittgenstein famously argued that talking of *identity over time* is, if not false, at least somewhat misleading: "Roughly speaking, to say of *two* things that they are identical is nonsense, and to say of *one* thing that it is identical with itself is to say nothing at all" (Wittgenstein, 1921/1961, p. 5.5303). This understanding applies to numerical identity conditions of basic material entities like stones, but seems too narrow in terms of personal identity. It goes without saying that it is impossible for a single person at two different points in time to be identical to itself in a strict logical sense; especially if taken into account that the human body's cells are constantly replaced. However, this does not seem to be the kind of identity that we are concerned about when we reflect upon personal identity in terms of caring for our own survival. It is closer to what David Wiggins refers to when he talks of the "conditions of persistence and survival through change" (Wiggins, 1967). An understanding of personal identity through change is, both from a pretheoretical point of view and after conceptual analysis, more compelling than to appeal to strict logical identity. For this reason, accounts of personal identity over time allow for persons to change and nonetheless hold on to a broad, i.e., not strict logical, notion of identity.

DIFFERENT CRITERIA OF PERSONAL IDENTITY

In the philosophical debate on personal identity, two main opposing strategies evolved in order to account for what is necessary and what is sufficient for a person to persist through time. Therein, personal identity is either based on a *Reductionist* or on a *Non-Reductionist* understanding.

According to reductionist theories, personal identity is reducible to more particular facts about persons and bodies. The approach is to describe a particular relation *R* that accounts for a person X to be identical to a later existing person Y, by virtue of X and Y being *R*-related. In other words: X is one and the same person as Y, if and only if X stands in relation *R* to Y. In principal, Relation *R* is believed to be empirically observable. However, there is major disagreement about what relation *R* consists in. That is to say, philosophers disagree about which particular ingredients determine the relation that constitutes personal identity over time. In the contemporary debate, most philosophers hold one or another form of a reductionist account; typically, either a form of physical/biological reductionism, or more often, a form of psychological reductionism. In what follows, we will discuss the merits and demerits of the most seminal versions of these criteria.

In contrast to reductionist theories of personal identity, non-reductionists believe that personal identity is not reducible to more particular facts about persons and/or bodies, but rather consists in a non-analyzable, or *simple*, further fact. This is why non-reductionist theories are also referred to as "simple views." Derek Parfit describes the notion of a further fact as "separately existing entities, distinct from our brains and bodies, and our experience" (Parfit, 1984, p. 445). Non-reductionists thus claim that personal identity consists in a special ontological fact, a Cartesian Ego or a soul; or stated in a less antiquated way, the view is that personal identity consists in a mental entity that is neither reducible to neural mechanisms in the human brain, nor to the way in which the human brain relates to its environment and thereby gives rise to consciousness.

In the contemporary discussion in philosophy of mind few philosophers advocate for non-reductionist accounts of personal identity because those accounts are, at least by the majority of philosophers, believed to be metaphysically contentious. It is argued that non-reductionists in the debate on personal identity take an obscure metaphysical belief and inflate it into a conceptual core conviction. We here refer to the term "metaphysical" explicitly in the way in which it is used in current philosophy of mind, and more particular, in the discussion on personal identity. This is not to ignore that metaphysics has very different nuances depending on the philosophical approach, and that it is hardly used in a non contentious way. In the case of personal identity, non-reductionists arguably presuppose a form of substance, or at minimum property dualism. Both these forms of dualism do not find many advocates in the contemporary discussion on personal identity. Substance dualism is a view in philosophy of mind according to which there are two essentially different *substances* in the world: material and immaterial substances. The mind is not just a collection of thoughts, but it is the substance itself that thinks, an immaterial substance over and above its material states. Property dualism is the view according to which there are two essentially different *properties* in the world. Properties—unlike

substances—are possessed by someone or something. Property dualists thus hold the view that immaterial properties like mental states are possessed by what is otherwise a purely material thing; for example, a brain.

Granting the aforementioned concerns about non-reductionism, we will not further elaborate on those accounts. Instead, we will focus on the most paradigmatic reductionist accounts of personal identity: the seminal versions of the *psychological* and the *bodily* criterion.

According to the psychological criterion of personal identity, X and Y is one and the same person at different points in time, if and only if, X stands in a *psychological continuity* relation to Y. You are the same person in the future (or past) as you are now if your current beliefs, memories, preferences and so on are linked by a chain of overlapping psychological connections. Among philosophers who advocate for psychological approaches to personal identity there is dispute over several issues: What mental features need to be inherited? What is the cause of psychological continuity, and how do its characteristics have to be? Must it be realized by some kind of brain continuity (cf. Northoff, 2004), or will "any cause" do? The any cause discussion is concerned about the yet counterfactual idea of whether personal identity that is realized by psychological continuity would still hold, even if this continuity would no longer be caused by the brain, but, for example, by a computer program. Another issue is whether a "non-branching clause" is needed, which ensures that psychological continuity holds to only one future person. Why this can become relevant will be explicated in what follows. We will also go over some of the other aforementioned issues hereafter.

Some agreement rests upon a notion of psychological continuity that has been put forward by Derek Parfit and can be seen as a standard account, according to which (**Table 2**).

Mere psychological connectedness does not suffice as a criterion of personal identity because it is subject to the "transitivity objection." The transitivity requirement of identity states that, if X is identical to Y, and Y is identical to Z, then X must also be identical to Z. Therefore, personal identity cannot consist in mere psychological connectedness. With the appeal to psychological continuity as overlapping chains of psychological connections, the transitivity objections is resolved, since it allows for indirect relations which ensure identity through time. For example, if you as and old man remember what you have done as a middle aged man, but fail to remember what you have done as a young boy, without overlapping chains of psychological connections between the old man and the young boy, you would no longer be identical to the young boy. Since this would violate the transitivity requirement of identity. However, if you as a middle aged man still remember what you have done as a young boy, then, by virtue of overlapping chains of psychological connections, you as an old man are still identical to the young boy, even though you don't have direct access to the young boy's memories anymore. The old man is one and the same person as the young boy because, broadly speaking, they are indirectly linked through the psychological states of the middle aged man. Here it becomes apparent that psychological continuity, particularly in the sense of persisting intentions, desires and other psychological features, not only hold backwards but, as it were, also forwards. When a person envisages herself into the future, she sees herself preserving certain intentions, desires and other psychological features. Only then can she see herself as the same person persisting through time.

Table 2 | Diachronic psychological criterion of personal identity.

We might appeal, either in addition or instead, to various psychological relations between different mental states and events, such as the relations involved in memory, or in the persistence of intentions, desires, and other psychological features. These relations together constitute what I call psychological *connectedness*, which is a matter of degree. Psychological *continuity* consists of overlapping chains of such connections (Parfit, 2007, p. 6).

According to the bodily criterion of personal identity, X and Y is one and the same person at different points in time, if and only if, X stands in a *bodily continuity* relation to Y. To put it plainly: you are the same person in the future (or past) as you are now (or have been earlier), as long as you continue to have the same body. A slightly modified version of the bodily criterion is *Animalism*; the view according to which you are the same being in the future (or past) as you are now (or have been earlier), as long as you are the same biological organism. Animalists usually deny the significance of personhood for the debate on personal identity. This is one reason animalists invoke in order to distinguish their criterion from bodily continuity criteria.

One might justifiably ask, what—in real life scenarios—is the discrepancy between psychological continuity and bodily continuity views of personal identity? Doesn't psychological continuity coincide with bodily continuity? The different criteria mainly (although not exclusively) start disagreeing in hypothetical cases. Puzzles such as Locke's famous "Prince and the Cobbler," are still widely discussed in the metaphysical debate on personal identity. Locke asks what would happen if the soul of a prince, carrying with it the consciousness of the prince's past life, were to enter the body of a cobbler. Locke suggests that as soon as the Cobbler deserted by his own soul, everyone would see that he was the same *person* as the prince, accountable only for the prince's actions. But, who would say it was, in Locke's term, the same man, i.e., human animal? With this thought experiment, Locke suggests that persons, unlike human animals, are only contingently connected to bodies. Locke further believes that what constitutes a person, and moreover the *same* person, is consciousness—by which he essentially means the awareness of one's thoughts and actions: "Nothing but consciousness can unite remote existences into the same person" (Locke, 1694/1975, p. 464). Referring to a man he had met who believed his soul had been the soul of Socrates, Locke asks: "If the man truly were Socrates in a previous life, why doesn't he remember any of Socrates' thoughts or actions?" Locke even goes so far as to say that if your little finger is cut off and consciousness should happen to go along with it, leaving the rest of the body, then that little finger would be the person—the same person that was, just before, identified with the whole body (cf. Locke, 1694/1975, pp. 459–460). Therefore, Locke and his modern day successors establish that wherever your mental life goes, that is where you as a person go as well.

Apart from thought experiments, in real life we might consider the case of permanent vegetative state patients to support

Locke's thought experiment, inasmuch as it shows that psychological continuity and bodily continuity do not always coincide. Psychological continuity is not necessarily in place whenever a human organism is around. This assertion does not, of course, imply any dualistic assumptions of immaterial sources of psychological continuity; it merely states that not every form of biological continuity of a human organism is sufficient to support psychological continuity. For all we know now, permanent vegetative state patients lack any higher-order mental features that could possibly constitute psychological continuity, albeit, they are biologically alive. Therefore, according to the psychological criterion of personal identity, there is no identity relation between a conscious person that later becomes a vegetative state patient. Advocates of the bodily criterion see things differently. In their view the identity relation still holds because there continues to be bodily continuity between the person that once had a mental life and the human organism that is now in a permanent vegetative state.

Despite all the difficulties within Locke's view, which cannot be discussed, let alone resolved here, the aforementioned puzzle cases, as well as the permanent vegetative state example, support the widely advocated psychological continuity theories of personal identity. Furthermore, our ordinary intuitions in these scenarios support psychological continuity rather than mere bodily/biological continuity as the criterion for personal identity over time.

The different psychological continuity theories, however, share a severe problem. Unlike identity, psychological continuity is not necessarily a one-one relation. For example, fission scenarios, either based on purely hypothetical cases or based on brain bisection (Corpus Callosotomy), as put forward, among others, by Thomas Nagel, show that psychological continuity does not follow the logic of an identity relation (Nagel, 1971). It is possible in principle, and in accordance with empirical evidence, that psychological continuity divides, and thus, that it holds to more than one person. [For an analysis of the empirical plausibility of different accounts of personal identity see Northoff (2001)]. Albeit, as David Lewis and others pointed out, identity is necessarily a one-one relation that can by definition only hold to itself; whereas psychological continuity is only contingently a one-one relation and may become one-many (Lewis, 1976). Therefore, as Bernard Williams took issue with, psychological continuity is unable to meet the metaphysical requirements of an account of personal identity, unless a non-branching clause is added which ensures that psychological continuity is a one-one relation (Williams, 1973). Nevertheless, the addition of such a non-branching clause is not fully convincing either. This is so, because, as Derek Parfit claimed, a non-branching clause has no impact on the intrinsic features of psychological continuity, and is therefore unable to preserve what we believe to be important in identity (Parfit, 1984). An identity relation can by definition apply to only one person. This leads Parfit to the conclusion that in the end, personal identity is neither here nor there, or as he famously puts it: "Identity is not what matters" after all because the importance we ascribe to it is merely contingent. It seems to be entirely dependent on psychological continuity, which, as mentioned before, is logically not an identity relation. When we are concerned with our survival, what we really should care about is, in Parfit's view, psychological continuity, whether or not it coincides with identity. [For a suggestion of how this problem can be tackled in terms of personal identity in practical reality see Wagner (2013). For a thoughtful critical discussion of Parfit's criterion see Teichert (2000)].

Hereafter, we will put forward the hypothesis that habits can serve to bridge the gap between synchronic and diachronic aspects of a person's life. In order to give a prospect of this hypothesis, we will briefly summarize the core points of personhood and personal identity that have been discussed up to this point.

As an interim result from the discussion of the constitutive features of personhood, it can be drawn the conclusion that a person is regarded as an agent that has certain mental, rather than singularly human features, wherein rationality is seen as the most fundamental feature. The discussion of the different theories of personal identity suggests that a form of psychological continuity, characterized by overlapping chains of psychological connections, is indispensable to account for the persistence of persons through time. Even though it can not account for all the metaphysical difficulties, in the relevant sense of everyday life, personal identity over time is created by links between present and past provided by autobiographical experience memories and other mental states. These links are seen as providing connections between two discrete, well-defined moments of consciousness. It is beyond the scope of this paper to make an attempt to resolve the ongoing debate on which criterion of personal identity is the most plausible. However, as the brief discussion has shown, we are sympathetic to the reductionist psychological approach which is a widely-held and well-defended view.

It becomes evident that in the discussion of personhood and personal identity a gap remains between the synchronic and the diachronic dimension of a person's life. Although psychological theories of personal identity are based on the assumption that it is a person, rather than a mere biological organism without mental states, who's identity over time is in question, it remains largely unclear how the transition between these timescales—that is, being a person at a discrete point in time, and persisting as a person through time—is mediated, both conceptually and empirically. In order to shed light on this temporal transition, we hereafter focus on habits and decision-making, and argue that therein a conceptually and empirically plausible bridge between personhood and personal identity is to be found.

HABITS AND DECISION-MAKING: A NEUROPHILOSOPHICAL HYPOTHESIS

What are habits? In philosophy of action, habits have been defined as a "pattern of a particular kind of behavior which is regularly performed in characteristic circumstances, and has become automatic for that agent due to this repetition" (Pollard, 2006, p. 57). Standard definitions in psychology are compatible with the philosophical view in the sense that they regard "automaticity and conditioning of repeated acts in stable contexts" (Wood et al., 2002, p. 1282) to be at the core of what habits are. An important feature that distinguishes habits from compulsive behavior is that, in the case of habitual behavior, the person has control over whether or not to perform the habitual action. Based

on this conception, habits can explain a vast amount of actions; even more than we would usually assume. This becomes obvious when we think about how much of our lives we spend exercising habits rather than subjecting our actions to deliberation. Starting each day with specific routines, for example getting dressed, brushing teeth, making coffee and so forth. What characterizes habitual behavior is its repetitiveness and automaticity. However, unlike reflexes—for which the same general characteristics apply—habits involve a previous and as the case may be more or less conscious and voluntary acquisition. That is to say, a habit is not something that just passively happens to a person; but rather it is a particular pattern of actions that once has been actively initiated by the person. In light of this, habits can conceptually be seen as a form of actions rather than mere movements. Needless to say, that the level of activeness in the acquisition of different habitual behaviors varies greatly.

Taken together, the criteria for habits extracted from the standard philosophical and psychological definitions are listed in **Table 3**.

To illustrate the different criteria of habitual behavior, let us consider an example of how habits develop accordingly to the above definition. In the case of running, both if performed professionally as well as in leisure sports, there is a conscious component to the acquisition of the habit to run. At some point, most likely consciously and voluntarily, the person decides to engage in running and to make it a habit by doing this repeatedly. By means of this repetition, let's say the runner decides to run three times a week, the very act of running becomes automatized. However, the involvement of building greater muscle tone that comes along with running is not the same as becoming automatic; rather it makes automaticity possible. That is, a runner becomes able to slowly raise the intensity of running according to the growth of muscle strength and thereby increasing his performance capacity. As a consequence, the runner doesn't have to concentrate anymore on the movements of his legs, arms, etc. while running, but can focus on something else. He could even let his mind wander, or think about something that is completely unrelated to running. The automatized act of running induces a form of learning and improvement in the motion sequence of running. This automaticity leads to a form of conditioning. The person feels the reward of doing sports, gets used to this reward, and gets thereby conditioned to stick to this behavior. It has to be noted here that the reward that comes with doing sports regularly is a feature which is, presumably, based on the voluntariness of engaging in this particular habit. It goes without saying that there are involuntary habits that do not involve reward. For example, slaving away in a mine and excavating stones can become automatic and thus arguably considered to be a habit; nonetheless it most likely does not involve reward. The act of running that occurs with increasing regularity in a well-specified and stable context, as for example in the case of using similar running tracks does further in habitualizing the act of running. Stable context are important in order to make it possible that the automatized act of running can be performed smoothly because the runner doesn't have to adjust to new situations. If, for example, a runner is used to running on tracks and, say, due to having no access to a track while on a trip, so he has to run in the forest, the very act of running might become less smooth because the runner has to adjust his movements to the new environment. Finally, the habit of running is subject to the runner's control. Whenever he decides not to engage in running anymore, for example because he caught a cold and wants to give his body some rest, he can simply decide to do so.

Table 3 | Criteria of Habits.

Component of Conscious Acquisition
Repetition
Automaticity
Conditioning
Stable Contexts
Control

Habits involve particular processes and different levels of decision-making. Following the above analysis, we will first consider the criteria for decision-making that have been examined in current neuroscience. Next, we will examine how these criteria relate to habits.

INTERNALLY AND EXTERNALLY GUIDED DECISION-MAKING WITHIN HABITS

In a recent neuroscientific review paper by Takashi Nakao et al., a distinction between "externally and internally guided decision-making" has been established (Nakao et al., 2012). According to the authors, "most experimental studies of decision-making have addressed situations in which one particular more or less-predictable answer is available" (Nakao et al., 2012, p. 1). It is assumed that in these situations there is one particular correct answer which is almost entirely dependent on external circumstances. Consequently, those kinds of decision processes have been called "externally guided decision-making." Let us consider an example. Imagine being at a crossroad at which the right-hand road leads to Turin and the left-hand one leads to Pisa. If the goal is to go to Turin, then there is only one correct answer to the decision of which road to take; the answer is entirely dependent on external criteria. The person has to take the right road.

In addition to externally guided decision-making, there are situations in which there is not one correct answer that is based on external circumstances according to which the person decides; but rather, the person has to draw almost entirely on internal resources to make a decision. In these kinds of situations, therefore, the answer depends on the person's own, internal preferences and not on external, circumstantial criteria. Consequently, Nakao et al. call this "internally guided decision-making." Consider again the example of the crossroad. If the goal is to go to the city you prefer (Turin or Pisa), then there is no externally guided right or wrong answer to the decision of which road to take; it is entirely up to the person's subjective preference whether to take the road to Turin or to Pisa.

In sum, the criteria for externally and internally guided decision-making that have been put forward by Nakao et al. are listed in **Table 4**.

There is empirical evidence in support of the distinction between internally and externally guided decision-making on a

Table 4 | Criteria of externally and internally guided decision-making.

Externally guided decision-making: The person has to decide mostly relying on externally determined factors. The decision has a single correct answer.
Internally guided decision-making: The person has to decide mostly relying on his/her own internal preferences. The decision has neither a correct nor an incorrect answer.

neural level. To test this distinction, Nakao et al. conducted a meta-analysis comparing studies on decision-making that rely on external cues (with high or low predictability of the subsequent gain, i.e., externally guided), with those where no external cues were presented (i.e., internally guided). Interestingly, externally guided decision-making studies yielded significantly stronger activity changes in lateral frontal and parietal regions. Whereas internally guided decision-making studies yielded significantly stronger activity changes in the midline regions; including pregenual anterior cingulate cortex, ventromedial prefrontal cortex, dorsomedial prefrontal cortex, posterior cingulate cortex, and precuneus (see also Northoff, 2014a,b). These data support the distinction between internally and externally guided decision-making on a neural level. The evidence shows that in different decision-making processes that can be characterized as externally and internally guided decisions different brain regions are activated. It has to be noted here that the neural processes underling internally and externally guided decision-making are bilaterally interdependent and reciprocally balanced. That is, activation in the midline regions during internally guided decision-making shows a negative correlation with lateral frontal and parietal regions. However, regardless of the form of decision-making, both regions show a proportional activation in each form of decision-making.

Granting the aforementioned distinction in decision-making, we now turn to ask the question to which degree the criteria of habits reflect internally and externally guided decision-making. Seen from this angle, we will again go through the example of running and examine the degree of externally and internally guided decision-making in the criteria of habits. In this regard, we will refer to elements of habits as "more externally" or "more internally" guided decisions. This, in accordance with the empirical data, suggests that the distinction between both levels of decision-making in the case of habits is not a principal difference, but rather a qualitative difference. It is a difference in levels of internally/externally guided decision-making on a continuum of decisions that range from being almost exclusively external (i.e., there is only one correct answer) to decisions that are almost exclusively internal (i.e., there is no right or wrong answer, only subjective preferences). Furthermore, the distinction between internally and externally guided decision-making in habitual behavior seems to be related to the level in which decisions are made more or less consciously or unconsciously. Concerning this matter, it is useful to distinguish between the process and the outcome of a decision in order to see how these levels are related. While the process of an externally guided decision can be rather unconscious, as for example, in how to adjust movements to certain environmental cues, the outcome of this unconscious process, namely the particular adjustments, can later become conscious and thus may become subject to internally guided deliberation. This gives some reason to suggest that externally guided decision-making is more associated with unconscious processing, whereas internally guided decision-making is more associated with conscious deliberation. Again, this has to be seen as a qualitative difference and not as an all-or-nothing matter.

While acquiring the habit of running, the conscious component in making the decision to run is mostly an internally guided decision, since the idea of engaging in running in the first place is subject to the person's preference. This is in line with the aforementioned assertion that the outcome of a decision, in this example the commitment to engage in the habit of running, is both internally guided and it occurs on a conscious level. Although, there is a more externally guided component to the decision to engage in running as well, that is, to engage in running rather than in, for example cycling, may be influenced by social factors such as the fact that your friends run as well, which is why you like the prospect of joining them. When running is performed repeatedly—in our example let's say the runner decides to run three times a week—the previously conscious component in the decision becomes rather unconscious. That is, the novelty of the decision to engage in running is lost over time. It is rather an externally guided unconscious process, a response to the external stimuli involved in running at specific times. The acquisition of the habit of running was initially a more internally guided conscious decision; however, due to its repetition it becomes a more externally guided unconscious component of habitual behavior. To put it differently, the internally guided decision to engage into running according to the person's preference for this particular sport becomes, due to its repetition, a more externally guided component because in the very act of running it are the external criteria (e.g., the weather conditions, the time schedule etc.) that the runner responds to and not the internal component of deciding which sport to get involved in. The same holds for the automatized component in the process of running. Thereby, the runner does not have to concentrate anymore on the movements of his legs, arms etc. while running, but can focus on something else. There is no conscious, preference dependent decision involved in the very movements of running, but rather an automatized response to external stimuli from the environment in which the running takes place. The component of automaticity in habitual behavior is thus a more externally guided decision-making process because it is merely subject to the environmental circumstances in running. For example, the conditions of the running track due to the weather, the equipment and so forth. Conditioning, on the contrary, is more of an internally guided decision-making process in habitual behavior, since it is based on the internal reward which is related to the preference decision that led to the acquisition of the specific habit in the first place. The element of habitual behavior in running in stable contexts seems to have both levels of internally and externally guided decision-making to it, since the decision to stick to stable contexts is based on the conscious acquisition of the habit to run and is thus more internally guided. Surely, internally guided decision-making is also influenced by the context in which it takes place; however a broader notion of context is meant here, i.e. the

social, the political context and so forth. Yet, what we are referring to in the realm of externally guided decision-making is a much narrower notion of context, namely the very concrete environmental conditions by which a decision is shaped. This is why the actualization of running that takes place within a stable context is more externally guided since it is a response to the contextual conditions itself and thus not subject to the runner's preference. The control that one has over the habit of running has also both internally and externally guided elements to it. Control is partly internally guided, because whether or not to continue engaging in the habit of running is based on the person's preference judgment; it is basically a subjective choice. However, this internally guided choice can be dependent on, or at least informed by, externally guided conditions; such as the earlier discussed example of deciding not to run anymore because you caught a cold. The decision to stop running in this case is externally guided to the extent that catching a cold determines whether or not you will physically be able to keep up the habit of running. The external component of catching a cold that influences the decision not to run is externally guided to the extent that it is out of the runner's immediate sphere of control. Whether or not he catches a cold is nothing the runner can do much about—apart from wearing the appropriate clothes according to the weather conditions and so forth. However, once the cold is there, it at least externally informs the decision not to run, because doing so would most likely lead to a worsening of the health condition, which in turn would be at odds with any prudential decision of a rational agent that takes his state of health seriously.

Taken together, the different criteria of habits reflect a balance between internally and externally guided decision-making. Habits, therefore, are neither purely internal nor purely external, but rather they reflect a specific balance between both forms of decision-making.

We now turn to ask what this balance of internally and externally guided components in habitual behavior can tell us about the different timescales that are involved therein. On the one hand, habits are actualized, or take place, at discrete points in time. On the other hand, by repeating the specific actions that take place at discrete points in time, habits take place over time.

INTERNALLY AND EXTERNALLY GUIDED DECISION-MAKING BALANCES SYNCHRONIC AND DIACHRONIC ELEMENTS OF HABITS

The actualization of habits manifests in discrete points in time which indicates a synchronic element of habits. The runner runs Tuesday at 7 PM. This decision-making process is more externally guided, since it largely relies on the external components at this particular point in time in which the decision takes place. For example, how are the weather conditions at this day and how do these conditions guide the decision to run, or influence how to prepare, e.g., to wear a rainjacket?

The repetitiveness of habits over time adds a diachronic element to the actualization of habits at discrete points in time. That is to say, by repeating the actualization of habitual behavior at discrete points in time, the habit takes place over time. The runner runs not only at a particular Tuesday at 7 PM, but he runs every Tuesday at 7 PM. This decision-making process is more internally guided, since it largely represents the person's subjective preference over time and thus involves a diachronic timescale. It is, however, not exclusively internally guided to decide to run every Tuesday at 7 PM, since the runner might only be able to run at 7 PM and not at 11 AM because his work schedule does not permit him to do so. On a neural level internally guided decision-making has been associated with the brain's intrinsic activity, as Nakao et al. point out: "Based on rest-stimulus interaction and the overlap between the network for internally guided decision-making with DMN [Default Mode Network], internally guided decision-making seems to be largely based on intrinsic brain activity" (Nakao et al., 2012, p. 12).

According to the previous analysis, we conclude that habits can be considered to reflect not only a balance between internally and externally guided decision-making, but also a balance between diachronic and synchronic timescales that are involved in the relevant decision-making processes. This means that decision-making in habits already implicates a particular balance between diachronic and synchronic aspects, thus linking two different temporal dimensions.

We now turn to ask what implications the above considerations of decision-making and timescales in habits have for the relation between the philosophical concepts of personhood and personal identity.

DISCUSSION: PHILOSOPHICAL IMPLICATIONS OF THE LINKAGE BETWEEN PERSONHOOD AND PERSONAL IDENTITY

The argument which we are going to put forward and discuss in what follows, looks in a semi-formalized way like this:

> Premise 1: Personhood is characterized in synchronic terms. In contrast, personal identity is characterized in diachronic terms.
> Premise 2: Habits are a form of ongoing, personalized decision-making processes that have both synchronic and diachronic timescales.
> Therefore: Habits link the synchronic and diachronic timescale of a person's life and thus bridge the gap between personhood and personal identity.

As the foregoing analysis suggests, there is reason to believe that habits are best conceptualized as the sum of personalized, both internally and externally guided decisions that we repeatedly make. This leads us to hypothesize that habits can be seen as the convergence between synchronic and diachronic aspects of a person's life, as illustrated in **Figure 1**. Personal habits reflect a personalized decision-making process that binds together the synchronic aspects of personhood and the diachronic aspects of personal identity. By so doing, habits, as based on the balance between internally and externally guided decision-making, have the potential to provide an empirically substantiated link between the philosophical concepts of personhood and personal identity. Despite the fact that the actualization of habits takes place synchronically, they nevertheless presuppose, for the possibility of their generation, time in a diachronic sense. Figuratively speaking,

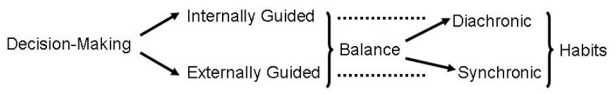

FIGURE 1 | Habits balance internally and externally guided decision-making and diachronic and synchronic timescales.

the temporal extension of personhood with the recruitment of personal identity is the necessary condition of possibility for the acquisition of habits. More specifically, the acquisition of habits rests both upon a form of rationality, and on psychological continuity, as examined in the accounts of personhood and personal identity. In order to explicate this claim in more detail, we now turn to ask why the acquisition of habits presupposes a form of rationality that has been claimed to be a constitutive condition for personhood, and how this is linked to psychological continuity.

As social psychologists point out, there are self-regulatory benefits of acquiring habits as a way of avoiding the stress, e.g. the time consumption of having to make decisions in similar situations over and over again (Armitage and Conner, 2001). As indicated in the examples given before, persons often rely on habits as an efficient mode of initiating and controlling routines in everyday life. The conscious acquisition of a habit itself, i.e. the conscious decision to keep up a certain pattern of action in stable contexts, therefore, relies on higher-order cognitive functions; namely, on rationality and self-reflectiveness. Once a habit is in place, it is relatively automated; there is no need anymore for a conscious guidance of the habitual behavior. The actualization of a habit is based on a previous, internally guided decision to engage in a particular habit, whereby the concrete performance of this habit becomes automated and is therefore no longer directly subject to self-reflective internally guided decision-making. But rather, the concrete decisions in the situation of performing the habit become responses to external stimuli. The very idea of habit-forming is to avoid the process of deliberative decision-making in recurrent situations for which a rational decision already has been formed. To acquire habits can thus, philosophically speaking, be seen as a form of "practical rationality." Practical rationality is generally described as the appropriate way of processing information through reasoning; furthermore, it is seen to be the nature of reasons for action and the norms for assessing acts or reasoning leading to action.

To illustrate the argument, consider again a sports example. As a rational agent, you know that it is healthy to do sports regularly. But, unless you purposely form a habit to do sports at specific times, each time doing sports comes to mind, it will bring up the same decision-making process again, and it may thus become difficult to motivate yourself repeatedly. If your goal is to stay healthy, consequently, it is both rational and efficacious to acquire the habit of doing sports. The rational acquisition of habits rests upon a, using Harry Frankfurt's vocabulary, *second-order volition*, i.e., the forming of a will about a will. It rests upon a form of self-reflective deliberation which has been claimed to be constitutive for being a person. Rather than making a rational decision at a specific point in time repetitively, habits are a way to

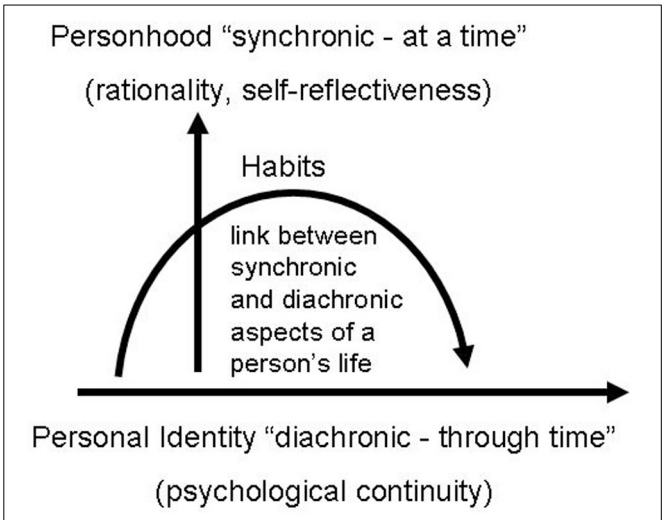

FIGURE 2 | Habits as a linkage between synchronic and diachronic aspects of a person's life.

make a rational decision over time and thus link synchronic and diachronic aspects of a person's life.

Conceptualizing the repeated intentional actualization of a certain behavior as a habit, however, is only plausible if the person who synchronically performs the particular action persists through time, thus becoming able to repeat the action. How habits bridge the gap between personhood and personal identity is illustrated in **Figure 2**.

In order to link present habitual behaviors with future ones, that is in order to establish habits, it is necessary that the person at the point of the actualization of the habit is psychological continuous with the person at another point of the actualization of the habit. Putting it more formally: if and only if synchronic person X at time t_1 is linked through psychological continuity (and is thus identical) with synchronic person Y at time t_2, an action can possibly become a habit. Seen in this way, acquiring a certain habit becomes a constitutive feature of what it is to be a particular person over time, i.e., what constitutes personal identity.

A person and her identity cannot be narrowly conceived as the synchronic state of psychological features and events alone, but rather a person's identity is inseparable from its familiar modes of behavior, in its familiar environment, which stretches back and forth in time. Habitual actions at a specific point in time emerge from conscious intentions or rather implicit guides that have been developed through past performance, thus linking together the synchronic and diachronic timescales of a person's life. This is true, even more so, if we believe that personal identity depends on the peculiar psychological aspects of a person that manifest in a unique pattern of thoughts and actions which persist through time.

Seeing habits in this light implicates some overlap with what Harry Frankfurt identifies as the constitutive features of being a particular person. Broadly speaking, the notion of distinctively caring about certain lifestyles presupposes the temporal persistence of a particular person. Frankfurt writes: "The outlook of a

person who cares about something is inherently prospective; that is, he necessarily considers himself as having a future" (Frankfurt, 1982, p. 260). Habits as deliberatively chosen patterns of behavior are a relevant part of what we care about in our lives and thus account for what it is to be a particular person persisting through time.

In his seminal work on human agency, Michael Bratman makes the case for three core features of agency that can help elucidating our hypothesis that habits bridge the gap between personhood and personal identity. Bratman writes: "We form prior plans and policies that organize our activity over time. And we see ourselves as agents who persist over time and who begin, develop, and then complete temporally extended activities and projects" (Bratman, 2000, p. 35). Accordingly, Bratman claims *reflectiveness*, *planfulness*, and the *conception of our agency as temporally extended* to be the core features of personhood. All of those features are to some relevant degree involved in the acquisition and performance of habits. Pertinent to the linkage between different timescales of a person's life is what Bratman calls "planning agency." By that he refers to future directed plans of actions that play basic roles in the organization and coordination of our activities over time; the significance of planning for habitual behavior, as discussed in, for example, the scheduling of running, is obvious. Although Bratman does not explicitly discuss habits, he acknowledges that planning typically concerns specific courses of action over time; accordingly he introduces the concept of "policies" as the "commitment [to] a certain kind of action on certain kinds of potentially recurrent occasions" (Bratman, 2000, p. 41). In discussing planfulness and reflectiveness, Bratman draws the attention to the seemingly problematic fact that "one might be reflective about one's motivation at any one time and yet not be a planner who projects her agency over time" (Bratman, 2000, p. 42). Here Bratman's account and our suggestion about habits become importantly connected to psychological continuity relations of personal identity. As mentioned before, psychological continuity does not only hold backwards, but also holds as forward-looking connections to planned habitual actions. That is, habitual behavior can be seen as the link between the forming of a prior intention, for example the plan to run Tuesday at 7 PM and the later execution of this intention. This is only possible if the person who forms an intention is psychological continuous with the person who later executes this intention. Interestingly, sticking with and executing prior plans is not only a passive, or, as it were, automatic psychological fact about persons, but, at the same time, it actively serves to ensure what might be called the "unity of a person over time." Psychological continuity is thus not only a prerequisite of habitual behavior, but sometimes also an intentional result of a person's activity. In Bratman's words: "[T]he characteristic stability of such intentions and policies normally induces relevant psychological continuities of intention and the like. In these ways our plans and policies play an important role in the constitution and support of continuities and connections characteristic of the identity of the agent over time" (Bratman, 2000, p. 47). Habits, similar to what Bratman calls policies, are thus grounded in their characteristic role of coordinating and organizing a person's identity over time in ways that both constitute and support psychological continuity.

CONCLUDING REMARKS

In this paper, we argued on empirically informed grounds that habits bridge the temporal gap between synchronic and diachronic timescales of a person's life, which are exemplified in the philosophical concepts of personhood and personal identity.

In order to substantiate this claim, we first analyzed the seminal concepts of personhood and personal identity in philosophy, thereby carving out the constitutive features of both concepts. According to this analysis, personhood is grounded foremost in rationality, and personal identity is constituted by psychological continuity.

In a next step, we suggested that habits, which are characterized as automatized and conditioned actions that are repeated in stable contexts, can be seen as a specific balance of internally and externally guided decision-making. For this purpose, we drew upon empirical evidence that supports the distinction between internally and externally guided decision-making. On a neuronal level, externally guided decision-making has been associated with lateral frontal and parietal regions. In contrast, internally guided decision-making has been associated with the midline regions. Furthermore, there is reason to believe that externally guided decision-making takes place largely on a synchronic timescale, whereas internally guided decision-making takes place largely on a diachronic timescale.

In a conclusive step, we analyzed how habitual behavior requires and supports both the constitutive features of personhood and personal identity. Based on this analysis, and complemented with what has been established before, namely that habits form a particular balance of internally and externally guided decision-making, we conclude that habits bridge the gap between personhood and personal identity. An empirically informed account of habits can link together what philosophically has so far mostly been described and analyzed separately, and it can therefore open a novel field of philosophical, or rather neurophilosophical investigations.

ACKNOWLEDGMENTS

We are grateful for financial support from CIHR, EJLB-CIHR, Michael Smith Foundation, and the Hope of Depression Foundation (HDRF/ISAN) to Georg Northoff and for a Postdoctoral Fellowship from the University of Ottawa IMHR to Nils-Frederic Wagner. We are thankful to the reviewers for their thoughtful comments that helped to improve earlier versions of this paper. We owe thanks to Jeffrey Robinson for editing the paper.

REFERENCES

Armitage, C. J., and Conner, M. (2001). Efficacy of the theory of planned behaviour: a meta-analytical review. *Br. J. Soc. Psychol.* 40, 471–499. doi: 10.1348/014466601164939

Bratman, M. (2000). Reflection, planning, and temporally extended agency. *Philos. Rev.* 109, 35–61. doi: 10.1215/00318108-109-1-35

Dennett, D. (1976). "Conditions of Personhood," in *The Identities of Persons*, ed A. Rorty, (Berkeley, CA: University of California Press), 175–196.

Dennett, D. (1996). *The Intentional Stance*. Cambridge, MA: MIT Press.

Frankfurt, H. (1971). Freedom of the will and the concept of a Person. *J. Philos.* 68, 5–20. doi: 10.2307/2024717

Frankfurt, H. (1988). *The Importance of What We Care About*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511818172

Frankfurt, H. (1982). The importance of what we care about. *Synthese* 53, 257–272. doi: 10.1007/BF00484902

Kant, I. (1785/2012). *Groundwork of the Metaphysics of Morals*. Cambridge: Cambridge University Press.

Kant, I. (1798/2012). *Lectures on Anthropology*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139028639

Korsgaard, C. (2009). *Self-constitution: Agency, Identity, and Integrity*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199552795.001.0001

Korsgaard, C. (1996). *The Sources of Normativity*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511554476

Lewis, D. (1976). "Survival and identity," in *The Identities of Persons*, ed A. Rorty (Berkeley, CA: University of California Press), 17–41.

Locke, J. (1694/1975). "Of Identity and Diversity," in *An Essay Concerning Human Understanding*, Chapter XXVII, ed P. Nidditch (Oxford: Clarendon Press), 311–333.

Nagel, T. (1971). Brain bisection and the unity of consciousness. *Synthèse* 22, 396–413. doi: 10.1007/BF00413435

Nagel, T. (1986). *The View From Nowhere*. Oxford: Oxford University Press.

Nakao, T., Ohira, H., and Northoff, G. (2012). Distinction between externally vs. internally guided decision-making: operational differences, meta-analytical comparisons and their theoretical implications. *Front. Neurosci*. 6:31. doi: 10.3389/fnins.2012.00031

Noonan, H. (2003). *Personal Identity*. 2nd Edn. London: Routledge.

Northoff, G. (2001). *Personale Identität und Operative Eingriffe in das Gehirn*. Paderborn: Mentis.

Northoff, G. (2004). Am I my brain? Personal identity and brain identity-a combined philosophical and psychological investigation in brain implants. *Philosophia Naturalis* 41, 257–282.

Northoff, G. (2014a). *Unlocking the Brain. Volume I: Coding*. New York, NY: Oxford University Press.

Northoff, G. (2014b). *Unlocking the Brain. Volume II: Consciousness*. New York, NY: Oxford University Press.

Nozick, R. (1981). *Philosophical Explanations*. Cambridge: Harvard University Press.

Parfit, D. (1971). Personal identity. *Philos. Rev.* 80, 3–27. doi: 10.2307/2184309

Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.

Parfit, D. (2007). "Is personal identity what matters?" in *The Ammonius Foundation* (South Plainfield, NJ), 1–32. Available online at: http://www.stafforini.com/txt/parfit_-_is_personal_identity_what_matters.pdf

Perry, J. (1972). Can the self divide? *J. Philos.* 69, 463–488. doi: 10.2307/2025324

Pollard, B. (2006). Explaining actions with habits. *Am. Philos. Q.* 43, 57–68.

Schechtman, M. (2010). Personhood and the practical. *Theor. Med. Bioethics* 31, 271–283. doi: 10.1007/s11017-010-9149-6

Schechtman, M. (2014). *Staying Alive–Personal Identity, Practical Concerns, and the Unity of a Life*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199684878.001.0001

Shoemaker, S. (1970). Persons and their pasts. *Am. Philos. Q.* 7, 269–285.

Shoemaker, S. (1984). "Personal identity: a materialist's account," in *Personal Identity*, eds S. Shoemaker and R. Swinburne (Oxford: Blackwell), 67–133.

Shoemaker, S. (1997). Self and substance. *Philos. Perspect.* 11, 283–319.

Shoemaker, S. (1999). Self, body, and coincidence. *Proc. Aristotelian Soc.* S73, 287–306. doi: 10.1111/1467-8349.00059

Singer, P. (1979). *Practical Ethics*. Cambridge: Cambridge University Press.

Teichert, D. (2000). *Personen und Identitäten*. Berlin; New York, NY: De Gruyter. doi: 10.1515/9783110802320

Wagner, N.-F. (2013). *Personenidentität in der Welt der Begegnungen*. Berlin; New York, NY: De Gruyter. doi: 10.1515/9783110336276

Wiggins, D. (1967). *Identity and Spatio-Temporal Continuity*. Oxford: Blackwell.

Williams, B. (1973). "The self and the future," in *Problems of the Self,* ed B. Williams, (Cambridge: Cambridge University Press), 46–64. doi: 10.1017/CBO9780511621253.006

Wittgenstein, L. (1921/1961). *Tractatus Logico Philosophicus*. New York, NY: The Humanities Press.

Wood, W., Quinn, J., and Kashy, D. (2002). Habits in everyday life: thought, emotion, and action. *J. Pers. Soc. Psychol.* 83, 1281–1297. doi: 10.1037/0022-3514.83.6.1281

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 23 March 2014; accepted: 02 May 2014; published online: 21 May 2014.
Citation: Wagner N-F and Northoff G (2014) Habits: bridging the gap between personhood and personal identity. Front. Hum. Neurosci. ***8****:330. doi: 10.3389/fnhum.2014.00330*
This article was submitted to the journal Frontiers in Human Neuroscience.

frontiers in
HUMAN NEUROSCIENCE

GENERAL COMMENTARY
published: 25 April 2014
doi: 10.3389/fnhum.2014.00255

Habit acquisition in the context of neuronal genomic and epigenomic mosaicism

Francisco J. Novo*

Biochemistry and Genetics, University of Navarra, Pamplona, Spain
**Correspondence: fnovo@unav.es*

Edited by:
Javier Bernacer, University of Navarra, Spain

Reviewed by:
Robbin Gibb, University of Lethbridge, Canada

Keywords: habits, epigenomics, mosaicism, metaplasticity, genomics

A commentary on

Epigenetic Priming of memory updating during reconsolidation to attenuate remote fear memories
by Gräff, J., Joseph, N. F., Horn, M. E., Samiei, A., Meng, J., Seo, J., et al. (2014). Cell 156, 261–276. doi: 10.1016/j.cell.2013.12.020

A recent paper (Gräff et al., 2014) shows that remote fear memories in mice can be stably attenuated with the administration of histone de-acetylase (HDAC) inhibitors during reconsolidation. This achieved persistent attenuation of remote memories, even though it is well established that the brief period of hippocampal neuroplasticity induced by recent memory recall is absent for remote memories. Apparently, such epigenetic intervention primed the expression of neuroplasticity-related genes.

This work comes shortly after the finding (McConnell et al., 2013) that individual neurons show an extraordinary degree of genomic mosaicism. Sequencing the genomes of single human frontal cortex neurons, these authors found that up to 41% of neurons contain at least one *de novo* copy-number variant (CNV) of at least one megabase in size. Segmental duplications have greatly expanded in African great apes (Marques-Bonet et al., 2009), and it is possible that increased retrotransposon activity during human neurogenesis also contributes to this striking diversity in CNV numbers in neuronal genomes (Singer et al., 2010).

Taken together, both studies support the notion that genomic and epigenomic mosaicism allows for the introduction of heritable changes at the single-cell level that promote neuronal plasticity, and thus help to explain how human actions can modify neural circuits involved in memory and learning.

(EPI) GENOMIC MOSAICISM AND SYNAPTIC PLASTICITY

The epigenomic basis of memory and learning is an active field of research in neuroscience (Mehler, 2008; Baker-Andresen et al., 2013). Long-term memory (LTM) formation requires the consolidation of short-term memories, so that these can be later recalled to participate in a wide range of behavioral responses such as making decisions based on previous knowledge (Puckett and Lubin, 2011). Studies about chromatin modifications in various brain regions have shown that learning experiences can trigger epigenetic changes that mediate synaptic long-term potentiation and contribute to LTM consolidation (Guo et al., 2011).

DNA methylation is a well-studied type of epigenetic modification. Cortical DNA methylation is one of the molecular mechanisms used by the brain to preserve remote memories (Miller et al., 2010) and regulates associative reward learning (Day et al., 2013). Changes in DNA methylation at specific genomic sites can modulate the expression of genes involved in synaptic plasticity and memory suppression, thus leading to memory consolidation. For example, knockout mice for methyltransferases DNMT1 or DNMT3A that lose DNMT activity in the hippocampus are unable to form new memories, indicating the importance of dynamic DNA methylation in the process of LTM formation (Feng et al., 2010). However, it is interesting that a number of CpGs differentially methylated in response to neuronal activity might not lead to stable changes in transcription, but rather prime the genome to respond to future stimuli. In the context of memory processing, experience-mediated variations in DNA methylation represent a type of genomic metaplasticity that could prime the transcriptional response and facilitate neuronal reactivation (Baker-Andresen et al., 2013).

In addition to DNA methylation, other epigenetic marks such as histone methylation and acetylation have been shown to play crucial roles in memory and learning processes (Mehler, 2008). For instance, certain histone methylation marks such as the tri-methylation of lysine 4 in histone 3 (H3K4me3) and the di-methylation of lysine 9 (H3K9me2), activate and repress gene transcription, respectively, in the hippocampus during fear–memory consolidation (Gupta et al., 2010).

In summary, experience-driven changes in various epigenetic marks could direct neuronal plasticity in several ways: regulating alternative splicing of specific genes, releasing transposable elements from transcriptional silencing, or creating bivalent chromatin domains that render genes poised for transcription (Baker-Andresen et al., 2013). Reactivation of transposable elements might be particularly relevant in the context of neuronal mosaicism, as it has been shown that L1 retrotransposons are transiently released from epigenetic suppression during neurogenesis so they can mobilize to different loci in individual cells. This would lead to genomic rearrangements that might enable different neurobiological processes,

including neural plasticity (Singer et al., 2010; Baillie et al., 2011).

HABITS AND (EPI) GENOMIC MOSAICISM

The genomic basis of neuronal plasticity and metaplasticity is particularly relevant in the context of human habits. From a neuroscientific perspective, habits arise from the repeated learning of associations between actions and their contextual features. In this regard, a fundamental issue in neuroscience will be the relationship between habit acquisition and neuronal (epi) genomic mosaicism in humans.

Recent advances in single-cell genomics and non-invasive imaging technologies suggest that significant developments will be achieved in the near future. Once neuronal circuits involved in habit learning are identified by imaging studies, the analysis of genomic and epigenomic neuronal mosaicism should reveal which changes facilitate (or result from) habit acquisition. This will require the development of techniques for the analysis of genomes and epigenomes in single-cells, and imaging technologies that capture epigenetic changes *in vivo*.

In this regard, single-cell genome sequencing is shedding new light into the genetic architecture and variability between cells, highlighting the dynamic nature of the genome (Blainey and Quake, 2014). Although single-cell epigenomics is still in its infancy, the use of a microfluidic platform has recently boosted efficiency and allowed the analysis of DNA methylation in six genes simultaneously in one cell (Lorthongpanich et al., 2013). Such advances will help to read the epigenomes of individual neurons obtained from brain surgery or post-mortem samples.

At the same time, new molecular imaging strategies are being implemented to monitor microRNA biogenesis and its post-transcriptional regulation, *in vivo* as well as *in vitro*, using several reporter systems such as fluorescent proteins, bioluminescent enzymes, molecular beacons, and/or various nanoparticles (Hernandez et al., 2013). For instance, an *in vivo* luciferase imaging system was used to monitor miR-221 biogenesis (Oh et al., 2013). Although non-invasive analysis of gene expression is still in the initial stages of development, molecular imaging of genomic and epigenomic changes might become a reality in a not-so-distant future. Then, it will be possible to design experiments to investigate how genomic and epigenomic mosaicism facilitate (or are influenced by) the acquisition of habits.

REFERENCES

Baillie, J. K., Barnett, M. W., Upton, K. R., Gerhardt, D. J., Richmond, T. A., De Sapio, F., et al. (2011). Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* 479, 534–537. doi: 10.1038/nature10531

Baker-Andresen, D., Ratnu, V. S., and Bredy, T. W. (2013). Dynamic DNA methylation: a prime candidate for genomic metaplasticity and behavioral adaptation. *Trends Neurosci.* 36, 3–13. doi: 10.1016/j.tins.2012.09.003

Blainey, P. C., and Quake, S. R. (2014). Dissecting genomic diversity, one cell at a time. *Nat. Methods* 11, 19–21. doi: 10.1038/nmeth.2783

Day, J. J., Childs, D., Guzman-Karlsson, M. C., Kibe, M., Moulden, J., Song, E., et al. (2013). DNA methylation regulates associative reward learning. *Nat. Neurosci.* 16, 1445–1452. doi: 10.1038/nn.3504

Feng, J., Zhou, Y., Campbell, S. L., Le, T., Li, E., Sweatt, J. D., et al. (2010). Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nat. Neurosci.* 13, 423–430. doi: 10.1038/nn.2514

Gräff, J., Joseph, N. F., Horn, M. E., Samiei, A., Meng, J., Seo, J., et al. (2014). Epigenetic Priming of memory updating during reconsolidation to attenuate remote fear memories. *Cell* 156, 261–276. doi: 10.1016/j.cell.2013.12.020

Guo, J. U., Ma, D. K., Mo, H., Ball, M. P., Jang, M. H., Bonaguidi, M. A., et al. (2011). Neuronal activity modifies the DNA methylation landscape in the adult brain. *Nat. Neurosci.* 14, 1345–1351. doi: 10.1038/nn.2900

Gupta, S., Kim, S. Y., Artis, S., Molfese, D. L., Schumacher, A., Sweatt, J. D., et al. (2010). Histone methylation regulates memory formation. *J. Neurosci.* 30, 3589–3599. doi: 10.1523/JNEUROSCI.3732-09.2010

Hernandez, R., Orbay, H., and Cai, W. (2013). Molecular imaging strategies for *in vivo* tracking of micrornas: a comprehensive review. *Curr. Med. Chem.* 20, 3594–3603. doi: 10.2174/09298673113202900005

Lorthongpanich, C., Cheow, L. F., Balu, S., Quake, S. R., Knowles, B. B., Burkholder, W. F., et al. (2013). Single-cell DNA-methylation analysis reveals epigenetic chimerism in preimplantation embryos. *Science* 341, 1110–1112. doi: 10.1126/science.1240617

Marques-Bonet, T., Kidd, J. M., Ventura, M., Graves, T. A., Cheng, Z., Hillier, L. W., et al. (2009). A burst of segmental duplications in the genome of the African great ape ancestor. *Nature* 457, 877–881. doi: 10.1038/nature07744

McConnell, M. J., Lindberg, M. R., Brennand, K. J., Piper, J. C., Voet, T., Cowing-Zitron, C., et al. (2013). Mosaic copy number variation in human neurons. *Science* 342, 632–637. doi: 10.1126/science.1243472

Mehler, M. F. (2008). Epigenetic principles and mechanisms underlying nervous system functions in health and disease. *Prog. Neurobiol.* 86, 305–341. doi: 10.1016/j.pneurobio.2008.10.001

Miller, C. A., Gavin, C. F., White, J. A., Parrish, R. R., Honasoge, A., Yancey, C. R., et al. (2010). Cortical DNA methylation maintains remote memory. *Nat. Neurosci.* 13, 664–666. doi: 10.1038/nn.2560

Oh, S. W., Hwang do, W., and Lee, D. S. (2013). *In vivo* monitoring of microRNA biogenesis using reporter gene imaging. *Theranostics* 3, 1004–1011. doi: 10.7150/thno.4580

Puckett, R. E., and Lubin, F. D. (2011). Epigenetic mechanisms in experience-driven memory formation and behaviour. *Epigenomics* 3, 649–664. doi: 10.2217/epi.11.86

Singer, T., McConnell, M. J., Marchetto, M. C., Coufal, N. G., and Gage, F. H. (2010). LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? *Trends Neurosci.* 33, 345–354. doi: 10.1016/j.tins.2010.04.001

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 26 February 2014; accepted: 07 April 2014; published online: 25 April 2014.

Citation: Novo FJ (2014) Habit acquisition in the context of neuronal genomic and epigenomic mosaicism. Front. Hum. Neurosci. ***8****:255. doi: 10.3389/fnhum.2014.00255*

This article was submitted to the journal Frontiers in Human Neuroscience.

frontiers in
HUMAN NEUROSCIENCE

OPINION ARTICLE
published: 18 September 2014
doi: 10.3389/fnhum.2014.00731

Is the philosophical construct of "habitus operativus bonus" compatible with the modern neuroscience concept of human flourishing through neuroplasticity? A consideration of prudence as a multidimensional regulator of virtue

Denis Larrivee[1]* and Adriana Gini[2]

[1] *Educational Outreach Office, Catholic Diocese of Charleston, Charleston, SC, USA*
[2] *Neuroradiology Division, Neuroscience Department, San Camillo-Forlanini Medical Center, Rome, Italy*
**Correspondence: sallar1@aol.com*

Edited by:
Jose Angel Lombo, Pontifical University of the Holy Cross, Italy

Reviewed by:
Jacek Debiec, University of Michigan, USA

Keywords: neuroplasticity, synapses, virtues, prudence, Aquinas

THE CLASSICAL AND THE CONTEMPORARY: NEUROPLASTICITY AND THE REEMERGENCE OF VIRTUE

Unlike ancient Greece where personal virtue was the route to fulfillment, modern man typically seeks to improve human well-being by external means, in a process known as the medicalization of society. The apparent novelty of recent proposals in psychological theory to develop character strength, therefore, lies in their reemphasis on a personal implementation of positive values (Peterson and Seligman, 2004). Among the factors contributing to a new look at self-determination has been the capacity for the neural substrate to selectively alter itself via neuroplasticity. Indeed, the confluence of past and contemporary thinking may presage a consideration of neurobiological instantiation within which virtuous behavior may be enhanced in accord with principles governing neuroplastic change.

But what are virtues and positive traits? And to what extent can these conceptions inform our growing understanding of the neural contribution to human behavior? Presupposed in such questions is a conceptual ground needed to define a corresponding empirical terrain (Bennett and Hacker, 2003), without which such information would lack coherence and conclusive power. Accordingly, positive psychology identifies a positive trait as a "disposition to act, desire, and feel" involving the exercise of judgment and leading to a recognizable human excellence' (Park et al., 2004). The concise, but more precise formula of Aquinas, "habitus operativus bonus," is similarly conceived (Hibbs, 1999). Anglicized, habitus connotes habit, often considered a compelling behavioral pattern reinforced through repetitive activity, but in an Aquinas context also evinces a freedom associated with the deployment of a skill acquired and honed through repeated engagement. Operativus, understood to mean operationally effective, connotes stability and continuity, a disposition to future performance. The third term, bonus, grants an orientational norm more precise than the analogous "recognizable excellence" and that Aquinas grounds in right reason and love of neighbor. Accordingly, we will employ the construct "habitus operativus bonus" rather than "positive traits" in the ensuing discussion.

A priori, habitus tacitly acknowledges a behavior's dependence on repetitive engagement. This acknowledgement has received much confirmation from empirical studies of patterned behavior; and many of the physiological, cellular, and molecular features have now been elucidated. Originally theorized by Hebb (1949) as an activity dependent synaptic strengthening, this interpretation was subsequently confirmed by Lomo's discovery (Lomo, 2003) of the long term potentiation effect (LTP). In the Hebbian scheme synaptic strength is enhanced by coincident, and repetitive, neural activity. The molecular details of this effect entail a host of short term, cell signaling and, when sufficiently stimulated, long term, transcriptional and cell restructuring mechanisms (Benfenati, 2007). The former involve an enhancement of Ca influx at both pre and post synaptic sites, together with a corresponding activation of Ca dependent protein kinases, lasting minutes to hours. The latter involve a wholesale restructuring of synaptic contacts that can potentiate enhanced synaptic efficiency for days and even months. A key mechanism in transcriptional up-regulation is the kinase mediated activation of the CREB set of activator and repressor proteins. The stabilization, and proliferation, of coordinated synaptic activity, thereby, increasingly routs information flow through select circuit pathways.

These observations confirm three conclusions that follow from the classic formulation. First, they show that habitual activity is needed to enhance synaptic strength. Second, the behavioral performance or skill is made more easily operative. The freedom spoken of by Aquinas is thus neurally provided for in the enhanced information flow through the behavioral circuit. Finally, the ease of flow facilitates

and so disposes, the circuit to similar operation in the future.

Yet, what underlies the selection of one circuit in preference of another? In the classical formulation stress is laid on the learned features of virtuous behavior. In contemporary neuroscience patterned behaviors presuppose a unique circuitry also selected through learning. Insight into the underlying mechanisms of learning has come from studies of circuits comprised of small numbers of neurons. Two of these, habituation and sensitization, are particularly well understood. Habituation, described as the progressive decrease in amplitude or frequency of an output in response to external stimuli, restricts information flow from irrelevant stimuli by reducing Ca influx needed for presynaptic vesicle release (Rankin et al., 2009). Sensitization reverses this effect, and thereby emphasizes the impact of relevant stimuli, through a cAMP kinase induced increase in Ca. Nor are these alone. Recent studies of underlying mechanisms of associative learning reveal that molecular switches, such as insulin isoforms, for example, can alter the behavioral pattern in a stimuli dependent manner (Ohno et al., 2014). Such mechanisms, observed in simple systems, are likely to constitute unitary learning modules broadly used for more complex learning programs. The repertoire of cellular mechanisms, in fact, highlights the rich potential for the choreography of patterned routines in large scale networks (Neville et al., 2010). Learned behavior for what are undoubtedly large scale networks have now been demonstrated for motor skill acquisition (Dayan and Cohen, 2012), clinical therapy (Cramer et al., 2011), stress related responsivity (Davidson and McEwen, 2013), and language learning (Hosoda et al., 2013). Indeed, all behaviors for which there is a demonstrable need for repeated or habitual performance are likely to be undergirded by such plastic mechanisms, including the execution of virtue.

PRUDENCE AND NEUROSCIENCE IN THE PURSUIT OF EXCELLENCE

The manner of the selection of one circuit over another, nonetheless, has raised the question of the valuation of a behavior that may lead to its selection. It is in this context that the orientational construct of Aquinas, bonus, is pertinent. In what manner, then, does the granting of normative weight to virtue bear on neural function? Complicating the issue of value is the matter of valence, defined as compelling loci within the focal space that provoke sustained interest for attainment. Humans, as do all species, possess appetitive desires distributed over a broad range of physiological and cognitive demands (Berridge and Kringelbach, 2008). Given that their salience can vary over a particularly broad range, it is clear that individual variation can become extraordinarily diverse. To accommodate a full spectrum of valences, therefore they must be ordered hierarchically. How so? This is done in two ways. First, neuroplastic mechanisms may be understood as an innate capacity to prioritize values dispositionally through circuit reiterations. Secondly, they may be invoked experientially or intentionally through higher order processes. Experience dependent learning of motor skills in the cerebellum, and conditioning dependent learning in the hippocampus, for example, both implement other cortical centers, that progressively shift in a learning dependent manner (Melia et al., 1996; Doyon et al., 2002).

The very diversity of valences, however, and the relative intensities to which salience is attributed, generate transformations as numerous in kind as the individuals in whom they are effected. Accordingly, it is to other dimensions that recourse must be made to order what will ultimately become dispositional preferences. In a Thomistic scheme value is established rationally, according to the dictates of practical reason, i.e., prudentia, and by conformity to an ordering principle, i.e.,"beatitude." The inclination to future performance, though, and the assessment of deviation from preferred behavior may be matters heavily influenced by neural architecture. There is, for example, a notable correspondence between habitual behavior, the computational properties of corrective learning algorithms, and the physiology/anatomy of the dopaminergic system, a correspondence that also appears to extend to goal directed behavior (Daw and Shohamy, 2008). Nonetheless, some values may be innate (Bloom, 2010), and others acquired (Stanley, 2008) with little or no conscious reflection. Babies, who lack a power for speech, but who can still indicate intentions through eye movement, are able to discriminate between various actions as to their moral worth. They know, for example, when an action is unjust, or altruistic. This has been interpreted to indicate a native capacity for goodness in its "infancy." Preconscious, implicitly acquired value, such as those studied with the implicit association test (IAT, 2014), likewise show that not all value is determined rationally.

Still, Aquinas' insistence on rational deliberation as the necessary, conscious precursor to normative assignments intrinsic to virtuous behavior, is receiving renewed neuroscientific interest. In value assignment studies of children, normative values were not conditioned by background attitudes, but rather by a structured rationale from which moral inferences were then drawn (Hussar and Harris, 2010). Moreover, the reflective process of deliberation is a manifestly and universal social tool (Bloom, 2010). Embedded in the recognition of such deliberation is the notion of its procedural development, according to logical inference and structured on grounding principles. Nevertheless, the practical reasoning spoken of by Aquinas, and by which prudence must be exercised, is very much in its infancy from the vantage of a neuroscientific understanding. Language, mental representations, syllogistic reasoning, insight, and relative judgments (Dadosky, 2014; Hauser et al., 2014) have stimulated theoretical discussion, but the empirical dimensions for which these concepts may antecede are at best correlative.

Perhaps most elusive is the manner in which the neural structure may be contributory to a state phenomenologically described by the orientational construct, bonus, the ordering principle by which normative assignments are inferred through rational and deliberative discourse. For Aquinas this first principle is self evident and non-deducible. Like habitus, this concept is also multitoned, expressing both the means that may be used, i.e., morality, as well as the goal, flourishing, or beatitude, that is to be attained. Such conceptions have been differentially interpreted neuroscientifically with most focus given to the

practical means. A large body of work has now been devoted to how a moral understanding may have originated, usually couched in terms of its evolutionary, social development (Hauser, 2014). Rudimentary notions of altruistic behavior have been observed in some species (Zwick and Fletcher, 2014) and mirror neurons, which have been inferred to grant a capacity for empathic associations, studied (Rizzolatti and Craighero, 2004). Still the notion of beatitudo of Aquinas with its connotations of meaning, fulfillment, and openness to infinite and transcendental being does not appear capable of resolution at anything less than an integrationist account of the whole neural platform. Some suggestion of this appears in discussions of the neural underpinnings of the self and of downwardly causative operations for which the entire platform is likely to be necessitated or mobilized for (Sanguineti, 2013), but for the most part is undefined. Its relevance for a large part of humanity, though, is undeniable, propelling ongoing investigation. Prudentia thus has many aspects, and enters into every other virtue (Aquinas, 2011), the "form" shaping each virtue.

So how are we to view the utility of "habitus operativus bonus," and the regulatory virtue of prudence in terms of conceptual schema for neuroscience? Born in a prebiological era Aquinas knew little of the functional elements of brain operation; yet he possessed an extraordinary analytical mind and a first person access to its events. Following a tradition traced to Aristotelian roots, Aquinas placed reason and will as the progenitors of behavior. Not so passion, Aquinas designation of emotion, despite his extensive and discrete analysis of the human emotional spectrum (Butera, 2010). Aquinas' observations, therefore, limited necessarily to those of whole systems, cannot be a guide to particular empirical events that underlie systems operations. Nevertheless, he recognized that such events are contributory, and in some cases determinative of their operation. It would have been no surprise that the virtue for which so much broad scale rational determination must be exercised would require so many subordinate neural circuits for its operation. And it would have been no surprise that to perfect its operation such circuits must be repeatedly deployed during learning, and in the process "disposed" and materially structured.

In his unified view of nature Aquinas offers landmarks circumscribing lower level events, the broad outlines of which serve in clarifying what the details must conform to, but not in identifying the materials by which the paths to these landmarks are structured. His integrationist accounts and objective philosophy of the purpose of brain operation, though, lay down a route of exploration likely to be increasingly relevant as neuroscience explores the global neural platform.

ACKNOWLEDGMENT

The authors wish to thank Leo White of Morgan State University for his gracious and extensive contribution in classical and ecclesiastical resources.

REFERENCES

Aquinas, T. (2011). *Aquinas' Moral, Political, and Legal Philosophy. Stanford Encyclopedia of Philosophy*. Available online at: http://plato.stanford.edu/entries/aquinas-moral-political [Accessed September 19, 2011]

Benfenati, F. (2007). Synaptic plasticity and the neurobiology of learning and memory. *Acta Biomed*. 78, 58–66.

Bennett, M. R., and Hacker, P. M. S. (2003). *Philosophical Foundations of Neuroscience*. Oxford: Blackwell Publishing.

Berridge, K. C., and Kringelbach, M. L. (2008). Affective neuroscience of pleasure: reward in humans and animals. *Psychopharmacology* 199, 457–480. doi: 10.1007/s00213-008-1099-6

Bloom, P. (2010). *The Moral Life of Babies. The New York Times*. Available online at: www.nytimes.com/2010/05/09/magazine/09babies-t.html

Butera, G. (2010). Thomas Aquinas and cognitive therapy: an exploration of the promise of the thomistic psychology. *Philosophy. Psychiatry Psychol.* 17, 347–366. doi: 10.1353/ppp.2010.0023

Cramer, S. C., Sur, M., Dobkin, B. H., O'Brien, C., Sanger, T. D., and Trojanowski, J. Q. (2011). Harnessing neuroplasticity for clinical applications. *Brain* 134(Pt 6), 1591–1609. doi: 10.1093/brain/awr039

Dadosky, J. (2014). Lonergan on wisdom. *Irish Theol. Q.* 79, 45–67. doi: 10.1177/0021140013509437

Davidson, R. J., and McEwen, B. S. (2013). Social influences on neuroplasticity: stress and interventions to promote well-being. *Nat. Neurosci.* 15, 689–695. doi: 10.1038/nn.3093

Daw, N. D., and Shohamy, D. (2008). The cognitive neuroscience of motivation and learning. *Soc. Cogn.* 26, 593–620. doi: 10.1521/soco.2008.26.5.593

Dayan, E., and Cohen, L. G. (2012). Neuroplasticity subserving motor skill learning. *Neuron* 72, 443–454. doi: 10.1016/j.neuron.2011.10.008

Doyon, J., Song, A. W., Karni, A., Lalonde, F., Adams, M. M., and Ungerleider, L. G. (2002). Experience-dependent changes in cerebellar contributions to motor sequence learning. *Proc. Natl. Acad. Sci. U.S.A.* 99, 1017–1022. doi: 10.1073/pnas.022615199

Hauser, M. (2014). *About the Moral Sense Test*. Available online at: http://wjh1.wjh.harvard.edu/~moral/learn.html

Hauser, M. D., Yans, C., Berwick, R. C., Tattersell, I., Ryan, M. J., Watermull, J., et al. (2014). The mystery of language evolution. *Front. Psychol.* 5:401. doi: 10.3389/fpsyg.2014.00401

Hebb, D. O. (1949). *The Organization of Behaviour*. New York, NY: John Wiley and Sons.

Hibbs, T. (1999). Aquinas, virtue, and recent epistemology. *Rev. Metaphys.* 52, 573–594.

Hosoda, C., Tanaka, K., Nariai, T., Honda, M., and Hanakawa, T. (2013). Dynamic neural network reorganization associated with second language vocabulary acquisition: a multimodal imaging study. *J. Neurosci.* 33, 13663–13672. doi: 10.1523/JNEUROSCI.0410-13.2013

Hussar, K. M., and Harris, P. L. (2010). Children who choose not to eat meat: a study of early moral decision-making. *Soc. Dev.* 19, 627–641. doi: 10.1111/j.1467-9507.2009.00547.x

Implicit Association Test. (2014). Available online at: http://en.m.wikipedia.org/wiki/implicit-association_test

Lomo, T. (2003). The discovery of long term potentiation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 617–620. doi: 10.1098/rstb.2002.1226

Melia, K. R., Ryabinin, A. E., Corodimas, K. P., Wilson, M. C., and Ledoux, J. E. (1996). Hippocampal-dependent learning and experience-dependent activation of the hippocampus are preferentially disrupted by ethanol. *Neuroscience* 74, 313–322. doi: 10.1016/0306-4522(96)00138-8

Neville, H., Stevens, C., and Pakulak, E. (2010). *Interacting Experiential and Genetic Effects on Human Neurocognitive Development. Human Neuroplasticity and Education*. Vatican: Pontifical Academy of Sciences, Scripta Varia117.

Ohno, H., Kato, S., Naito, Y., Kunitomo, H., Tomioka, M., and Iino, Y. (2014). Role of synaptic phosphatidylinositol 3-kinase in a behavioral learning response in *C. elegans*. *Science* 345, 313–317. doi: 10.1126/science.1250709

Park, N., Peterson, C., and Seligman, M. E. P. (2004). Strengths of character and well-being. *J. Soc. Clin. Psychol.* 23, 603–619. doi: 10.1521/jscp.23.5.603.50748

Peterson, C., and Seligman, M. E. P. (2004). *Character Strengths and Virtues*. New York, NY: Oxford University Press.

Rankin, C. H., Marshland, S., McSweeney, F. K., Wilson, D. A., Wu, C. F., Thompson, R. F., et al. (2009). Habituation revisited: an updated and revised description of the behavioral characteristics of habituation. *Neurobiol. Learn. Mem.* 92, 135–142. doi: 10.1016/j.nlm.2008.09.012

Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Ann. Rev. Neurosci.* 27, 169–192. doi: 10.1146/annurev.neuro.27.070203.144230

Sanguineti, J. J. (2013). "Can the self be considered a cause?," in *Brains Top Down: Is Top-down Causation Challenging Neuroscience*, eds G.

Auletta, I. Colage, and M. Jeannerod (London: World Scientific Publishing), 121–142.
Stanley, D, Phelps, E., and Banji, J. (2008). The neural basis of implicit attitudes. *Curr. Dir. Psychol. Sci.* 17, 164–170. doi: 10.1111/j.1467-8721.2008.00568.x
Zwick, M., and Fletcher, J. A. (2014). Levels of altruism. *Biol. Theory* 9, 100–107. doi: 10.1007/s13752-013-0145-8

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 27 March 2014; accepted: 31 August 2014; published online: 18 September 2014.
Citation: Larrivee D and Gini A (2014) Is the philosophical construct of "habitus operativus bonus" compatible with the modern neuroscience concept of human flourishing through neuroplasticity? A consideration of prudence as a multidimensional regulator of virtue. Front. Hum. Neurosci. ***8****:731. doi: 10.3389/fnhum.2014.00731*
This article was submitted to the journal Frontiers in Human Neuroscience.

frontiers in
HUMAN NEUROSCIENCE

OPINION ARTICLE
published: 02 September 2014
doi: 10.3389/fnhum.2014.00682

A dynamic systems view of habits

Nathaniel F. Barrett*

Mind-Brain Group, Institute for Culture and Society, University of Navarra, Pamplona, Spain
**Correspondence: nbarrett@unav.es*

Edited by:
Jose Angel Lombo, Pontifical University of the Holy Cross, Italy

Reviewed by:
Scott Kelso, Florida Atlantic University, USA

Keywords: dynamic systems, habits, multistability, hysterisis, learning

This paper explores some of the insights offered by a dynamic systems approach into the nature of habits. "Dynamic systems approach" is used here as an umbrella term for studies of cognition, behavior, or development as systems of elements that change over time (e.g., Thelen and Smith, 1994, 2006), while "dynamical systems" is reserved for studies that use differential equations to describe time-based systems (e.g., Schöner and Kelso, 1988; Tschacher and Dauwalder, 2003). The following discussion draws primarily from the coordination dynamics research of Kelso (1995, 2012), which stems from Haken's theory of synergetics (1977, 2003). However, the view of habits presented here is more of an interpretive application than a literature review, as the work on which it draws does not address habits explicitly. Perhaps this is because conventional notions of habit are too broad and loose to be captured succinctly in dynamic terms. Dynamical studies of human behavior have focused on more specific capacities such as motor coordination (Thelen et al., 1987), perception (Tuller et al., 1994), and learning (Kostrubiec et al., 2012). Yet this variety of applications suggests that the scope of the dynamic approach overlaps significantly with the domain of habits, so that dynamic concepts could be used to challenge and refine our conventional notions of habitual behavior. Accordingly, the goal of this paper is to raise questions about the nature of habits rather than present a comprehensive scientific theory.

For a dynamic systems approach, stability is "the central concept" (Schöner and Kelso, 1988, p. 1515). The "essential issues are the stability of the system, as indexed by the behavior of some collective measure of the multiple components, and the changes in stability over time" (Thelen and Smith, 2006, p. 289). Intuitively, it would seem that the characteristic stability or stabilities of a system—its preferred states—are its habits. But the connection between stable states and habitual behavior is not as straightforward as it appears. The preferred states of a dynamic system are not simply "built in"; rather they depend on the interactive dynamics of the system's components as well as the interactive couplings of the system with its environment. The following discussion explores the implications of four features of dynamic system stability for our understanding of habit: (1) stability is relative to timescale, and system stabilities at different timescales are interdependent; (2) the attractor landscape describing the characteristic stabilities of a system can be altered by various control parameters, including situational parameters; (3) systems can have multiple stabilities, such that the stability they exhibit at any given time may depend on their recent history; (4) learning processes tend to affect a whole cluster of interrelated stabilities and not just one stability in isolation.

In light of these features of dynamic stability, it seems that there is no straightforward way to map conventional notions of personal habits onto stabilities of the human person considered as a nested dynamic system of body, brain, and environment (Chiel and Beer, 1997). Should we consider as habits only the "intrinsic" stabilities of brain and body, regardless of the variety of behaviors that can arise from these stabilities in different situations? Or should we only consider as habits those patterns of behavior that are regularly observed within a certain type of situation, regardless of how differently these patterns might be assembled at the body-brain level? From the dynamic perspective, stability and change are ubiquitous features at every level or spatiotemporal scale of description. Thus, it seems arbitrary to apply the term "habit" only to one kind or level of stability, and perhaps this explains why the term is seldom used in dynamic systems literature. Yet it could be argued that this multilevel complexity is an advantage, as it can be used to challenge conventional notions of habit in interesting ways. Let us suppose that for any level of human behavior that can be described as a dynamic system, the stabilities or preferred states of that system—its "attractor landscape"—are at least analogous to habits, and should be considered as such. What is revealed by this broader, dynamic perspective?

First, this view calls into question the usual timescale of habits, which is typically restricted to stable features of personality and behavior on the timescale of months or years (Lewis, 2000). From a dynamic perspective, these stabilities are in principle no different from stabilities at faster and slower timescales. Moreover, different timescales of stability are interrelated: while the habits of any given timescale are shaped by the "deeper" habits of a slower timescale, they also can lead to changes at this deeper level. In other words, the "force of habit" is not one-way: habits are shaped by the behaviors that they themselves constrain. For example, mood is shaped by habits of personality, while a string of similar moods can lead to changes of personality. And though it may seem strange to think of moods as temporary emotional habits on the timescale of hours or days, their way of shaping and being shaped by emotional states on the

timescale of seconds or minutes is analogous to the relationship between personality and mood. Likewise one can treat emotional states as habits that contribute to the attractor landscapes for thoughts and sensorimotor activity at an even faster timescale, while personality itself can be seen as evolving over the very "deep" landscape of developmental habits (Thelen and Smith, 2006). Thus, a dynamic view opens up a wider range of timescales across which the concept of habit might apply. But more importantly, even if we choose to restrict habit to just one of these levels, the interaction of timescales suggests that our understanding of any one level should draw upon at least two neighboring levels (Kelso, 1995), if not the entire nested system.

Second, a dynamic view of habit will include regular variations of the attractor landscape that occur in relation to changing parameters. In some cases these variations can be represented by a bifurcation diagram that shows how the attractor landscape changes in relation to a single control parameter (Kelso and Engstrom, 2006, pp. 124–137). Most human behaviors, however, require that multiple parameters are taken into account. Now, provided that the important variables and parameters for describing characteristic variations of a certain behavior for an individual can be determined (a very difficult task, in most cases), one could, in theory, construct a comprehensive "habit topology" for that behavior: a map of how that behavior's preferred states change in relation to various parameters. Notice that, depending on the parameters involved, some stabilities of this habit topology will be visited by the system more or less regularly. For instance, in the case of quadruped motion, assuming that the parameter of speed varies regularly over its natural range, a quadruped (e.g., horse) regularly visits the various gaits (walk, trot, gallop) that make up the preferred states of its habit topology (Hoyt and Taylor, 1981; Schöner and Kelso, 1988, p. 1516). However, for some behaviors, especially those that are sensitive to multiple parameters, certain regions of the habit topology can remain "hidden" because the required values of the relevant control parameters are rarely if ever encountered. Imagine, for example, that a person who never dances might be found, on one occasion, happily dancing the night away. Conventionally speaking, this behavior is not habitual. But from a dynamic view, it is hard to say. Perhaps on that occasion the person encountered just the right combination of circumstances—excellent mood, pleasurable company, Afro-Cuban music, fantastic mojitos, etc.—that made, for that person, dancing a very deep (and enjoyable) stability. If these circumstances will regularly facilitate the same behavior for that person, cannot we say that dancing is habitual for them in those circumstances? The point here is not to insist on this characterization, but to question our conventional understanding of habits, which typically involves assumptions about "normal" circumstances. Are habitual behaviors only rightfully considered as such if we regularly encounter the circumstances that facilitate their expression? If not, how would we determine the unexpressed or latent habits of a person, independently of circumstances?

Third, it is important to consider that the attractor landscape of a particular behavior commonly has multiple stabilities even within restricted parameter values (Schöner and Kelso, 1988, p. 1518). This phenomenon of multistability implies that "habitual behavior" may not be always the same even when all relevant variables are the same. In such cases, which stability describes the habit? The stability into which the system enters may depend on its recent history or on a symmetry breaking. For example, as represented by the much-studied HKB model (Chemero, 2009), coordinated finger movement—waving the two index fingers in time with a metronome, where tempo or speed is the main control parameter—exhibits bistability at slower speeds and monostability at faster speeds (Kelso, 1995). Within the slower, bistable regime, which of the two stable states the system exhibits may depend on its historical trajectory: for instance, if the system has just entered the bistable regime from the monostable regime, it will remain in the preferred state of the latter. Although coordinated finger movement may not seem representative of human behavior, coordination dynamics and its telltale characteristics—such as multistability—have been found in a wide range of behavioral and neural systems (Schöner and Kelso, 1988; Kelso, 2012), suggesting an important implication: perhaps what a person presently exhibits as habitual behavior within certain circumstances does not uniquely characterize what is habitual for that person *even within those circumstances*, but must also be understood as a result of a particular historical trajectory. The importance of history is also indicated by the phenomenon of hysteresis (Haken, 2003, pp. 8–9), i.e., when the state or behavior of a system is influenced by the residual stability of an antecedent regime. Thus, certain behaviors might persist even after the attractor landscape has changed so that they are no longer "habitual" within the current context.

Finally, if we define learning in terms of alterations to the attractor landscape for a particular behavior such that some states are newly stabilized while others are destabilized (Kostrubiec et al., 2012), the interrelatedness of habits as implied by the previous three points indicates that habits are rarely altered in a piecemeal fashion: learning affects entire clusters of habits that are composed of shared components. That is, learning affects an entire "habit space," and not just individual habits in isolation (Kelso, 1995, pp. 159–186). This further complicates the picture of personal habits, as well as the kinds of learning strategies we should adopt to change them. For instance, it suggests that certain habitual behaviors that are difficult to target can be altered indirectly by focusing on other, related behaviors. It also suggests that in many cases—especially where a complex task demands a finely articulated set of behaviors—training should focus on shaping the overall "habit space" rather than each individual behavior. Moreover, as implied by the definition of learning just given, the dynamical approach suggests that habits are a necessary condition for learning and not just a product—contrary to the so-called "blank slate" model (Kostrubiec et al., 2012).

In conclusion, we see that a fairly simple definition of habit in terms of dynamic stability yields a number of insights that question our conventional notion of personal habits as slow-changing, context-independent, uniquely repetitive, and discrete behaviors. At the same time, a

dynamical view can give an account of the relative fixity of behaviors that we take to be habitual in the conventional sense. However, by showing that this fixity is dynamically and relationally constituted, a dynamical view reveals the elaborate context in which all habits are embedded. Moreover, human habits appear to be instances of a very general pattern: all living systems exhibit self-organized stabilities that reduce their degrees of freedom, producing robust but flexible repertoires of behavior.

REFERENCES

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, MA: MIT Press.

Chiel, H. J., and Beer, R. D. (1997). The brain has a body: adaptive behavior emerges from interactions of nervous system, body, and environment. *Trends Neurosci.* 20, 553–557. doi: 10.1016/S0166-2236(97)01149-1

Haken, H. (1977). *Synergetics: An Introduction*. Heidelberg: Springer-Verlag.

Haken, H. (2003). "Intelligent behavior: a synergetic view," in *The Dynamical Systems Approach to Cognition*, eds W. Tschacher and J.-P. Dauwalder (Singapore: World Scientific), 3–16.

Hoyt, D. A., and Taylor, C. R. (1981). Gait and the energetics of locomotion in horses. *Nature* 292, 239–240. doi: 10.1038/292239a0

Kelso, J. A. S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Cambridge: MIT Press.

Kelso, J. A. S. (2012). Multistability and metastability: understanding dynamic coordination in the brain. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 906–918. doi: 10.1098/rstb.2011.0351

Kelso, J. A. S., and Engstrom, D. A. (2006). *The Complementary Nature*. Cambridge, MA: MIT Press.

Kostrubiec, V., Zanone, P.-G., Fuchs, A., and Kelso, J. A. S. (2012). Beyond the blank slate: routes to learning new coordination patterns depend on the intrinsic dynamics of the learner—experimental evidence and theoretical model. *Front. Hum. Neurosci.* 6:222. doi: 10.3389/fnhum.2012.00222

Lewis, D. (2000). "Emotional self-organization at three time scales," in *Emotion, Development, and Self-Organization: Dynamic Systems Approaches to Emotional Development*, eds M. D. Lewis and I. Granic (Cambridge: Cambridge University Press), 37–69.

Schöner, G., and Kelso, J. A. S. (1988). Dynamic pattern generation in behavioral and neural systems. *Science* 239, 1513–1520. doi: 10.1126/science.3281253

Thelen, E., Skala, K. D., and Kelso, J. A. S. (1987). The dynamic nature of early coordination: evidence from bilateral leg movements in young infants. *Dev. Psychol.* 23, 179–186. doi: 10.1037/0012-1649.23.2.179

Thelen, E., and Smith, L. B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: Bradford/MIT Press.

Thelen, E., and Smith, L. B. (2006). "Dynamic systems theories," in *Handbook of Child Psychology, Vol. I: Theoretical Models of Human Development, 6th Edn.*, ed W. Damon (Hoboken, NJ: John Wiley & Sons), 258–312.

Tschacher, W., and Dauwalder, J.-P. (eds.). (2003). *The Dynamical Systems Approach to Cognition: Concepts and Empirical Paradigms Based on Self-Organization, Embodiment, and Coordination Dynamics*. Singapore: World Scientific.

Tuller, B., Case, P., Ding, M., and Kelso, J. A. S. (1994). The nonlinear dynamics of categorical perception. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 3–16.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 April 2014; accepted: 14 August 2014; published online: 02 September 2014.

Citation: Barrett NF (2014) A dynamic systems view of habits. Front. Hum. Neurosci. ***8****:682. doi: 10.3389/fnhum.2014.00682*

This article was submitted to the journal Frontiers in Human Neuroscience.

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 08 August 2014
doi: 10.3389/fnhum.2014.00590

Modeling habits as self-sustaining patterns of sensorimotor behavior

Matthew D. Egbert*[1] *and* ***Xabier E. Barandiaran*[2,3]**

[1] *Embodied Emotion, Cognition and (Inter-)Action Lab, School of Computer Science, University of Hertfordshire, Hatfield, UK*
[2] *Department of Philosophy, University School of Social Work, UPV/EHU, University of the Basque Country, Spain*
[3] *Department of Philosophy, IAS-Research Center for Life, Mind, and Society, UPV/EHU University of the Basque Country, Spain*

Edited by:
Javier Bernacer, University of Navarra, Spain

Reviewed by:
Paul Williams, Cognitive Science Program at Indiana University, USA
Takashi Ikegami, The University of Tokyo, Japan

****Correspondence:***
Matthew D. Egbert, Embodied Emotion, Cognition and (Inter-)Action Lab, School of Computer Science, University of Hertfordshire, College Lane, Hatfield, Herts AL10 9AB, UK
e-mail: mde@matthewegbert.com

In the recent history of psychology and cognitive neuroscience, the notion of habit has been reduced to a stimulus-triggered response probability correlation. In this paper we use a computational model to present an alternative theoretical view (with some philosophical implications), where habits are seen as self-maintaining patterns of behavior that share properties in common with self-maintaining biological processes, and that inhabit a complex ecological context, including the presence and influence of other habits. Far from mechanical automatisms, this organismic and self-organizing concept of habit can overcome the dominating atomistic and statistical conceptions, and the high temporal resolution effects of situatedness, embodiment and sensorimotor loops emerge as playing a more central, subtle and complex role in the organization of behavior. The model is based on a novel "iterant deformable sensorimotor medium (IDSM)," designed such that trajectories taken through sensorimotor-space increase the likelihood that in the future, similar trajectories will be taken. We couple the IDSM to sensors and motors of a simulated robot, and show that under certain conditions, the IDSM conditions, the IDSM forms self-maintaining patterns of activity that operate across the IDSM, the robot's body, and the environment. We present various environments and the resulting habits that form in them. The model acts as an abstraction of habits at a much needed sensorimotor "meso-scale" between microscopic neuron-based models and macroscopic descriptions of behavior. Finally, we discuss how this model and extensions of it can help us understand aspects of behavioral self-organization, historicity and autonomy that remain out of the scope of contemporary representationalist frameworks.

Keywords: sensorimotor, self-maintaining patterns-of-behavior, mental-life, habits, meso-scale modeling

1. INTRODUCTION

Our mental life is populated by myriads of often covert, fluid and inconspicuous patterns of behavior that have slowly grown on us, continuously sustained by repetition and scaffolded by reliable environmental structures. Looking left or right before crossing the road, lacing your shoes, or simply walking can be understood as nested complexes of sensorimotor coordination patterns, entrained by a history of subtle self-reinforcement, a history of *habit*.

That habit is "second nature" was well understood by Greek philosophers; i.e., that in contrast to the nature of vegetative function, psychological nature was made of history-dependent ecological (i.e., agent-environment relational) entities in which physiological aspects of the organism (brain and body) were intertwined, through practice, with environmental resources, forming "natural" structures of behavior. In this sense, James stated that "animals are bundles of habit" (James, 1890, p.104) and considered habits to be the building block of the main object of psychology (and neuroscience): "the Science of Mental Life" (James, 1890, p.1). For a time, habits were the cornerstone of psychology (and some early neuroscientific intuitions) until the rise of cognitivism and the conception of the mind as computational processing of internal representations (see Barandiaran and Di Paolo, 2014).

Unfortunately, the rise of computational representationalism in neuroscience relegated the concept of habits to mere stimulus-triggered response automatisms, far removed from the contemporary intellectualist interest in the rational, linguistic or conscious processes that are nowadays seen as the epitome of human cognition. And yet, cognitive and neural sciences have been witnessing a paradigmatic change for the last two decades, moving away from the computer metaphor and becoming increasingly aware of the role of sensorimotor interaction for neural function (Engel et al., 2013), of self-organization in brain dynamics (Kelso, 1995; Freeman, 2001), plasticity and multiscale dynamics (Hurley and Noë, 2003), or the role of embodiment for cognition (Maturana and Varela, 1980; Pfeifer et al., 2007; Chemero, 2009).

The goal of this paper is to provide a simulation model that works as an illustration and a proof of concept for a theoretical reappraisal of a notion of habit that challenges some of the contemporary assumptions and limitations, both in behavioral neuroscience and cognitive science. This is why

we provide considerable philosophical, historical and theoretical background. It allows us to frame the value and contribution of the model and to deliver an insightful theoretical interpretation of the results. The use of simulation models with theoretical goals follows the tradition of Cybernetics, Artificial Life and Cognitive Science where opaque conceptual relationships (between micro and macro, between mechanisms and behavior, philogeny and ontogeny, etc.) can be disclosed and elaborated. Relatively simple (compared to natural systems) computational models can help shifting strong philosophical assumptions (Dennett, 1994; Di Paolo et al., 2000; Barandiaran and Moreno, 2006). In particular, this paper explores the idea of habits as embodied sensorimotor life-forms, extending upon several contemporary trends in cognitive and neural science that take self-organizing and self-sustaining living processes as the root of cognitive capacities (in opposition to the abstract and functionally disembodied foundations of representational computationalism) (Damasio, 2003; Di Paolo, 2003; Barandiaran, 2008; Thompson, 2010). We shall identify life-like properties of habit at the meso-scale defined by sensorimotor contingencies and coordination dynamics (O'Regan and Noë, 2001; Noë, 2006; Buhrmann et al., 2013): that is, below the macroscopic level of modeling but above the microscopic level of neuro-synaptic activity. It is at this mesoscopic level that a first approximation to a continuous-time, plastic and embodied conception of habit can be adequately investigated using simulations of simple robots that, through plastic sensorimotor controllers, explore and exploit their embodied interaction with their environment thereby making possible the emergence and self-organization of habits.

In the next sections we introduce the wider background and motivation for this work, with a short historical introduction to the notion of habit and its reappraisal in the context of contemporary neuro and cognitive sciences. We then introduce a new modeling paradigm for habits: a node-based *iterant deformable sensorimotor medium*. We couple this medium to a robots body, situated in 1D and 2D environments and we show how it supports the sensorimotor imprinting of habits and their spontaneous formation, maintenance and development. We also point out some possible extensions of our model, together with some reflexion upon the advantages and possibilities of a habit-based robotics modeling framework, before concluding with some general discussion about the nature of habits, the autonomy of behavior and its link with neurodynamic identity, autonomy and freedom.

1.1. HABITS: FROM ARISTOTLE TO NEUROSCIENCE

The notion of "habit" was once (and for a very long time) a central element of psychological and behavioral theory; either as a unit of behavioral organization or as a mechanism of association of ideas, impressions, or other psychological units of analysis. From Aristotle in the 4th century BC to Clark Hull in the late 40 s, throughout Hume, Hegel, Lamarck, William James, Dewey, Allport, Thorndike, Skinner, Merleau-Ponty or Piaget (see Barandiaran and Di Paolo, 2014 for a general overview) they all gave a privileged status to the notion of habit in psychological, behavioral or neural theory. With behaviorism, however, the philosophical and conceptual diversity and complexity of the concept of habit collapsed down to the notion of a stimulus-response probability correlation and the theoretical relevance of the concept diminished radically with the rise of cognitivism and the introduction of representations into the center of psychological theorizing. Today, the mind is "officially" made out of *representations* and made by *computations*, but for a long time before that, it was made out of *habits* and by *habit*.

The first scientific formulation of a habit as a self-reinforcing repetitive pattern of behavior might be attributed to Thorndike's *Law of Exercise* which states that:

> Any response to a situation will, other things being equal, be more strongly connected with the situation in proportion to the number of times it has been connected with that situation and to the average vigor and duration of the connections. (Thorndike, 1911, p. 244)

Previously, similar formulations (albeit more speculative and without explicit experimental basis) were made by Hartley, James, and other associationists. Almost as early as the XVIIIth century (Hartley, 1749; Buckingham and Finger, 1997), the notion of habit was closely associated with neuronal properties. It took the strong epistemological standards that logical-positivism imposed upon psychology for behaviorism to completely give up on internal mechanisms and center habit research on purely externalist grounds, avoiding any interpretation of the internal brain mechanisms that could sustain them. But, from their early conception, these theories found a material basis for habit on the plasticity of nervous "vibrations" or pathways, to be much later developed into a scientifically mature hypothesis about synaptic plasticity on what is now widely known as "Hebb's rule." But this neuronal principle soon became almost exclusively applied within an informational or representational framework in cognitive neuroscience (Hebb, 1949) and the sensorimotor and embodied development of this principles still remains relatively under-explored.

Despite the displacement toward more sensorimotor and interaction-centered dynamical and embodied approaches to cognition (Kelso, 1995; Thompson and Varela, 2001; Chemero, 2009), and despite the recent emphasis on the relationship between life and mind in neuroscience (Damasio, 2003; Thompson, 2010), the notion of habit has attracted little attention so far. And yet, this concept holds the potential to become a blending category between the biological and the psychological. Habits have the capacity to become a theoretical building block for an organicist conception of mind that makes justice to the recent focus on sensorimotor and embodied approaches (Di Paolo, 2003) while it avoids the problems that the concepts of information and representation have been shown to face in contemporary cognitive science (Hutto and Myin, 2012). In fact, if we are to take mental life as the main object of study of human (and animal) neuroscience, it is worth considering the deep analogy with life that the notion of habit makes possible in the realm of psychology and behavioral neuroscience: just as self-sustaining, far-from-equilibrium dissipative structures, such as auto-catalytic metabolic chemistry, have been considered an essential building block of minimal living organization (Nicolis and Prigogine, 1977; Kauffman, 2000; Virgo, 2011), so could we explore the possibility of self-sustaining, "far-from-equilibrium," dissipative

sensorimotor patterns as the most basic building blocks of mental life (Barandiaran, 2007, 2008)[1]. What different forms of life share (at the most basic or fundamental level) is the presence of spontaneously emerging self-organized patterns (Bedau, 1997), and habits can be conceived as a paradigmatic example of these. They can be conceived as precarious, self-maintaining "mental life-forms" that can persist through repetition in the space of behavioral neuro-dynamics.

Ever since Hebb's work and the rise of computationalism, theoretical neuroscience has made considerable progress through the use of computer simulations of neural dynamics and the use of robots to embody and test different theoretical principles (Grey Walter, 1950; Ruppin, 2002; Edelman, 2007). Current embodied and situated simulation techniques (Beer, 2003; Froese and Ziemke, 2009) might help a reappraisal of a richer conception of habits that takes their sensorimotor lifelike properties as a departure point. But how can habits, as behavioral life-forms, be modeled? What is the simplest and most direct (yet open-ended) implementation for a robot controller capable to display spontaneous habit formation, self-maintenance and evolution?

1.2. MODELING HABITS, A NEW APPROACH

Historical and contemporary attempts to model and formalize habits (Hull, 1950; Sutton and Barto, 1998; Dezfouli et al., 2012) share some of the following features: (a) they assume a probabilistic stimulus-response approach with a discretized set of stimuli and responses, (b) they assume a neural network level of implementation and/or (c) they implement an explicit and decoupled reward system (i.e., sensorimotor coupling is modulated by a reward function that is independent from sensorimotor dynamics, that is, they are dependent on the result of actions but not on the very dynamics of behavior). Here, instead, we attempt a modeling approach that departs from a different set of assumptions: (a) we leave aside how habit formation and activation might be supported by neural networks and different forms of synaptic plasticity, and develop the model directly at a mesoscopic level of sensorimotor dynamics, (b) we assume a continuous sensorimotor space (i.e., we do not accept a discretized or prespecified input or output spaces in the form of symbolic input or pre-defined action outputs); and, (c) the system allows for the self-organization of macroscopic patterns of sensorimotor coordination by repetition. In a nutshell, we model directly at a mesoscopic level of continuous sensorimotor contingencies or coordination dynamics (Noë, 2006; Buhrmann et al., 2013) with a plastic controller that is shaped by the very trajectories of the sensorimotor flow.

In this paper we identify micro, macro and mesoscopic levels of modeling of habits. The micro-meso-macroscale distinction can be applied to a variety of phenomena, and, in turn, to each level of modeling we might be interested in. So, for instance, Freeman (2000) identifies the microscopic level of modeling for neurodynamics with individual neuronal activity and the macroscopic level with behavioral or cognitive states and focuses his research on a mesoscopic level of brain regions (as identified by EEG signals) [2]. For the case of habit modeling, the most widespread macro level is the level of functionally distinguishable and discretizable stimuli and responses (e.g., food colors or spatial landmarks as stimuli and eating or ignoring the food, turning left or right as macroscopic descriptions of the response). The microscopic level of modeling of habits might correspond to a neuronal level of implementation, where different sensory or effector neurons, for example, strengthen their connection with an interneuron following Hebb's rule or some other synaptic strengthening process. Interestingly, most of habit modeling frameworks assume a one-to-one mapping between the macroscopic and microscopic levels of description/modeling, such that specific environmental features or stimuli correspond to a specific neurons or ensembles of neurons, and the same goes for reinforcers and responses (e.g., a neuron might represent the action of turning left or the reward value of an action outcome). What we mean by a mesoscopic level of modeling for habits is one that is above the neuronal details yet below the macroscopic discretized and individualized stimulus and response units. Our goal is to develop a modeling framework where those macroscopic units emerge as unified patterns out of a continuous sensorimotor flow by means of iterating reinforcement processes without explicit neuronal assumptions.

Thus we propose a sensorimotor architecture that permits patterns of sensorimotor contingencies to self-organize in a manner analogous to the way in which human trails are formed in nature (Helbing et al., 1997): the more the path is used, the more grass struggles to grow; the less grass, the more likely for a human to choose that path, so the more the path is used the more likely it will be used again. For the exploratory purpose of this paper, we take habits to be instances of a similarly self-reinforcing process; the more frequently a pattern of behavior (i.e., sensorimotor coordination trajectory) is performed, the more likely it will be repeated in the future. With this idea in mind we take the following working definition of habit: "a self-sustaining pattern of sensorimotor coordination that is formed when the stability of a particular mode of sensorimotor engagement is dynamically coupled with the stability of the mechanisms generating it" (Barandiaran, 2008, p. 281) and we add the property of reinforcement by repetition.

To capture this kind of self-organization of sensorimotor trajectories in a computational model, we developed the notion of an *Iterant Deformable Sensorimotor Medium* (IDSM). The IDSM is a construct that plays a role similar to the grass in the above metaphor; it is imprinted by paths taken through

[1] Biological life has also been reduced or studied through the exclusive lenses of information theory and representation; and the debate around the origins and definition of life suffers a parallel divide between the so called "replication-first" and "metabolism-first" schools of thought, the former advocating for genes or replicators as informational templates, the latter advocating for a network of far-from-equilibrium chemical reactions (Szathmáry, 2000; Shapiro, 2006)

[2] But the very neuronal level (what Freeman identifies as microscopic level) could also in turn be divided into its own micro-meso-macro levels, molecular mechanisms constituting the micro level, neuronal input-output dynamics constituting the macro level and intermediate levels being those that include statistical aspects of the molecular level (e.g., chemical dynamics) and the spatial configuration of the neuron to generate a specific action potential. For each case the level of detail, the spatiotemporal scales, the degree of abstraction or generality might determine what micro, meso and macro means.

it, and it influences subsequent paths such that they are similar to those that have been taken in the past. Similar to how an imprintable ground, such as grass, is necessary for self-reinforcing trail-formation, the IDSM makes possible the existence of self-reinforcing sensorimotor trajectories.

A *sensorimotor space* defines all possible sensory and motor states of an agent, where each point indicates a single state of every motor and sensor of the agent. An organism (e.g., a bacteria) with a single photoreceptor and a single flagellar motor (that can rotate clockwise or counter-clockwise) has a 2D sensorimotor space where an organism with three chemoreceptors and five muscles has an 8D sensorimotor space.

A *sensorimotor medium* defines, for each sensorimotor state (i.e., for each point in the sensorimotor space), what the next motor state will be. A sensorimotor medium is *deformable* when the mapping between the sensorimotor state and the next motor state (or the rate of change of the motors) changes in time in a state-dependent manner. This deformation could be plastic (where deformations are conserved) or elastic (where deformations tend to recover the original shape of the medium). And we call a deformable sensorimotor medium *iterant* when deformations caused by trajectories reinforce the pathways taken by those trajectories, that is, when iterations or repetitions of the trajectories through the sensorimotor space increase the likelihood of subsequent trajectories being similar. This way we get to the notion of *Iterant Deformable Sensorimotor Medium* (IDSM): a mapping between current sensorimotor state and the next motor state that is modified so as to reinforce or strengthen those trajectories that are iterant or repetitive. We can think of an IDSM as similar to a river's drainage basin (that both channels the future flow of water and, at the same time, is molded by it) or the trail formation example above: the more a trajectory is taken, the "stronger" it becomes, i.e., the higher the tendency of similar states to fall into the same pathway and the harder for this trajectories to deviate from the previously traversed course.

To our knowledge no previous attempts have yet been made to model behavior with an IDSM. The rise of situated robotics in the 90 s (Brooks, 1991; Steels, 1993) was centered on subsumption architectures where specialized control circuits gave rise, in embodied interaction with the environment, to specific behavioral patterns. Neural network controllers (Ruppin, 2002; Edelman, 2007) and more specifically Continuous Time Recurrent Neural Networks (Beer, 2003), and particularly the work with plastic CTRNNs (Di Paolo, 2000, 2003) came closer to our notion of IDSM, but they don't quite capture the properties of iterant deformation we want to explore, in particular, they do not sufficiently facilitate the explorations of habits as self-maintaining patterns of behavior.

There are many ways that an IDSM could be mathematically formulated and computationally implemented. We have experimented with several such architectures. The model presented below remains an experimental and preliminary design, but one that already presents interesting dynamics demonstrating the idea of habits as self-sustaining behavioral patterns, and allowing us to view habit-formation, habit-maintenance, and habit-based behavior from a richer dynamical perspective than the classical stimulus-response, reinforcement learning or various neural network models.

2. MODEL

For the purpose of this paper we take habits to be patterns of behavior (i.e., sensorimotor coordination) that are reinforced by their repetition. To model these properties in a sensorimotor-focused framework, we developed an Iterant Deformable Sensorimotor Medium (IDSM), a plastic, self-modifying dynamical system that when coupled to a robots sensors and motors, (1) causes the robot to repeat behaviors that it has performed in the past, and (2) allows for the reinforcement of patterns of behavior through repetition, such that the more frequently and recently a pattern of behavior has been performed, the more likely it is to be performed again in the future. The remainder of this section explains in technical detail how we implemented an IDSM. Then, in Section 3, we present a series of experiments where the IDSM controls a simulated robot. In these experiments self-maintaining mechanisms of behavior emerge that share properties in common with living systems, and in this way the IDSM is demonstrated as a useful model for investigating habits seen as self-maintaining patterns of behavior.

The IDSM operates by developing and maintaining a history of sensorimotor dynamics. This history takes the form of many "nodes," where each node describes the SM-velocity at a SM-state at some point in the past. As the agent behaves, and its SM-state changes, nodes are added, such that a record is constructed of how sensors and motors have changed for various SM-states during the system's history. These are used in a continuous, dynamical framework to determine future motor-actions such that when a familiar SM-state is encountered, the IDSM produces behavior that is similar to the behavior that was performed when the agent was previously in a similar situation.

More formally, each node is a tuple of two vectors and a scalar, $N = \langle \boldsymbol{p}, \boldsymbol{v}, w \rangle$, where $\boldsymbol{p}$ indicates the SM-state associated with the node (referred to as the node's "position" in SM-space), $\boldsymbol{v}$ indicates a velocity of SM-change, and the scalar, w indicates the "weight" of the node, a value that partially determines the overall influence of the node as described below (**Table 1** provides a list of all symbols with brief descriptions). We shall refer to

Table 1 | Symbols and brief descriptions.

Symbol	Description
$\boldsymbol{x}$	Current SM-state
$N_{\boldsymbol{p}}$	SM-state associated with node N (in normalized SM-space coordinates)
$N_{\boldsymbol{v}}$	SM-velocity indicated by node N (in normalized SM-space coordinates)
N_w	Weight of node N
$d(\boldsymbol{x}, \boldsymbol{y})$	Distance function between two SM-states
$\omega(N_w)$	Function describing how the weight of a node scales its influence
$\phi(\boldsymbol{y})$	Function describing the local density of nodes of SM-state $\boldsymbol{y}$

these components using a subscript notation, where the position, SM-velocity, and weight of node N are written as $N_{\boldsymbol{p}}$ and $N_{\boldsymbol{v}}$ and N_w, respectively.

2.1. CREATION AND MAINTENANCE OF NODES

As a robot controlled by the IDSM moves through SM-states, new nodes are created recording the SM-velocities experienced at different SM-states. More formally, when a new node is created, its "position," $N_{\boldsymbol{p}}$ is set to the current SM-state; its "velocity," $N_{\boldsymbol{v}}$ is set to the current rate of change in each SM-dimension, and its weight, N_w is set to 0 (note that slightly unconventionally, in this model a weight of 0 does not mean that the node is ineffectual, but rather that is "neutral," i.e., neither stronger nor weaker than when initially created). The two vector terms ($N_{\boldsymbol{p}}$ and $N_{\boldsymbol{v}}$) are calculated in a normalized sensorimotor space, where the range of all sensor and motor values are linearly scaled to lie, in each dimension, between 0 and 1.

New nodes are only added when the density of nodes near the current SM-state, as described by the function ϕ, is less than a threshold value, $\phi(\boldsymbol{x}) < k_t = 1$. This density function, ϕ, can be thought of as a measure of how many nodes there are near to the SM-state $\boldsymbol{x}$, and how heavily weighted those nodes are. It is calculated by summing a non-linear function of the distance from every node to the current SM-state $d(N_{\boldsymbol{p}}, \boldsymbol{x})$, scaled by a sigmoidal function of the node's weight $\omega(N_w)$, as described in Equations (1–3) and **Figure 1**.

$$\phi(\boldsymbol{x}) = \sum_N \omega(N_w) \cdot d(N_{\boldsymbol{p}}, \boldsymbol{x}) \tag{1}$$

$$\omega(N_w) = \frac{2}{1 + \exp(-k_\omega N_w)} \tag{2}$$

$$d(N_{\boldsymbol{p}}, \boldsymbol{x}) = \frac{2}{1 + \exp(k_d ||N_{\boldsymbol{p}} - \boldsymbol{x}||^2)} \tag{3}$$

$$k_d = 1000; \qquad k_\omega = 0.0025$$

After a node is created, its weight changes according to differential Equation (4), where the first term represents a steady degradation of the node's influence, and the second term represents a strengthening of the node that occurs when the current SM-state is close to the node's position. This latter term allows for the self-reinforcement/self-maintenance of patterns of behavior, such that when SM-states are revisited, the nodes there are reinforced and thus, patterns of behavior that are repeated are more likely to persist than those that only occur once.

$$\frac{\mathrm{d}N_w}{\mathrm{d}t} = -1 + r(N, \boldsymbol{x}) \tag{4}$$

$$r(N, \boldsymbol{x}) = 10 \cdot d(N_{\boldsymbol{p}}, \boldsymbol{x}); \tag{5}$$

2.2. NODES INFLUENCE THE MOTOR-STATE

A short period of time after creation (10 simulated time-units), nodes are activated, meaning that they are added to the pool of nodes that influence the motor state. If this delay were absent, any newly created nodes would more strongly influence the next SM-velocity than the nodes created during previous SM-trajectories, which would prevent the system from accomplishing the desired SM-trajectory reinforcement described above. Every activated node influences the motor state, but at any one time only a subset of these will have a substantial influence, for the influence of a node is scaled non-linearly by its distance from the current SM-state by the same distance function used in ϕ above. The influence of each node is also scaled by its weight, and thus nodes that are close to the current SM-state and nodes with higher weights have a greater influence. We shall look into the influence of node weight in greater detail in a moment, but first let us look at how the nodes influence the SM-state.

The influence of a node upon the motors can be broken down into two factors: a "velocity" factor and an "attraction" factor. The velocity factor (Equation 6) is simply the motor components of the $N_{\boldsymbol{v}}$ vector. The attraction factor (Equation 7), is slightly more complicated. It is a "force" that draws the system toward the node. This tends to result in a motion in SM-space toward regions of SM-space that are familiar, i.e., for which there is a higher density of nodes. **Figure 2** provides a visualization of the influence of a single, activated node, located at $N_{\boldsymbol{p}} = (0.5, 0.5)$ with $N_{\boldsymbol{v}} = (0, 0.1)$ in a hypothetical 2-motor, 0-sensor IDSM. Because $N_{\boldsymbol{v}}$ is exactly vertical in this example, all horizontal motion is

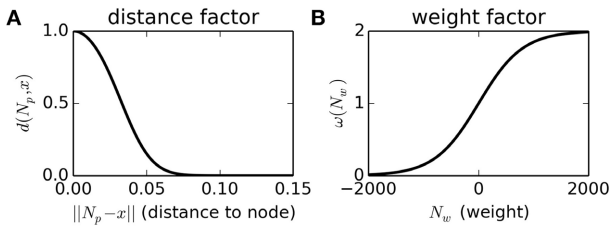

FIGURE 1 | Non-linear functions used to calculate the node-density of a SM-state, and to scale the influence of nodes by their proximity to the current SM-state (Plot A) and by their weight (Plot B). See main text for details.

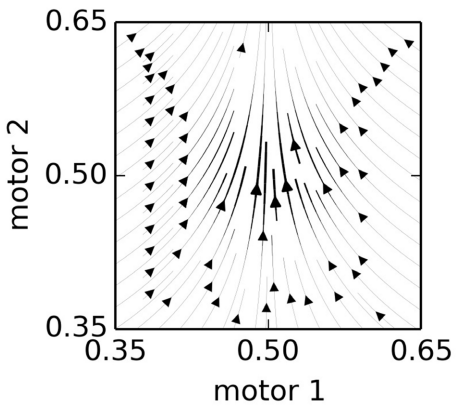

FIGURE 2 | The influence of a single node. This plot shows the combined influence of single node, located at $N_{\boldsymbol{p}} = (0.5, 0.5)$ with $N_{\boldsymbol{v}} = (0, 0.1)$ in a hypothetical 2-motor, 0-sensor IDSM. The $N_{\boldsymbol{v}}$ is exactly vertical, so all horizontal motion is due to the attraction factor, and vertical motion is due to the velocity factor. See Equations (6–9) and main text for details.

due to the "attractive force" of the node. The attraction influence draws the SM-state toward the node and the velocity influence pushes the SM-state away from the node. To prevent the attraction influence from interfering with the velocity influence, the component of the attraction influence that is parallel to the node's velocity vector is removed [as described by the Γ function used in Equations (7 and 10) and defined in Equation (8)].

To calculate the total influence of the IDSM upon the motor state, the velocity and attraction influences of every node are scaled by the node's weight and distance to the SM-state (Equations 6 and 7), and then these are all summed before being scaled by the density of the nodes at the current SM-state (Equation 9) such that the influence of all the nodes is averaged and not cumulative. Obviously, the IDSM only has direct control of its motors and the sensor-components of the SM-state are determined by the systems interaction with its environment. Accordingly, the superscript-$\boldsymbol{\mu}$ notation in the equations below indicates where we are only using the motor-components of the indicated vector terms.

$$V(\boldsymbol{x}) = \sum_N \omega(N_w) \cdot d(N_{\boldsymbol{p}}, \boldsymbol{x}) \cdot N_{\boldsymbol{v}}^{\boldsymbol{\mu}} \tag{6}$$

$$A(\boldsymbol{x}) = \sum_N \omega(N_w) \cdot d(N_{\boldsymbol{p}}, \boldsymbol{x}) \cdot \Gamma(N_{\boldsymbol{p}} - \boldsymbol{x}, N_{\boldsymbol{v}})^{\boldsymbol{\mu}} \tag{7}$$

$$\Gamma(\boldsymbol{a}, N_{\boldsymbol{v}}) = \boldsymbol{a} - \boldsymbol{a} \cdot \frac{N_{\boldsymbol{v}}}{||N_{\boldsymbol{v}}||} \tag{8}$$

$$\frac{\mathrm{d}\boldsymbol{\mu}}{\mathrm{d}t} = \frac{V(\boldsymbol{x}) + A(\boldsymbol{x})}{\phi(\boldsymbol{x})} \tag{9}$$

The repetition of terms in Equations (6,7) allows us to combine and simplify Equations (6–9) into the following more concise formulation:

$$\frac{\mathrm{d}\boldsymbol{\mu}}{\mathrm{d}t} = \frac{1}{\phi(\boldsymbol{x})} \sum_N \left(\omega(N_w) \cdot d(N_{\boldsymbol{p}}, \boldsymbol{x}) \cdot \left(\underbrace{N_{\boldsymbol{v}}}_{\text{Velocity}} + \underbrace{\Gamma(N_{\boldsymbol{p}} - \boldsymbol{x}, N_{\boldsymbol{v}})}_{\text{Attraction}} \right)^{\boldsymbol{\mu}} \right) \tag{10}$$

Figure 3 provides a visualization of how the weight of a node impacts its influence in a hypothetical 2-motor, 0-sensor IDSM. To generate this figure, we manually added four nodes in relative proximity, and plotted the flow field generated by the influence of these nodes. Each plot shows the field with the weight of the rightmost node set to the value indicated at the top of the figure.

Figure 4 provides a visualization of the influence of many nodes. To generate this plot, we simulated a IDSM-controlled robot with two motors and no sensors. For 20 time-units we (externally) assigned its motor state ($\boldsymbol{m}_1, \boldsymbol{m}_2$) according to the following time-dependent equations,

$$\boldsymbol{m}_1 = 0.75 \cdot \cos\left(\frac{2\pi}{10}t\right); \boldsymbol{m}_2 = 0.75 \cdot \sin\left(\frac{2\pi}{10}t\right) \tag{11}$$

and then generated stream plots indicating the motor trajectories that would be taken if the IDSM were "frozen" at $t = 20$ (i.e., if the weights of nodes did not change and no new nodes were added). The left and center plots show how the velocity and attraction influences affect different sensorimotor states if the other influence were absent, and the rightmost plot shows the combination of the two influences. At $t = 30$, we randomized the two motor values to the state indicated by the star, and allowed the IDSM to control the motor states. The blue trajectory shows that the IDSM returned the robot to the motor behavior that it was externally forced to perform at the start of the trial. In the next section, we will see this capability of the IDSM in more detail.

3. EXPERIMENTS AND RESULTS

3.1. RECREATING PREVIOUS SENSORIMOTOR BEHAVIOR

To elaborate upon how the IDSM maintains a history of previous SM-trajectories and how it uses these records to recreate previously performed patterns of behavior, we now present a scenario involving a simple IDSM-controlled robot. In this scenario, the robot first undergoes a training phase, where it is driven to perform a specific behavior, and then a free action phase where the IDSM has control of the robots motors and it recreates the patterns of behavior performed during the training phase.

The robot is embedded in a one-dimensional environment with a single point light-source located at the origin. It has a single motor that allows it to move forward or backward and a single non-directional light sensor. The robot's velocity, $\dot{x}$, is equivalent to the state of its motor $m \in [-1, 1]$. The activation of the light sensor is inversely proportional to the square of the distance between the robot and the light according to the following equation $s = \frac{1}{1+x^2}$. The robot has one sensor and one motor, so its SM-space is two-dimensional.

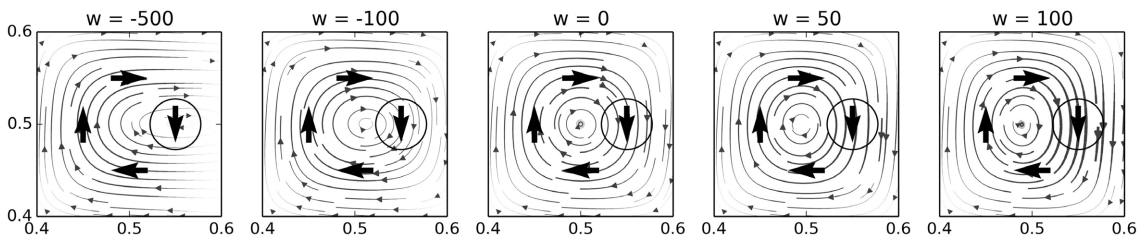

FIGURE 3 | Nodes with lower weights have less influence on system-dynamics. These plots show how the influence of a node decreases with its weight. Each plot shows the dynamics of the same in the same 2-motor, 0-sensor IDSM with four activated nodes, each given a weight (N_w) of 0, except for the circled node on the right, which has the weight indicated at the top of each plot.

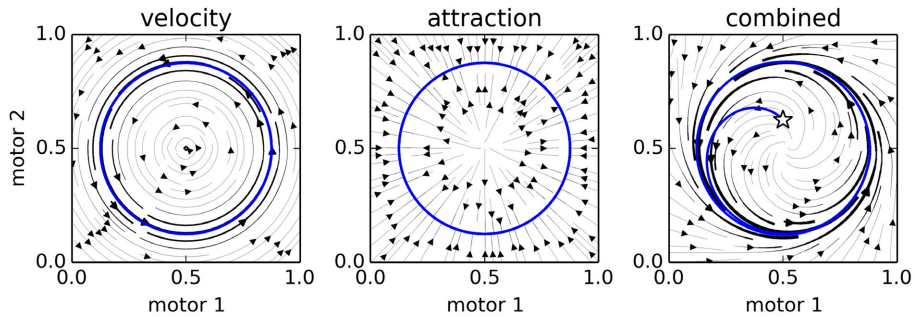

FIGURE 4 | Three snapshots of the 2-Motor IDSM as a fixed dynamical system. The left plot indicates the influence of the velocity term, the central plot indicates the influence of the attraction factor, and the right plot indicates the combination of the two. In the final plot, a randomly selected initial condition (star) is shown to have a trajectory (blue curve) that approaches the trained cycle of motor activity (gray circle).

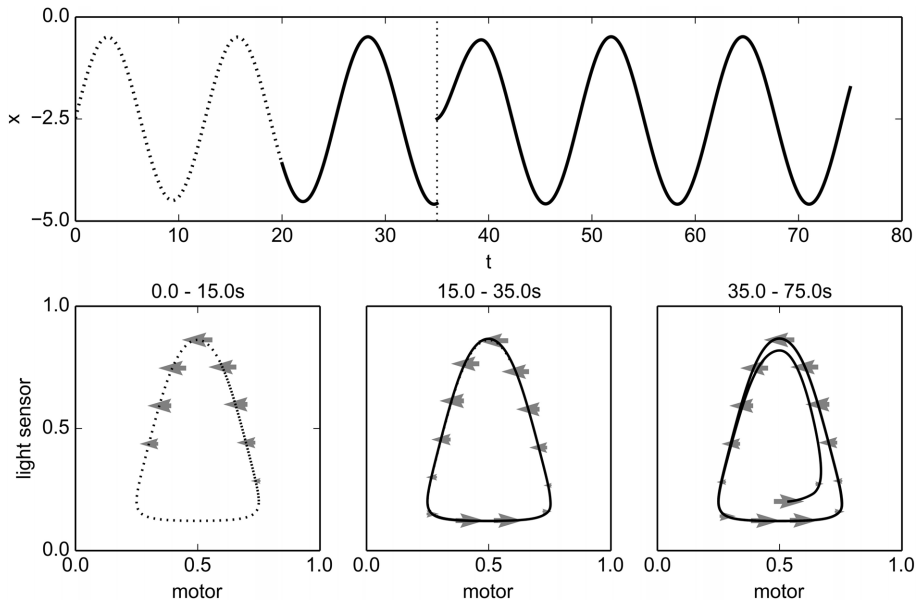

FIGURE 5 | Training and performance of an oscillatory behavior. The top plot shows the position of the robot, and the bottom three plots indicate SM-trajectories and the motor components of activated IDSM-nodes (arrows) for different time-periods in normalized SM-space. See main text for details.

We start with the robot located at $x = -2.5$. For the first 20 time-units of the simulation, the motor is not controlled by the IDSM, but is instead determined by the training controller, which sets the motor state according to the time-dependent equation $m = \cos(t/2)/2$. This causes the robot to move back and forth, but remain on one side of the light. The physical position and sensorimotor trajectory during this training phase are plotted as dotted curves in **Figure 5**. As the robot moves through the training trajectory, the IDSM adds nodes to its record, describing the change in SM-state for experienced SM-states. The motor component of activated nodes are shown as gray arrows in the SM-plots of **Figure 5**, with only every 25th node plotted for clarity.

At $t = 20$, the training phase ends, and we give control of the motors to the IDSM. We can see in **Figure 5** that the robot continues to perform a behavior that is very similar to the pattern of behavior experienced during the training regime, oscillating at approximately the same amplitude, frequency and distance from the light. How does this occur? During the training phase, several nodes were created describing how the SM-state changes for various encountered SM-states. After training ends and the IDSM takes control of the motors, the velocity-factor of these nodes causes the motors to change in response to the SM-state in the same way that they changed when in a similar SM-state experienced during training. Simultaneously, the attraction-factor pulls the system toward SM-states that it has experienced before. This latter influence attracts the system toward familiar SM-states so that potentially, if the system finds itself in an unfamiliar SM-state, it would modulate its motors in such a way that it is more likely to return to a familiar SM-state. It also can correct an SM-trajectory in the sense that when perturbations or deviations from the trained SM-trajectory occur, the attraction-factor can compensate for them, allowing for the pattern of activity to

recur (perhaps in a slightly different form and provided that the environment continues to allow the SM-trajectory) and thus the pattern of behavior is somewhat robust to varied environments. These influences of the attraction factor are demonstrated in the simulation at $t = 35$, when we relocated the robot to its starting location and the although after the perturbation the robot is at a new SM-state (see bottom-right plot in **Figure 5**), the robot rapidly returns to the trained behavior, oscillating at the same amplitude and frequency and distance from the light.

There are many possible patterns that could be trained and that would remain stable. During our experimentation we observed that the system could be trained to oscillate at a different distance from the light source, or to move in oscillations of larger or smaller magnitude (details not presented). However, the IDSM cannot be trained to re-enact *any* pattern of behavior. For instance, it would be impossible for the IDSM to recreate a behavior that varies completely independently of the SM-state. An example of this would be a training phase that consisted of oscillating at 33 Hz in front of the light at one amplitude for 10 s and then oscillating at the same frequency, but a different amplitude for the next 10 s. The switch between amplitudes is a function of time and it is independent of the sensorimotor-state, in that it does not always occur at a specific sensorimotor state, and that sensorimotor states where it does occur do not always correspond to a switch. Without a modification to the IDSM, such as the addition of a sensory-state variable that indicates the passage of time, the IDSM would be unable to recreate that behavior as the switch from one oscillation to the other could not be encoded into the IDSM. Several factors determine which patterns of behavior can be re-enacted and which can not: the update rules of the IDSM, the form of the environment and its relationship with the form of the body of the robot, i.e., how its motors change the robots interaction with its environment thereby influencing the activation of its sensors. If any of these were to change, for instance, if the light were mobile, or if there were no light at all, or if the robot were simulated as having inertia, etc., the set of possible stable trainable patterns would be different.

3.2. TRAINING FUNCTIONAL HABITS

In a further demonstration of the dynamical properties of the IDSM, we shall now show that when it is coupled to an environment through the sensors and motors of a simulated robot, it can be trained to have self-maintaining patterns of behavior ("habits") and that these habits can be functional, in the sense that they can accomplish a task. To do this, we shall use a slightly more complicated IDSM-controlled robot that is embedded in a two-dimensional spatial environment, with two directional light sensors and two independently driven motorized wheels. The motion of the robot is determined by the differential equations $\dot{x} = \cos(\alpha)(m_l + m_r)$; $\dot{y} = \sin(\alpha)(m_l + m_r)$; $\dot{\alpha} = 2(m_r - m_l)$, where x,y is the robots spatial position, $\alpha \in [-\pi, \pi]$ is the robots orientation and $m_l \in [-1, 1]$ and $m_r \in [-1, 1]$ are the robots left and right motor speeds. The robot's directional light sensors are located at $x + r \cdot \cos(\alpha + \beta)$,$y + r \cdot \sin(\alpha + \beta)$, where $r = 0.25$ is the robot's radius and $\beta = \pm\pi/3$ is the angular offset of the sensors from α, the heading of the robot (see **Figure 6**), and the activation of each sensor is determined by

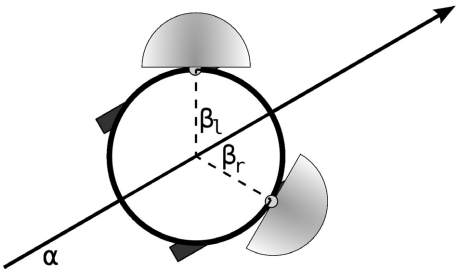

FIGURE 6 | Robot with two motors and two directional light sensors.

$$s = \frac{(\boldsymbol{b} \cdot ||\boldsymbol{c}||)^+}{1 + D^2}, \tag{12}$$

where $\boldsymbol{b} = [\cos(\alpha + \beta), \sin(\alpha + \beta)]$ is a unit vector indicating the direction that the sensor is facing, $\boldsymbol{c}$ is the vector from the sensor to the light, which is placed at $(x = 0, y = 0)$, and D is the distance from the sensor to the light. The arena is of width 4, with periodic boundary conditions. The robot has two motors and two sensors, and thus a four-dimensional sensorimotor space.

We used Braitenberg vehicle-inspired controllers (Braitenberg, 1986) to train the IDSM-controller to produce two different phototactic (light-seeking) behaviors and a photophobic behavior. The motor activity for these trained behaviors all involve a fairly direct motor response to sensory input. In the "simple-phototaxis" case, the connection is inverse and ipsilateral, resulting in a motion of the robot toward the light that slows to a stop as it approaches the light. The "sinusoidal-phototaxis" behavior, employs the same equations as simple-phototaxis, but with the addition of time-dependent sinusoidal functions that cause the robot to wiggle back and forth as it approaches the light. Finally, the "photophobic" behavior involves equations similar to those used in the simple-phototaxis case, but with contralateral rather than ipsilateral connections between sensors and motors. This results in a steady forward motion that turns away from the light whenever the robot approaches it. The equations below describe the target left and right motor values (χ_l, χ_r) given sensory input values (σ_l, σ_r) for the three behaviors, which are limited to lie in the range $[-1.0, 1.0]$ and then used to update the left and right motors (m_l, m_r) to approach these target values in a smooth transition according to Equation (19).

Simple phototaxis:

$$\chi_l = 1 - 1.5\sqrt{\sigma_l} \tag{13}$$

$$\chi_r = 1 - 1.5\sqrt{\sigma_r} \tag{14}$$

Sinusoidal-phototaxis:

$$\chi_l = 1 - 1.5\sqrt{\sigma_l} + \sin(2t)/2 \tag{15}$$

$$\chi_r = 1 - 1.5\sqrt{\sigma_r} - \sin(2t)/2 \tag{16}$$

Photophobia:

$$\chi_l = 1 - 1.5\sqrt{\sigma_r} \tag{17}$$

$$\chi_r = 1 - 1.5\sqrt{\sigma_l} \tag{18}$$

Motor update:

$$\frac{\mathrm{d}\boldsymbol{m}}{\mathrm{d}t} = (\boldsymbol{\chi} - \boldsymbol{m}) \tag{19}$$

Similar to the previous experiment, the motor-state of the robot is determined by one of the above sets of training equations for the first 100 time-units, and after this training phase, the robot enters a free-action phase, where the motor state is determined entirely by the IDSM. To train the robot from a variety of initial conditions and to demonstrate the system's behavior after training, every 50 time-units, the robot is relocated to a random position and assigned a random motor-state.

Figure 7, depicts the spatial trajectories of IDSM-controlled robots trained with the controllers described above. The square frames show the spatial trajectories of the robot during the time-period indicated at the top of the column, with the filled circles indicating the final position of the robot before a relocation took place. Plotted underneath these is a bar-chart indicating the mean distance of the robot from the light (located at the center of the arena). It is clear from evaluating the trajectories and the final location of the robots plotted in **Figure 7** that the IDSM has been substantially influenced by the pattern it was exposed to during training. Both the two forms of phototaxis training result in robots that tends to approach the light and the photophobe training results in a robot that tends to avoid it. Moreover, the way that these behaviors are performed is similar in the way that it accomplishes the behavior; compare the sinusoidal approach engendered by the sinusoidal-phototactic training agent to the more direct approach to the light performed by the agent trained with the simple-phototaxis algorithm.

In this scenario, we have the first clear example of a self-maintaining pattern of behavior, i.e., a habit. To understand why the pattern of behavior is self-maintaining, we must consider the weight of the nodes, what causes these weights to change (Equation 4), and how the influence of the node is affected by the weight [**Figure 1** and Equations (6–10)]. The weight of every node steadily degrades (according to the first term in Equation 4). This degradation can be counteracted by reinforcement which occurs when the SM-state is close to N_p, the node's position (second term of Equation 4). In the absence of reinforcement, the nodes created during training would have degraded to the point of being quite ineffectual and any new or reinforced nodes would override the originally trained behavior. But, the nodes influence behavior such that the SM-space near to those nodes is repeatedly revisited, thereby reinforcing the nodes such that even after a period of time longer than the non-reinforced effective "life-span" of the nodes, the nodes and the behavior itself persist.

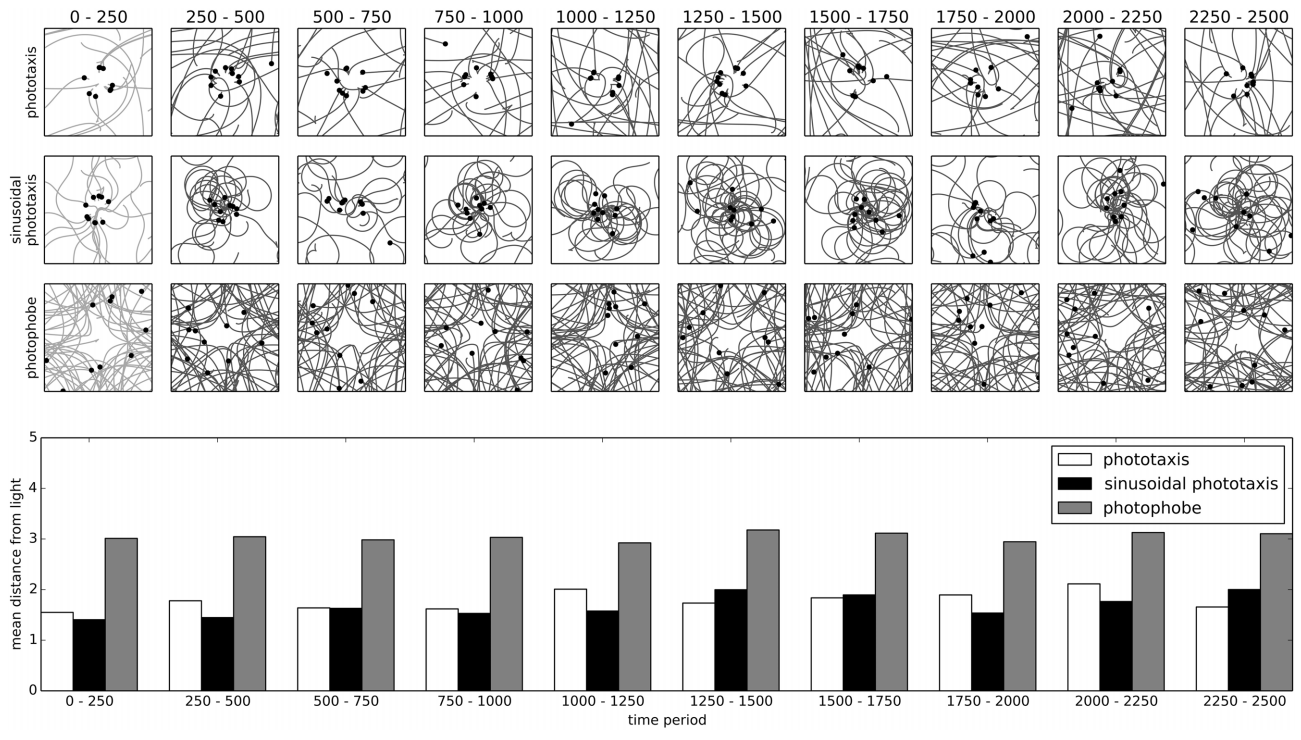

FIGURE 7 | Training of phototactic and photophobic behaviors and the long term evolution of each of the trained behaviors. The square frames show the spatial trajectories taken by a robot trained with the behavior indicated to the left of the row, during the time indicated at the top of the column. Robots are relocated to a random position and assigned a random motor-state every 50 time-units. The light is fixed at the center of the arena. The bar chart shows the mean distance of the robot from the light for each behavior during each indicated time-period.

In the long term, the IDSM-controlled robots fall into apparently robust behavior that do not show any signs of changing. There are many influences that determine which patterns of behavior can become self-maintaining habits, and that influence the robustness of these habits. These include many of the factors that we mentioned when discussing the factors that determine which patterns of behavior are trainable: the form of the IDSM, the presence of other habits, the form of the environment and the sensorimotor contingencies, etc. Determining the likely habits, or evaluating the robustness of an existing habit is complex task. In the next section we make a first step in this direction by investigating the habits that form from an randomly initialized IDSM.

3.3. EMERGENCE OF SELF-ORGANIZED HABITS

In this section, we show that with a randomly initialized IDSM, patterns of SM-activity form that interact with the environment in a self-stabilizing manner such that habits emerge. We shall show that these habits are not purely random behaviors, but relate to the environment, body and sensorimotor contingencies of the agent, in that they involve repetitive structured patterns that exploit agent-environment regularities.

In this experiment, the robot and environment are identical to those of the previous experiment. There is, however, no training phase. Instead, we randomly initialized the IDSM with 5000 nodes. These nodes were generated by performing 100 random walks in the 4-dimensional SM-space, each starting from a random location within the SM-space and with subsequent loci calculated according to the following equation, $\boldsymbol{l}_{i+1} = \boldsymbol{l}_i + \boldsymbol{r}$, where the components of $\boldsymbol{r}$ are selected from a flat distribution $[-0.05, 0.05]$ and where any components that would take l_i out of the normalized SM-volume are inverted. Nodes were added at each locus of the walk l_i with $N_{\boldsymbol{p}}$ set to l_i, $N_{\boldsymbol{v}}$ set to $I_{i+1} - I_i$, and $N_w = 0$. This random initialization of the IDSM is intended at this stage as minimal-assumption, stand-in for other mechanisms that would scaffold the formation of habits, such as reflexive behavior, or parental scaffolding, etc.

The experiment consists of a sequence of trials, where for each trial we observe the pattern of behavior that the robot falls into after having had its sensorimotor state and position randomized. Each trial starts with the robot being placed at a random location within the arena, with its motors set to random values selected from the flat distribution $[-1, 1]$. The IDSM then controls the motors of the robot for 100 time-units, and we record the sensorimotor and spatial trajectories. At the end of the experiment, we categorized the trials by hand, by comparing plots of the spatial trajectories taken during the last 25 time-units of the trial. This was accomplished by looking at the spatial trajectories plotted in **Figure 8A** and selecting by hand which trajectories seemed similar to each other. Five categories were identified, and colored red, green, blue, magenta and cyan. **Figures 8B** and **9** show the sensorimotor trajectories for the same trials as plotted in **Figure 8A**.

From the randomly initialized IDSM, self-maintaining patterns of behavior emerge, where the robot repeats behavioral motifs such as the square-with-rounded-corners motion of the robot around the light seen in red in **Figure 8A**. These patterns

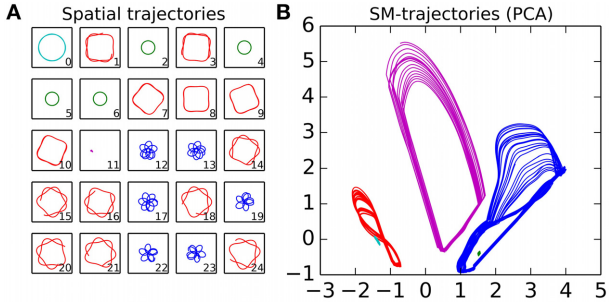

FIGURE 8 | Spatial and sensorimotor trajectories of habits that have emerged from a randomly initialized IDSM. The spatial plots (Plot **A**) indicate the spatial trajectories taken by the agent during the last 25% of the trial indicated in the lower right corner. Plot **(B)** shows a PCA dimensional reduction projection of the sensorimotor trajectories for these same trajectories, with colors used to group those trials that have a similar spatial trajectory.

are repeated and although they take their form in part from the random initialization of the nodes, they are not entirely random in that they relate to the environment. Notice, for instance, how each of the spatial trajectories keep the light within a fixed range of distances. The agent plotted in **Figures 8**, **9** has a set of habits that keep it close to the light, but other randomly initialized agents had one or more habits that kept it away from the light, or a set of habits where some habits kept the robot close to the light and other(s) kept it away from the light.

Habits are not always attractors in the IDSM plus body plus world system. Or, put another way: although the robot does sometimes fall into self-maintaining patterns of behavior that will last forever, there are also habits of repetitive behavior that naturally transition into another habit. For instance, in a randomly initialized IDSM (not plotted) we have observed behaviors where the robot turns in a tight loop, but each time through the loop, moves slightly closer to the light. Eventually, due to the motion toward the light, the robot enters a new region of SM-space, and a different set of nodes, perhaps a habit, take over.

4. DISCUSSION

4.1. HABITS AS SELF-SUSTAINING SENSORIMOTOR STRUCTURES

Following the tradition of defining life in terms of self-organized autonomous processes (Varela, 1979; Maturana and Varela, 1980; Kauffman, 2000; Ruiz-Mirazo and Moreno, 2004; Egbert et al., 2009, 2010) we have used our computational model to develop and investigate a view of habits, seen as self-maintaining patterns of behavior that share properties in common with the self-maintaining metabolic chemistry of living systems. Both habits and metabolism are self-maintaining, precarious, dissipative structures that rely upon cyclic processes to persist and, in both cases, the processes of self-maintenance are contingent upon the existence of an appropriate environment. Specifically, metabolism (understood as a network of far-from-equilibrium chemical reactions) relies upon an external energy-matter gradients and habits rely upon sensorimotor-contingency structures. The environment makes possible the necessary flow of matter and energy for dissipative chemical organizations. Similarly, it is

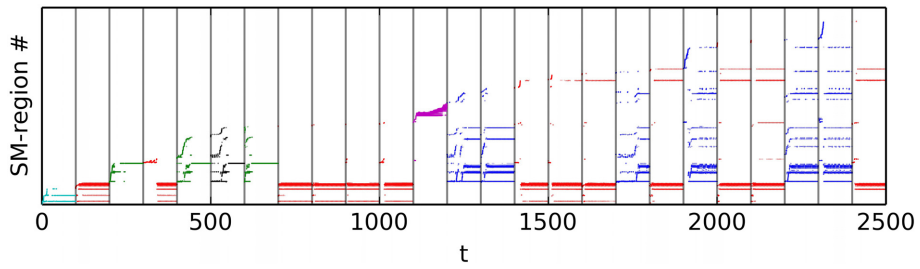

FIGURE 9 | Exploration and re-visitation of sensorimotor regions in habits that have emerged from a randomly initialized IDSM. To generate this alternative view of the sensorimotor trajectories displayed in **Figure 8**, we subdivided the SM-space into a $10 \times 10 \times 10 \times 10$ lattice and assigned a region ID number to each hypercube in order that they were visited. We then plot the region ID number of the current SM-state against time. Colors correspond to those used in **Figure 8**.

the environment that provides the structure for the sensorimotor flow that is necessary for the maintenance of habits. Where basic autonomy is made of an organized set of dissipative, far-from-equilibrium chemical reactions (Ruiz-Mirazo and Moreno, 2004), cognitive autonomy is made of habits (Barandiaran, 2007, 2008). The habits are dissipative structures, not in the thermodynamic sense (there are no thermodynamics in the model) but in the closely related dynamical systems sense that the IDSM dynamics are irreversible or non-conservative (Nicolis and Prigogine, 1989). This is clear when we recognize that any existing habit only persists via processes of reinforcing re-enactment of the pattern of behavior. In the absence of this, all of the nodes in the IDSM degrade and all patterns eventually cease to exist. Similar to how Benard-cells disappear when a source of heat is removed, habits disappear when the enactment of behavior is prevented. In this sense, like chemical and physical dissipative systems are thermodynamically open, the IDSM and the structures that are therein created are open to a "sensorimotor flow" that they, together with the structure of body and environment, make possible.

In our model, the formation of new nodes and their modification and reinforcement, is determined by the system's behavior in an environment. Structured collections of nodes are reinforced while others cease to have influence and thus, habits emerge and are sustained by the behavior they create, in a circular self-organized manner. It is in this sense that habits can be considered to be some kind of mental or sensorimotor life-forms. And thus, to say it with Di Paolo, "[w]e may invest our robots not with *life*, but with the mechanisms for acquiring a *way of life*, that is, with habits." (Di Paolo, 2003, p. 32).

In the node-based IDSM, a habit should not be confused with the collection of nodes that partially constitutes it. A habit also includes the repeated enactment of the sensorimotor correlations, for the nodes are only part of the self-maintaining system, i.e., part of the network of processes that maintains and is maintained by their influence. This is made evident when we observe that if a pattern of behavior is environmentally (or historically, due to the paths taken by the robot) prevented from being performed, then the nodes would not be reinforced, the behavior would not be recreated and the whole self-maintaining system that is the habit would cease to exist. The habit does not stand "purely in the head," but its conditions for existence extend out into body and environment, involving internal mechanisms (modeled as nodes in the IDSM) and interaction with the world through sensorimotor behavior.

The formation and conservation of habits, on our model, is implicitly constrained by several factors: (i) the properties of the IDSM; (ii) sensorimotor contingencies, which are in turn determined by the form of the environment and the robot's embodiment; (iii) the historical process and current structure of the habit; and (iv) the history and present form of *other* habits. The first two of these are fixed, in the sense that they are predefined and static throughout the course of a simulation. The last two are emergent and dynamic. Put another way: in most cases, habits are constrained but not determined by factors (i) and (ii); for almost any IDSM and any sensorimotor environment (Buhrmann et al., 2013), there are many possible meta-stable forms that a habit could take. But, once a habit has formed, the set of possible future, or concurrent habits shrinks. Again, this is reminiscent of a untouched pasture where, as animals walk through it, paths are carved in the grass, decreasing the variety of paths taken in the future.

The phototaxis training experiment (**Figure 7**), where the history of the agent influences its long term future, shows how the habits in the IDSM are historical processes. The IDSM is deterministic, and yet when coupled to an embodied robot situated in a minimal environment, it provides us with a model of a rich form of behavioral development where *the present actions of the robot are intricately and richly influenced by a long and detailed history of its sensorimotor flow*. It is not just that the robot will turn left as it approaches the light if it has done that in the past, but more that the behaviors that it has performed in the distant past have influenced and constrained the behaviors that has performed in the more recent past, which influence the behaviors it performs now, and which habits will form or be destroyed, etc.

Instead of the mind relying upon computations of internal representations of the external world, we can see how interesting behaviors can emerge through a sort of "resonance" between the plastic IDSM, the robot's body and the environment. To be precise, in our model, the agent is not resonating with the environment in the conventional sense of the term "resonance" as applied to oscillation. Yet the interaction between the IDSM and the embodied, situated robot can be considered as a kind

of resonant relationship, where complex patterns of behavior dynamically adapt until they are entrained with the environment through reliable interactions; and we see how an agent can accomplish adapted structured behavior without any isomorphic mapping or representational relationship with the environment. In this sense we can see habits as adapted to their embodied habitats.

Just as there are a variety of ways in which living organisms can be more or less adaptive, habits can also have different degrees of adaptivity. Here we do not refer to the influence of the habit upon the adaptivity of the robot that it controls, but rather the adaptability of the habit *itself*, i.e., the habit's ability to persist in a variety of conditions. Some habits may be mildly adaptive, increasing the chances that they will reoccur in the future. Others might be more impressively adaptive, modifying parts of their organization such that they persist even when faced with radical changes in their environment, but we have not yet explored the adaptivity of habits in detail and this remains future work.

Habits can be beneficial or detrimental to the "host" organism upon which they operate. And they can also influence the viability of *other* habits. Just as is the case in ecosystems of biological organisms, some habits might compete, while others might be symbiotic, each increasing the chances of the other's persistence. How could this occur? In the most simple case, the presence of a habit can influence what other habits can or will emerge and what form they will take. For instance, a behavior that prevents the robot from ever approaching the light will prevent it from exploring the SM-states where the light sensor is highly activated, preventing those habits from forming. Similarly, the *absence* of a habit can be necessary for certain other habits to form.

The question remains open as to whether a single habit is sufficient to speak of genuine autonomy and agency in the sensorimotor domain or a full self-regulating ecology of interrelated habits is required instead (Barandiaran, 2007, 2008). Further variations and experiments with more complex environments, higher dimensional IDSMs or the addition of internal variables into the IDSM can be used to make progress in these and other directions. Still, the habits in the model share properties with real habits, and they bear some significance upon human neuroscience and the notions of sensorimotor identity, autonomy, agency, and, ultimately, freedom.

Most of the contemporary attention on human freedom is put on the deliberative capacity of humans to represent the consequences of their actions and take decisions accordingly. Within this standard and widespread position, habits, as the residue of the behaviorist conception of mind, are found marginalized as mere stimulus-triggered response probabilities, that at best play a supportive role to our more impressive rational and deliberative capacities. In the view taken here, the embodied brain is seen as supporting a complex ecology of habits that can grow in complexity, adaptivity and coherence in a path-dependent historical manner, where the behavioral identity of the agent (the topology of the IDSM) is both the cause and effect of the behavior. Habits emerge and are sustained by the behavior they create, in a circular self-organized manner, similar to other self-organizing aspects of life. Our model opens up a way to re-position habits, understood as sensorimotor neuro-ecological life-forms, back at the center of the debate over our autonomy and agency.

4.2. A FRAMEWORK FOR HABIT MODELING AND HABIT-BASED ROBOTICS

In this paper we have only just started to investigate the various factors that influence the form of the habits. A great deal of work remains to understand how the form of the environment, or interactions with other agents can scaffold the creation of new habits or modification of existing habits, together with the inclusion of aditional, non-sensorimotor, dimensions to the IDSM. As part of the ALIZ-E project, we are currently investigating how habits can be influenced by essential variables (such as blood-sugar) (Ashby, 1952), and in particular how homeostatic adaptation can be accomplished in a system involving essential variables, hormonal regulation and habit-based behavior (Avila-Garcia and Cañamero, 2004; Egbert and Cañamero, 2014). The goal is to better understand how good and bad habits can form, and to look into methods for helping to transform unhealthy habits into healthy habits. We are looking into questions such as: How could habit formation be biased to perform behavior that performs well at maintaining blood sugar within a healthy range? How do unhealthy habits form and how can they be restructured into healthy habits, in particular in the context of the behavioral management of diabetes (Lewis and Cañamero, 2014)? How does environment modulate the formation of habits? In particular how can interaction with other agents scaffold the formation of new habits and the modification of existing habits? and how might fixed "instinctual" or "reflexive" behaviors scaffold the formation of habits? At this stage, we are intentionally avoiding the investigation of explicit reward or punishment mechanisms. We are instead focusing on how the form of the IDSM, body (sensors and motors) and world result in particular patterns of behavior being more or less likely to self-stabilize into habits.

There also remains a great deal of work to be done to better understand the influence of the model parameters and alternative designs to the IDSM. To carry this out it will be necessary to develop new measures and visualization tools for categorizing and describing habits. In this paper we investigated IDSM systems with two and four SM-dimensions. As the number of SM-dimensions grows, it should be increasingly difficult for the system to return to previously experienced SM-states. Alternative SM-distance metrics may help and perhaps, the influence of sensorimotor contingencies, reliable structures in the environment, and the influence of habits upon subsequent habit formation may mean that this is not be as big a problem as it initially appears. Otherwise, this challenge may be addressed by using more sophisticated plasticity rules. For instance, in the current implementation, although each node stores the SM-velocity, only the motor components of N_v are used. In future extensions, the sensory components could also be used in a more sophisticated reinforcement rule, where nodes that cause changes in sensory state similar to change experienced in the past are more reinforced than those that do not. It will also be interesting to investigate how the scaling of the SM-dimensions can be accomplished in a self-regulatory manner. Finally, it remains to be explored how

additional non-sensorimotor dimensions can be added to the IDSM, together with delayed reinforcement and richer timescale deformations.

This research connects to, by now, classical developments in the neuroscience of habits, where habits are seen as purely stimulus-triggered responses that are not modulated or modified in response to a behavior's outcome (Dickinson, 1985). The paradigmatic example is the result of behavioral training of a rat toward water sources where the salt deficient rodent is incapable of selecting the route to the most saline water and selects the most familiar or repetitive route instead. This is contrasted with *action-oriented behavior*, where the performance of an action is sensitive to different motivational values (e.g., salt deficiency) or revaluations of the outcome of the behavior and manipulations of the contingency that the action will have the desired outcome (e.g., lower or more variable probability of finding water in one of the routes). According to two recent reviews of habits Yin and Knowlton (2006); Graybiel (2008), these two operationally defined categories of behavior (habitual, stimulus-response or S-R, and instrumental, action-outcome sensitive or A-O) have been thought of as being supported by different brain regions, both in rodents (Balleine and Dickinson, 1998) and humans (Valentin et al., 2007), that underlie two different forms of learning. Breaking with this view, recent developments in experimental neuroscience give reason to believe that these two systems are more integrated than previously thought, and moreover that it is not clear how they (or their underlying mechanisms) are related to one another. The neuroscience has opened the door to the more not-yet-understood interaction between habits and A-O behavior and therefore also for the possibility that habits are not just about "off-loading cognitive work," but might have an ongoing influence on even action-oriented behaviors. Our dynamical sensorimotor model, unlike discrete action-selection or S-R-probabilities based models, allows us to further investigate these ideas. A mesoscopic level of modeling, where dynamic sensorimotor reinforcement (as we modeled here) coupled to additional dimensions and internal dynamics such as blood-sugar levels (Egbert and Cañamero, 2014), might help exploring the transition and interaction between S-R and A-O forms of behavior. In this sense, the habit-based robotic modeling framework we presented here might help neuroscientist to fill the need for " (. . .) dynamic models in which activity can occur simultaneously in multiple cortico-basal ganglia loops, not move in toto from one site to another, and models in which, as the learning process occurs, activity patterns change at all these sites." (Graybiel, 2008, pp. 337–389).

5. CONCLUSIONS

In this paper we have provided a proof of concept and a modeling framework for a new conception of habits. We have introduced the very notion and one possible instance of an *iterant deformable sensorimotor medium* and shown its capacity as a medium that supports sensorimotor imprinting and the spontaneous formation, transformation and evolution of self-maintaining patterns of behavior, i.e., *habits*. Unlike previous habit modeling attempts, we opted for a mesoscopic, continuous-time dynamic modeling, where habits do not presuppose a specific set of discrete stimuli to be linked (by reinforcement or repetition) to a given probability of triggering a specific response (from a set of available actions). As a result, it is the fine-grained sensorimotor contingency dynamics (that the embodiment and history of the agent make possible) that define the emergence and self-maintenance of habits, giving rise to a complex morphology of habits within a specific body and world. This modeling framework affords for a deeper conception of habits, where mental life emerges from a sensorimotor substrata that makes possible the development of an increasingly complex ecology of self-sustaining *sensorimotor* life-forms.

There have been calls for non-computationalist and non-intellectualist approaches to mind and even an explicit call for habit-based robotics (Noë, 2009, pp. 97–98). We believe that further development of the IDSM modeling framework could assist on bringing forth a set of theoretical suggestions for enactive approaches to human cognition and neuroscience (Varela et al., 1974; Di Paolo, 2003; Barandiaran, 2004; Noë, 2006; Thompson, 2010). In contrast to standard engineering principles (where functionally specific robotic performance is the goal) or classical neuro-cognitive assumptions (where the use of internal representations is the dominating modeling assumptions), habit-based robotics (in the sense we explored along this paper) can open up the way to target behavioral phenomena that often fall out of general attention: history dependent identity formation, the mutual shaping between an agent's sensorimotor identity and the sensorimotor environment it inhabits, etc.

Piaget's approach to cognitive development considered higher cognitive capacities to stir from the tendency to maximally equilibrate sensorimotor habits, progressively stratified in the form of schemas (see Di Paolo et al., 2014 for a dynamical interpretation of these ideas). It shows that habits need not be understood as opposed to higher cognitive capacities but as their pre-condition and continuous support. Human freedom is not only about the deliberative reflexion upon our actions, but about their re-inscription, through practice and repetition, into the "invisible" web of habits that constitutes our identity. Developing a modeling framework that is suited to this conception of habit puts us closer to attain a deeper conception of human freedom and identity, one that acknowledges habits as the necessary origin of neuro-cognitive capacities and as the necessary end of incorporating our virtuous ways of coping with the world back into the second nature of habitual behavior.

ACKNOWLEDGMENTS

Matthew Egbert's contributions to this research were funded by the European Commission as part of the ALIZ-E project (FP7-ICT-248116). Xabier Barandiaran's work was funded by the eSMCs: Extending Sensorimotor Contingencies to Cognition project, FP7-ICT-2009-6 no: IST-270212. Research project "Autonomia y Niveles de Organizacion" financed by the Spanish Government (ref. FFI2011-25665) and IAS-Research group funding IT590-13 from the Basque Government. The authors would also like to thank Eran Agmon as well as Lola Cañamero and the other members of the Embodied Emotion, Cognition and (Inter-)Action Lab for discussions of the research presented above. The opinions expressed are solely the authors.

REFERENCES

Ashby, W. R. (1952). *Design for a Brain: The Origin of Adaptative Behaviour*. 2nd Edn. London: J. Wiley

Avila-Garcia, O., and Cañamero, L. (2004). "Using hormonal feedback to modulate action selection in a competitive scenario," in *Proceedings of the 8th International Conferece on Simulation of Adaptive Behavior (SAB 2004)*, (Cambridge, MA: MIT Press), 243–252.

Balleine, B. W., and Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37, 407–419. doi: 10.1016/S0028-3908(98)00033-1

Barandiaran, X. (2004). "Behavioral adaptive autonomy. A milestone in the alife route to AI," in *Proceedings of the 9th International Conference on Artificial Life* (Cambridge, MA), 514–521.

Barandiaran, X., and Moreno, A. (2006). "ALife models as epistemic artefacts," in *Artificial Life X : Proceedings of the Tenth International Conference on the Simulation and Synthesis of Living Systems*, eds L. Rocha, L. Yaeger, M. Bedau, D. Floreano, R. Goldstone and A. Vespignani (Cambridge, MA: The MIT Press Bradford Books), 513–519.

Barandiaran, X. E. (2007). "Mental life: conceptual models and synthetic methodologies for a post-cognitivist psychology," in *The World, the Mind and the Body: Psychology after Cognitivism*, eds B. Wallace, A. Ross, J. Davies, and T. Anderson (Exeter: Imprint Academic), 49–90.

Barandiaran, X. E. (2008). *Mental Life: a Naturalized Approach to the Autonomy of Cognitive Agents.* PhD Thesis, Gipuzkoa, Spain: University of the Basque Country (UPV-EHU), Donostia - San Sebastián.

Barandiaran, X. E., and Di Paolo, E. A. (2014). A genealogical map of the concept of habit. *Front. Hum. Neurosci.* 8:522. doi: 10.3389/fnhum.2014.00522

Bedau, M. A. (1997). Emergent models of supple dynamics in life and mind. *Brain Cogn.* 34, 5–27. doi: 10.1006/brcg.1997.0904

Beer, R. D. (2003). The dynamics of active categorical perception in an evolved model agent. *Adapt. Behav.* 11, 209–243. doi: 10.1177/1059712303114001

Braitenberg, V. (1986). *Vehicles: Experiments in Synthetic Psychology*. Cambridge, MA: MIT Press.

Brooks, R. A. (1991). Intelligence without representation. *Art. Intell.* 47, 139–159. doi: 10.1016/0004-3702(91)90053-M

Buckingham, H. W., and Finger, S. (1997). David hartley's psychobiological associationism and the legacy of aristotle. *J. Hist. Neurosci.* 6, 21–37. doi: 10.1080/09647049709525683

Buhrmann, T., Paolo, E. A. D., and Barandiaran, X. (2013). A dynamical systems account of sensorimotor contingencies. *Front. Psychol.* 4:285. doi: 10.3389/fpsyg.2013.00285

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, MA: The MIT Press.

Damasio, A. R. (2003). *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*. Orlando, FL: Harcourt.

Dennett, D. (1994). Artificial life as philosophy. *Art. Life* 1, 291–292. doi: 10.1162/artl.1994.1.3.291

Dezfouli, A., Balleine, B. W., Dezfouli, A., and Balleine, B. W. (2012). Habits, action sequences and reinforcement learning, habits, action sequences and reinforcement learning. *Eur. J. Neurosci.* 35, 1036–1051. doi: 10.1111/j.1460-9568.2012.08050.x

Di Paolo, E. A. (2000). "Homeostatic adaptation to inversion of the visual field and other sensorimotor disruptions," in *From Animals to Animats 6: Proceedings of the 6th International Conference on the Simulation of Adaptive Behavior* (Cambridge, MA), 440–449.

Di Paolo, E. A. (2003). "Organismically-inspired robotics: homeostatic adaptation and teleology beyond the closed sensorimotor loop," in *Dynamical Systems Approaches to Embodiment and Sociality*, eds K. Murase, and Asakura (Adelaide:Advanced Knowledge International), 19–42.

Di Paolo, E. A., Barandiaran, X. E., Beaton, M., and Buhrmann, T. (2014). Learning to perceive in the sensorimotor approach: Piaget's theory of equilibration interpreted dynamically. *Front. Hum. Neurosci.* 8:551. doi: 10.3389/fnhum.2014.00551

Di Paolo, E. A., Noble, J., and Bullock, S. (2000). "Simulation models as opaque thought experiments," in *Seventh International Conference on Artificial Life*, (Cambridge, MA:MIT Press), 497–506.

Dickinson, A. (1985). Actions and habits: The development of behavioural autonomy. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* 308, 67–78. doi: 10.1098/rstb.1985.0010

Edelman, G. M. (2007). Learning in and from brain-based devices. *Science* 318, 1103–1105. doi: 10.1126/science.1148677

Egbert, M., and Cañamero, L. (2014). "Habit-based regulation of essential variables," in *Artificial Life 14: Proceedings of the Fourteenth International Conference on the Synthesis and Simulation of Living Systems*, eds H. Sayama, J. Rieffel, S. Risi, R. Doursat, H. and Lipson (Cambridge, MA: MIT Press), 168–175.

Egbert, M. D., Barandiaran, X. E., and Di Paolo, E. A. (2010). A minimal model of metabolism-based chemotaxis. *PLoS Comput. Biol.* 6:e1001004. doi: 10.1371/journal.pcbi.1001004

Egbert, M. D., Di Paolo, E. A., and Barandiaran, X. E. (2009). "Chemo-ethology of an adaptive protocell: sensorless sensitivity to implicit viability conditions," in *Advances in Artificial Life, Proceedings of the 10th European Conference on Artificial Life, ECAL* (Berlin: Springer), 242–250.

Engel, A. K., Maye, A., Kurthen, M., and Knig, P. (2013). Where's the action? the pragmatic turn in cognitive science. *Trends Cogn. Sci.* 17, 202–209. doi: 10.1016/j.tics.2013.03.006

Freeman, W. J. (2000). *Neurodynamics: An Exploration in Mesoscopic Brain Dynamics: An Exploration in Mesoscopic Brain Dynamics*. Berlin: Springer. doi: 10.1007/978-1-4471-0371-4

Freeman, W. J. (2001). *How Brains Make Up Their Minds*. 1st Edn. New York, NY: Columbia University Press.

Froese, T., and Ziemke, T. (2009). Enactive artificial intelligence: investigating the systemic organization of life and mind. *Art. Intel.* 173, 466–500. doi: 10.1016/j.artint.2008.12.001

Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. *Annu. Rev. Neurosci.* 31, 359–387. doi: 10.1146/annurev.neuro.29.051605.112851

Grey Walter, W. (1950). An imitation of life. *Sci. Am.* 182, 42–45. doi: 10.1038/scientificamerican0550-42

Hartley, D. (1749). *Observations on Man, his Frame, his Duty, and his Expectations*. London: S. Richardson.

Hebb, D. (1949). *The Organization of Behavior: A Neuropsychological Theory*. New York, NY: Psychology Press.

Helbing, D., Keltsch, J., and Molnár, P. (1997). Modelling the evolution of human trail systems. *Nature* 388, 47–50. doi: 10.1038/40353

Hull, C. L. (1950). Behavior postulates and corollaries–1949. *Psychol. Rev.* 57, 173–180. doi: 10.1037/h0062809

Hurley, S. L., and Noë, A. (2003). Neural plasticity and consciousness: reply to block. *Trends Cogn. Sci.* 7:342. doi: 10.1016/S1364-6613(03)00165-7

Hutto, D. D., and Myin, E. (2012). *Radicalizing Enactivism: Basic Minds Without Content*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9780262018548.001.0001

James, W. (1890). *The Principles of Psychology*. New York, NY: Cosimo, Inc. doi: 10.1037/11059-000

Kauffman, S. (2000). *Investigations*. New York, NY: Oxford University Press.

Kelso, J. A. S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Cambridge, MA: MIT Press.

Lewis, M., and Cañamero, L. (2014). "An affective autonomous robot toddler to support the development of self-efficacy in diabetic children," in *Accepted at the 23rd Annual IEEE International Symposium on Robot and Human Interactive Communication (IEEE RO-MAN 2014)*.

Maturana, H. R., and Varela, F. J. (1980). *Autopoiesis and Cognition: The Realization of The Living*. Dordrecht: Springer. doi: 10.1007/978-94-009-8947-4

Nicolis, G., and Prigogine, I. (1977). *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order Through Fluctuations*. New York, NY: Wiley.

Nicolis, G., and Prigogine, I. (1989). *Exploring Complexity: An Introduction*. New York, NY: W H Freeman.

Noë, A. (2006). *Action in Perception*. Cambridge, MA: The MIT Press.

Noë, A. (2009). *Out of Our Heads: Why You Are Not Your Brain, and Other Lessons from the Biology of Consciousness*. New York, NY: Farrar, Straus and Giroux.

O'Regan, J. K., and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* 24, 939–1031. doi: 10.1017/S0140525X01000115

Pfeifer, R., Lungarella, M., and Iida, F. (2007). Self-organization, embodiment, and biologically inspired robotics. *Science* 318, 1088–1093. doi: 10.1126/science.1145803

Ruiz-Mirazo, K., and Moreno, A. (2004). Basic autonomy as a fundamental step in the synthesis of life. *Art. Life* 10, 235–259. doi: 10.1162/1064546041255584

Ruppin, E. (2002). Evolutionary autonomous agents: A neuroscience perspective. *Nat. Rev. Neurosci.* 3, 132–141. doi: 10.1038/nrn729

Shapiro, R. (2006). Small molecule interactions were central to the origin of life. *Q. Rev. Biol.* 81, 105–126. doi: 10.1086/506024

Steels, L. (1993). The artificial life roots of artificial intelligence. *Art. Life* 1, 75–110. doi: 10.1162/artl.1993.1.1_2.75

Sutton, R. S., and Barto, A. G. (1998). *Reinforcement Learning*. Cambridge, MA: MIT Press.

Szathmáry, E. (2000). The evolution of replicators. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* 355, 1669–1676. doi: 10.1098/rstb.2000.0730

Thompson, E. (2010). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge, MA: Belknap.

Thompson, E., and Varela, F. J. (2001). Radical embodiment: neural dynamics and consciousness. *Trends Cogn. Sci.* 5, 418–425. doi: 10.1016/S1364-6613(00)01750-2

Thorndike, E. (1911). *Animal Intelligence: Experimental Studies*. New York, NY: Transaction Publishers.

Valentin, V. V., Dickinson, A., and ODoherty, J. P. (2007). Determining the neural substrates of goal-directed learning in the human brain. *J. Neurosci.* 27, 4019–4026. doi: 10.1523/JNEUROSCI.0564-07.2007

Varela, F., Maturana, H., and Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *Biosystems* 5, 187–196. doi: 10.1016/0303-2647(74)90031-8

Varela, F. J. (1979). *Principles of Biological Autonomy*. New York, NY: North Holland.

Virgo, N. D. (2011). *Thermodynamics and the Structure of Living Systems*. Thesis, Brighton: University of Sussex.

Yin, H. H., and Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nat. Rev. Neurosci.* 7, 464–476. doi: 10.1038/nrn1919

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 April 2014; accepted: 16 July 2014; published online: 08 August 2014.
Citation: Egbert MD and Barandiaran XE (2014) Modeling habits as self-sustaining patterns of sensorimotor behavior. Front. Hum. Neurosci. ***8****:590. doi: 10.3389/fnhum.2014.00590*
This article was submitted to the journal Frontiers in Human Neuroscience.

frontiers in Human Neuroscience

GENERAL COMMENTARY
published: 20 April 2015
doi: 10.3389/fnhum.2015.00209

Corrigendum: Modeling habits as self-sustaining patterns of sensorimotor behavior

Matthew D. Egbert[1]* *and Xabier E. Barandiaran*[2,3]

[1] *Embodied Emotion, Cognition and (Inter) Action Lab, School of Computer Science, University of Hertfordshire, Hatfield, UK,*
[2] *Department of Philosophy, University School of Social Work, UPV/EHU, University of the Basque Country, Spain,*
[3] *IAS-Research Center for Life, Mind, and Society, UPV/EHU University of the Basque Country, Spain*

Keywords: sensorimotor, self-maintaining patterns-of-behavior, mental-life, habits, meso-scale modeling

A corrigendum on

Modeling habits as self-sustaining patterns of sensorimotor behavior
by Egbert, M. D., and Barandiaran, X. E.(2014) Front. Hum. Neurosci. 8:590. doi: 10.3389/fnhum.2014.00590

In this article, Equation 8 is incorrectly written:

$$\Gamma(a, N_v) = a - a \cdot \frac{N_v}{||N_v||}, \tag{1}$$

The correct equation is:

$$\Gamma(a, N_v) = a - \left(a \cdot \frac{N_v}{||N_v||}\right) \frac{N_v}{||N_v||}, \tag{2}$$

where the right-hand-side of the equation represents the vector a with its component parallel to N_v removed.

In the paper, this function is used to calculate the "attraction" influence of nodes. The vector a is the difference between the node's position (N_p) and the current sensorimotor state (x), and Γ removes the component of a that is parallel to N_v.

All of the simulations presented in the original paper were performed with the correct evaluation of $\Gamma(a, N_v)$. This was only a typo in the manuscript.

OPEN ACCESS

Edited and reviewed by:
Javier Bernacer, University of Navarra, Spain

***Correspondence:**
Matthew D. Egbert, mde@matthewegbert.com

Received: *11 March 2015*
Accepted: *30 March 2015*
Published: *20 April 2015*

Citation:
Egbert MD and Barandiaran XE (2015) Corrigendum: Modeling habits as self-sustaining patterns of sensorimotor behavior. Front. Hum. Neurosci. 9:209. doi: 10.3389/fnhum.2015.00209

Acknowledgments

Matthew Egbert's contributions to this research were funded by the European Commission as part of the ALIZ-E project (FP7-ICT-248116). Xabier Barandiaran's work was funded by the eSMCs: Extending Sensorimotor Contingencies to Cognition project, FP7-ICT-2009-6 no: IST-270212. Research project "Autonomia y Niveles de Organizacion" financed by the Spanish Government (ref. FFI2011-25665) and IAS-Research group funding IT590-13 from the Basque Government. The authors would also like to thank Eran Agmon as well as Lola Cañamero and the other members of the Embodied Emotion, Cognition and (Inter-)Action Lab for discussions of the research presented above. The opinions expressed are solely the authors.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in
HUMAN NEUROSCIENCE

OPINION ARTICLE
published: 27 May 2014
doi: 10.3389/fnhum.2014.00341

Pre-dispositional constitution and plastic disposition: toward a more adequate descriptive framework for the notions of habits, learning and plasticity

Francisco Güell*

Mind–Brain Group, Institute for Culture and Society, Universidad de Navarra, Pamplona, Spain
**Correspondence: fguell@unav.es*

Edited by:
Jose Angel Lombo, Pontifical University of the Holy Cross, Italy

Reviewed by:
Claudia Castro Batista, Federal University of Rio de Janeiro, Brazil

Keywords: habits, motor cognition, associative learning, mirror neurons, schizophrenia, autistic disorder, neural plasticity, neural migration

Behavioral studies and neurobiological models of mental illnesses can be used to inform theories of mind and action. In this paper I use specific aspects of some paradigmatic cases in order to establish what I consider to be a useful distinction for the analysis of human action and, more specifically, the delimitation of habitual action.

Patient SM is a well-known case of Urbach-Wiethe disease—one of only 300 cases reported in the reference literature—that was submitted to a decade of investigations (Adolphs et al., 1994, 2005). This syndrome, also known as lipoid proteinosis, produces dermatological lesions as well as calcifications in regions of the brain, often affecting the amygdaloid region (Siebert et al., 2003; Bahadir et al., 2006). While her basic perception, memory, and language skills are essentially normal, SM has nearly complete bilateral destruction of the amygdala, and her social behavior is indiscriminately trusting and friendly (Adolphs et al., 1994). Ten years of research showed an intriguing impairment in her ability to recognize fear in facial expressions, due to a lack of spontaneous fixation on the eyes when viewing faces (Adolphs et al., 2005). The research showed that in control patients, spontaneous fixation is directed principally to the eyes and the mouth, tracking the regions of the face that allow one to distinguish facial expressions. However, patient SM spontaneously focused on the nose, thereby missing necessary information for judging emotions. What is so interesting to point out is that when given explicit instructions ("look at the eyes of this person") SM had no problem focusing on the eyes and recognizing emotions but, surprisingly, after a decade of treatment, SM was not able to learn the habit of looking at the eyes of the face spontaneously.

Another case of interest for action theory can be found in studies of subjects in the autistic spectrum (Klin et al., 2003; Boria et al., 2009; Gallese, 2009; Kana et al., 2014), particularly studies of deficits in the functioning of the "mirror mechanism" (Antonietti, 2013). This deficit appears related to other deficits such as atypical visual processing and encoding of social stimuli, as well as imitative behavior and the ability to share attention (for review, see Gallese et al., 2013). Atypical brain development has been identified in cases of autism; specifically, the neural organization in areas involved in social cognition, facial expression, and facial recognition, as well as in areas associated with the mirror mechanism, appears to be related to the functional architecture that characterizes the atypical development of the autistic spectrum (Cauda et al., 2011; Gallese et al., 2013).

Another paradigmatic case that is informative for theories of mind-based action is schizophrenia (Synofzik et al., 2010; Leube et al., 2012; Mausbach et al., 2013). Multiple investigations suggest that schizophrenia and other neuropsychiatric disorders are associated with deficits in mirror neurons (Enticott et al., 2008; Mehta et al., 2013) and with interneuron dysfunction (Marin, 2012). For now, let us focus on the second dysfunction. Interneurons regulate the activity of pyramidal cells, largely through inhibitory mechanisms, and one of the functions of pyramidal cells is to maintain cerebral patterns associated with perception and memory. Migration of the interneurons during the development of the nervous system is fundamental to this function, as it determines the final positioning of the neurons, thereby establishing the basis for correct wiring of neural circuitry (Marín, 2013). It has been demonstrated that schizophrenics possess mutations in certain genes that affect the migration of the intercortical neurons during embryonic development (Valiente and Marin, 2010). In addition, for decades we have known that there is a correlation between schizophrenia and the alteration of visual perception and eye movements. In fact, just recently it has been shown that simple tests for the detection of abnormal eye movements can discriminate cases of schizophrenia from controls with exceptional accuracy (Benson et al., 2012).

Of the genes that are involved in the disrupted tangential migration of cortical neurons, NRG1, ERBB4, GRIN1, DISC1, and DTNBP1 (Marin, 2012), four of them are involved in the expression of molecules related with visual structures, and one of them is related to early visual processing: NRG1 is expressed in the cornea (Brown et al., 2004) and one of its mutations (rs3924999) affects spatial accuracy on the anti-saccade (AS) task (Schmechtig et al., 2010) and is associated with auditory P300 in schizophrenia (Kang et al., 2012);

ERBB4 (especially, rs7598440) is associated with 8 endophenotypes, including AS abnormality and smooth pursuit eye movement (Greenwood et al., 2011; Baea et al., 2012); two N-methyl-D-aspartate receptor subunits (NMDARs) encoded by the gene GRIN1 belong to the ionotropic glutamate receptors, which play key roles in neuronal communication in the retina (Fana et al., 2013). DTNBP1 affects the expression of Dysbindin (Benson et al., 2001), a protein whose deficit is associated with early visual deficits in schizophrenia (Donohoea et al., 2008).

Now, to understand the import of these and other cases, it is helpful to introduce a distinction between *constitutional pre-disposition* and *dispositional plasticity*. The neural architecture of SM pre-disposes her to look at faces in a certain way that resists training; likewise it has been shown that the constitutional pre-disposition of schizophrenia does not permit modification through the acquirement of new habits, and the same can be said for autistic individuals. In all of these cases, an atypical neural organization constitutionally predisposes the subject to perceive the world in a specific way. For example, it seems that schizophrenics cannot perceive the kinds of optical illusions normally perceived by healthy individuals. It is important to note that this pre-disposition does not need to be understood in terms of genetic determinism. Instead, studies point toward changes at the epigenetic level that affect neuronal plasticity (Fagiolini et al., 2009; Baker-Andresen et al., 2013), and there is an increasing number of examples of post-natal experiences that are affected at this level (Woldemichael et al., 2014). In addition, environmental factors play a crucial role in the formation of constitutional pre-disposition. For example, the lack of interneuronal migration is also caused by fetal exposure to cocaine (Valiente and Marin, 2010), and it has been demonstrated that individuals possessing susceptibility alleles in genes, such as DISC1, express psychiatric phenotypes only when these genetic variants occur in a propitious genome and when certain environmental pre-natal factors come into play (Abazyan et al., 2012).

Now let us turn to *dispositional plasticity*. This refers to the plastic dimension of the organic substrate, a plasticity that makes possible the modulation and function of biological structures. Activity and environmental stimuli continually modify the disposition of the subject, permitting the subject to obtain, inter alia, a certain tone of skin, to develop muscles or to "perfect" the organism on the most basic motor level through the repetition of a task. However, this perfection or specialization can also occur at the perceptual level (such as in the case of an oenophile) or at the level of higher functions (e.g., enhanced memory capacity). Habits, from this perspective, can be considered as actions that regulate the subject's disposition so as to facilitate a task and make others possible. In short, habits, and generally the repetition of tasks, adjust dispositions to act, and they do this thanks to the plastic character of the organic substrate.

The relative incapacity to regulate the constitutional pre-disposition (once consolidated) does not mean that there is no effective treatment for a subject with a specific constitutional pre-disposition. It is well known that many mental diseases can be treated but, according to the model presented here, such treatments do not modify the constitutional pre-disposition. Instead, what treatments do is compensate for the deficits of a certain constitution (for example, by supplying a neurotransmitter) or establish behavioral strategies that minimize effects on the subject's behavior. For instance, note that when SM looks at the eyes of a person because she is asked to do so, she is not modifying her pre-disposition; she is simply fixing her gaze on a certain point voluntarily, just as she would if she were to read this article. However, the fact that the constitutional pre-disposition cannot be regulated (regulation occurs only at the level of dispositional plasticity) does not mean that it cannot be damaged: continuous consumption of drugs can affect dispositional plasticity and, sooner or later (depending on the constitutional pre-disposition), damage the constitutional level as well.

The proposed distinction may be useful for explaining the risk factors related to cancer, as the constitutional predispositions for developing cancer are varied (indeed, we need to keep in mind that there are as many constitutional pre-dispositions as there are subjects). A constitutional pre-disposition for cancer does not necessarily mean that the subject is going to develop cancer; at the same time, the role of dispositional plasticity helps us to understand the importance of certain external factors which, by affecting the plasticity level, can function as the "trigger" for the appearance of cancer. The concept of constitutional pre-disposition also offers theoretical support for evidence that some ethnic groups are particularly vulnerable to certain diseases (Helgadottir et al., 2006; Ng et al., 2012) and may explain differences of organic reaction to certain therapies or drug use as a function of ethnicity (Marsha et al., 1999; Ono et al., 2013).

Let us consider a further example that shows how the proposed distinction can help to frame current debates over the genetic basis of behavior. Going back to the topic of mirror neurons, there is some debate as to whether the associated neural network is genetically inherited (innate) or if it is the product of associative learning. According to the terms just introduced, this choice between genes and learning is oversimplified; in addition to the genetic dimension and associative learning, we should also keep in mind the epigenetic dimension and the importance of environmental factors. Indeed, the importance of epigenetics has now been well established from the "evo-devo" standpoint (Ferrari et al., 2013). To address this added complexity, we can formulate the question more precisely as follows: does associative learning change the constitutional disposition or act on a level of plasticity, regulating the disposition of the subject?

In conclusion, I believe that the distinction between constitutional predisposition and dispositional plasticity offers a conceptual framework that can help place theories of mind and action into its developmental context, throw light on current debates, and offer an interpretative key for results arising from research. This distinction allows us to place the notion of habit within the broadest context of human action and thereby better understand its scope. In closing, it is important to clarify that the statement that some aspects of human behavior cannot be changed should not be taken as a deterministic argument against human freedom. It is merely an expression

of the universal and widely recognized experience of human limitations.

ACKNOWLEDGMENTS

This work was supported by Obra Social La Caixa. I am grateful to Institute for Culture and Society (Universidad de Navarra).

REFERENCES

Abazyan, B., Nomura, J., Kannan, G., Ishizuka, K., Tamashiro, K. L., Nucifora, F., et al. (2012). Prenatal interaction of mutant DISC1 and immune activation produces adult psychopathology. *Biol. Psychiatry* 68, 1172–1181. doi: 10.1016/j.biopsych.2010.09.022

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., and Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature* 433, 68–72. doi: 10.1038/nature03086

Adolphs, R., Tranel, D., Damasio, H., and Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature* 15, 669–672. doi: 10.1038/372669a0

Antonietti, A. (2013). Mirroring mirror neurons in an interdisciplinary debate. *Conscious. Cogn.* 22, 092–1094. doi: 10.1016/j.concog.2013.04.007

Baea, J. S., Pasajea, C. F., Parkb, B.-L., Cheongb, H. S., Kima, J.-H., Kima, J. Y., et al. (2012). Genetic association analysis of ERBB4 polymorphisms with the risk of schizophrenia and SPEM abnormality in a Korean population. *Brain Res.* 1466, 146–151. doi: 10.1097/YPG.0b013e32836130a9

Bahadir, S., Cobanoglu, U., Kapicioglu, Z., Kandil, S. T., Cimsit, G., Sönmez, M., et al. (2006). Lipoid proteinosis: a case with ophthalmological and psychiatric findings. *J. Dermatol.* 33, 215–218. doi: 10.1111/j.1346-8138.2006.00049.x

Baker-Andresen, D., Ratnu, V. S., and Bredy, T. W. (2013). Dynamic DNA methylation: a prime candidate for genomic metaplasticity and behavioral adaptation. *Trends Neurosci.* 36, 3–13. doi: 10.1016/j.tins.2012.09.003

Benson, M. A., Newey, S. E., Martin-Rendon, E., Hawkes, R., and Blake, D. J. (2001). Dysbindin, a novel coiled-coil-containing protein that interacts with the dystrobrevins in muscle and brain. *J. Biol. Chem.* 276, 24232–24241. doi: 10.1074/jbc.M010418200

Benson, P. J., Beedie, A. A., Shephard, E., Giegling, I., Rujescu, D., and St. Clair, D. (2012). Simple viewing tests can detect eye movement abnormalities that distinguish schizophrenia cases from controls with exceptional accuracy. *Biol. Psychiatry* 72, 716–724. doi: 10.1016/j.biopsych.2012.04.019

Boria, S., Fabbri-Destro, M., Cattaneo, L., Sparaci, L., Sinigaglia, C., Santelli, E., et al. (2009). Correction: intention understanding in autism. *PLoS ONE* 4: e5596. doi: 10.1371/journal.pone.0005596

Brown, D. J., Lin, B., and Holguin, B. (2004). Expression of neuregulin 1, a member of the epidermal growth factor family, is expressed as multiple splice variants in the adult human cornea. *Invest. Ophthalmol. Vis. Sci.* 45, 3021–3029. doi: 10.1167/iovs.04-0229

Cauda, F., Geda, E., Sacco, K., D'Agata, F., Duca, S., Geminiani, G., et al. (2011). Grey matter abnormality in autism spectrum disorder: an activation likelihood estimation meta-analysis study. *J. Neurol. Neurosurg. Psychiatry* 82, 1304–1313. doi: 10.1136/jnnp.2010.239111

Donohoea, G., Derek, W. M., De Sanctisb, P., Magnob, E., Montesib, J. L., Garavanb, H. P., et al. (2008). Early visual processing deficits in dysbindin-associated schizophrenia. *Biol. Psychiatry* 63, 484–489. doi: 10.1016/j.biopsych.2007.07.022

Enticott, P. G., Hoy, K. E., Herring, S. E., Johnston, P. J., Daskalakis, Z. J., and Fitzgerald, P. B. (2008). Reduced motor facilitation during action observation in schizophrenia: a mirror neuron deficit? *Schizophr. Res.* 102, 116–121. doi: 10.1016/j.schres.2008.04.001

Fagiolini, M., Jensen, C. L., and Champagne, F. A. (2009). Epigenetic influences on brain development and plasticity. *Curr. Opin. Neurobiol.* 19, 207–221. doi: 10.1016/j.conb.2009.05.009

Fana, W., Xinga, Y., Zhongb, Y., Chena, C., and Shena, Y. (2013). Expression of NMDA receptor subunit 1 in the rat retina. *Acta Histochem.* 15, 42–47. doi: 10.1016/j.acthis.2012.03.005

Ferrari, P. F., Tramacere, A., Simpson, E. A., and Iriki, A. (2013). Trends Mirror neurons through the lens of epigenetics. *Trends Cogn. Sci.* 17, 450–457. doi: 10.1016/j.tics.2013.07.003

Gallese, V. (2009). Motor abstraction: a neuroscientific account of how action goals and intentions are mapped and understood. *Psychol. Res.* 73, 486–498. doi: 10.1007/s00426-009-0232-4

Gallese, V., Rochat, M. J., and Berchio, C. (2013). The mirror mechanism and its potential role in autism spectrum disorder. *Dev. Med. Child Neurol.* 55, 15–22. doi: 10.1111/j.1469-8749.2012.04398.x

Greenwood, T. A., Lazzeroni, L. C., Murray, S. S., Cadenhead, K. S., Calkins, M. E., Dobie, D. J., et al. (2011). Analysis of 94 candidate genes and 12 endophenotypes for schizophrenia from the consortium on the genetics of schizophrenia. *Am. J. Psychiatry* 168, 930–946. doi: 10.1176/appi.ajp.2011.10050723

Helgadottir, A., Manolescu, A., Helgason, A., Thorleifsson, G., Thorsteinsdottir, U., Gudbjartsson, D. F., et al. (2006). A variant of the gene encoding leukotriene A4 hydrolase confers ethnicity-specific risk of myocardial infarction. *Nat. Genet.* 38, 68–74. doi: 10.1038/ng1692

Kana, R., Libero, L., Hu, C., Deshpande, H., and Colburn, J. (2014). Functional brain networks and white matter underlying theory-of-mind in autism. *Soc. Cogn. Affect. Neurosci.* 9, 98–105. doi: 10.1093/scan/nss106

Kang, C., Yang, X., Xu, X., Liu, H., Su, P., and Yang, J. (2012). Association study of neuregulin 1 gene polymorphisms with auditory P300 in schizophrenia. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 159B, 422–428. doi: 10.1002/ajmg.b.32045

Klin, A., Jones, W., Schultz, R., and Volkmar, F. (2003). The enactive mind, or from actions to cognition: lessons from autism. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 345–360. doi: 10.1098/rstb.2002.1202

Leube, D., Straube, B., Green, A., Blümel, I., Prinz, S., Schlotterbeck, P., et al. (2012). Possible brain network for representation of cooperative behavior and its implications for the psychopathology of schizophrenia. *Neuropsychobiology* 66, 24–32. doi: 10.1159/000337131

Marin, O. (2012). Interneuron dysfunction in psychiatric disorders. *Nat. Rev. Neurosci.* 13, 107–120. doi: 10.1038/nrn3155

Marín, O. (2013). Cellular and molecular mechanism controlling the migration of neocortical interneurons. *Eur. J. Neurosci.* 38, 2019–2029. doi: 10.1111/ejn.12225

Marsha, S., Collie-Duguida, E. S. R., Lib, T., Liuc, X., and McLeoda, H. L. (1999). Ethnic variation in the thymidylate synthase enhancer region polymorphism among caucasian and asian populations. *Genomics* 58, 310–312. doi: 10.1006/geno.1999.5833

Mausbach, B. T., Moore, R. C., Davine, T., Cardenas, V., Bowie, C. R., Ho, J., et al. (2013). The use of the theory of planned behavior to predict engagement in functional behaviors in schizophrenia. *Psychiatry Res.* 205, 36–42. doi: 10.1016/j.psychres.2012.09.016

Mehta, U. M., Thirthalli, J., Basavaraju, R., Gangadhar, B. N., and Pascual-Leone, A. (2013). Reduced mirror neuron activity in schizophrenia and its association with theory of mind deficits: evidence from a transcranial magnetic stimulation study. *Schizophr. Bull.* doi: 10.1093/schbul/sbt155. [Epub ahead of print].

Ng, S. C., Tsoi, K. K. F., Kamm, M. A., Xia, B., Wu, J., Chan, F. K. L., et al. (2012). Genetics of inflammatory bowel disease in Asia: systematic review and meta-analysis. *Inflamm. Bowel Dis.* 18, 1164–1176. doi: 10.1002/ibd.21845

Ono, C., Kikkawa, H., Suzuki, A., Suzuki, M., Yamamoto, Y., Ichikawa, K., et al. (2013), Clinical impact of genetic variants of drug transporters in different ethnic groups within and across regions. *Pharmacogenomics* 14, 1745–1764. doi: 10.2217/pgs.13.171

Schmechtig, A., Vassos, E., Kumari, V., Hutton, S. B., Collier, D. A., Morris, R. G., et al. (2010). Association of neuregulin 1 rs3924999 genotype with antisaccades and smooth pursuit eye movements. *Genes Brain Behav.* 9, 621–627. doi: 10.1111/j.1601-183X.2010.00594.x

Siebert, M., Markowitsch, H. J., and Bartel, P. (2003). Amygdala, affect and cognition: evidence from 10 patients with urbach-wiethe disease. *Brain* 126, 2627–2637. doi: 10.1093/brain/awg271

Synofzik, M., Thier, P., Leube, D. T., Schlotterbeck, P., and Lindner, A. (2010). Misattributions of agency in schizophrenia are based on imprecise predictions about the sensory consequences of one's actions. *Brain* 133, 262–271. doi: 10.1093/brain/awp291

Valiente, M., and Marin, O. (2010). Neural migration mechanism in development and disease. *Curr. Opin. Neurobiol.* 20, 68–78. doi: 10.1016/j.conb.2009

Woldemichael, B. T., Bohacek, J., Gapp, K., and Mansuy, I. M. (2014). Epigenetics of memory and plasticity. *Prog. Mol. Biol. Transl. Sci.* 122, 305–340. doi: 10.1016/B978-0-12-420170-5.00011-8

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 26 March 2014; paper pending published: 24 April 2014; accepted: 05 May 2014; published online: 27 May 2014.

Citation: Güell F (2014) Pre-dispositional constitution and plastic disposition: toward a more adequate descriptive framework for the notions of habits, learning and plasticity. Front. Hum. Neurosci. ***8****:341. doi: 10.3389/fnhum.2014.00341*

This article was submitted to the journal Frontiers in Human Neuroscience.

frontiers in
HUMAN NEUROSCIENCE

OPINION ARTICLE
published: 17 June 2014
doi: 10.3389/fnhum.2014.00432

A dialogical conception of *Habitus*: allowing human freedom and restoring the social basis of learning

Kleio Akrivou* and Lorenzo Todorow Di San Giorgio

Leadership, Organisations and Behaviour, Henley Business School, University of Reading, Reading, UK
**Correspondence: k.akrivou@henley.reading.ac.uk*

Edited by:
Javier Bernacer, University of Navarra, Spain

Reviewed by:
Jose Ignacio Murillo, University of Navarra, Spain
John Cromby, Loughborough University, UK

Keywords: consciousness, conversation, dialogue, habitus, human development, processual self, learning, neuroscience

BOURDIEU'S CONSTRUCT OF HABITUS AND CRITIQUE

The current sociological understanding of habitus expressed in the work of Pierre Bourdieu as its key academic construct emphasizes the importance of habits for human action. Bourdieu understands human behavior as fundamentally cultural; rejecting behaviorist view of cognition and action as related to a stimulus—response chains, the French sociologist posits instead that human action emanates from internalized habits (Swartz, 2002). Bourdieu, utilizes the concepts of field and *habitus*—the latin word for habit- (1990/1977), to critically analyze the microsphere of society and economy elevated as habituated human action (Bourdieu, 2005).

Bourdieu understands action as habituated, related to the unconscious reproduction of external social fields. This allows *habitus* to be this foundational concept it is in Bourdieu's reflexive sociology, examining societies as socio-economic relations and classes, rather than actors. His emphasis on the social context's influence on human habits being internalized relatively unconsciously is important, as it looks to action-interaction sequences as habituated, opposing behaviorist (Swartz, 2002) models of human action as simple stimulus-response chains, or objectivist accounts. Bourdieu's *habitus* emphasizes two things relevant to human action: (a) opposing consciousness to habit(us), which continues the modern sociological tradition from Weber (Camic, 1986) and, (b) the socially learnt (externally located) nature of habit(us). A habit is viewed as an unconscious principle of action, a deeply internalized set of dispositions, schemas and ways of knowing (Swartz, 2002) which locate habits in a cultural, economic, or social field.

In our opinion, the main problem in Bourdieu's view of *habitus* is that it largely accounts for human action being reproductive of an existing field, rather than transformative. Because the Bourdieuian *habitus* is theorized as an adopted "thrown way of being" in the world (Akrivou and Bradbury-Huang, 2014 forthcoming) it blocks human freedom with social bonds, as action is posited to emerge directly from the internationalization of norms of relational exchange in the outside field(s) of practice.

Our critique regarding Bourdieu's (quintessential) sociological *habitus* is that its conception explains processes accounting for human behavior regulation to carry forward existing conventions and rules, reproductive of existing social bonds; it is less mindful to processes of critical questioning or transformation of an existing status quo and the role of human cognition in generating action which can interrupt and interrogate the field.

To explain our critique further, even when Bourdieu accepts that actors can generate new action, he understands the "new" as habits from earlier socialization. Our main concern with Bourdieu's view is that looking to *habitus* as outside introjection means that individual *habitus* is in the best case "an active residue of (one's) past" (Swartz, 2002, p. 63S) which is Bourdieu's view! But we wish to critique this as it theoretically misses to account for the possibility for human freedom, which can be appreciated by reference to other understandings of *habitus* (Aristotle's for example).

A DIALOGICAL CONCEPTION OF HABITUS, SUPPORTED BY NEUROSCIENCE, AND PSYCHOLOGICAL THEORY ON HUMAN DEVELOPMENT

We believe that our critique of Bourdieu's *habitus* enables us to argue, that, in the frame of a dialogical conception *habitus* can be compatible with the social basis of human freedom and learning. Bourdieu's *habitus* defines it as site of replication of social bonds and boundaries, unless we revise his conception with a generative less deterministic structure, which protects the possibility of new habits emerging from agency. A dialogical conception understands human agency to be simultaneously part of a field of practice (and earlier socialization), and open to a gradual co-creation of novel action. The latter, emerging from an intentional conversational engagement practice between acting agents, can release a *new experience* of in-betweeness cognition, as dialog is a reciprocal "mode of communication that builds mutuality through awareness of others and as an instance of unfolding interaction" (Eisenberg and Goodall, 1993; Bohm, 1996; Putnam and Fairhurst, 2001, p. 116; Ballantyne, 2004). A dialogical conception of habitus can be compatible with the social basis of human freedom and learning, and core philosophical theories illustrate how dialog and conversation can gradually catalyze new *habitus*.

Buber's and Gadamer's **ethics of dialog** are relevant to our argument. Buber understands human freedom to emerge from locating oneself ethically in genuine relationships of a reciprocal "world of relation" (Buber, 1970, p. 56), with another fellow human. For Buber, the difference between "I-Thou" and "I-It" is not in the nature of the object to which we relate, but in the binding relationship itself (Levinas, 1989; Buber, 2002). Responding ethically to "Thou" replaces a passive response habit, an unreflective reproduction of external sets of relations and previously learnt dispositions. To keep dialogical ways of responding, one must engage in shared reflection to how to mindfully develop a "quality of genuine relationship in which partners are mutually unique as whole…this deep bonding is contained neither in one, nor the other, nor in the sum of both- but becomes really present between them" (Kramer, 2003, p. 15).

Gadamer also sees dialog as the *process* fostering a gradual mutual development of a shared gradually binding quality of relatedness. Gadamer notes that "to conduct a conversation means to allow oneself to … be caught up in something larger, which is neither subjective, nor objective, neither totally relative or fixed, … but … a structural unity … being conducted by the subject matter to which the partners in the dialog are oriented" (Gadamer, 1965, p. 367). Conversation partners gradually become less preoccupied with safe habits and engage in reaching a shared truth (White, 1994) with regard to how to proceed in a shared quest for *die Sache -or subject matter of inquiry* (Gadamer and Lawrence, 1982; Kelly, 1988; Gadamer, 1989, p. 383). The relationship becoming gradually a binding *"play of persons,"* the bond being conversation itself (White, 1994).

We argue that, a dialogical conception of habitus releases human freedom. Based on the previous analysis, engagement in dialogic habitus gradually forms a semi-autonomous zone (Akrivou and Bradbury-Huang, 2014) of action, which can generate new ways of knowing, while it also converses with habitus of the outside field of practice. In this argument, the idea of human freedom is meaningless outside the conversational practice; instead the necessity challenges of dialogic habitus requires to transcend the conventional assumption of independently autonomous rational agency to engage in the conversational "structural unity" (Gadamer, 1965; Akrivou and Bradbury-Huang, 2014) with a specific other fellow human. Developing the argument here the necessity challenges of dialogic ethics as a way to help address the critique of Bourdieu's *habitus*.

This argument can be supported by advances in neuroscience and psychological theory. It may seem new to many Westerners the idea of dyadic conversational structures (an I-thou structural unity) being the locus of consciousness as the sole or primary arbiters of social action; rather than each individual solitary independent autonomous rational processing (Akrivou and Bradbury-Huang, 2014). This supports revising the conception of human brain and cognitive processing, opposing a human brain operating via top-down predictions about sensory inputs and fully predicts the sensory information being received (Benacer and Murillo, 2012). Conversation is dynamically releasing new shared cognition pathways as one gradually learns to listen, feel, respond and engage in thoughtful responsiveness to a specific other actor.

The implication of such view of the locus of social action means that each human being bears the possibility of freedom (and the responsibility) to reflect what one brings forth in the world of relations and how. Once a dialogical conception is present in a given social field of action between inter-dependent agents *habitus* can be compatible with the social basis of human freedom and learning. A dialogical habitus opposes many Western philosophical theories emphasis on detached, autonomous scientific rationality. It instead supports neuroscience research that we are endowed with a brain adapted (Gazzaniga et al., 2002) to parcel out reality as separable units of an ever-changing flow of experience. We learn that solving self occurs mainly in the prefrontal cortex, with the emotional self-arising from the amygdala (Lewis and Todd, 2007). Any momentarily active aspects of the self-engage a fraction of the brain's networks (Gusnard et al., 2001; Legrand and Ruby, 2009). Contrary to our deeply and psychosomatically held belief in ourselves as "distinct individuals," many personal aspects happen automatically such as our heart beating. "In effect" summarizes Hanson, "subjectivity arises from the inherent distinction between this body and that world (2009, p. 210)." Indeed, Koch and Tsuchiya (2006) have found that diminishing habitual self-consciousness yields more positive results for the performer. Western neuroscience is pointing to the tantalizing fact that subjectivity is a way to structure experience but is not necessarily linked to individual persona (Hanson, 2009, p. 212).

Neuroscience is also supported by insights from human learning (Kolb, 1984; Maturana and Varela, 1987; Varela et al., 1991) and human development theory (Dewey, 1929; Werner, 1948; Harvey et al., 1961; Rogers, 1961; Schroder et al., 1967; Loevinger, 1976; Kohlberg and Ryncarz, 1990) on superior human cognitive moral maturity capacity. This is seen grounded in human freedom to choose both to be moral and **how** to engage in qualitative ways; "how" refers to a certain quality of cognitive processing which transcends subjectivity and engages in fluid, mutual inter-subjective ways of knowing (Rogers, 1961; Kegan, 1994). This quality of experiencing cognition is possible once a person freely "gives up" the safety of one's autonomous self-authorship on the basis of solitary reason—an ideal of conventional moral maturity based on the (Piaget, 1962) theory of development toward formal operations relying on abstract knowing (Flavell, 1963; Loevinger, 1976).

Rather than a mechanical (monological), crystallized cognitive map already stored in the brain in the form of abstract schemas, enabling a purely adaptive cognitive processing on the basis of habituated knowing how to respond to a set of outside stimuli, being in conversation intervenes to change the very way a previously taken for granted form of action can be experienced and dynamically practiced anew. Transcending its very reliance on formal operational thinking, the **processual self** emerges from within a dialogical habitus experiencing, an organic

way of being complete (integrated) *in situ* from within the process of narrative relational responsiveness (Akrivou, 2008), whereby one engages in the experience of relating genuinely with another human being as ground rather than a figure (James, 1979; Kohlberg and Ryncarz, 1990; Gendlin, 1997). This gradually develops a diverse set of the brain's cognitive pathways, as Bradbury- illustrates (in press) we learn "over time, to skillfully be with experience."

The emergence of previously unthought degrees of freedom generates novel action and social learning from within conversational fields itself rather than previously known habituated response schemas (Akrivou and Bradbury-Huang, 2014). This idea can be illustrated by the language of co-emergence in enactivist theories of human learning (Maturana and Varela, 1987; Varela et al., 1991). A dialogic process of narrative consciousness replaces cultural tools, rules, conventions and language as mechanisms for action regulation. It is instead a dynamic view of human cognition, a socially responsive brain which intentionally self regulates itself to context and other human beings own responses (Lewis and Todd, 2007). This conception of *habitus* generates meaningful novel action, binding one's own conscious attention and other actors' causal intervention responses in the process of shared conversational learning (Baker et al., 2002).

In conclusion, in the frame of a dialogical conception supported by psychological and neuroscientific findings, *habitus* can be compatible with the social basis of human freedom and learning. A dialogical conception of *habitus* may allow for habitus counter-intuitive to Bourdieu (Akrivou and Bradbury-Huang, 2014) which can be compatible with the social basis of human freedom and learning. It is closer to Aristotle's idea that rational agents (ought to) remain conscious of which habits to embrace and an active role of human agents being a consequence of engaging with virtuous habits. Perhaps, our argument helps bring Bourdieu's habitus closer to Aristotle's inquiry on the significance on human intentional action for a social world capable for virtue.

REFERENCES

Akrivou, K. (2008). *Differentiation and Integration in Adult Development: the Influence of Self Complexity and Integrative Learning on Self Integration.* Doctoral dissertation, Case Western Reserve University.

Akrivou, K., and Bradbury-Huang, H. (2014). Educating integrated catalysts: transforming business schools toward ethics and sustainability. *Acad. Manage. Learn. Edu*. 21, 2014. doi: 10.5465/amle.2012.0343

Baker, A. C., Jensen, P. J., and Kolb, D. A. (2002). *Conversational Learning: An Experiential Approach to Knowledge Creation.* Westport, CT: Quorum Books.

Ballantyne, D. (2004). Dialogue and its role in the development of relationship specific knowledge. *J. Bus. Ind. Mark.* 19, 114–123. doi: 10.5465/amle.2012.0343

Benacer, J., and Murillo, J. I. (2012). An incomplete theory of the mind. *Front. Psychol. Gen.* 3:418. doi: 10.3389/fpsyg.2012.00418

Bohm, D. (1996). *On Dialogue*. London: Roudedge.

Bourdieu, P. (2005). *The Social Structures of the Economy.* Cambridge: Polity Press.

Bradbury, H. (in press). Collaborative selflessness: toward an experiential understanding of the emergent "responsive self" in a caregiving context. *J. Appl. Behav. Sci.* doi: 10.1177/0021886313502729

Buber, M. (1970). *I and Thou.* W. Kaufmann, Trans. New York, NY: Scribner.

Buber, M. (2002). *Between Man and Man.* New York, NY: Routledge.

Camic, C. (1986). The matter of habit. *Am. J. Sociol.* 91, 1039–1087.

Dewey, J. (1929). *Human Nature and Conduct*. New York, NY: Modern Library.

Eisenberg, E., and Goodall, H. Jr. (1993). *Organizational Communication: Balancing Creativity and Constraint*. New York, NY: St. Martin's Press.

Flavell, J. (1963). *The Developmental Psychology of Jean Piaget.* New York, NY: Van Nostrand Reinhold Co.

Gadamer, H.-G. (1965). *Truth and Method.* New York, NY: Crossroad.

Gadamer, H. G. (1989). *Truth and Method*. Transl. Joel Weinsheimer and Donald G. Marshall. New York, NY: Continuum.

Gadamer, H. G., and Lawrence, F. G. (1982). *Reason in the Age of Science*: Cambridge University Press.

Gazzaniga, M., Ivry, R., and Mangun, G. (2002). *Cognitive Neuroscience: The Biology of the Mind, 2nd Edn.* New York, NY; London: W. W. Norton & Co.

Gendlin, E. T. (1997). *Experiencing and the Creation of Meaning: A Philosophical and Psychological Approach to the Subjective*. Evanston, Il: Northwestern University Press.

Gusnard, D. A., Abuja, E., Shulman, G. I., and Raichle, M. E. (2001). Medial prefrontal cortex and self referential mental activity: relation to a default mode of brain function. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4259–4264. doi: 10.1073/pnas.071043098

Hanson, R. (2009). *Buddha's Brain: The Practical Neuroscience of Happiness, Love and Wisdom*. Oakland, CA: New Harbinger Publications.

Harvey, O. J., Hunt, D., and Schroeder, H. (1961). *Conceptual Systems and Personality Organization*. New York, NY: John Wiley.

James, W. (1979). *The Will to Believe and Other Essays in Popular Philosophy.* Vol. 6, Cambridge, MA: Harvard University Press.

Kegan, R. (1994). *In Over our Heads: The Mental Demands of Modern Life*. Cambridge, MA; London: Harvard University Press.

Kelly, M. (1988). Gadamer and philosophical ethics. *Man World* 21, 327–346.

Koch, C., and Tsuchiya, N. (2006). Attention and consciousness: two distinct brain processes. *Trends Cogn. Sci.* 11, 16–21. doi: 10.1016/j.tics.2006.10.012

Kohlberg, L., and Ryncarz, R. A. (1990). "Beyond justice reasoning: moral development and consideration of a seventh stage," in *Higher Stages of Human Development,* eds C. N. Alexander and E. L. Langer (London: Oxford University Press), 191–207.

Kolb, D. A. (1984). *Experiential Learning: Experience as the Source of Learning and Development.* Vol. 1. Englewood Cliffs, NJ: Prentice-Hall.

Kramer, K. (2003). *Martin Buber's I and Thou: Practicing Living Dialogue*. New York, NY: Paulist Press.

Legrand, D., and Ruby, P. (2009). What is self specific? Theoretical investigation and critical review of neuroimaging results. *Psycholog. Rev.* 116, 252–282. doi: 10.1037/a0014172

Levinas, E. (1989). Martin Buber and the theory of knowledge. *Levinas Read.* 59–74.

Lewis, M. D., and Todd, R. M. (2007). The self regulating brain: cortical-subcortical feedback and the self development of intelligent action. *Cogn. Dev.* 22, 406–430. doi: 10.1016/j.cogdev.2007.08.004

Loevinger, J. (1976). *Ego Development: Conceptions and Theories*. San Francisco, CA: Jossey Bass, Inc.

Maturana, H., and Varela, F. (1987). *The Tree of Knowledge*. Boston, MA: Shambhala.

Piaget, J. (1962). *The Moral Judgement of the Child.* New York, NY: Collier Books.

Putnam, L. L., and Fairhurst, G. T. (2001). "Discourse analysis in organizations," in *The New Handbook of Organizational Communication*, eds F. Jablin and L. Putnam (London: Sage Publications), 78–136.

Rogers, C. R. (1961). *On Becoming a Person*. Boston, MA: Houghton Mifflin.

Schroder, H. M., Driver, M. J., and Streufert, S. (1967). *Human Information Processing*. New York, NY: Holt, Rinehart and Winston.

Swartz, D. L. (2002). The sociology of habit: the perspective of Pierre Bourdieu. *Occup. Ther. J. Res.* 22, 61S–69S.

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and*

Human Experience. Cambridge, MA: MIT Press.

Werner, H. (1948). *Comparative Psychology of Mental Development*. Chicago, IL: Follett Publishing.

White, K. W. (1994). "Gans-georg gadamer's philosophy of language: a constitutive-dialogic approach to interpersonal understanding," in *Interpretive Approaches to Interpersonal Communication,* eds K. Carter and M. Presnell (New York, NY: State University of New York Press), 94–95.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 20 February 2014; accepted: 28 May 2014; published online: 17 June 2014.
Citation: Akrivou K and Di San Giorgio LT (2014) A dialogical conception of Habitus: allowing human freedom and restoring the social basis of learning. Front. Hum. Neurosci. ***8****:432. doi: 10.3389/fnhum.2014.00432*
This article was submitted to the journal Frontiers in Human Neuroscience.

frontiers in
HUMAN NEUROSCIENCE

OPINION ARTICLE
published: 29 April 2014
doi: 10.3389/fnhum.2014.00261

Conceptual mappings and neural reuse

***Cristóbal Pagán Cánovas*[1]* and *Javier Valenzuela Manzanares*[2]**

[1] *Institute for Culture and Society, University of Navarra, Pamplona, Spain*
[2] *English Department, University of Murcia, Murcia, Spain*
**Correspondence: cpaganc@unav.es*

Edited by:
Jose Angel Lombo, Pontifical University of the Holy Cross, Italy

Reviewed by:
Mark Bernard Turner, Case Western Reserve University, USA

Keywords: conceptual mappings, conceptual blending, neural theory of language, neural reuse, metaphor, SNARC effect, material anchors for conceptual blends, mental timeline

TOO MUCH NEURAL REUSE, EVEN FOR A METAPHORICAL BRAIN

Conceptual mappings are correspondances between *conceptual domains* (SPACE, TIME, FORCE, EMOTION, etc.) or between entities within the same conceptual domain. Through mapping, we project inferences, elements, and relations from one mental configuration to another. Sets of mappings can become entrenched, creating powerful cognitive habits. For example, across many cultures around the world, temporal relations are conceived by means of spatial relations, in language ("Saturday is almost here"), artifacts (timelines, calendars, sundials), or gesture (Núñez and Sweetser, 2006). Some of the mappings for this template are: duration is spatial extent, events are landmarks, or time is motion along a path.

From over 30 years of conceptual mappings research emerges a picture of the conceptual system as a set of mapping habits. Instead of a static repository of concepts, we have a dynamic network connecting mental structures. Mapping is not exceptional: it is the norm. It is through mapping that concepts are formed, learned, and developed creatively. These ideas have boosted the interest in the most remarkable manifestations of mapping in language and thought: metaphor, metonymy, analogy, counterfactuals, etc.

Metaphor has received far more attention than all the other phenomena combined. Researchers in Conceptual Metaphor Theory (CMT) (Lakoff and Johnson, 1980; Lakoff, 1993), have identified many sets of cross-domain mappings underlying conventional metaphorical expressions: TIME IS SPACE, LIFE IS A JOURNEY, ANGER IS HEAT, EVENTS ARE ACTIONS, etc. According to CMT, conceptual metaphors are static, ontological, fixed sets of partial correspondances between two conceptual domains. Metaphor transfers inferences from the source domain, more concrete or better structured, to the target domain, which is more abstract or less delineated. A system of thousands of metaphorical mappings constitutes the main mechanism for abstract thought in the human mind.

From the nineties, the semantic postulates of CMT have been used to develop the Neural Theory of Language (NTL) (Gallese and Lakoff, 2005; Feldman, 2008; Lakoff, 2008). In NTL, conceptual metaphors are replaced by neural mappings, combinations of simple neural circuits that carry out conceptual mappings. The major function of all this neural binding is the reuse of sensorimotor brain mechanisms for new roles in language and reasoning. This is congenial with the *grounding* of abstract concepts in perceptual experience (Barsalou, 1999, 2008), called *embodiment* in conceptual mappings research (Johnson, 1987, 2008; Lakoff, 1987; Lakoff and Johnson, 1999). The overarching idea of CMT-NTL is a "metaphorical mind-brain," based on direct, binary transfer across domains. The transfer is carried out through static cognitive habits, which are implemented by neural circuitry connecting pairs of brain areas. For the brain, metaphor is once more privileged over all other manifestations of mapping, just like it was for the mind. A good example of the rapidly growing interest in the neuroscience of metaphor is the *Frontiers* issue about the topic, currently being edited by Vicky T. Lai and Seana Coulson.

But the metaphorical brain seems quite insufficient to account for the pervasiveness and complexity of neural reuse. In a recent *BBS* target article, Michael Anderson (2010) shows that what is going on in the brain dwarfs the predictions of embodiment or CMT-NTL. Statistics run on thousands of fMRI studies indicate that even fairly small brain regions are typically reused in multiple tasks, with even higher reuse probabilities if a region is involved in perception or action (Anderson et al., 2010).

Neural reuse is ubiquitous and dynamic, and many of its results cannot be explained as domain-structuring inheritance. Anderson examines, among others, the following examples: the SNARC effect (a mental number line with magnitudes increasing from left to right), the correlation between finger representation and numerical cognition, the interaction of word and gesture, or the phonological loop in working memory. These cases present no metaphorical projection, and some of them involve more than two components. Rather than direct transfer of information, a given system seems to be reused for a non-primary purpose because it happens to have a function or structure that are appropriate for the particular cognitive task at hand. As Anderson claims, we need a broader theoretical framework, able to account for those individual phenomena as well as for the general prevalence of neural reuse.

FROM TRANSFER TO EMERGENCE: THE NETWORK MODEL OF CONCEPTUAL INTEGRATION

Can conceptual mappings research provide such a framework? For one thing,

the CMT-NTL model is certainly not the only one in the field. CMT and the pervasiveness of mapping were the point of departure of Gilles Fauconnier's Mental Space Theory (Fauconnier, 1985, 1997), later developed by Fauconnier and Mark Turner into Conceptual Integration Theory, or Blending Theory (BT) (Turner, 1996, 2014; Fauconnier and Turner, 2002).

Beyond the common ground with CMT (Fauconnier and Lakoff, 2013), BT introduces significant innovations. Mental spaces are not vast domains such as TIME or SPACE, but small conceptual packets that flexibly combine entrenched structures and contextual information, for local purposes of thought and action. Mappings are established through structural or functional correspondances between input spaces in a generic mental space. The participation of more than two inputs is quite typical. Selectively projected to a blended space or *conceptual blend*, elements from the inputs interact, typically producing emergent structure, which cannot be accounted for by direct transfer between domains. Inferences can take place in the blend, but also be projected back to the inputs, which can be modified as the process unfolds. The mappings, the emergence of novel structure, the adjustment of the inputs, and everything else going on is guided by universal governing principles and competing optimality principles, by the functional requirements of the particular network of mental spaces, dictated by context and goals, and by the creativity of the individual or group who are striving to make the most of it all.

The overarching picture that results from BT bears important differences with that of CMT-NTL. Advanced blending underlies all manifestations of mapping, including metaphor, which is just one more surface product of this species-defining capacity for integrating disparate mental components into new, meaningful wholes. The human brain is a bubble chamber of mental spaces, constantly building new integration networks, and a culture is an even larger bubble chamber (Fauconnier and Turner, 2002, p. 321–322). Just like evolution—and neural reuse—, blending is opportunistic: it reuses whatever is functionally suitable, right there and then. As it happens with the natural selection of living organisms, the process of trial and error in conceptual and neural reuse never stops. Through it, minds and cultures select the few integrations that are really useful, anchor them by means of symbolic procedures, and pass them on to the next generation. To become productive habits, both generic templates of conceptual integration and patterns of neural reuse need to find an adequate niche within the general system (about the notion of *neural niche*, see: Anderson, 2010, and the commentary by Atsushi Iriki therein; Iriki and Sakura, 2008; Iriki and Taoka, 2012).

ONE EXAMPLE: OPPORTUNISTIC REUSE IN THE NUMBER LINE AND THE TIMELINE

In blending as in neural reuse, a given item, once identified as potentially useful, is integrated into the network under construction. If necessary and possible, the item is adjusted for optimization in its new function. If it works, the item is kept in the network, although it still remains available for its older functions. Networks and their components are discarded and entrenched in a dynamic, extremely agile process. What is going on here is not direct transfer of structure, but rather the construction of a new whole with old pieces. The novel properties are not borrowed from the structures being reused, but result from their performance in a new network.

Among other examples, Anderson (2010) illustrates this with the spatial-numerical association of response codes (SNARC) effect, that is, a mental number line in which numerals are arrayed from left to right, in order of increasing magnitude (Dehaene et al., 1993). As Anderson explains, there is no metaphoric mapping or perceptual grounding here: in sensorimotor experience, magnitude may increase with height (the MORE IS UP metaphor), but not laterally, and certainly not in the direction of writing. Numerals do not inherit the structure of the spatial shifting mechanism: the left-to-right line has been picked opportunistically, and integrated with numeral magnitude, simply because the resulting blend meets the requirements of the task. We could add to Anderson's argument by pointing out that the mental number line, as a symbolic device, needed considerable cultural time to emerge: it was only invented in seventeenth-century Europe, although awareness of potential correspondances between numbers and spatial relations dates back to Babylon (Núñez, 2009). It took thousands of years for the pattern to find an appropriate niche, alongside a representational format that would ensure its transmission.

The number line is what gets called a *material anchor for a conceptual blend* (Hutchins, 2005). A perceptual structure is used as an input in the integration process. In the blend, perceptual relations are fused with conceptual relations. Now consider a very similar case. Varied psycholinguistic evidence shows that processing temporal expressions causes the automatic activation of a mental timeline, also running from left to right in cultures with that writing system (Torralbo et al., 2006; Weger and Pratt, 2008; Santiago et al., 2010). Blending theorists have revised the TIME IS SPACE metaphor, and shown that it is a complex network that produces a motion scene with special rules and constraints, designed to facilitate the representation of time: all observers are on the same spot, all objects move along the same path, and spatial relations can even be modified by the emotional attitude of the observer: "Monday is almost here, but Friday is so far away" (Fauconnier and Turner, 2008). The straight line is particularly useful for anchoring this blended scene. The result is a graphic representation with novel properties, which allow us to see diachrony at a glance, to divide it easily into periods, to represent events as dots, etc.

The timeline is a generic integration template that blends at least four inputs: time and time measures, spatial extent, objects, and events (Coulson and Cánovas, 2013). Spatial shifting from left to right is absent from all these components, but it happens to facilitate the task immensely, and thus it is imported to the blend, for local purposes. The pattern is not functional as a metaphor in language, where past and future are not on the left or right, but it is extremely productive in gesture (Casasanto and Jasmin, 2012), where the lateral axis is more easily available. Again, the timeline has a long history of failed cultural representations behind it (Rosenberg and Grafton, 2010), but, once the blending template found its niche, it is reused time and again, and can even be adapted for representing complex emotions and creating sophisticated poetic effects (Cánovas

and Jensen, 2013). Metaphor is useful, but not enough to understand the timeline: a broader framework of reuse and integration is needed.

WHAT KIND OF MODEL WE NEED

BT researchers have indeed identified many recurrent patterns and theorized about them, but we still lack a general framework for generic integration templates (some work along those lines: Fauconnier and Turner, 2002; Fauconnier, 2009; Cánovas, 2010, 2011; Turner, 2014). A fruitful interaction with research on neural reuse can impose further constraints and requisites than those observed in the semantic or semiotic analyses. A model of conceptual mapping habits fully compatible with neural reuse may include the following:

- Network thinking (Mitchell, 2006) rather than direct binary transfer.
- Flexibility in the activation, selection, and integration of conceptual and neural patterns.
- Focus on emergence.
- Emphasis on competing optimality principles, e.g., a left-to-right straight line leaves aside many relevant aspects of time or magnitude, but its functionality is privileged.
- Detailed examination of how context and goals, including cultural diachrony, shape the process of integration.
- A model of entrenchment not based on ontological projection, but on the idea of "attaining a niche" through instance-based learning and context-sensitive usage.

REFERENCES

Anderson, M. L. (2010). Neural reuse: a fundamental organizational principle of the brain michael. *Behav. Brain Sci.* 33, 1–69. doi: 10.1017/S0140525X10000853

Anderson, M. L., Brumbaugh, J., and Aysu Şuben, A. (2010). "Investigating functional cooperation in the human brain using simple graph-theoretic methods," in *Computational Neuroscience*, (Springer) 31–42. doi: 10.1007/978-0-387-88630-5_2. Available online at: http://link.springer.com/chapter/10.1007/978-0-387-88630-5_2

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660. doi: 10.1017/S0140525X99002149

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. doi: 10.1146/annurev.psych.59.103006.093639

Cánovas, C. P. (2010). Erotic emissions in greek poetry: a generic integration network. *Cogn. Semiotics* 6, 7–32. doi: 10.3726/81610_7

Cánovas, C. P. (2011). The genesis of the arrows of love: diachronic conceptual integration in greek mythology. *Am. J. Philol.* 132, 553–579. doi: 10.1353/ajp.2011.0044

Cánovas, C. P., and Jensen, M. F. (2013). Anchoring time-space mappings and their emotions: the timeline blend in poetic metaphors. *Lang. Lit.* 22, 45–59. doi: 10.1177/0963947012469751

Casasanto, D., and Jasmin, K. (2012). *The Hands of Time: Temporal Gestures in English Speakers*. Available online at: www.degruyter.com/view/j/cog.2012.23.issue-4/cog-2012-0020/cog-2012-0020/cog-2012-0020.xml

Coulson, S., and Cánovas, C. P. (2013). Understanding timelines. *J. Cogn. Semiotics* 5, 198–219. doi: 10.1515/cogsem.2013.5.12.198

Dehaene, S., Bossini, S., and Giraux, P. (1993). The mental representation of parity and number magnitude. *J. Exp. Psychol. Gen.* 122, 371–396. doi: 10.1037/0096-3445.122.3.371

Fauconnier, G. (1985). *Mental Spaces: Aspects of Meaning Construction in Natural Language*. Cambridge, MA: Cambridge University Press.

Fauconnier, G. (1997). *Mappings in Thought and Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139914220

Fauconnier, G. (2009). Generalized integration networks. *New Dir. Cogn. Linguist*. 147–160.

Fauconnier, G., and Lakoff, G. (2013). On metaphor and blending. *Cogn. Semiotics* 5, 393–399.

Fauconnier, G., and Turner, M. (2002). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York, NY: Basic Books.

Fauconnier, G., and Turner, M. (2008). "Rethinking metaphor," in *The Cambridge Handbook of Metaphor and Thought*, ed R. W Gibbs (Cambridge, MA: Cambridge University Press), 57–66. doi: 10.1017/CBO9780511816802.005

Feldman, J. A. (2008). *From Molecule to Metaphor: A Neural Theory of Language*. Cambridge, MA: MIT Press.

Gallese, V., and Lakoff, G. (2005). The brain's concepts: the role of the sensory-motor system in conceptual knowledge. *Cogn. Neuropsychol.* 22, 455–479. doi: 10.1080/02643290442000310

Hutchins, E. (2005). Material anchors for conceptual blends. *J. Pragmatics* 37, 1555–1577. doi: 10.1016/j.pragma.2004.06.008

Iriki, A., and Sakura, O. (2008). The neuroscience of primate intellectual evolution: natural selection and passive and intentional niche construction. *Philos. Trans. R. Soc. B Biol. Sci.* 363, 2229–2241. doi: 10.1098/rstb.2008.2274

Iriki, A., and Taoka, M. (2012). Triadic (ecological, Neural, Cognitive) niche construction: a scenario of human brain evolution extrapolating tool use and language from the control of reaching actions. *Philos. Trans. R. Soc. B Biol. Sci.* 367, 10–23. doi: 10.1098/rstb.2011.0190

Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago, IL: University Of Chicago Press.

Johnson, M. (2008). *The Meaning of the Body: Aesthetics of Human Understanding*. Reprint. Chicago, IL: University Of Chicago Press.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things*. 1997th Edn. Chicago, IL: University Of Chicago Press.

Lakoff, G. (1993). "The contemporary theory of metaphor." in *Metaphor and Thought*, ed A. Ortony (Cambridge, CA: Cambridge University Press). doi: 10.1017/CBO9781139173865.013

Lakoff, G. (2008). "The neural theory of metaphor." in *The Cambridge Handbook of Metaphor and Thought*, ed R. W Gibbs, Jr (Cambridge: Cambridge University Press), 17–38. doi: 10.1017/CBO9780511816802.003

Lakoff, G., and Johnson, M. (1980). *Metaphors We Live By*. Chicago, IL: University Of Chicago Press.

Lakoff, G., and Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York, NY: Basic Books.

Mitchell, M. (2006). Complex systems: network thinking. *Artif. Intell.* 170, 1194–1212. doi: 10.1016/j.artint.2006.10.002

Núñez, R. (2009). Numbers and arithmetic: neither hardwired nor out there. *Biol. Theory* 4, 68–83. doi: 10.1162/biot.2009.4.1.68

Núñez, R. E., and Sweetser, E. (2006). With the future behind them: convergent evidence from aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cogn. Sci.* 30, 401–450. doi: 10.1207/s15516709cog0000_62

Rosenberg, D., and Grafton, A. (2010). *Cartographies of Time: A History of the Timeline*. New York, NY: Princeton Architectural Press.

Santiago, J., Román, A., Ouellet, M., Rodríguez, N., and Pérez-Azor, P. (2010). In hindsight, life flows from left to right. *Psychol. Res.* 74, 59–70. doi: 10.1007/s00426-008-0220-0

Torralbo, A., Santiago, J., and Lupiáñez, J. (2006). Flexible conceptual projection of time onto spatial frames of reference. *Cogn. Sci.* 30, 745–757. doi: 10.1207/s15516709cog0000_67

Turner, M. (1996). *The Literary Mind: The Origins of Thought and Language*. New York, NY: Oxford University Press.

Turner, M. (2014). *The Origin of Ideas: Blending, Creativity, and the Human Spark*. New York, NY: Oxford University Press.

Weger, U. W., and Pratt, J. (2008). Time flies like an arrow: space-time compatibility effects suggest the use of a mental timeline. *Psychon. Bull. Rev.* 15, 426–430. doi: 10.3758/PBR.15.2.426

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 27 March 2014; accepted: 09 April 2014; published online: 29 April 2014.

Citation: Cánovas CP and Manzanares JV (2014) Conceptual mappings and neural reuse. Front. Hum. Neurosci. ***8****:261. doi: 10.3389/fnhum.2014.00261*

This article was submitted to the journal Frontiers in Human Neuroscience.

frontiers in
HUMAN NEUROSCIENCE

OPINION ARTICLE
published: 21 May 2014
doi: 10.3389/fnhum.2014.00312

The role of consciousness in triggering intellectual habits

Javier Sánchez-Cañizares*

Mind-Brain Project, Institute for Culture and Society (ICS), University of Navarra, Pamplona, Spain
**Correspondence: js.canizares@unav.es*

Edited and reviewed by:
Jose Angel Lombo, Pontifical University of the Holy Cross, Italy

Keywords: consciousness and truth, intellectual habits, inhibitory control, post-error slowing, attentional orientation

Why does a theoretical physicist describe a system from the point of view of its mathematical symmetries instead of performing a numerical simulation? The researcher chooses the strategy that should provide the most relevant information. Nevertheless, the different strategies have not always been available. Scientific discovery has actually happened in history thanks to the creativity of a good number of thinkers, whose insights proved to be decisive for the developing of new branches of science. New ways of confronting well-posed problems may eventually become intellectual habits of generations of scientists, but such habits can only develop after checking the validity of these new perspectives. What the history of science shows is somehow reproduced in the learning process. "Knowing-about" forms the heart of standard education: students can learn and be tested on it. But success in examinations gives little indication of whether that knowledge can be employed when required, which is the essence of "knowing-to." When must a specific intellectual habit be called upon? Knowing-to has to do with the conscious use of cognitive habits. It implies a conscious judgment of an upper-level truth about the problem. The practice of reflection is a means to help students improve their knowing-to act in the moment because the triggering situation for the enactment of a new behavioral schema must be conscious. Thus being explicit about one's own thinking improves mathematics teaching and learning (Lim and Selden, 2009). The "aha" moment is experienced by someone who learns a new strategy to tackle a problem, develops a new intellectual habit and knows when to use it. This paper comments on neuroscientific support of the "aha" moment through the inhibition mechanisms of the brain.

Inhibitory control is an executive process involved in attention, self-regulation, and consciousness. Intelligence is closely tied to the ability to inhibit a misleading behavior, judgment, or strategy, and inhibition is precisely the cognitive mechanism that should allow one to redirect attention toward logically relevant issues. It seems to be crucial in order to validate and activate a new mode of thinking. Houdé's group experimentally showed that the biased (spatial) to logical shift in the way of solving a logical problem with geometrical objects of different colors and shapes is a specific consequence of executive training in matching-bias inhibition. The relevant point is that inhibition allows subjects to redirect attention to the logically correct shapes, a shift process in which the activated brain networks can change radically in the same subjects depending on their ability to inhibit a misleading strategy (Houdé et al., 2000). Posterior-to-anterior reconfiguration of the activated brain regions brought about by inhibitory control might be the neural correlate of human abstraction, the ability to break away from perceptual biases during cognitive development. Houdé's training helps the subject to be conscious of the implicit mistake he or she is making and, therefore, what he or she must do to avoid the trap.

Recent empirical work clarifies the specificity of high order cognitive process enacting inhibition. Overall, how errors impact the processing of subsequent stimuli and in turn shape behavior remains unresolved. However, the literature documents two main mechanisms when correcting errors: bottom-up, automatic and top-down, controlled forms of inhibition (Spierer et al., 2013). Actually, many results are interpreted in terms of a shift from a fast automatic to a slow controlled form of inhibitory control induced by the detection of errors, which could have been caused by an attentional modulation (Manuel et al., 2012). Some experiments on post-error slowing in subjects support the view that outcome expectancy (not accuracy) is essential for such effect. Post-error slowing is caused by attentional orienting to unexpected events and not by a strategic adjustment of cognitive nature (Núñez Castellar et al., 2010). Then, there seems to be a qualitative criterion for the subject to decide when accuracy is important because expectancy is not fulfilled. In short, the subject has expectancy. He or she looks for an *adequacy* with the input signal. And he or she needs attentional reorientation when this adequacy is not satisfied.

Intentional inhibitory control exists whenever the subject's creativity and checking of the truth (adequacy) is required in a new problem. Deliberate inhibition is necessary even if the subject has the proper *knowledge about*, but does not *know when to* use it. In other words, inhibition is an effect of *detecting error and shifting to a new cognitive strategy* in which awareness is necessary for shifting from fast to slow forms of control. The role of consciousness is related to the inhibition of the common (habitual) cognitive strategy, which would allow for the activation and recruiting of the brain areas involved in a new type of reasoning. But the specific "judgment of truth" to accept or reject a new strategy turns out to be an exclusive feature of consciousness. This explains, for instance, why the validation of a mathematical generalization cannot initially be a habit. Theoretical scientists must beforehand *judge* the relevant strategy for

tackling a problem and then make *conscious* use of an intellectual habit—which might be new in the case of new theoretical discoveries—to try to solve it. To sum up, human consciousness—as something different from a pure brain state—is required in order to establish the validity of a new theoretical perspective. Once this is reached, new habits may be at work. Consciousness mediates between the unconscious formation of new ideas and the development of new habits, which need check of the new ideas' adequacy in order to be prompted. Therefore, consciousness is the precursor for the activation of intellectual habits.

ACKNOWLEDGMENT

The author acknowledges financial support by "Obra Social La Caixa."

REFERENCES

Houdé, O., Zago, L., Mellet, E., Moutier, S., Pineau, A., Mazoyer, B., et al. (2000). Shifting from the perceptual brain to the logical brain: the neural impact of cognitive inhibition training. *J. Cogn. Neurosci.* 12, 721–728. doi: 10.1162/089892900562525

Lim, K., and Selden, A. (2009). "Mathematical habits of mind," in *Proceedings of the 31st Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, eds S. L. Swars, D. W. Stinson, and S. Lemons-Smith (Atlanta, GA: Georgia State University), 1576–1583.

Manuel, A. L., Bernasconi, F., Murray, M. M., and Spierer, L. (2012). Spatio-temporal brain dynamics mediating post-error behavioral adjustments. *J. Cogn. Neurosci.* 24, 1331–1343. doi: 10.1162/jocn_a_00150

Núñez Castellar, E., Kühn, S., Fias, W., and Notebaert, W. (2010). Outcome expectancy and not accuracy determines posterror slowing: ERP support. *Cogn. Affect. Behav. Neurosci.* 10, 270–278. doi: 10.3758/CABN.10.2.270

Spierer, L., Chavan, C. F., and Manuel, A. L. (2013). Training-induced behavioral and brain plasticity in inhibitory control. *Front. Hum. Neurosci.* 7:427. doi: 10.3389/fnhum.2013.00427

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 22 March 2014; accepted: 27 April 2014; published online: 21 May 2014.
Citation: Sánchez-Cañizares J (2014) The role of consciousness in triggering intellectual habits. Front. Hum. Neurosci. ***8****:312. doi: 10.3389/fnhum.2014.00312*
This article was submitted to the journal Frontiers in Human Neuroscience.

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 24 February 2014
doi: 10.3389/fnhum.2014.00072

No horizontal numerical mapping in a culture with mixed-reading habits

Neda Rashidi-Ranjbar[1,2], Mahdi Goudarzvand[1], Sorour Jahangiri[1], Peter Brugger[3,4] and Tobias Loetscher[5]*

[1] Faculty of Medicine, Alborz University of Medical Sciences, Karaj, Iran
[2] Department of Cognitive Science, University of Trento, Trento, Italy
[3] Department of Neurology, University Hospital Zurich, Zurich, Switzerland
[4] Zurich Center for Integrative Human Physiology (ZIHP), University of Zurich, Zurich, Switzerland
[5] School of Psychology, Flinders University, Adelaide, SA, Australia

Edited by:
Jose Ignacio Murillo, University of Navarra, Spain

Reviewed by:
Matthias Hartmann, University of Bern, Switzerland
Sylvie Chokron, CNRS, Université Paris-Descartes and Fondation Ophtalmologique Rothschild, France

***Correspondence:**
Tobias Loetscher, School of Psychology, Flinders University, GPO Box 2100, Adelaide, SA 5001, Australia
e-mail: tobias.loetscher@alumni.ethz.ch

Reading habits are thought to play an important role in the emergence of cultural differences in visuo-spatial and numerical tasks. Left-to-right readers show a slight visuo-spatial bias to the left side of space, and automatically associate small numbers to the left and larger numbers to the right side of space, respectively. A paradigm that demonstrated an automatic spatial-numerical association involved the generation of random numbers while participants performed lateral head turns. That is, Westerners have been shown to produce more small numbers when the head was turned to the left compared to the right side. We here employed the head turning/random number generation (RNG) paradigm and a line bisection (LB) task with a group of 34 Iranians in their home country. In the participants' native language (Farsi) text is read from right-to-left, but numbers are read from left-to-right. If the reading direction for text determines the layout of spatial-numerical mappings we expected to find more small numbers after right than left head turns. Yet, the generation of small or large numbers was not modulated by lateral head turns and the Iranians showed therefore no association of numbers with space. There was, however, a significant rightward shift in the LB task. Thus, while the current results are congruent with the idea that text reading habits play an important role in the cultural differences observed in visuo-spatial tasks, our data also imply that these habits on their own are not strong enough to induce significant horizontal spatial-numerical associations. In agreement with previous suggestions, we assume that for the emergence of horizontal numerical mappings a congruency between reading habits for words and numbers is required.

Keywords: cross-cultural, random number generation, mental number line, embodied numerical cognition, automatic processing, line bisection, visuo-motor behavior

INTRODUCTION

Our thoughts, perception and actions are shaped by the culture in which we live. Our way of thinking, for example, depends on the social systems we grew up with. That is, East Asians tend to reason in a holistic way, while Westerners exhibit a more analytical thinking style (Nisbett et al., 2001). Different cultural habits also modulate how we perceive things. Italians, for example, judge soccer goals more beautiful when presented with a left-to-right compared to right-to-left trajectory, whereas Arabic speakers show the opposite directional bias (Maass et al., 2007). Our cultural background might also determine motor actions, such as whether we preferably turn our head to the left or right side for kissing somebody on the lips (Shaki, 2013).

Differences in reading directions are thought to play an important role in the emergence of cultural differences in visuo-spatial tasks (Kazandjian and Chokron, 2008). When bisecting horizontal lines, for example, left-to-right readers bisect slightly to the left of the line's true center. Right-to-left readers, on the other hand, have been reported to misplace the bisection mark to the right of the line's center (Chokron and Imbert, 1993; Chokron et al., 1997). Similarly, reading direction predicts whether one attends to rightward or leftward features of chimeric faces (Vaid and Singh, 1989), while writing direction can determine whether the trajectory of an apparent motion is perceived as moving to the left or right side (Tse and Cavanagh, 2000).

Reading direction might not only modulate performance in visuo-spatial tasks, but may also influence the way numbers are represented and processed. Preliminary evidence for such influences was reported by Dehaene et al. (1993). In a series of experiments the authors first established that French readers spontaneously mapped small numbers to left and larger numbers to right-sided response codes (the SNARC effect). In their Experiment 7, the authors then showed that a group of Iranians, who had immigrated to France, showed a weaker SNARC effect than French participants. Intriguingly, the time since immigration was related to the direction of the SNARC effect. Iranians with

a longer exposure to left-to-right reading direction tended to show a regular SNARC effect, while those Iranians with less familiarity with this reading direction tended to show a reversed SNARC effect -with larger numbers being associated with the left hand. The finding in this study implied a congruency between reading direction for *words* and the representational layout of small to large numbers. It is important to note here that the Iranians' native language, Farsi, is a mixed-reading language. That is, words in Farsi are written/read from right-to-left, but numerals from left-to-right. Therefore the above experiment suggests that the reading direction for words, and not the one for numerals, determines the mapping between numbers and space.

Subsequent studies provided further evidence for a link between the direction of number representations and reading habits (see Göbel et al., 2011 for a review). Zebian (2005) showed, for example, that Arabic monolingual right-to-left readers associate large and small numbers with the left and right sides of space, respectively. This reversed SNARC effect was significantly reduced in bilingual Arabic participants fluent in right-to-left and left-to-right reading languages. It has been suggested that being fluent in languages with opposite reading habits could weaken spatial-numerical associations (Göbel et al., 2011). Importantly, links between reading direction and spatial-numerical mappings are not restricted to SNARC paradigms, but are also found with other paradigms tapping into spatial-numerical representations, such as bisection tasks (e.g., Kazandjian et al., 2010).

Research on the effects of reading habits also provides ample evidence that the direction of spatial-numerical mapping is flexible and hinges on recently processed stimuli. Bilingual Russian-Hebrew readers, for example, showed a regular SNARC effect after reading a left-to-right Cyrillic script, but they exhibited a significantly reduced effect after reading a right-to-left Hebrew script (Shaki and Fischer, 2008; see also Fischer et al., 2010). In the same vein, Hung et al. (2008) demonstrated that the orientation of the mental number line depends on the task's context. Chinese readers mapped Arabic numerals on a left-to-right oriented number line, but associated Chinese number words with a vertical, top-to-bottom oriented number line. That is, depending on the format of the numerical notation the spatial-numerical associations differed (Hung et al., 2008).

A wide range of different paradigms have been used to investigate spatial-numerical interactions in Western cultures (see Dehaene and Brannon, 2011). One of those paradigms simply requires participants to generate sequences of random numbers (Loetscher and Brugger, 2007). Studies using random number generation (RNG) paradigms have demonstrated that Westerners implicitly associate the generation of small and large numbers with the left and right side of space, respectively (Hartmann et al., 2012; Vicario, 2012; Di Bono and Zorzi, 2013; Grade et al., 2013). It has been shown, for example, that participants tend to shift their gaze slightly leftward when randomly naming a small number. Rightward gaze shifts, on the other hand, are accompanied with the generation of larger numbers (Loetscher et al., 2010). An analogous pattern of results is found when participants generate random numbers while performing lateral head turns. That is, more small numbers are produced when the head is turned to the left compared to right side turns (Loetscher et al., 2008).

In light of the above findings it is surprising that RNG tasks have never been used to asses spatial mappings of numbers in cultures with right-to-left reading habits. The goal of the current research was to fill this gap. We set out to investigate the spatial representations of numbers in Iranians with an RNG paradigm. For this purpose we replicated the head turning paradigm used by Loetscher et al. (2008). As in the original study, participants were required to rhythmically turn the head from one side to the other while generating random numbers. If the reading direction for words determines the layout of spatial-numerical mappings we expected to find more small numbers after right than left head turns. Such a finding would imply that Iranians code smaller numbers to the right and larger number to the left side of space—the opposite pattern reported by Loetscher et al. (2008) in Western participants. An alternative prediction is that Iranians will show no effect of lateral head turns on the magnitude of generated numbers. Writing/reading directions differ in their native language, Farsi, for words (right-to-left) and numerals (left-to-right). These two opposing habits may cancel one another out. Support for the prediction of a null-finding derives from a study conducted with Israeli participants. Hebrew is also a mixed-reading language, with opposite reading directions for words and numbers, and the participants did not exhibit a reliable spatial association for numbers in a SNARC paradigm (Shaki et al., 2009). Finally, there is also the possibility that small/left and large/right associations as for Westerners are found. This would be an indication that the reading direction for numerals, and not words, is dominant in determining the orientation of the mental number line.

In addition to assessing spatial-numerical association, we also measured visuo-spatial biases in a manual line bisection task (LB task). Based on the previous literature we here expected to find a slight bias to the right of the line's true center (Chokron and Imbert, 1993; Chokron et al., 1997). Comparing the performances in the visuo-spatial and numerical tasks allowed us to comment on the effect of reading habits in these tasks.

MATERIALS AND METHODS

PARTICIPANTS

Thirty-four Iranian men with a mean age of 24 (SD = 7.5) participated in this study. The 25 right and 9 left-handed participants (Chapman and Chapman, 1987) were mostly students and without history of neuropsychiatric or neurological disorder (Campbell, 2000). The higher representation of left-handed participants (26%) in our sample than the proportion of left-handers found in the general population (around 10%, Nicholls et al., 2013) was due to a selection bias. Initially, it was planned to recruit an equal number of right-handers and left-handers for the current experiment. However, this proved to be unachievable due to the difficulty of recruiting left-handed participants. Nevertheless, given the relatively high number of left-handed participants we incorporated handedness as a factor in the analyses.

The native language for all participants was Farsi. All participants had regular English classes in school. They all described themselves as "beginners" and not fluent in any language with a left-to-right reading direction.

The study was approved by the Medical Sciences Ethics Committee of the Alborz University.

TASKS

The RNG task was as described in Loetscher et al. (2008). Participants were asked to name numbers between 1 and 30 in a sequence as random as possible. With their eyes closed, participants generated a new number every 2 s. The speed of generation was controlled with a metronome running at 0.5 Hz. As in the original study there were two counterbalanced conditions. In the baseline condition, 40 responses were generated while the head was kept straight ahead. In the head-turning condition, participants performed rhythmic head turns to the left and right side, respectively. After participants turned their head about 80° to one side they named a number and then started to turn the head to the opposite side again. The rhythmic head turns continued until a total of 80 numbers, 40 for either direction, were recorded by the examiner. Numbers between 1 and 15 represent "small" numbers in the number space ranging from 1 to 30, those from 16 to 30 represent "large" numbers. The dependent variable was the number of "small" numbers generated.

In the LB task, participants were asked to bisect nine horizontal lines using a pen with their dominant hand. Each line was presented on a separate A4 sheet and measured 160 mm. The dependent variable was the average deviation from the lines' true center in mm—with positive values indicating a rightward deviation and negative values a leftward deviation.

RESULTS

The number of "small" numbers was submitted to a repeated-measure ANOVA with *Condition* (left turn, baseline, right turn) as a within-subjects factor and *Handedness* (left, right) as a between-subjects factor. The analysis revealed neither a main effect for *Condition* ($F_{(2,64)} = 0.17$, $p = 0.85$) nor for *Handedness* ($F_{(1,32)} = 0.32$, $p = 0.57$). The interaction between *Condition* and *Handedness* was not significant either ($F_{(2,64)} = 0.73$, $p = 0.49$). Due to its theoretical importance for the current study we also directly compared the number of generated "small" numbers during left and right head turns. Paired *t*-tests revealed no significant differences in the number of "small" numbers between lateral head turns (all participants: $t_{(33)} = 0.65$, $p = 0.52$; only right-handers: $t_{(24)} = 0.91$, $p = 0.37$; only left-handers: $t_{(8)} = -0.21$, $p = 0.84$).

One sample *t*-tests were conducted to investigate whether there was a bias for naming too many "small" numbers in any of the three conditions. As there was no handedness effect in the ANOVA, data were collapsed across this factor for this analysis. The number of "small" numbers generated did not differ from the expected value of 20.0 in any of the three conditions ($t_{(33)} < 1.61$, $p > 0.11$, see **Figure 1**). Also, the average of small numbers generated across the three conditions was not significantly different from 20.0 ($t_{(33)} = 1.21$, $p = 0.23$).

The subjective midpoint in the LB task was shifted 0.93 mm (SEM = 0.35) to the right side of the lines' true center. A one sample *t*-test comparing the participants' mean deviation to 0 indicated that there was a significant rightward bias ($t_{(33)} = 2.67$, $p < 0.02$) for the 34 participants. The performance in the LB task differed between left (mean deviation: −0.22 mm, SEM = 0.13) and right-handers (mean deviation: 1.34 mm, SEM = 0.44; $t_{(32)} = 2.08$, $p < 0.05$). One sample *t*-tests showed that right-handers deviated significantly to the right of the true center ($t_{(24)} = 3.02$, $p < 0.01$) and that there was no LB bias for left-handers ($t_{(8)} = -1.79$, $p > 0.11$).

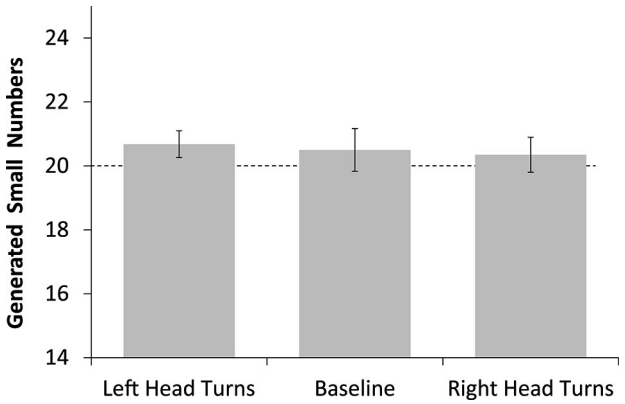

FIGURE 1 | Number of generated "small" numbers after head turns to the left, a baseline condition (head facing straight ahead), and after head turns to the right. The dashed line indicates chance expectation.

The bias in the LB task was not related to the average magnitude of generated "small" numbers across all three conditions ($r = 0.21$, $p = 0.23$).

DISCUSSION

The study aimed to investigate the spatial mappings of numbers in a culture in which words are read and written from right to left, but numerals from left to right ("mixed-reading habit"). A paradigm that revealed an automatic mapping of small and large numbers to left and right head turns respectively in Westerners (Loetscher et al., 2008) was applied to 34 Iranian participants. In contrast to Westerners, the generation of small or large numbers by Iranians was not modulated by lateral head turns. That is, there was no association of numbers with space, and hence, no evidence for an embodied representation of numbers (Fischer and Brugger, 2011).

The lack of an automatic mapping of numbers in space is in agreement with the few studies that investigated spatial-numerical associations in Iranians. Dehaene et al. (1993), for example, found a weakened SNARC effect in Iranians who had immigrated to France. Our study corroborates these findings by showing that no associations are found when data is collected in Iran, with participants who have been less exposed to Western culture than those in the studies which rely on immigrated participants. Nonetheless, all our participants had some interaction with Western culture. While these interactions were probably less extensive than in previous studies (e.g., Dehaene et al., 1993), we cannot rule out the possibility that they were sufficient to affect the association between numbers and space in the current task. The current study design does not allow disentangling the effects of Western culture exposure and native reading habits on the results. It is noteworthy, however, that even when exposed to Western cultures on a daily

basis, native right-to-left readers continue to show specific spatial biases in mental representations (Maass and Russo, 2003).

The reading directions of words (right-to-left) and numbers (left-to-right) differ in Farsi. Our working hypothesis is that these opposite reading habits lead to the disappearance of any preferred lateral association of numbers along the horizontal mental number line. Hebrew readers also use opposite reading directions for words and numbers, and these readers also lacked reliable spatial-numerical associations in a SNARC paradigm (Shaki et al., 2009). It seems reasonable to propose therefore that horizontal associations between numbers and space might only become significant if the reading directions of words and numbers are consistent (Shaki et al., 2009).

Although cultures with mixed-reading directions do not evidence a significant horizontal representation of numbers, it is important to point out that this does not imply the lack of any spatial-numerical mappings in these cultures. The current null-finding, for example, might be the consequence of two conflicting horizontal mappings that cancel each other out. While this idea needs to be further investigated, it has previously been shown that participants with mixed-reading directions (monolingual Israelis) exhibit a radial spatial-numerical mapping when response buttons in a SNARC paradigm were placed in a radial instead of the conventional horizontal arrangement (Shaki and Fischer, 2012). This first demonstration of a spatial-numerical mapping in a mixed-reading culture corroborates the idea that these mappings are flexible and can vary within the same participant depending on the situational context and task demands (Bachtold et al., 1998; Hung et al., 2008; van Dijck et al., 2009; Fischer et al., 2010; Shaki and Fischer, 2012). It seems noteworthy that task demands not only affect spatial-numerical mappings, but also mappings in other dimensions such as space and words (Thornton et al., 2013), or numbers and time (Nicholls et al., 2011). The observation of mappings between word meaning ("moon") and space ("upper visual space"), for example, is contingent on task demands as it depends on the arrangement of response buttons (Thornton et al., 2013).

Participants' handedness only affected performance in the visuo-motor LB task, but not in the RNG task. This finding is consistent with previous research. Differences between left and right-handers in LB tasks are commonly observed (Sampaio and Chokron, 1992; Jewell and McCourt, 2000), while handedness seems to be unrelated to spatial-numerical associations (Dehaene et al., 1993; Fischer, 2008).

The observed rightward shift in the LB task is analogous to that described in previous studies assessing visuo-spatial biases in right-to-left reading cultures (Chokron and Imbert, 1993; Chokron et al., 1997), but opposite to the leftward shift found in left-to-right reading cultures (Jewell and McCourt, 2000). Thus, the current results are consistent with the suggestion that reading habits for text play an important role in the cultural differences observed in visuo-spatial tasks (Kazandjian and Chokron, 2008; Kazandjian et al., 2010). However, our data also suggest that these habits on their own are not strong enough to induce significant horizontal spatial-numerical associations. In accord with the conclusions of Shaki et al. (2009) we assume that for the emergence of horizontal spatial mappings a congruency between reading habits for text and numbers is required.

REFERENCES

Bachtold, D., Baumuller, M., and Brugger, P. (1998). Stimulus-response compatibility in representational space. *Neuropsychologia* 36, 731–735. doi: 10.1016/s0028-3932(98)00002-5

Campbell, J. J. (2000). "Neuropsychiatric assessment," in *Textbook of Geriatric Neuropsychiatry*, eds C. E. Coffey and J. L. Cummings (Washington, DC: American Psychiatric Publishing), 109–124.

Chapman, L. J., and Chapman, J. P. (1987). The measurement of handedness. *Brain Cogn.* 6, 175–183. doi: 10.1080/08856559.1933.10533126

Chokron, S., Bernard, J. M., and Imbert, M. (1997). Length representation in normal and neglect subjects with opposite reading habits studied through a line extension task. *Cortex* 33, 47–64. doi: 10.1016/s0010-9452(97)80004-4

Chokron, S., and Imbert, M. (1993). Influence of reading habits on line bisection. *Brain Res. Cogn. Brain Res.* 1, 219–222. doi: 10.1016/0926-6410(93)90005-p

Dehaene, S., Bossini, S., and Giraux, P. (1993). The mental representation of parity and number magnitude. *J. Exp. Psychol. Gen.* 122, 371–396. doi: 10.1037/0096-3445.122.3.371

Dehaene, S., and Brannon, E. (2011). *Space, Time and Number in the Brain: Searching for the Foundations of Mathematical Thought.* London: Academic Press.

Di Bono, M., and Zorzi, M. (2013). The spatial representation of numerical and non-numerical ordered sequences: insights from a random generation task. *Q. J. Exp. Psychol. (Hove)* 66, 2348–2362. doi: 10.1080/17470218.2013.779730

Fischer, M. H. (2008). Finger counting habits modulate spatial-numerical associations. *Cortex* 44, 386–392. doi: 10.1016/j.cortex.2007.08.004

Fischer, M. H., and Brugger, P. (2011). When digits help digits: spatial-numerical associations point to finger counting as prime example of embodied cognition. *Front. Psychol.* 2:260. doi: 10.3389/fpsyg.2011.00260

Fischer, M. H., Mills, R. A., and Shaki, S. (2010). How to cook a SNARC: number placement in text rapidly changes spatial-numerical associations. *Brain Cogn.* 72, 333–336. doi: 10.1016/j.bandc.2009.10.010

Göbel, S. M., Shaki, S., and Fischer, M. H. (2011). The cultural number line: a review of cultural and linguistic influences on the development of number processing. *J. Cross. Cult. Psychol.* 42, 543–565. doi: 10.1177/0022022111406251

Grade, S., Lefèvre, N., and Pesenti, M. (2013). Influence of gaze observation on random number generation. *Exp. Psychol.* 60, 122–130. doi: 10.1027/1618-3169/a000178

Hartmann, M., Grabherr, L., and Mast, F. W. (2012). Moving along the mental number line: interactions between whole-body motion and numerical cognition. *J. Exp. Psychol. Hum. Percept. Perform.* 38, 1416–1427. doi: 10.1037/a0026706

Hung, Y., Hung, D., Tzeng, O., and Wu, D. (2008). Flexible spatial mapping of different notations of numbers in Chinese readers. *Cognition* 106, 1441–1450. doi: 10.1016/j.cognition.2007.04.017

Jewell, G., and McCourt, M. E. (2000). Pseudoneglect: a review and meta-analysis of performance factors in line bisection tasks. *Neuropsychologia* 38, 93–110. doi: 10.1016/s0028-3932(99)00045-7

Kazandjian, S., Cavézian, C., Zivotofsky, A. Z., and Chokron, S. (2010). Bisections in two languages: when number processing, spatial representation and habitual reading direction interact. *Neuropsychologia* 48, 4031–4037. doi: 10.1016/j.neuropsychologia.2010.10.020

Kazandjian, S., and Chokron, S. (2008). Paying attention to reading direction. *Nat. Rev. Neurosci.* 9, 965. doi: 10.1038/nrn2456-c1

Loetscher, T., Bockisch, C. J., Nicholls, M. E., and Brugger, P. (2010). Eye position predicts what number you have in mind. *Curr. Biol.* 20, R264–R265. doi: 10.1016/j.cub.2010.01.015

Loetscher, T., and Brugger, P. (2007). Exploring number space by random digit generation. *Exp. Brain Res.* 180, 655–665. doi: 10.1007/s00221-007-0889-0

Loetscher, T., Schwarz, U., Schubiger, M., and Brugger, P. (2008). Head turns bias the brain's internal random generator. *Curr. Biol.* 18, R60–R62. doi: 10.1016/j.cub.2007.11.015

Maass, A., Pagani, D., and Berta, E. (2007). How beautiful is the goal and how violent is the fistfight? spatial bias in the interpretation of human behavior. *Soc. Cogn.* 25, 833–852. doi: 10.1521/soco.2007.25.6.833

Maass, A., and Russo, A. (2003). Directional bias in the mental representation of spatial events: nature or culture? *Psychol. Sci.* 14, 296–301. doi: 10.1111/1467-9280.14421

Nicholls, M. E., Lew, M., Loetscher, T., and Yates, M. J. (2011). The importance of response type to the relationship between temporal order and numerical magnitude. *Atten. Percept. Psychophys.* 73, 1604–1613. doi: 10.3758/s13414-011-0114-x

Nicholls, M. E. R., Thomas, N. A., Loetscher, T., and Grimshaw, G. M. (2013). The Flinders Handedness survey (FLANDERS): a brief measure of skilled hand preference. *Cortex* 49, 2914–2926. doi: 10.1016/j.cortex.2013.02.002

Nisbett, R. E., Choi, I., Peng, K., and Norenzayan, A. (2001). Culture and systems of thought: holistic versus analytic cognition. *Psychol. Rev.* 108, 291–310. doi: 10.1037//0033-295x.108.2.291

Sampaio, E., and Chokron, S. (1992). Pseudoneglect and reversed pseudoneglect among left-handers and right-handers. *Neuropsychologia* 30, 797–805. doi: 10.1016/0028-3932(92)90083-x

Shaki, S. (2013). What's in a kiss? spatial experience shapes directional bias during kissing. *J. Nonverbal Behav.* 37, 43–50. doi: 10.1007/s10919-012-0141-x

Shaki, S., and Fischer, M. H. (2008). Reading space into numbers—a cross-linguistic comparison of the SNARC effect. *Cognition* 108, 590–599. doi: 10.1016/j.cognition.2008.04.001

Shaki, S., and Fischer, M. H. (2012). Multiple spatial mappings in numerical cognition. *J. Exp. Psychol. Hum. Percept. Perform.* 38, 804–809. doi: 10.1037/a0027562

Shaki, S., Fischer, M. H., and Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychon. Bull. Rev.* 16, 328–331. doi: 10.3758/pbr.16.2.328

Thornton, T., Loetscher, T., Yates, M. J., and Nicholls, M. E. (2013). The highs and lows of the interaction between word meaning and space. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 964–973. doi: 10.1037/a0030467

Tse, P. U., and Cavanagh, P. (2000). Chinese and Americans see opposite apparent motions in a Chinese character. *Cognition* 74, B27–B32. doi: 10.1016/s0010-0277(99)00065-7

Vaid, J., and Singh, M. (1989). Asymmetries in the perception of facial affect: is there an influence of reading habits? *Neuropsychologia* 27, 1277–1287. doi: 10.1016/0028-3932(89)90040-7

van Dijck, J. P., Gevers, W., and Fias, W. (2009). Numbers are associated with different types of spatial information depending on the task. *Cognition* 113, 248–253. doi: 10.1016/j.cognition.2009.08.005

Vicario, C. M. (2012). Perceiving numbers affects the internal random movements generator. *Sci. World J.* 2012:347068. doi: 10.1100/2012/347068

Zebian, S. (2005). Linkages between number concepts, spatial thinking and directionality of writing: the SNARC effect and the reverse SNARC effect in English and Arabic monoliterates, biliterates and illiterate Arabic speakers. *J. Cogn. Cult.* 5, 165–190. doi: 10.1163/156853705406 8660

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 09 October 2013; accepted: 29 January 2014; published online: 24 February 2014.

Citation: Rashidi-Ranjbar N, Goudarzvand M, Jahangiri S, Brugger P and Loetscher T (2014) No horizontal numerical mapping in a culture with mixed-reading habits. Front. Hum. Neurosci. ***8****:72. doi: 10.3389/fnhum.2014.00072*

The article was submitted to the journal Frontiers in Human Neuroscience.

frontiers in
HUMAN NEUROSCIENCE

OPINION ARTICLE
published: 08 August 2014
doi: 10.3389/fnhum.2014.00614

Behavioral duality in an integrated agent

Ivan Martinez-Valbuena and Javier Bernacer*

Mind-Brain Group, Institute for Culture and Society, University of Navarra, Pamplona, Spain
**Correspondence: jbernacer@unav.es*

Edited by:
Jose Angel Lombo, Pontifical University of the Holy Cross, Italy

Reviewed by:
Ignacio Morón, University of Granada, Spain

Keywords: Kahneman, habits, goal-directed actions, consciousness, cognitive control, prefrontal cortex

Humans can consolidate and carry out habits other animals cannot. This statement is mainly sustained by the fact that humans have a unique cognitive control of their actions: we can let our attention fade away to perform automatic tasks more efficiently, we can detect if there has been an unexpected problem in their implementation, and we can regain conscious control of the action if necessary. We tend to dichotomize this cognitive process into two "systems," namely goal-directed versus habitual (Dickinson, 1985), conscious versus unconscious (Crick and Koch, 1998), or slow versus fast (Kahneman, 2011). If we just put it in those terms, these two ways of tackling the challenges of a changing environment seem to be present in non-human animals. However, all dichotomies imply a difficulty to deal with: the regulation of the transition between the two systems. Is this carried out by a third element, or regulated by one of the systems? Could it be more convenient to view it as a continuum, rather than a dichotomy? In any case, we believe this transition has a level of complexity in humans that makes it qualitatively different from its analog in animals. In fact, this "cognitive bridge" might be a major feature to characterize a reliable behavior, since a particular task or problem is more efficiently tackled when the transition between the two systems is more adequate. Moreover, the integrity of this link could be an indicator to detect prodromal psychiatric conditions, as it has been suggested for slips-of-action (Gillan et al., 2011).

In order to justify these ideas, we will focus first on Kahneman's distinction between systems 1 and 2 (Kahneman, 2011). On the one hand, System 1 is responsible for making decisions rapidly. The purpose of this system is to give us an assessment of the environment around us as quickly as possible so that we are able to respond as fast as possible. To perform this task, System 1 follows general rules or guidelines (heuristics). In all, System 1 is intended to help us make decisions more quickly, and is very useful (let's say "just fine") in most cases. However, one of the characteristics of these decisions is the lack of voluntary control, what is a problem considering this system is responsible of many of the decisions and judgments we make. Given to its "automatic" nature, System 1 also has biases and systematic errors that are likely to happen in some situations.

On the other hand, System 2 acts when a problem which System 1 has no solution for arises. System 2, apparently, can take control of the whole process at any time. It is somewhat triggered by some external or internal alarm that draws its attention and makes it take "conscious" control of the situation. One of the problems of this system is that it is lazy and can be easily exhausted. Therefore, it usually accepts the decisions of System 1 without monitoring them. One proof of System 2's negligence is what Kahneman calls WYSIATI ("What You See Is All There Is"), a general rule that "facilitates the achievement of coherence and of the cognitive ease that causes us to accept a statement as true." System 1 easily gets that coherence, and System 2 usually allows it to jump to conclusions and act. In different sets of experiments, Kahneman demonstrates that humans are not good at all with statistics or handling mathematics; in his opinion, this is because humans simplify judgments to make them more understandable and deal with them just through heuristics that System 1 can handle. This general view of humans as poor rational decision-makers is also supported by other authors (see, for example, Ariely, 2008).

In our opinion, this division of human cognition into two systems fits well with the usual opposition between goal-directed versus habitual systems (Dickinson, 1985). In general, goal-directed actions are viewed as conscious, flexible, and sensitive to outcome devaluation, whereas habits are mainly unconscious, rigid and insensitive to changes in the value of the outcome. The features of goal-directed and habit systems were mainly drawn from studies in animals. The typical experiment about this subject consists on teaching the contingency between an instrumental action (for example, a lever press) and a reward to the animal (Adams and Dickinson, 1981). At the beginning, the animal's behavior is goal directed, and it performs the action to obtain the reward. However, this behavior becomes "habitual" (in this context, a motor routine) after many repetitions. When that happens, the value of the reward is transferred to the lever press itself: even though the outcome is devalued (gets the animal sick) or the animal is sated, it keeps pressing the lever. This is why habits have been opposed to goal-directed behavior.

A quick look suggests that habits and goal-directed actions are intimately related to Systems 1 and 2, respectively. This is also supported by the identification of the goal-directed system with a model-based reinforcement learning scheme, since it can be viewed "in terms of sophisticated, computationally demanding, prospective planning, in which a decision tree of

possible future states and actions is built using a learned internal model of the environment" (Dolan and Dayan, 2013). The habitual system, on the other hand, follows a model-free scheme, which "is computationally efficient, since it replaces computation (i.e., the burdensome simulation of future states) with memory (i.e., stored discounted values of expected future reward); however, the forward-looking nature of the prediction error makes it statistically inefficient" (Dolan and Dayan, 2013). Following these analogies between systems, we can assume that some actions that at the beginning fall under the domain of System 2 might be transferred to System 1 through learning, like goal-directed actions become habits through experience.

Concerning the neural bases of these systems, the striatum and its cortical afferents and –indirect– target areas in the cortex play a major role. It is widely accepted that the cognitive part of the striatum –caudate nucleus and anterior putamen– are involved in the planning and execution of goal-directed actions, together with the prefrontal cortex (Balleine et al., 2007). On the other hand, the sensorimotor striatal aspects –mainly the posterior putamen– and the supplementary motor area of the cortex are particularly active when the agent is performing a well-learned action (Miyachi et al., 2002; Ashby et al., 2010). Furthermore, the activity of the neurons in these areas follows a "chunked" pattern: they are mainly active at certain stages of the motor routine (for example at the beginning and the end of the sequence, when a particular switch or turn is needed, etc), and this activity is reduced in the rest of the motor sequence (Graybiel, 1998). Although some authors question a sharp neuroanatomical basis of Kahneman's Systems 1 and 2 (Gold and Shadlen, 2007), our train of thought in this manuscript suggests that the more reflective System 2 should be based in the prefrontal cortex –both dorsal and ventral–, and the cognitive regions of the basal ganglia. Likewise, the more automatic System 1 would lie on motor and premotor cortical regions, as well as on the sensorimotor aspects of those subcortical nuclei.

This neuroscientific framework identifies the habit system with automaticity, rigidity and unconsciousness; however, we are intending to challenge this view in past and forthcoming contributions (Bernacer and Gimenez-Amaya, 2013; Bernacer et al., 2014). In a nutshell, we propose to view the phenomenon of action from the point of view of the agent as a whole, and not from an isolated movement. Hence, it could be more convenient to understand System 1 –or habits– as a resource of System 2, rather than as opposed systems in competition. Whereas a motor routine (i.e., what is commonly called "habit" in neuroscience) implies the sequential and unconscious performance of movements, they usually pursue the goal set by the agent. In fact, the more engrained the routine is, the easier for the agent to achieve that goal. Furthermore, the agent can consciously stop or correct the movement at any point, since the habit releases the higher cognitive regions of the brain to improve the performance of the action. A very simple example of this is a tennis service, which should be "goal directed" to place the ball wherever the player wants. It involves a set of movements such as throwing the ball upwards, moving the feet, putting the arm back, etc. Only when these motor routines are learned correctly, the player is able to concentrate on other aspects of the service such as the speed, spin, or exploiting the weaknesses of the receiver. This can be also exemplified with other kinds of habits such as driving, playing an instrument, tackling a mathematical problem, and so on. They all suggest that "automatic" routines are governed by higher cognitive functions to better achieve a particular goal.

Kahneman's Systems 1 and 2 allow as well this release of consciousness from everyday decisions to focus on more complicated situations. As Kahneman himself and other authors defend (Ariely, 2008), the problem arises when System 2 is rarely used or either system is applied to inadequate situations. However, we believe that the most effective agent does not exclusively rely on System 2, but efficiently uses all resources of "each system" in the right situation and, more importantly, carries out an appropriate transition between them. That is, in our opinion, a "rational" agent. This could be also said about goal-directed and habitual systems. Moreover, we believe that this transition between systems is subject to learning, and it can be performed more effectively through experience.

If we understand these fragmentations of cognition as independent systems in competition, we encounter an important problem: is there an additional mechanism in charge of the transition between systems, or is this regulated by one of the systems itself? If the first option were true, we would find the difficulty of defining the nature –both conceptually and anatomically– of a "third system" qualitatively different than the other two. This would lead us to an *ad infinitum* process –the need of a fourth element to regulate the activity of the third, and so on–, and therefore we believe this hypothesis should be rejected. Considering the second option, it seems that only the highly cognitive System 2 could be in charge of leading the transition between systems, which in turn dissolves a rigorous separation in two systems. The role of System 2 in leading the transition is clear when the agent decides to regain conscious control of a task generally performed in an unconscious manner. In this sense, the interaction of the orbitofrontal cortex with either the cognitive or sensorimotor aspects of the striatum plays a central role in shifting between goal-directed actions and motor routines (Gremel and Costa, 2013). In other situations, an external cue such as an error may set the alarm for System 2 to retake control of the action. Regarding this, the anterior cingulate cortex has been reported to be active in highly-conflictive decision making situations (Goñi et al., 2011); for that reason, some authors relate this cortical area with error monitoring (Carter, 1998; Botvinick et al., 2004). A recent report suggests a new model of reinforcement learning and conflict monitoring, which involves a wide network including different areas of the cortex (posterior parietal, precentral, anterior cingulate and prefrontal) and the basal ganglia (Zendehrouh et al., 2013).

To sum up, this opinion article suggests viewing Kahneman's systems as analogous to the goal-directed/habits dichotomy in order to improve the understanding of some aspects of human cognition. Further, we believe a strict separation between systems in competition is problematic, since

System 2 is always in charge of governing the interplay between systems: therefore, System 1 –or habits– should be understood as a resource of System 2. This view could shed some light on the understanding of habits as a source of learning, plasticity and freedom for the agent. Finally, an inappropriate cognitive control of habits could be an indicator of certain psychiatric conditions.

ACKNOWLEDGMENTS

The authors appreciate the suggestions of Doctors Guell, Blanco, Murillo and Barrett in the preparation of the manuscript. Our research is supported by Obra Social La Caixa.

REFERENCES

Adams, C., and Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *Q. J. Exp. Psychol. Sect. B Comp. Physiol. Psychol.* 33, 109–121.

Ariely, D. (2008). *Predictably Irrational*. New York, NY: HaperCollins.

Ashby, F. G., Turner, B. O., and Horvitz, J. C. (2010). Cortical and basal ganglia contributions to habit learning and automaticity. *Trends Cogn. Sci.* 14, 208–215. doi: 10.1016/j.tics.2010.02.001

Balleine, B. W., Delgado, M. R., and Hikosaka, O. (2007). The role of the dorsal striatum in reward and decision-making. *J. Neurosci.* 27, 8161–8165. doi: 10.1523/JNEUROSCI.1554-07.2007

Bernacer, J., Balderas, G., Martinez-Valbuena, I., Pastor, M. A., and Murillo, J. I. (2014). The problem of consciousness in habitual decision making. *Behav. Brain Sci.* 37, 21–22. doi: 10.1017/S0140525X13000642

Bernacer, J., and Gimenez-Amaya, J. (2013). "On habit learning in neuroscience and free will," in *Is Science Compatible With Free Will?* eds A. Suarez and P. Adams (New York, NY: Springer), 177–193.

Botvinick, M. M., Cohen, J. D., and Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends Cogn. Sci.* 8, 539–546. doi: 10.1016/j.tics.2004.10.003

Carter, C. S. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science* 280, 747–749. doi: 10.1126/science.280.5364.747

Crick, F., and Koch, C. (1998). Consciousness and neuroscience. *Cereb. Cortex* 8, 97–107. doi: 10.1093/cercor/8.2.97

Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Philos. Trans. R. Soc. B Biol. Sci.* 308, 67–78. doi: 10.1098/rstb.1985.0010

Dolan, R. J., and Dayan, P. (2013). Goals and habits in the brain. *Neuron* 80, 312–325. doi: 10.1016/j.neuron.2013.09.007

Gillan, C. M., Papmeyer, M., Morein-Zamir, S., Sahakian, B. J., Fineberg, N. A., Robbins, T. W., et al. (2011). Disruption in the balance between goal-directed behavior and habit learning in obsessive-compulsive disorder. *Am. J. Psychiatry* 168, 718–726. doi: 10.1176/appi.ajp.2011.10071062

Gold, J. I., and Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.* 30, 535–574. doi: 10.1146/annurev.neuro.29.051605.113038

Goñi, J., Aznárez-Sanado, M., Arrondo, G., Fernández-Seara, M., Loayza, F. R., Heukamp, F. H., et al. (2011). The neural substrate and functional integration of uncertainty in decision making: an information theory approach. *PLoS ONE* 6:e17408. doi: 10.1371/journal.pone.0017408

Graybiel, A. M. (1998). The basal ganglia and chunking of action repertoires. *Neurobiol. Learn. Mem.* 70, 119–136. doi: 10.1006/nlme.1998.3843

Gremel, C. M., and Costa, R. M. (2013). Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nat. Commun.* 4, 2264. doi: 10.1038/ncomms3264

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York, NY: Farrar, Straus and Giroux.

Miyachi, S., Hikosaka, O., and Lu, X. D. A.-S. (2002). Differential activation of monkey striatal neurons in the early and late stages of procedural learning. *Exp. Brain Res.* 146, 122–126. doi: 10.1007/s00221-002-1213-7

Zendehrouh, S., Gharibzadeh, S., and Towhidkhah, F. (2013). Modeling error detection in human brain: A preliminary unification of reinforcement learning and conflict monitoring theories. *Neurocomputing* 103, 1–13. doi: 10.1016/j.neucom.2012.04.026

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 June 2014; paper pending published: 05 July 2014; accepted: 22 July 2014; published online: 08 August 2014.

*Citation: Martinez-Valbuena I and Bernacer J (2014) Behavioral duality in an integrated agent. Front. Hum. Neurosci. **8**:614. doi: 10.3389/fnhum.2014.00614*

This article was submitted to the journal Frontiers in Human Neuroscience.

frontiers in
HUMAN NEUROSCIENCE

HYPOTHESIS AND THEORY ARTICLE
published: 26 June 2014
doi: 10.3389/fnhum.2014.00457

Model averaging, optimal inference, and habit formation

Thomas H. B. FitzGerald*, Raymond J. Dolan and Karl J. Friston

Wellcome Trust Centre for Neuroimaging, UCL Institute of Neurology, University College London, London, UK

Edited by:
Javier Bernacer, University of Navarra, Spain

Reviewed by:
Vincent De Gardelle, Université Paris Descartes, France
Samuel Joseph Gershman, Princeton University, USA

****Correspondence:***
Thomas H. B. FitzGerald, Wellcome Trust Centre for Neuroimaging, UCL Institute of Neurology, University College London, 12 Queen Square, London, WC1N 3BG, UK
e-mail: thomas.fitzgerald@ucl.ac.uk

Postulating that the brain performs approximate Bayesian inference generates principled and empirically testable models of neuronal function—the subject of much current interest in neuroscience and related disciplines. Current formulations address inference and learning under some assumed and particular model. In reality, organisms are often faced with an additional challenge—that of determining which model or models of their environment are the best for guiding behavior. Bayesian model averaging—which says that an agent should weight the predictions of different models according to their evidence—provides a principled way to solve this problem. Importantly, because model evidence is determined by both the accuracy and complexity of the model, optimal inference requires that these be traded off against one another. This means an agent's behavior should show an equivalent balance. We hypothesize that Bayesian model averaging plays an important role in cognition, given that it is both optimal and realizable within a plausible neuronal architecture. We outline model averaging and how it might be implemented, and then explore a number of implications for brain and behavior. In particular, we propose that model averaging can explain a number of apparently suboptimal phenomena within the framework of approximate (bounded) Bayesian inference, focusing particularly upon the relationship between goal-directed and habitual behavior.

Keywords: predictive coding, Bayesian inference, habit, interference effect, active inference

INTRODUCTION

The idea, first articulated by Helmholtz, that agents perform inference based on a generative model of the world, is the subject of much recent interest in theoretical and experimental neuroscience (Gregory, 1980; Dayan et al., 1995; Rao and Ballard, 1999; Summerfield and Egner, 2009; Friston, 2010; Clark, 2012). In this framework, given a particular model of the world, an agent needs to perform both *inference* about hidden variables and *learning* about the parameters and hyperparameters of the model (**Figure 1**)—processes that are the focus of much recent study (Friston, 2010; Moran et al., 2013). An equally important consideration however, is determining what model an agent should use in the first place (Hoeting et al., 1999; Penny et al., 2007). This gives rise to an additional tier of uncertainty to those customarily treated in the neuroscientific literature (Yu and Dayan, 2005; Bach and Dolan, 2012)—uncertainty over models. Establishing the best model to use is a pressing concern because, in many situations, the causal structure governing the phenomena of interest is unknown or context dependent (Acuña and Schrater, 2010; Penny et al., 2013). A Bayesian agent needs to consider its own uncertainty about which model is best, and make inferences about evidence for different models, a process known as *model comparison* (**Figure 1**).

Despite its manifest importance, how the brain adjudicates among models has received little study thus far (though see Courville et al., 2005; Gershman and Niv, 2012; Penny et al., 2013). We first briefly describe Bayesian model comparison (a fuller account is given in the Supplementary Material, Appendix), noting that it depends upon model evidence, which can be approximated using neurobiologically plausible predictive coding schemes (Friston, 2005; Bastos et al., 2012). Crucially, model evidence can be decomposed into an accuracy component—reflecting how well the model predicts observed data—and a (penalizing) complexity component reflecting the computational cost of the model. Thus, Bayes optimal agents seek both to maximize the accuracy of their predictions *and* to minimize the complexity of the models they use to generate those predictions (Jefferys and Berger, 1992). This allows us to formalize heuristic explanations about selection among different models, based on resource costs or their relative reliability (Daw et al., 2005), within a simple and Bayes optimal framework.

The optimal way in which the predictions of different models can be traded off against one another is given by Bayesian model averaging. It is thus highly plausible that this operation is implemented by the brain. We discuss this, together with the relationship between Bayesian model averaging and a related procedure—Bayesian model selection. We then discuss anatomical and behavioral implications of model averaging, and consider several examples of phenomena that can be parsimoniously accounted for by invoking inference over models as a key component of cognitive function. In particular, we focus on the process of habit formation, where, with repeated experience, agents come to rely on simpler models to govern behavior (Dolan and Dayan, 2013). Casting cognition and behavior in this light allows us to reconcile the manifest advantages of performing optimal inference with apparently contradictory phenomena such as bounded rationality (Simon, 1972; Camerer et al., 2004), interference effects (Stroop, 1935; Tucker and Ellis, 2004), and the

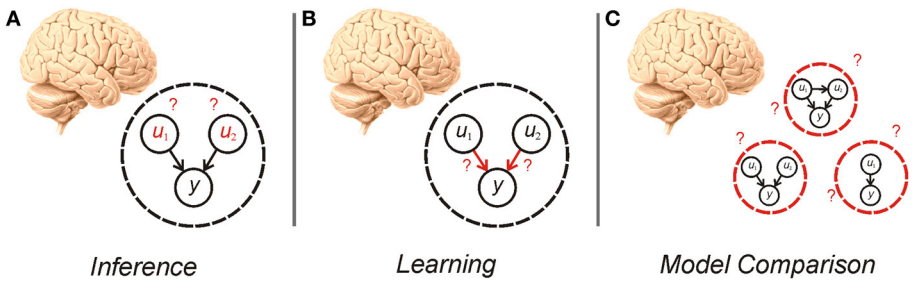

FIGURE 1 | Cartoon illustrating inference (A), learning (B), and model comparison (C). Inference requires an agent to alter its beliefs about the causes ($\boldsymbol{u_1}$, $\boldsymbol{u_2}$) of sensory data ($\boldsymbol{y}$) to maximize model evidence (minimize surprise). Learning also involves the maximization of model evidence, this time through adjustment of the parameters of the model (the mapping between hidden causes and observations). Model comparison involves averaging over—or selecting from—alternative models that can be used for inference and learning.

formation of apparently goal-insensitive habitual behaviors (Yin and Knowlton, 2006).

MODEL EVIDENCE AND MODEL COMPARISON

ESTIMATING THE EVIDENCE FOR A MODEL

We start by outlining the calculations necessary to perform Bayesian model comparison. (these issues are treated more fully in the Supplementary Material, Appendix). First, it is necessary to define a model space containing the set of models $\{m_i : i = 1, \ldots, I\}$ that are to be compared. Now, given a set of observations y, it follows from Bayes theorem that the posterior distribution $p(m_i|y)$ over the set of models is given by:

$$p\left(m_i \,|\, y\right) \propto p\left(y \,|\, m_i\right) p\left(m_i\right) \tag{1}$$

This means that model comparison depends on two quantities, the prior probability of the model $p\left(m_i\right)$, which we will assume here to be equal across models, and the model evidence $p\left(y \,|\, m_i\right)$. This is a key result because the model evidence $p\left(y \,|\, m_i\right)$ is exactly the quantity that is maximized by approximate Bayesian inference and learning. Thus, any agent that performs inference and learning using a particular model of the world necessarily evaluates (implicitly or explicitly) the exact quantity necessary to compare it with other models.

The central importance of model evidence for comparing different models has another important consequence that it is useful to highlight here. Because the model evidence (and approximations to it such as the variational free energy or Bayesian information criterion) contain accuracy and (penalizing) complexity terms (see Supplementary Material, Appendix), the posterior probability of different models also reflects a trade-off between accuracy and complexity. This means that agents will tend to favor simple models, provided they are accurate and, as we shall argue below, this can provide a normative explanation for processes such as habit formation.

Scoring models on more than just the accuracy of their predictions may at first glance seem paradoxical, but in fact the use of a complexity penalty (sometimes called an "Occam factor") is crucial for optimal inference. This is because it prevents overfitting, a situation where an overly complex model becomes sensitive to noise in the data, limiting its generalization or predictive power for future observations [for a clear discussion of this see (Bishop, 2006) Chapters 1 and 3]. From another perspective, minimizing complexity corresponds to the principle of Occam's razor, where parsimony mandates postulating no more degrees of freedom than are required by the evidence (Jefferys and Berger, 1992).

MODEL AVERAGING AND MODEL SELECTION

We now turn to the question of how an agent should use information from multiple models of its environment. The optimal way in which it can use the predictions of different models is to create a weighted average, with the weight determined by the posterior probability $p(m_i|y)$ of each model (**Figure 2**). This is known as Bayesian model averaging (Hoeting et al., 1999; Attias, 2000; Penny et al., 2007). Intuitively, model averaging is optimal because it uses all available information, weighted according to its reliability, and in this sense it is closely related to optimal integration of information within a single model (Ernst and Banks, 2002). Furthermore, it properly accommodates uncertainty over models in situations where there is no predominant model to call on.

Bayesian model averaging is often contrasted with Bayesian model selection, in which only the best model is used (Stephan et al., 2009). This is suboptimal, but provides a close approximation to model averaging when one model is strongly favored over the rest. In fact, model averaging can always be converted into model selection, as can be seen by changing the softmax parameter implicit in Bayesian model averaging (see Supplementary Material, Appendix). In other words, if one is sufficiently sensitive to differences in model evidence, Bayesian model averaging and selection will yield the same results. This raises the fascinating possibility that, under appropriate conditions, agents can vary the sensitivity of the model comparison they perform (see Model Averaging and Perception). This sensitivity also represents a potential computational phenotype underlying individual differences in normal and pathological behavior.

FREE ENERGY AND PREDICTIVE CODING

For certain cases, such as linear Gaussian models, the model evidence can be calculated analytically, but in general its computation is intractable. This necessitates approximate inference, most commonly implemented either using variational methods

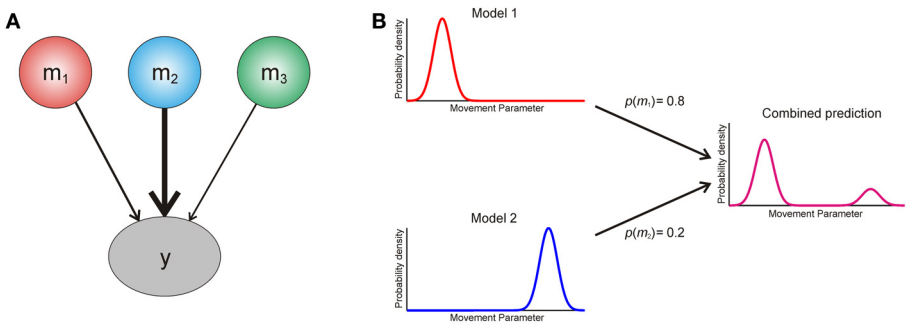

FIGURE 2 | (A) Graphical illustration of Bayesian model averaging. To generate a single Bayes optimal prediction about data *y*, the predictions of three models m_{1-3} are weighted according to their posterior probabilities [see Equation (A5)]. Here model two has the largest posterior probability, and thus its prediction is weighted most strongly. **(B)** Cartoon explaining interference effects using model comparison. An agent entertains two models of the world, which make different predictions about the probability of making an action based on some movement parameter (x axis). The model probabilities for these are $\boldsymbol{p}(\boldsymbol{m_1}) = \mathbf{0.8}$ and $\boldsymbol{p}(\boldsymbol{m_2}) = \mathbf{0.2}$ respectively, and the resulting weighted prediction (magenta) shows an interference effect based on this weighted averaging [see Equation (A5)].

or sampling [for example Markov Chain Monte Carlo or particle filtering (Bishop, 2006)]. We focus on variational inference here, because it is fast and can (in principle) be implemented within neuronal architectures (Mumford, 1992; Friston, 2005), making it a plausible account of brain function (Friston, 2005; Friston et al., 2013). Here, the model evidence is approximated by the variational free energy, which is minimized during learning and inference (**Figure 1**). It is easy to see (see Supplementary Material, A4 "Free Energy and Model Averaging") that model comparison can be performed simply by minimizing the variational free energy across a set of models, suggesting that it could be implemented by the brain.

The most popular and developed account of how the brain might perform variational inference is predictive coding, using hierarchical generative models embodied in the hierarchical structure of the brain (Mumford, 1992; Rao and Ballard, 1999; Friston, 2005, 2008; Bastos et al., 2012) (see Supplementary Material, A5 "Hierarchical Models and Predictive Coding"). Here, model comparison is performed by minimizing the precision-weighted sum of squared prediction errors across a set of models. On this account, if the brain entertains different models of its environment, then these need to make converging top-down predictions of representations in the cortical hierarchy. In some cases, this target might be in primary sensory areas, but it also seems likely that different models may make convergent predictions about higher level representations (the presence or absence of whole objects, for example). A plausible candidate mechanism for weighting the predictions of different models is modulation of the synaptic efficacy of their top-down predictions, either through synchronous gain or through neuromodulators like dopamine. This is an important implementational issue, and one we hope to consider more fully in future work—especially in light of the somewhat surprising finding that at the level of behavior dopamine boosts the influence of complex models at the expense of simpler ones (Wunderlich et al., 2012b).

In summary, we are suggesting that representations at any level of a hierarchical (predictive coding) model are optimized using top-down predictions that represent a Bayesian model average. These predictions are simply the posterior predictions of any given model weighted by posterior beliefs about the model *per se*—beliefs that are directly related to the free energy of each model.

RELATED WORK

A similar approach to the Bayesian model comparison and averaging described here has been employed in the context of supervised learning in mixture of expert models (Jacobs et al., 1991a,b; Jordan and Jacobs, 1994). These consist of a set of expert networks, the outputs of which are weighted by a gating network and combined according to some fixed rule (Jacobs, 1995), which can then be used for classification. Our proposal also bears some resemblance to the MOSAIC model for motor behavior proposed by Kawato and colleagues (Haruno et al., 2001). In MOSAIC, agents are equipped with multiple control modules, which consist of paired forward (predictor) and inverse (controller) models. The weights ("responsibility") assigned to each module depend upon the accuracy of the forward model predictions in a particular context, and are implemented as prior probabilities according to Bayes rule (Haruno et al., 2001). Motor commands are then the responsibility weighted sum of the outputs of the set of inverse models, and—in situations where more than one control module is assigned a significant responsibility—this may produce similar interference effects to those described above. Compared with both these approaches (at least as they are typically formulated), Bayesian model averaging has the advantage that it considers model evidence, rather than simply model accuracy, and thus meets the demands of optimal inference. In the specific domain of motor control, we note that active (Bayesian) inference formulations require only a single generative model, rather than paired inverse and forward models (Friston, 2011).

Bayesian model averaging itself has been considered in theories of Bayesian conditioning (Courville et al., 2003, 2005); in which models with different numbers of latent causes are entertained by the agent—and their predictions weighted according to the evidence for the different models as in Equation (A5). An interesting and related approach is taken by Gershman and Niv (2012) where

instead of averaging the predictions of different models, agents implement a Bayesian non-parametric model (Rasmussen and Ghahramani, 2002; Gershman and Blei, 2012), whose complexity adjusts automatically to the data in hand. These proposals are very close in spirit to the idea presented here, and we note their ability to account for a number of phenomena that are difficult to explain using traditional conditioning models like Rescorla-Wagner learning (Courville et al., 2003, 2005; Gershman and Niv, 2012). It has also recently been proposed that spatial cognition can be explained using approximate Bayesian inference (Penny et al., 2013). In this context, different models correspond to different environments, and thus model comparison can be used as a natural way to perform inference about which environment an agent finds itself in Penny et al. (2013).

MODEL AVERAGING AND THE BRAIN

Here, we briefly consider the implications of Bayesian model averaging for neuroanatomy and development. Much more can (and needs) to be said about this, but our principal focus here is on cognition and behavior, so we will restrict ourselves to some key points:

ANATOMY AND DEVELOPMENT

If agents entertain several models of their environment, in many cases these are likely to co-exist within the same anatomical region. For example, one might imagine that—on encountering a new maze—the hippocampus contains models with many different spatial structures (Blum and Abbott, 1996; Penny et al., 2013), or in other situations that the prefrontal cortex models and compares the evidence for different rules simultaneously (Wallis et al., 2001; Koechlin and Summerfield, 2007). It also seems likely however, given the degree of functional specialization seen in the brain (Zeki et al., 1991)—which itself may arise as a result of approximate Bayesian inference (Friston, 2005; Friston et al., 2013)—that model averaging may call on models encoded in different brain structures (Daw et al., 2005; Graybiel, 2008). One instance of this may underlie the distinction between goal-directed and habitual behavior (Yin and Knowlton, 2006), which we consider in more detail below (for detailed review see Dolan and Dayan, 2013). Another (perhaps related) example might be the apparent competition between hippocampal (largely spatial) and striatal (largely cue-based) mechanisms during instrumental learning (Lee et al., 2008). In general, given that the space of possible models for any situation is potentially uncountable, it makes sense that both evolution and prior experience should act to narrow the space of models entertained, and that particular constraints, such as what features of the environment are considered in the model, should be instantiated in different structures. One can thus think of the brain as performing *selective model averaging* (Heckerman, 1998).

The need to consider different models of the world also provides an interesting perspective on neurodevelopment. Analogous to the way in which model parameters are thought be learnt during development (Fiser et al., 2010; Berkes et al., 2011), one might hypothesize that the posterior distribution over models $p\left(m_i \mid y\right)$ becomes increasingly peaked, as learning the best models proceeds. One might further suppose that some form of Occam's window is applied by the brain, in which models below a certain posterior probability are discarded entirely (Madigan and Raftery, 1994). This makes sense in terms of metabolic and other costs and might, in part, explain the decline in cortical volume that occurs with normal ageing (Salat et al., 2004)—since over time agents come to entertain fewer and fewer models. Different degrees of sculpting model space (or else differences in the number or types of models entertained) might then explain regional differences in synaptic regression, such as the observation that neurodevelopmental regression is most pronounced in the prefrontal cortex (Salat et al., 2004). Recently, synaptic regression during sleep has been portrayed in terms of model optimization. In this context, the removal of unnecessary or redundant in synaptic connections (model parameters) minimizes free energy by reducing model complexity (Hobson and Friston, 2012).

FREE ENERGY AND RESOURCE COSTS

A widely invoked constraint on the type and complexity of models that animals might build of the world is that imposed by resource or complexity costs. This fits comfortably with minimizing variational free energy—that necessarily entails a minimization of complexity (under accuracy constraints). The link between minimizing thermodynamic free energy and variational free energy has again been discussed in terms of complexity minimization—in the sense that thermodynamic free energy is minimized when complexity is minimized (Sengupta et al., 2013): neuronal activity is highly costly from a metabolic point of view (Laughlin et al., 1998) and for any given phenotype, only a certain volume of neurons (and space) are available within the central nervous system. It is fairly easy to see that—under plausible assumptions about how generative models are implemented neuronally—there will be a high degree of correlation between the complexity of a model and the resource costs of implementing it. Heuristically, having a larger number of models or model parameters would require a larger network of neurons to encode it, which will induce both metabolic and anatomical costs. Another heuristic follows if we assume that the brain uses a predictive coding scheme with explicit biophysical representation of prediction errors. In this context, minimizing the variational free energy will serve to reduce overall neuronal activity (prediction error) and hence metabolic demands. This is because predictive coding minimizes prediction errors throughout the models hierarchy.

While other factors are undoubtedly going to influence the computational cost to an organism of implementing a particular model (there is likely, for example, to be a complex interplay between complexity and different types of cost like time and space costs), there is likely to be a strong relationship between complexity costs (as assessed by the variational free energy) and the metabolic costs to an organism (Sengupta et al., 2013).

MODEL AVERAGING AND MULTIPLE-SYSTEMS MODELS OF DECISION-MASKING

A recurring theme in theoretical approaches to human decision-making is that multiple mechanisms are involved in control of behavior, and there is a considerable body of evidence in support of such ideas (Kahneman, 2003; Summerfield et al., 2011; Dolan and Dayan, 2013). We suggest that rather than entirely separate systems competing for control of behavior, the phenomena

motivating this tradition can be captured by a view in which anatomically and functionally dissociable networks embody different types of model [which will often have different hierarchical depths—and hence complexity (Kiebel et al., 2008)]. Instead of simple competition behavior can be thought of as resulting from Bayesian model averaging over the predictions of different models. This perspective provides a way to ease the tension between the insight (which goes back at least as far as Plato's tripartite soul) that multiple motivations can be discerned in human behavior, and the manifest advantages of being able to act in a unitary and coherent fashion, particularly if this is approximately Bayes-optimal. We discuss this briefly below, focusing particularly on the interplay between simple and complex models in the control of behavior.

HABITUAL AND GOAL-DIRECTED BEHAVIOR

It is well established that animals exhibit both goal-directed behavior, in which action selection is flexible and sensitive to anticipated outcomes, and habitual behavior that is stereotyped and elicited directly by a preceding stimulus or context (Yin and Knowlton, 2006; Graybiel, 2008; Dolan and Dayan, 2013). It has also been shown that the neural substrates of these behaviors are at least partially dissociable (Adams and Dickinson, 1981; Hatfield and Han, 1996; Pickens et al., 2003; Izquierdo et al., 2004; Yin et al., 2004).

Broadly speaking, two mechanisms have been proposed to explain the emergence of habitual behavior. The first posits the existence of separate "model-free" and "model-based" reinforcement learning schemes in different parts of the brain (the dorsolateral striatum and prefrontal cortex) (Daw et al., 2005) that support habitual and goal-directed behavior respectively (Dolan and Dayan, 2013). Which of these two systems controls behavior is determined by their relative uncertainties (Daw et al., 2005), and the emergence of habitual behavior over time results from the model-free system having an asymptotically lower uncertainty than the model-based system. A second hypothesis (though one rarely spelled out explicitly) is that the existence of habits reflects a need to minimize some form of computational, metabolic or attentional cost (Moors and De Houwer, 2006). Once an action has been repeated many times, it comes to be elicited automatically by a particular stimulus or context, removing the need for costly deliberation (these explanations may not be entirely separate from one another, since, as pointed out by one of our reviewers, one reason for the presence of significant noise in the model-based system could be the resource cost of performing complex searches).

Both these hypotheses have much to recommend them, but neither provides a wholly satisfactory account of habit formation. To take the "arbitration by uncertainty" hypothesis first; while the insight that different models of the environment should be traded off against one another—through the accuracy of their predictions—is important, this seems insufficient to explain a transition to habitual behavior in many situations. More specifically, in most (if not all) habit learning experiments, the environment that the agent has to represent is extremely simple (pressing a lever to gain a food pellet, knowing whether to turn left or right in a cross maze). In such contexts it seems *prima facie* implausible that explicit cognitive representations induce a sufficiently large degree of uncertainty so as to be dominated by simple ones [we note that the transition to habitual behavior in Daw et al.'s simulations requires that an arbitrary noise component be used to inflate the uncertainty of the model-based scheme (Daw et al., 2005)]. We suggest that differential uncertainty alone is insufficient to provide a satisfying account of the emergence of habitual behavior. The "cost" hypothesis, by contrast, is inadequate as things stand, because it does not specify in what situations the increased resources necessary for an explicit representation of the environment are justified (or conversely, when the cost of extra complexity is too high).

An alternative hypothesis is that habit formation comes about as the result of Bayesian model averaging *between* simple (hierarchically shallow) models and more complicated ones involving richer (hierarchically deep) and more flexible representations of the environment (Kiebel et al., 2008; Wunderlich et al., 2012a). The critical observation is that in Bayesian model comparison models are scored according to both their accuracy and complexity. This means that whilst initially behavior is based largely upon complex models, that are able to generate accurate predictions based on little or no experience, over time simpler models come to predominate, provided their predictions are sufficiently accurate. This will be the case in the stable environments that support habit formation (**Figure 3**). Bayesian model averaging therefore provides a principled framework that incorporates the insights of both uncertainty- and cost-based explanations, and remedies their defects. On the one hand, model comparison explains why habit formation occurs even in very simple environments that are unlikely to induce significant uncertainty in explicit cognitive representations. The use of simple models will always be favored by the brain, provided those models are accurate *enough*. Informally, this may explain why it is so difficult to suppress learnt habits and other forms of simple stimulus-response behaviors, such as the tendency to approach appetitive stimuli and avoid aversive ones (Guitart-Masip et al., 2011). Very simple models have a very low complexity cost, which means they do not have to be especially accurate in order to be selected for prescribing behavior. On the other hand, invoking model comparison allows us to precisely specify the currency in which different models should be traded off against one another, and provide (in theory at least) a precise account of when increased complexity is justified by increased accuracy, and *vice versa*.

What then, would constitute evidence for the model averaging hypothesis? The strongest grounds, perhaps, are those already described—the extensive body of work characterizing the emergence of habitual behavior, which seems best captured by a view that makes allowance for both model accuracy and model complexity. However, some important recent work using model-based neuroimaging also provides strong support for our hypothesis (Daw et al., 2011; Lee et al., 2014). Both these studies involve asking subjects to perform moderately complex learning tasks, where behavior reflected a combination of both simple (stimulus-response or model-free like) and more complicated (action-outcome or model-based like) models of the environment. Similar findings have been reported by Wunderlich et al. (2012b), Otto et al. (2013) and Smittenaar et al. (2013). In the

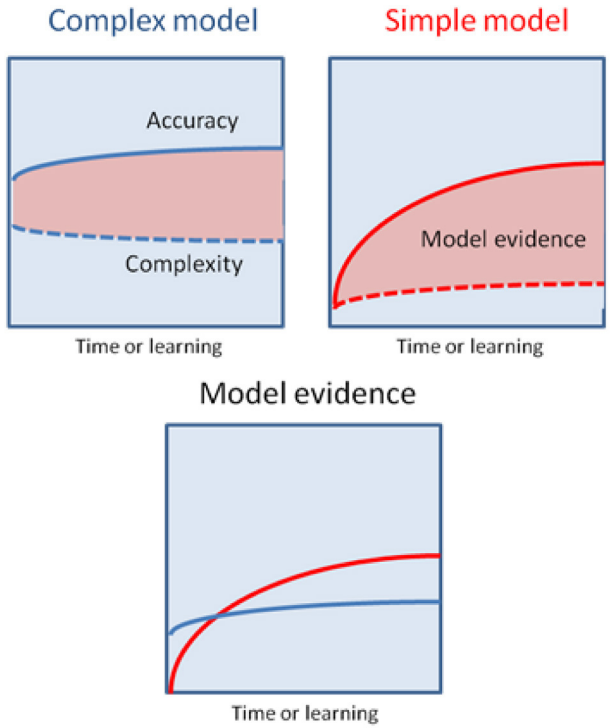

FIGURE 3 | This schematic illustrates the possibility that a more complex model may have greater model evidence at the start of learning but will then give way to a simpler model as their parameters are optimized. The upper panels show the learning-related improvement in accuracy and complexity for a complex model (left panel) and a simple model (right panel). The model evidence is shown as the difference (pink areas). The more complex model explains the data more accurately but with a greater complexity cost, that is finessed during the learning. Conversely, the simpler model will always have a lower accuracy but can (with learning) attain greater model evidence—and thereby be selected by Bayesian model averaging as time proceeds and the active inference becomes habitual.

context of such tasks, model averaging makes two clear predictions. The first is that the control of behavior will be biased toward simple models, once the effects of uncertainty are accounted for. The second is that because the predictions of simple and complex models are unified, there should be evidence of unified (and appropriate weighted) prediction error signals in the brain.

It turns out that both these predictions are borne out by the experimental data. The behavioral modeling presented in Lee et al. strongly suggests that subjects show a bias toward relying on simple models over complex ones (the model-free system over the model-based one in the terminology they employ) (Lee et al., 2014). This is exactly what one would expect if both complexity and accuracy are taken into account. (Daw et al. did not report the results of any similar analysis). Turning to the second prediction Lee et al. report evidence that value signals derived from simple and complex models are integrated in a contextually appropriate way in the ventromedial prefrontal cortex (Lee et al., 2014). Equally importantly, rather than finding separate prediction error signals at outcome presentation for the simple and complex models, Daw et al. instead reported an integrated signal in the ventral striatum, with the strength of expression of the different prediction errors correlated with the relative influence they had over behavior (Daw et al., 2011). Both these findings are precisely in accord with the view that the predictions of simple and complex models are subject to Bayesian model averaging during decision-making. Clearly, the explanation for habit formation on offer is a hypothesis that will need to be tested using simulations and empirical studies; for example, using devaluation paradigms of the sort addressed in Daw et al. (2005)—as suggested by one of our reviewers.

HABITS AND BEHAVIORAL FLEXIBILITY

The view of habit formation presented here is also consistent with recent discussions that have stressed the flexibility of habitual behavior, and the complex relationship between habitual and goal-directed action (Bernácer and Giménez-Amaya, 2013; Bernácer et al., 2014). Although habitual behavior results from the use of hierarchically shallow models that do not include information about the higher order goals of an organism, they can, under appropriate conditions, instantiate complex links between external stimuli and behavior of the type manifest when performing tasks like driving or playing the piano, rather than just simple stimulus-response mappings. Using shallow models to perform a particular task also frees up neuronal circuits at deeper hierarchical levels, potentially enabling them to be employed in other tasks. Thus, whilst habit formation reduces the flexibility of behavior on a particular task, it may simultaneously increase the overall behavioral repertoire available to the agent. For example, whilst it is difficult for people in the early stages of learning to drive to simultaneously hold a conversation, experienced drivers find this easy. This raises the interesting possibility that, rather than always being antithetical to goal-directed behavior, considered from the perspective of the entire agent, habit formation often enables it. A Bayesian perspective also provides an explanation for how habitual behaviors can be at the same time apparently unconscious and automatic, and yet also rapidly become subject to conscious awareness and goal-directed control when something unexpected occurs (if the brake pedal of the car suddenly stops working, for example) (Bernácer et al., 2014). This occurs because the shallow model generating habitual control of behavior suddenly becomes a poor predictor of current and future sensory information, necessitating the switch to a more complex, flexible model.

INTERFERENCE EFFECTS, AFFORDANCES, AND PAVLOVIAN RESPONSES

It has been well documented that human behavior, across a wide variety of domains, shows evidence of what are usually called "interference effects" (Stroop, 1935; Simon et al., 1990; Tipper et al., 1997; Tucker and Ellis, 2004; Guitart-Masip et al., 2011). Typically, these are manifest when subjects are asked to make responses based on one attribute or dimension of a stimulus, but show behavioral impairments, such as slower responding or increased error rates, that can be attributed to a different attribute. Examples of this include the affordance compatibility effect (Tucker and Ellis, 2004), the "Pavlovian" tendency to approach appetitive and avoid aversive stimuli (Dayan, 2008; Guitart-Masip et al., 2011; Huys et al., 2011) and the effect

of distractors during reaching (Tipper et al., 1997; Welsh and Elliott, 2004). A closely related phenomenon is that of task switching effects, where subjects' performance is impaired immediately after being asked to swap between performing different tasks (Monsell, 2003).

These effects are generally considered to result from the existence of multiple mechanisms for controlling action (or alternatively, task sets) engaged in more or less blind competition (Dayan, 2008), a scenario virtually guaranteed to produce suboptimal behavior. The arguments presented here suggest another possibility; namely, that such phenomena are the manifestation of agents pursuing a model averaging strategy that is in general optimal, but produces suboptimal behavior in the context of non-ecological experiments (**Figure 2**). There is a natural parallel with perceptual illusions here, since these result from the application of generally appropriate prior beliefs to situations designed such that these beliefs are inappropriate (Weiss et al., 2002; Shams et al., 2005; Brown and Friston, 2012). To return to the affordance competition and Pavlovian bias effects mentioned above, it seems reasonable to suppose that subjects simultaneously call on a model of their environment induced by the (non-ecological) task demands, and an entrenched (and simpler) model linking stimulus properties like object affordances and stimulus valence to behavioral responding. Since the predictions of these models are averaged, the influence of the simpler models is suppressed, but not entirely attenuated, producing characteristic effects on behavior (**Figure 3**). This is a hypothesis we will consider more fully in future work. Task switching effects can also naturally be explained, on the hypothesis that models that have recently provided accurate predictions have been accorded a higher posterior probability that is only partially suppressed during switching.

MODEL AVERAGING IN OTHER COGNITIVE DOMAINS

We now turn to considering the consequences of, and evidence for, Bayesian model comparison and averaging in other areas of cognition. We confine our discussion to a small number of examples but we suspect that these ideas may have much broader applicability to other cognitive domains (and perhaps beyond (Friston, 2010, 2012)).

MODEL AVERAGING AND PERCEPTION

In certain contexts, perception does indeed show the hallmark of model averaging, namely integration between the predictions of different plausible models. Famous examples of this include the McGurk and ventriloquist effects (McGurk and MacDonald, 1976; Bertelson et al., 2000), in which distinct representations (for example of phonemes in the McGurk effect) are fused into a single percept that is a combination of the two. However, there is also a large literature describing multistability in perception, for example in the face-vase illusion and the Necker cube (Sterzer et al., 2009). Here distinct hypotheses about the world clearly alternate rather than co-existing (Dayan, 1998; Hohwy et al., 2008). A natural explanation for this in the framework we have suggested here is that agents perform apply model averaging with a high sensitivity parameter (see Supplementary Material, A2 "Bayesian Model Averaging"). This effectively implements Bayesian model selection, and ensures that only the predictions of a single preferred model are used. Other explanations are also possible, for example that multistability results from sampling from different models (Gershman et al., 2012) or, as suggested by one of our reviewers, from strong negative covariance between the prior probabilities of different models.

It is unclear precisely why—in some contexts—perception should exhibit integration, and in others multistability, but one attractive possibility is that this is determined by the extent to which an integrated percept is, in itself, plausible. Thus the fused percepts produced by the McGurk and ventriloquist illusions reflect plausible hidden states of the world. By contrast, the intermediate state of a Necker cube, or Rubin's face-vase illusion would be implausible, if not impossible; suggesting that in these contexts agents should preclude perceptual integration by increasing the strictness of their model comparison.

EXPERIENCE AND BOUNDED RATIONALITY

Although in some (particularly perceptual) contexts, human behavior closely approximates the best possible performance (Ernst and Banks, 2002), in many situations it falls well short of this, giving rise to the suggestion that humans are bounded rational decision-makers (Simon, 1972; Kahneman, 2003) rather than perfectly rational; particularly when it comes to economic choice. Bounded rationality means that decision-making is as good as possible, given constraints of one kind or another. A phenomenon is found in theories of social interaction, where it has been shown that humans are able to consider only a (perhaps surprisingly) limited number of levels of recursion on interpersonal choice tasks (Stahl and Wilson, 1995; Camerer et al., 2004; Yoshida et al., 2008; Coricelli and Nagel, 2009).

These specific examples illustrate a more general point. If models are weighted or chosen according to their evidence rather than simply their accuracy, then one should not necessarily expect agents to perform tasks with extremely high levels of accuracy even if they are Bayes optimal. This is because approximate Bayesian inference naturally introduces bounded rationality, since it trades off accuracy (rationality) against complexity (cost). On this view, there are two key determinants of whether agents employ complex models (and hence approximate ideal behavior on tasks where these are necessary). The first is the amount of experience the agent has with a particular task or environment. More experience (equivalent to collecting a large data set in a scientific experiment) allows the increased accuracy of its predictions to outweigh the complexity penalty of a complex model (Courville et al., 2003). The second determinant is the gain in accuracy per observation associated with using the more complex model. This picture fits, at least approximately with what is actually observed in human behavior, where near-ideal performance is often observed in perceptual tasks (which presumably employ models that are used extremely frequently) and suboptimal performance more typically seen in tasks such as abstract reasoning, which are performed less often.

This perspective relates to recent work showing that bounded rationality can be derived from a free energy formulation, where model complexity is introduced by the need to process information in order to perform inference (Ortega and Braun, 2013). Model comparison, as performed by gradient ascent on

variational free energy, supplements this insight by explaining how the Bayes-optimal model of the environment arises.

OTHER ISSUES

WHERE DOES THE MODEL SPACE COME FROM?

One issue we have not touched on is how models are created in the first place. This is a deep and challenging topic, whose proper consideration falls outside the scope of this piece. One easy answer is that the space of possible models is constrained by phylogeny and thus ultimately by natural selection, which can itself be thought of in terms of free energy minimization (Kaila and Annila, 2008). From the perspective of neuroscience, this is at the same time true and unsatisfying. To understand how new models are generated within the lifetime of an organism (and *a fortiori* on the timescale of laboratory experiments), it is interesting to consider structure learning (Heckerman, 1998; Needham et al., 2007; Braun et al., 2010; Tenenbaum et al., 2011). Structure learning deals with the problem of how to infer dependencies between hidden variables, and allows inferences to be drawn about both the specific model structure (Heckerman, 1998; Tenenbaum et al., 2011) and the general structural form (for example a ring, tree or hierarchy) (Kemp and Tenenbaum, 2008) most appropriate for a dataset. From our perspective, this is simply the problem of Bayesian model selection applied to probabilistic graphical models. This approach has been used with remarkable success to explore inductive learning and concept acquisition (Tenenbaum et al., 2011). The issue of how to select the correct hidden variables in the first place has been less well explored, at least in cognitive science (though see Collins and Koechlin, 2012; Gershman and Niv, 2012), but one solution to this problem is provided by Bayesian non-parametric models that entertain, in principle, an infinite model space (Rasmussen and Ghahramani, 2002; Gershman and Blei, 2012).

A clear prediction of structure learning models is that previously acquired structures may be utilized on novel tasks, as manifested by "learning to learn," where new tasks with the same structure as previously experienced ones are learnt faster. This pattern of behavior has been repeatedly demonstrated in animal experiments (Harlow, 1949; Schrier, 1984; Langbein and Siebert, 2007), as well as those involving human children and adults (Duncan, 1960; Hultsch, 1974; Brown and Kane, 1988; Halford et al., 1998; Acuña and Schrater, 2010), as usefully reviewed in Braun et al. (2010). The same phenomenon has also been rediscovered recently by memory researchers, and described in terms of cognitive schema (Tse et al., 2007; van Kesteren et al., 2012). This means that, given the constraints of their phenotype, adult organisms are likely to have already acquired a large number of possible structures (Kemp and Tenenbaum, 2008; Tenenbaum et al., 2011), which they can use to model the world, and model comparison can thus proceed considering only this reduced model space.

SIGNATURES OF MODEL COMPARISON

An interesting practical question is how we distinguish between separate models, and different parts of a single more complicated model. This is particularly pertinent, because as we have discussed elsewhere (see Supplementary Material, A4 "Free Energy and Model Averaging"), performing variational inference on model probabilities effectively involves embedding them within a larger hierarchical model. On one level, this question is a philosophical one, but in the context of specific cognitive or neuronal hypotheses we take it that what is useful to consider as separate models will generally be fairly clear in terms of functional anatomy [for example, the anatomical dissociation between the neuronal mechanisms underlying goal-directed and habitual behavior discussed earlier (Yin and Knowlton, 2006)]. More concretely, we can point to the fact that complexity plays a key role in adjudicating among different models, but not when weighting different kinds of information within a model (Deneve and Pouget, 2004), and suggest that if behavior shows clear evidence of a bias toward using simple models (as in habit-formation), then this is evidence that model evidence is being used to optimize behavior.

ACTIVE SAMPLING AND MODEL COMPARISON

Although—for the sake of simplicity—we have only considered static models in our theoretical discussion, the principles outlined can be easily extended to incorporate extended timeframes and dynamics by minimizing the path-integral of the variational free energy (or the action) over time (Feynman, 1964; Friston, 2008; Friston et al., 2008). Given a particular model, this leads naturally to active sampling of the world in such a way as to minimize uncertainty about its parameters (hypothesis testing) (Friston et al., 2012a). In the context of uncertainty over models, a similar process should occur; with agents actively sampling sensory data in order to disambiguate which model of the world (hypothesis) is best [a beautiful example of this is Eddington's test of general relativity using gravitational lensing (Dyson et al., 1920)]. This notion is supported by recent work showing that in a sequential decision-making context, human subjects trade off reward minimization against gaining information about the underlying structure of the task (Acuña and Schrater, 2010).

MODEL COMPARISON AND PSYCHOPATHOLOGY

A number of psychiatric disorders are associated with symptoms such as delusions and hallucinations which seem likely to reflect dysfunctional models of their environment (Fletcher and Frith, 2008; Adams et al., 2013; Brown et al., 2013). In some cases this might be the product of pathological learning of the parameters of particular models, but it is also conceivable that impairments in the ability to adequately compare models (to make or utilize inferences about model probabilities) might underlie some deficits. This is also a promising area for future study.

SUMMARY

In this paper we suggest, based on both theoretical grounds and consideration of behavioral and neuroscientific evidence, that the brain entertains multiple models of its environment, which it adjudicates among using the principles of approximate Bayesian inference. We discussed these principles, which can be implemented in a neurobiologically plausible way using predictive coding (Friston, 2005). Finally, we argue that a number of disparate behavioral and neuroscientific observations are well explained by invoking Bayesian model averaging, focusing particularly on habitual vs. goal-directed control, and why simple

models often prevail over more sophisticated ones. We anticipate that this perspective may be useful for hypothesis generation and data interpretation across a number of fields treating both normal function and psychiatric disease.

ACKNOWLEDGMENTS

This work was supported by Wellcome Trust Senior Investigator Awards to Karl J. Friston [088130/Z/09/Z] and Raymond J. Dolan [098362/Z/12/Z]; The Wellcome Trust Centre for Neuroimaging is supported by core funding from Wellcome Trust Grant 091593/Z/10/Z.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fnhum.2014.00457/abstract

REFERENCES

Acuña, D. E., and Schrater, P. (2010). Structure learning in human sequential decision-making. *PLoS Comput. Biol.* 6:e1001003. doi: 10.1371/journal.pcbi.1001003

Adams, C., and Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *Q. J. Exp. Psychol.* 33, 109–121. doi: 10.1080/14640748108400816

Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D., and Friston, K. J. (2013). The computational anatomy of psychosis. *Front. Psychiatry* 4:47. doi: 10.3389/fpsyt.2013.00047

Attias, H. (2000). A variational Bayesian framework for graphical models. *Adv. Neural Inf. Process. Syst.* 12, 209–215.

Bach, D. R., and Dolan, R. J. (2012). Knowing how much you don't know: a neural organization of uncertainty estimates. *Nat. Rev. Neurosci.* 13, 572–586. doi: 10.1038/nrn3289

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76, 695–711. doi: 10.1016/j.neuron.2012.10.038

Beal, M. J. (2003). *Variational Algorithms for Approximate Bayesian Inference*. Ph.D. thesis, University of London, London.

Beal, M. J., and Ghahramani, Z. (2003). The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures. *Bayesian Stat.* 7, 453–464.

Berkes, P., Orban, G., Lengyel, M., and Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science* 331, 83–87. doi: 10.1126/science.1195870

Bernácer, J., Balderas, G., Martinez-Valbuena, I., Pastor, M. A., and Murillo, J. I. (2014). The problem of consciousness in habitual decision making. *Behav. Brain Sci.* 37, 21–22. doi: 10.1017/S0140525X13000642

Bernácer, J., and Giménez-Amaya, J. (2013). "On habit learning in neuroscience and free will," in *Is Science Compatible with Free Will?* eds A. Suarez and P. Adams (New York, NY: Springer New York), 177–193. doi: 10.1007/978-1-4614-5212-6_12

Bertelson, P., Vroomen, J., de Gelder, B., and Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Percept. Psychophys.* 62, 321–332. doi: 10.3758/BF03205552

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York, NY: Springer.

Blum, K., and Abbott, L. (1996). A model of spatial map formation in the hippocampus of the rat. *Neural Comput.* 8, 85–93. doi: 10.1162/neco.1996.8.1.85

Braun, D. A., Mehring, C., and Wolpert, D. D. M. (2010). Structure learning in action. *Behav. Brain Res.* 206, 157–165. doi: 10.1016/j.bbr.2009.08.031

Brown, A., and Kane, M. (1988). Preschool children can learn to transfer: learning to learn and learning from example. *Cogn. Psychol.* 20, 493–523.

Brown, H., Adams, R. A., Parees, I., Edwards, M., and Friston, K. (2013). Active inference, sensory attenuation and illusions. *Cogn. Process* 14, 411–427. doi: 10.1007/s10339-013-0571-3

Brown, H., and Friston, K. J. (2012). Free-energy and illusions: the cornsweet effect. *Front. Psychol.* 3:43. doi: 10.3389/fpsyg.2012.00043

Camerer, C. F., Ho, T.-H., and Chong, J.-K. (2004). A cognitive hierarchy model of games. *Q. J. Econ.* 119, 861–898. doi: 10.1162/0033553041502225

Clark, A. (2012). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477

Collins, A., and Koechlin, E. (2012). Reasoning, learning, and creativity: frontal lobe function and human decision-making. *PLoS Biol.* 10:e1001293. doi: 10.1371/journal.pbio.1001293

Coricelli, G., and Nagel, R. (2009). Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9163–9168. doi: 10.1073/pnas.0807721106

Courville, A., Daw, N. D., and Touretzky, D. S. (2005). Similarity and discrimination in classical conditioning: a latent variable account. *Adv. Neural Inf. Process. Syst.* 17, 313–320.

Courville, A., Daw, N. D., Touretzky, D. S., and Gordon, G. J. (2003). Model uncertainty in classical conditioning. *Adv. Neural Inf. Process. Syst.* 16, 977–984.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., and Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69, 1204–1215. doi: 10.1016/j.neuron.2011.02.027

Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711. doi: 10.1038/nn1560

Dayan, P. (1998). A hierarchical model of binocular rivalry. *Neural Comput.* 10, 1119–1135. doi: 10.1162/089976698300017377

Dayan, P. (2008). "The role of value systems in decision making," in *Better than Conscious? Decision Making, the Human Mind, and Implications for Institutions*, eds C. Engel and W. Singer (Frankfurt: MIT Press), 51–70.

Dayan, P., Hinton, G. E., Neal, R., and Zemel, R. (1995). The Helmholtz machine. *Neural Comput.* 7, 889–904. doi: 10.1162/neco.1995.7.5.889

Deneve, S., and Pouget, A. (2004). Bayesian multisensory integration and cross-modal spatial links. *J. Physiol. Paris* 98, 249–258. doi: 10.1016/j.jphysparis.2004.03.011

Dolan, R. J., and Dayan, P. (2013). Goals and habits in the brain. *Neuron* 80, 312–325. doi: 10.1016/j.neuron.2013.09.007

Duncan, C. (1960). Description of learning to learn in human subjects. *Am. J. Psychol.* 73, 108–114. doi: 10.2307/1419121

Dyson, F., Eddington, A., and Davidson, C. (1920). A determination of the deflection of light by the sun's gravitational field, from observations made at the total eclipse of May 29, 1919. *Philos. Trans. R. Soc. Lond. A* 220, 291–333. doi: 10.1098/rsta.1920.0009

Efron, B., and Morris, C. (1973). Stein's estimation rule and its competitors—an empirical Bayes approach. *J. Am. Stat. Assoc.* 68, 117–130. doi: 10.1080/01621459.1973.10481350

Ernst, M. O., and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433. doi: 10.1038/415429a

Feldman, H., and Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Front. Hum. Neurosci.* 4:215. doi: 10.3389/fnhum.2010.00215

Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47. doi: 10.1093/cercor/1.1.1

Feynman, R. P. (1964). "The principle of least action," in *The Feynman Lectures on Physics*, Vol. 2, eds R. P. Feynman, R. B. Leighton, and M. Sands (Reading, MA: Addison-Wesley), 19-1–19-14.

Fiser, J., Berkes, P., Orbán, G., and Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. *Trends Cogn. Sci.* 14, 119–130. doi: 10.1016/j.tics.2010.01.003

Fletcher, P., and Frith, C. (2008). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat. Rev. Neurosci.* 10, 48–58. doi: 10.1038/nrn2536

Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622

Friston, K. (2008). Hierarchical models in the brain. *PLoS Comput. Biol.* 4:e1000211. doi: 10.1371/journal.pcbi.1000211

Friston, K. (2011). What is optimal about motor control? *Neuron* 72, 488–498. doi: 10.1016/j.neuron.2011.10.018

Friston, K. (2012). A free energy principle for biological systems. *Entropy* 14, 2100–2121. doi: 10.3390/e14112100

Friston, K., Adams, R., Perrinet, L., and Breakspear, M. (2012a). Perceptions as hypotheses: saccades as experiments. *Front. Psychol.* 3:151. doi: 10.3389/fpsyg.2012.00151

Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., and Penny, W. (2007). Variational free energy and the Laplace approximation. *Neuroimage* 34, 220–234. doi: 10.1016/j.neuroimage.2006.08.035

Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., and Dolan, R. J. (2013). The anatomy of choice: active inference and agency. *Front. Hum. Neurosci.* 7:598. doi: 10.3389/fnhum.2013.00598

Friston, K., Trujillo-Barreto, N., and Daunizeau, J. (2008). DEM: a variational treatment of dynamic systems. *Neuroimage* 41, 849–885. doi: 10.1016/j.neuroimage.2008.02.054

Friston, K. J. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Friston, K. J., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., et al. (2012b). Dopamine, affordance and active inference. *PLoS Comput. Biol.* 8:e1002327. doi: 10.1371/journal.pcbi.1002327

Gershman, S., and Blei, D. (2012). A tutorial on Bayesian nonparametric models. *J. Math. Psychol.* 56, 1–12. doi: 10.1016/j.jmp.2011.08.004

Gershman, S. J., and Niv, Y. (2012). Exploring a latent cause theory of classical conditioning. *Learn. Behav.* 40, 255–268. doi: 10.3758/s13420-012-0080-8

Gershman, S. J., Vul, E., and Tenenbaum, J. B. (2012). Multistability and perceptual inference. *Neural Comput.* 24, 1–24. doi: 10.1162/NECO_a_00226

Graybiel, A. (2008). Habits, rituals, and the evaluative brain. *Annu. Rev. Neurosci.* 31, 359–387. doi: 10.1146/annurev.neuro.29.051605.112851

Gregory, R. L. (1980). Perceptions as hypotheses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 290, 181–197. doi: 10.1098/rstb.1980.0090

Guitart-Masip, M., Fuentemilla, L., Bach, D. R., Huys, Q. J. M., Dayan, P., Dolan, R. J., et al. (2011). Action dominates valence in anticipatory representations in the human striatum and dopaminergic midbrain. *J. Neurosci.* 31, 7867–7875. doi: 10.1523/JNEUROSCI.6376-10.2011

Halford, G., Bain, J., Maybery, M., and Andrews, G. (1998). Induction of relational schemas: common processes in reasoning and complex learning. *Cogn. Psychol.* 35, 201–245. doi: 10.1006/cogp.1998.0679

Harlow, H. (1949). The formation of learning sets. *Psychol. Rev.* 56, 51–65. doi: 10.1037/h0062474

Haruno, M., Wolpert, D. M., and Kawato, M. (2001). MOSAIC model for sensorimotor learning and control. *Neural Comput.* 13, 2201–2220. doi: 10.1162/089976601750541778

Hatfield, T., and Han, J. (1996). Neurotoxic lesions of basolateral, but not central, amygdala interfere with Pavlovian second-order conditioning and reinforcer devaluation effects. *J. Neurosci.* 16, 5256–5265.

Heckerman, D. (1998). "A tutorial on learning with Bayesian networks," in *Learning in Graphical Models,* ed M. I. Jordan (Dordrecht: Kluwer Academic), 301–354.

Hobson, J. A., and Friston, K. J. (2012). Waking and dreaming consciousness: neurobiological and functional considerations. *Prog. Neurobiol.* 98, 82–98. doi: 10.1016/j.pneurobio.2012.05.003

Hoeting, J., Madigan, D., Raftery, A., and Volinsky, C. (1999). Bayesian model averaging: a tutorial. *Stat. Sci.* 14, 382–417.

Hohwy, J., Roepstorff, A., and Friston, K. (2008). Predictive coding explains binocular rivalry: an epistemological review. *Cognition* 108, 687–701. doi: 10.1016/j.cognition.2008.05.010

Hultsch, D. (1974). Learning to learn in adulthood. *J. Gerontol.* 29, 302–309. doi: 10.1093/geronj/29.3.302

Huys, Q. J. M., Cools, R., Gölzer, M., Friedel, E., Heinz, A., Dolan, R. J., et al. (2011). Disentangling the roles of approach, activation and valence in instrumental and Pavlovian responding. *PLoS Comput. Biol.* 7:e1002028. doi: 10.1371/journal.pcbi.1002028

Izquierdo, A., Suda, R., and Murray, E. (2004). Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *J. Neurosci.* 24, 7540–7548. doi: 10.1523/JNEUROSCI.1921-04.2004

Jacobs, R. A. (1995). Methods for combining experts' probability assessments. *Neural Comput.* 7, 867–888. doi: 10.1162/neco.1995.7.5.867

Jacobs, R. A., Jordan, M. I., and Barto, A. G. (1991a). Task decomposition through competition in a modular connectionist architecture: the what and where vision tasks. *Cogn. Sci.* 15, 219–250. doi: 10.1207/s15516709cog1502_2

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., and Hinton, G. E. (1991b). Adaptive mixtures of local experts. *Neural Comput.* 3, 79–87. doi: 10.1162/neco.1991.3.1.79

Jefferys, W., and Berger, J. (1992). Ockham's razor and Bayesian analysis. *Am. Sci.* 80, 64–72.

Jordan, M. I., and Jacobs, R. A. (1994). Hierarchical mixtures of experts and the EM algorithm. *Neural Comput.* 6, 181–214. doi: 10.1162/neco.1994.6.2.181

Kahneman, D. (2003). Maps of bounded rationality: psychology for behavioral economics. *Am. Econ. Rev.* 93, 1449–1475. doi: 10.1257/000282803322655392

Kaila, V. R., and Annila, A. (2008). Natural selection for least action. *Proc. R. Soc. A Math. Phys. Eng. Sci.* 464, 3055–3070. doi: 10.1098/rspa.2008.0178

Kemp, C., and Tenenbaum, J. B. (2008). The discovery of structural form. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10687–10692. doi: 10.1073/pnas.0802631105

Kiebel, S. J., Daunizeau, J., and Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Comput. Biol.* 4:e1000209. doi: 10.1371/journal.pcbi.1000209

Koechlin, E., and Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends Cogn. Sci.* 11, 229–235. doi: 10.1016/j.tics.2007.04.005

Langbein, J., and Siebert, K. (2007). Learning to learn during visual discrimination in group housed dwarf goats (Capra hircus). *J. Comp. Psychol.* 121, 447–456. doi: 10.1037/0735-7036.121.4.447

Laughlin, S., van Steveninck, R., and Anderson, J. (1998). The metabolic cost of neural information. *Nat. Neurosci.* 1, 36–41. doi: 10.1038/236

Lee, A., Duman, R., and Pittenger, C. (2008). A double dissociation revealing bidirectional competition between striatum and hippocampus during learning. *Proc. Natl. Acad. Sci. U.S.A.* 105, 17163–17168. doi: 10.1073/pnas.0807749105

Lee, S. W., Shimojo, S., and O'Doherty, J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. *Neuron* 81, 687–699. doi: 10.1016/j.neuron.2013.11.028

MacKay, D. (1992). Bayesian interpolation. *Neural Comput.* 4, 415–447. doi: 10.1162/neco.1992.4.3.415

Madigan, D., and Raftery, A. (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *J. Am. Stat. Assoc.* 89, 153. doi: 10.1080/01621459.1994.10476894

Markov, N. T., and Kennedy, H. (2013). The importance of being hierarchical. *Curr. Opin. Neurobiol.* 23, 194–187. doi: 10.1016/j.conb.2012.12.008

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0

Monsell, S. (2003). Task switching. *Trends Cogn. Sci.* 7, 134–140. doi: 10.1016/S1364-6613(03)00028-7

Moors, A., and De Houwer, J. (2006). Automaticity: a theoretical and conceptual analysis. *Psychol. Bull.* 132, 297–326. doi: 10.1037/0033-2909.132.2.297

Moran, R. J., Campo, P., Symmonds, M., Stephan, K. E., Dolan, R. J., and Friston, K. J. (2013). Free energy, precision and learning: the role of cholinergic neuromodulation. *J. Neurosci.* 33, 8227–8236. doi: 10.1523/JNEUROSCI.4255-12.2013

Mumford, D. (1992). On the computational architecture of the neocortex. *Biol. Cybern.* 66, 241–251. doi: 10.1007/BF00198477

Needham, C. J., Bradford, J. R., Bulpitt, A. J., and Westhead, D. R. (2007). A primer on learning in Bayesian networks for computational biology. *PLoS Comput. Biol.* 3:e129. doi: 10.1371/journal.pcbi.0030129

Ortega, P., and Braun, D. (2013). Thermodynamics as a theory of decision-making with information-processing costs. *Proc. R. Soc. A* 469. doi: 10.1098/rspa.2012.0683

Otto, A. R., Gershman, S. J., Markman, A. B., and Daw, N. D. (2013). The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychol. Sci.* 24, 751–761. doi: 10.1177/0956797612463080

Penny, W. D., Mattout, J., and Trujillo-Barreto, N. (2007). "Bayesian model selection and averaging," in *Statistical Parametric Mapping: The Analysis of Functional Brain Images,* eds K. J. Friston, J. T. Ashburner, S. J. Kiebel, T. E. Nichols, and W. D. Penny (London: Elsevier), 454–470.

Penny, W. D., Zeidman, P., and Burgess, N. (2013). Forward and backward inference in spatial cognition. *PLoS Comput. Biol.* 9:e1003383. doi: 10.1371/journal.pcbi.1003383

Pickens, C. L., Saddoris, M. P., Setlow, B., Gallagher, M., Holland, P. C., and Schoenbaum, G. (2003). Different roles for orbitofrontal cortex and basolateral amygdala in a reinforcer devaluation task. *J. Neurosci.* 23, 11078–11084.

Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580

Rasmussen, C., and Ghahramani, Z. (2002). Infinite mixtures of Gaussian process experts. *Adv. Neural Inf. Process. Syst.* 14, 881–888.

Salat, D. H., Buckner, R. L., Snyder, A. Z., Greve, D. N., Desikan, R. S. R., Busa, E., et al. (2004). Thinning of the cerebral cortex in aging. *Cereb. Cortex* 14, 721–730. doi: 10.1093/cercor/bhh032

Schrier, A. (1984). Learning how to learn: the significance and current status of learning set formation. *Primates* 25, 95–102. doi: 10.1007/BF02382299

Sengupta, B., Stemmler, M. B., and Friston, K. J. (2013). Information and efficiency in the nervous system-a synthesis. *PLoS Comput. Biol.* 9:e1003157. doi: 10.1371/journal.pcbi.1003157

Shams, L., Ma, W. J., and Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *Neuroreport* 16, 1923–1927. doi: 10.1097/01.wnr.0000187634.68504.bb

Simon, H. (1972). Theories of bounded rationality. *Decis. Organ.* 1, 161–176.

Simon, J., Proctor, R., and Reeve, T. (1990). "The effects of an irrelevant directional cue on human information processing," in *Stimulus–Response Compatibility: An Integrated Perspective,* eds R. Proctor and T. Reeve (Amsterdam: Elsevier), 31–86.

Smittenaar, P., FitzGerald, T. H. B., Romei, V., Wright, N., and Dolan, R. J. (2013). Disruption of dorsolateral prefrontal cortex decreases model-based in favour of model-free control in humans. *Neuron* 80, 914–919. doi: 10.1016/j.neuron.2013.08.009

Stahl, D. O., and Wilson, P. W. (1995). On players' models of other players: theory and experimental evidence. *Games Econ. Behav.* 10, 218–254. doi: 10.1006/game.1995.1031

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., and Friston, K. J. (2009). Bayesian model selection for group studies. *Neuroimage* 46, 1004–1017. doi: 10.1016/j.neuroimage.2009.03.025

Sterzer, P., Kleinschmidt, A., and Rees, G. (2009). The neural bases of multistable perception. *Trends Cogn. Sci.* 13, 310–318. doi: 10.1016/j.tics.2009.04.006

Stroop, J. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643–662. doi: 10.1037/h0054651

Summerfield, C., Behrens, T. E., and Koechlin, E. (2011). Perceptual classification in a rapidly changing environment. *Neuron* 71, 725–736. doi: 10.1016/j.neuron.2011.06.022

Summerfield, C., and Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends Cogn. Sci.* 13, 403–409. doi: 10.1016/j.tics.2009.06.003

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. (2011). How to grow a mind: statistics, structure, and abstraction. *Science* 331, 1279–1285. doi: 10.1126/science.1192788

Tipper, S. P., Howard, L. A., and Jackson, S. R. (1997). Selective reaching to grasp: evidence for distractor interference effects. *Vis. Cogn.* 4, 1–38. doi: 10.1080/713756749

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., et al. (2007). Schemas and memory consolidation. *Science* 316, 76–82. doi: 10.1126/science.1135935

Tucker, M., and Ellis, R. (2004). Action priming by briefly presented objects. *Acta Psychol. (Amst.)* 116, 185–203. doi: 10.1016/j.actpsy.2004.01.004

van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., and Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends Neurosci.* 35, 211–219. doi: 10.1016/j.tins.2012.02.001

Wallis, J., Anderson, K., and Miller, E. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature* 411, 953–956. doi: 10.1038/35082081

Weiss, Y., Simoncelli, E. P., and Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nat. Neurosci.* 5, 598–604. doi: 10.1038/nn0602-858

Welsh, T., and Elliott, D. (2004). Movement trajectories in the presence of a distracting stimulus: evidence for a response activation model of selective reaching. *Q. J. Exp. Psychol. A* 57, 1031–1057. doi: 10.1080/02724980343000666

Wunderlich, K., Dayan, P., and Dolan, R. J. (2012a). Mapping value based planning and extensively trained choice in the human brain. *Nat. Neurosci.* 15, 786–791. doi: 10.1038/nn.3068

Wunderlich, K., Smittenaar, P., and Dolan, R. (2012b). Dopamine enhances model-based over model-free choice behavior. *Neuron* 75, 418–424. doi: 10.1016/j.neuron.2012.03.042

Yin, H., and Knowlton, B. (2006). The role of the basal ganglia in habit formation. *Nat. Rev. Neurosci.* 7, 464–476. doi: 10.1038/nrn1919

Yin, H. H., Knowlton, B. J., and Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur. J. Neurosci.* 19, 181–189. doi: 10.1111/j.1460-9568.2004.03095.x

Yoshida, W., Dolan, R. J., and Friston, K. J. (2008). Game theory of mind. *PLoS Comput. Biol.* 4:e1000254. doi: 10.1371/journal.pcbi.1000254

Yu, A. J., and Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron* 46, 681–692. doi: 10.1016/j.neuron.2005.04.026

Zeki, S., Watson, J., and Lueck, C. (1991). A direct demonstration of functional specialization in human visual cortex. *J. Neurosci.* 11, 641–649.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 04 March 2014; accepted: 04 June 2014; published online: 26 June 2014.
Citation: FitzGerald THB, Dolan RJ and Friston KJ (2014) Model averaging, optimal inference, and habit formation. Front. Hum. Neurosci. ***8****:457. doi: 10.3389/fnhum.2014.00457*
This article was submitted to the journal Frontiers in Human Neuroscience.

frontiers in
HUMAN NEUROSCIENCE

OPINION ARTICLE
published: 20 June 2014
doi: 10.3389/fnhum.2014.00449

Procedural skills and neurobehavioral freedom

Nerea Crespo-Eguílaz, Sara Magallón and Juan Narbona*

Pediatric Neurology Unit and Department of Education, University of Navarra, Pamplona, Spain
**Correspondence: jnarbona@unav.es*

Edited by:
Jose Angel Lombo, Pontifical University of the Holy Cross, Italy

Reviewed by:
Gerry Leisman, O.R.T.-Braude College of Engineering, Israel

Keywords: attention, central coherence, developmental coordination disorder, implicit memory, operational habit, pragmatics, procedural learning, social communication disorder

INTRODUCTION

Procedural learning (PL) is a part of implicit memory (Shiffrin and Schneider, 1977) and is based on brain subsystems of associative cortex and its connections with basal ganglia and cerebellum (Squire, 1992). PL gives the human individual a gain in freedom: automatic healthy cognitive and motor skills help to save an important amount of conscious work in daily routines and in effortful cognitive and/or motor action (Treisman and Gelade, 1980; Kahneman et al., 1983). In this way, attention can be focused on quick understanding, central coherence awareness, problem-solving processes and social accuracy. Human procedural skills and executively-controlled aspects of action intersect and cooperate with each other (Leisman et al., 2014). Useful procedural automatisms are basically acquired during childhood and youth, but also over the whole course of life, by means of incidental experience and by formal education. PL enhances the natural potentialities (i.e., predispositions) of the agent for a suitable unfolding of his or her operations. From this point of view, acquired automatisms could be included among operational habits in the interface between perceptual-motor and cognitive-volitional human activities.

DEVELOPMENTAL COORDINATION DISORDER: AN EXPANDED VIEW

There is a child population for which operational habit learning is particularly difficult. Clumsiness, disproportionate to general development, is the most evident characteristic of individuals with this developmental condition, which has been labeled *developmental dyspraxia* and, more recently, *developmental coordination disorder* (DCD). At present, the most widely accepted definition of DCD in childhood comes from the Diagnostic and Statistical Manual of Mental Disorders, IVth and 5th editions (APA-American Psychiatric Association, 2000, 2013) and from the International Classification of Diseases-11th edition draft (World Health Organisation, 2013). DCD is essentially regarded as a disturbance of motor coordination, which consequently is substantially below that expected given the child's age and intelligence, but this is not due to a general medical condition (e.g., cerebral palsy) and does not meet the criteria for a pervasive developmental disorder; if mental retardation is present, motor difficulties exceed those expected for the level of mental development. DCD causes disruption of daily living activities and academic achievement. DCD is estimated to affect to 2–8% of schoolchildren (Kadesjö and Gillberg, 1998; Crespo-Eguílaz and Narbona, 2009; Lingam et al., 2009; Missiuna et al., 2011).

Young people with DCD have a characteristic slowness in daily routines. They are disproportionately unskilled not only for motor actions, as can be measured using *ad-hoc* scales and batteries (Bruininks and Bruininks, 2005; Henderson and Sugden, 2007; Wilson et al., 2009) but also for quick perceptual management of complex visuospatial information and motor imagery (Noten et al., 2014). Moreover DCD has a high comorbidity with attention deficit / hyperactivity disorder (ADHD), and with social communication (language pragmatic) disorder (Gillberg, 2003; Crespo-Eguílaz and Narbona, 2009; Crespo-Eguílaz et al., 2012; American Psychiatric Association, DSM-5, 2013; World Health Organisation, IDC-11 draft, 2013). As a consequence, affected children and adolescents typically behave in a naïve manner, and their social use of language is frequently inaccurate (Volden, 2004; Crespo-Eguílaz and Narbona, 2009; Brossard-Racine et al., 2011; Westendorp et al., 2011; Blank et al., 2012). All these impairments have a significant negative impact on activities of daily living, such as, dressing, handwriting, sports, and social exchanges (Blank et al., 2012). Depression, anxiety, and risk of bullying by peers are significantly more frequent in children with DCD and those with comorbid DCD and ADHD vs. typical controls (Zwicker et al., 2012; Missiuna et al., 2014).

PROCEDURAL LEARNING IN CHILDREN WITH DEVELOPMENTAL COORDINATION DISORDER: SOME EXPERIMENTAL EVIDENCE

A variety of dysfunctions of neural loops relating prefrontal, secondary premotor and parietal cortices, with basal ganglia and cerebellum, have been proposed (Bo and Lee, 2013; Leisman et al., 2014) to explain the physiopathology of DCD.

In a continuous task with implicit visual sequences, schoolchildren with DCD learn poorly relative to typically developing children. Children with DCD demonstrated a general learning of visuo-perceptive task demands that was comparable to that of controls, but they failed to learn anticipation of implicit visuo-motor sequences. Interestingly, a sequence recall test, administered after the whole task, indicated some awareness of the repeating sequence pattern (Gheysen et al., 2011). By contrast, using the same paradigm, Lejeune et al. (2013) found no

evidence of a difference in performance between children with DCD and typically developing children.

In order to assess PL in children with ADHD, DCD and reading learning disorder (RD), Magallón et al. (submitted) tested the children with two implicit / procedural learning tasks using the Purdue pegboard (Gardner and Broman, 1979) and an adaptation of the mirror drawing task (Milner et al., 1968). Participants aged 6–12 years old were classified into four groups matched for gender, age and severity of ADHD symptoms: 19 children with ADHD only, 30 children with DCD+ADHD, 48 children with RD+ADHD, and 90 typically developing children. All participants accomplished three consecutive trials of each of the two tasks and a delayed fourth trial following a verbal interference task. Typical-for-age scoring measures of performance were compared (Student's t) within trials and between groups. The baseline results of the DCD+ADHD group were significantly lower than those of the other groups. Nevertheless, after three repetitions of the two tasks, DCD+ADHD children improved their efficiency and reached that of the baseline of both the non-DCD clinical groups. This learned performance was retained at the delayed fourth trial. However, the percentage improvement obtained by DCD children was lower than that of the other two clinical groups and controls in all the trials.

Another study (Crespo-Eguílaz et al., 2012) addressed the ability of schoolchildren to quickly grasp and verbally explain "nonsense" in complex figurative pictures: a chimeric figure and an absurd scene. Only 11.3% of schoolchildren with DCD and with DCD+ADHD resolved the tasks accurately, whereas 87.5% of controls and ADHD-alone children did these two central coherence function tasks successfully (chi-square test: $p < 0.01$).

As mentioned above, children with DCD+ADHD also usually have difficulties integrating inputs of complex visual or verbal information. As a consequence, they struggle to get the whole picture, miss relevant clues in social contexts, have problems dealing with inference, and fail to make sense of figurative language, jokes, narratives and adapted conversation. These psycholinguistic difficulties are reminiscent of the characteristics of Social (pragmatic) Communication Disorder (SCD) as defined in DSM-5 and in ICD-11 draft. So, we might ask whether DCD is typically comorbid with SCD. An alternative explanation would be that pragmatic difficulties are a part of DCD. To investigate this question, a Spanish translation of the Children's Communication Checklist-CCC (Bishop, 1998) was given to the parents of children aged 6–12 years who were divided into five groups: those with DCD+ADHD, those with ADHD only, those with SCD, those with high functioning autism spectrum disorder (HFASD), and those with typical development (Narbona et al., in press). The five groups were matched for mental age and gender. The results suggest that communication difficulties in children with DCD+ADHD are qualitatively different, more severe and have a larger impact on social relationships than those shown by children with ADHD only. On the other hand, the pragmatic difficulties in children with DCD+ADHD are milder than those defining SCD and HFASD. Moreover the HFASD group showed unusual, restricted and stereotyped interests. In contrast, DCD+ADHD and SCD groups do not have a characteristic restriction of interests, and their basic social motivation and abilities are preserved, apart from the linguistic difficulties. These results are in accordance with recent research reviews (Gibson et al., 2013; Norbury, 2014).

Pragmatic difficulties may be present in children with ADHD, developmental coordination disorder, autism spectrum disorders, Williams syndrome, spina bifida with hydrocephalus, cerebral palsy, etc. (Holck et al., 2009); thus pragmatic difficulties can be either a component of several large behavioral phenotypes or an isolated communication disorder (i.e., SCD, as it has been recently proposed in DSM-5 and in ICD-11 draft). Given that children with DCD most frequently have pragmatic difficulties, it would seem that these are not comorbid but constitute a component of DCD related to the failure to grasp visuospatial clues useful in evaluating social appropriateness. In contrast, pragmatic communication difficulties of autistic persons are included in their social/intersubjective pervasively disordered abilities (Norbury, 2014).

DISCUSSION AND CONCLUDING REMARKS

We propose that the core dysfunction in DCD affects procedural learning. PL deals not only with motor skills but also with fast perceptive integration, cognitive routines and socially accurate habits. As a consequence, children with DCD are characterized by slowness not only for motor tasks but also for awareness of relevant cognitive and social clues, which causes difficulties in contextualizing information and in social relationships with peers. Children with DCD do have normal intersubjective skills and a normal desire to communicate with other people, in contrast to children with autistic spectrum disorders (Norbury, 2014). The above-mentioned experimental results on procedural learning of visual sequences, of mirror drawing, of motor manual skills and of quick verification of central coherence, suggest that a basic neuropsychological dysfunction of procedural learning may be the central problem in DCD, with its frequent association to social communication disorder. This basic PL dysfunction seems to be intrinsic to DCD and independent of attention deficit: the experiments took account of attention deficit by considering a group of subjects with ADHD alone.

A limitation of the above experimental studies is that the tasks were highly specific. Similar studies with larger samples, with more diverse and ecological tasks, and with greater number of trials (to justify the assumption of long-term learning), are necessary.

Children with DCD can improve their motor and cognitive performance by repetition. Therefore, we suggest that this developmental condition does not imply an absolute inability, but a poor natural disposition, to learn motor and/or cognitive facilitating strategies. Assuming, as indicated by the research findings, that the core dysfunction lies in automation, an appropriate approach to help affected children would be to base intervention on repetition of the skills needed by each individual patient in his or her everyday ecological context and taking account of personal motivations and preserved abilities (for example, language for auto-instructions).

The persistent nature of DCD in around one-half of individuals first diagnosed in childhood (Cantell et al., 1994) emphasizes the importance of occupational therapy intervention in youth. The majority of approaches to intervention fit into two main categories. The "process or deficit approaches" aim to remedy some underlying process deficit with intervention targeted at a neural structure (Polatajko and Cantin, 2005). By contrast, the "functional skill approaches" work on teaching the activities of daily living that the child needs to be able to carry out. Recent meta-analyses demonstrate that the latter category of approaches produces the best therapeutic effect (Blank et al., 2012; Smits-Engelsman et al., 2013). Intervention designs should be addressed not only to the training of neurophysiological procedural circuitry but also to respond to motivations of each subject and to enhance generalization of newly acquired skills and good habits for managing significant cues of daily life, social relationships, and schooling (Polatajko and Cantin, 2005; Sugden and Dunford, 2007). The P4C model (Missiuna et al., 2011) emphasizes the partnership of the occupational therapist with educators and parents to change the life and daily environment of a child; the model focuses on capacity building through collaboration and coaching in context and includes whole class instruction, dynamic performance analysis, and monitoring response to intervention.

Neurobiological habits can be viewed as constrictions of dispositional resources of the agent. Such a perspective on operational habits is, perhaps, more appropriate for so-called "bad" or pathological habits, i.e., obsessions, tics, movement disorders etc. In this article, however, we have emphasized a positive, healthy view of habits because the functions of the human brain are precisely orchestrated on the basis of a huge number of beneficial automatisms that allow us to perform fluently the complex cognitive and motor activities of daily life. Psychoeducation can help young people suffering from DCD to become physically more adept and to liberate their potential for complex thinking, for planning of practical actions and for evaluating the social appropriateness of their behavior.

ACKNOWLEDGMENTS

This work has been supported by grants from the Government of Navarra and the Fundación Fuentes Dutor (Pamplona, Spain) for research in developmental neurology and neuropsychology (2009–2013).

REFERENCES

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV TR, 4th Edn*. Arlington, VA: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5, 5th Edn*. Washington, DC: American Psychiatric Association.

Bishop, D. V. M. (1998). Development of the Children's Communication Checklist (CCC): a method for assessing qualitative aspects of communicative impairment in children. *J. Child Psychol. Psychiatry* 39, 879–891. doi: 10.1017/S0021963098002832

Blank, R., Smits-Engelsman, B., Polatajko, H., and Wilson, P. (2012). European Academy for Childhood Disability: recommendations on the definition, diagnosis and intervention of developmental coordination disorder (long version). *Dev. Med. Child Neurol.* 54, 54–93. doi: 10.1111/j.1469-8749.2011.04171.x

Bo, J., and Lee, C. M. (2013). Motor skill learning in children with developmental coordination disorder. *Res. Dev. Disabil.* 34, 2047–2055. doi: 10.1016/j.ridd.2013.03.012

Brossard-Racine, M., Majnemer, A., and Shevell, M. (2011). Exploring the neural mechanisms that underlie motor difficulties in children with Attention Deficit Hyperactivity Disorder. *Dev. Neurorehabil.* 14, 101–111. doi: 10.3109/17518423.2010.547545

Bruininks, R. H., and Bruininks, B. D. (2005). *Bruininks-Oseretsky Test of Motor Proficiency, 2nd Edn.* Minneapolis, MN: Pearson Assessment.

Cantell, M. H., Smyth, M. M., and Ahonen, T. P. (1994). Clumsiness in adolescence: educational, motor, and social outcomes of motor delay detected at 5 years. *Adapt. Phys. Activ.* 11, 115–129.

Crespo-Eguílaz, N., and Narbona, J. (2009). Trastorno de aprendizaje procedimental: características neuropsicológicas. *Rev. Neurol.* 49, 409–416.

Crespo-Eguílaz, N., Narbona, J., and Magallón, S. (2012). Disfunción de la coherencia central en niños con trastorno de aprendizaje procedimental. *Rev. Neurol.* 55, 513–519.

Gardner, R. A., and Broman, M. (1979). The Purdue Pegboard normative data on 1334 school children. *J. Clin. Child Psychol.* 8, 156–162. doi: 10.1080/15374417909532912

Gheysen, F., Van Waelvelde, H., and Fias, W. (2011). Impaired visuo-motor sequence learning in developmental coordination disorder. *Res. Dev. Disab.* 32, 749–756. doi: 10.1016/j.ridd.2010.11.005

Gibson, J., Adams, C., Elaine Lockton, E., and Jonathan Green, J. (2013). Social communication disorder outside autism? A diagnostic classification approach to delineating pragmatic language impairment, high functioning autism and specific language impairment. *J. Child Psychol. Psychiatry* 54, 1186–1197. doi: 10.1111/jcpp.12079

Gillberg, C. (2003). Deficits in attention, motor control, and perception: a brief review. *Arch. Dis. Child.* 88, 904–910. doi: 10.1136/adc.88.10.904

Henderson, S. E., and Sugden, D. A. (2007). *Movement Assessment Battery for Children, 2nd Edn.* London: The Psychological Corporation.

Holck, P., Nettelbladt, U., and Sandberg, A. D. (2009). Children with cerebral palsy, spina bifida and pragmatic language impairment: differences and similarities in pragmatic ability. *Res. Dev. Disabil.* 30, 942–951. doi: 10.1016/j.ridd.2009.01.008

Kadesjö, B., and Gillberg, C. (1998). Attention deficits and clumsiness in Swedish 7-year-old children. *Dev. Med. Child Neurol.* 40, 796–804. doi: 10.1111/j.1469-8749.1998.tb12356.x

Kahneman, D., Treisman, A., and Burkell, J. (1983). The cost of visual filtering. *J. Exp. Psychol. Hum. Percept. Perform.* 9, 510–522. doi: 10.1037/0096-1523.9.4.510

Leisman, G., Braun-Benjamin, O., and Melillo, R. (2014). Cognitive-motor interactions of the basal ganglia in development. *Front. Syst. Neurosci.* 8:16. doi: 10.3389/fnsys.2014.00016

Lejeune, C., Catale, C., Willems, S., and Meulemans, T. (2013). Intact procedural motor sequence learning in developmental coordination disorder. *Res. Develop. Disab.* 34, 1974–1981. doi: 10.1016/j.ridd.2013.03.017

Lingam, R., Hunt, L., Golding, J., Jongmans, M., and Emond, A. (2009). Prevalence of developmental coordination disorder using the DSM-IV at 7 years of age: a UK population-based study. *Pediatrics* 123, 693–700. doi: 10.1542/peds.2008-1770

Milner, B., Corkin, S., and Teuber, H. L. (1968). Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of H.M. *Neuropsychologia* 6, 317–338. doi: 10.1016/0028-3932(68)90021-3

Missiuna, C., Cairney, J., Pollock, N., Russell, D., Macdonald, K., Cousins, M., et al. (2011). A staged approach for identifying children with developmental coordination disorder from the population. *Res. Dev. Disabil.* 32, 549–559. doi: 10.1016/j.ridd.2010.12.025

Missiuna, C., Cairney, J., Pollock, N., Campbell, W., Russell, D. J., Macdonald, K., et al. (2014). Psychological distress in children with developmental coordination disorder and attention-deficit hyperactivity disorder. *Res. Dev. Disabil.* 35, 1198–1207. doi: 10.1016/j.ridd.2014.01.007

Narbona, J., Crespo-Eguílaz, N., Sánchez-Carpintero, R., and Magallón, S. (in press). Diagnóstico diferencial del trastorno de la comunicación social. Utilidad del Children's Communication Checklist-CCC (Abstract). *Rev. Neurol.* 58 (Suppl. 3).

Norbury, C. F. (2014). Practitioner review: social (pragmatic) communication disorder conceptualization, evidence and clinical implications. *J. Child Psychol. Psychiatry* 55, 204–216. doi: 10.1111/jcpp.12154

Noten, M., Wilson, P., Ruddock, S., and Steenbergen, B. (2014). Mild impairments of motor imagery skills in children with DCD. *Res. Dev. Disabil.* 35, 1152–1159. doi: 10.1016/j.ridd.2014.01.026

Polatajko, H. J., and Cantin, N. (2005). Developmental coordination disorder (dyspraxia): an overview of the state of the art.

Semin. Pediatr. Neurol. 12, 250–258. doi: 10.1016/j.spen.2005.12.007

Shiffrin, R. M., and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychol. Rev.* 54, 127–190. doi: 10.1037/0033-295X.84.2.127

Smits-Engelsman, B. C. M., Blank, R., Van Der Kaay, A., Mosterd-Van Der Meijs, R., Vlugt-Van Den Brand, E., Polatajko, H. J., et al. (2013). Efficacy of interventions to improve motor performance in children with developmental coordination disorder: a combined systematic review and meta-analysis. *Dev. Med. Child Neurol.* 55, 229–237. doi: 10.1111/dmcn.12008

Squire, L. R. (1992). Declarative and non declarative memory: multiple brain systems supporting learning and memory. *J. Cogn. Neurosci.* 4, 232–243. doi: 10.1162/jocn.1992.4.3.232

Sugden, D. A., and Dunford, C. D. (2007). The role of theory empiricism and experience in intervention for children with movement difficulties. *Disabil. Rehabil.* 29, 3–11. doi: 10.1080/09638280600947542

Treisman, A. M., and Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psychol.* 12, 97–136. doi: 10.1016/0010-0285(80)90005-5

Volden, J. (2004). Nonverbal learning disability: a tutorial for speech-language pathologist. *Am. J. Speech Lang. Pathol.* 13, 128–141. doi: 10.1044/1058-0360(2004/014)

Westendorp, M., Hartman, E., Houwen, S., Smith, J., and Visscher, C. (2011). The relationship between gross motor skills and academic achievement in children with learning disabilities. *Res. Dev. Disabil.* 32, 2773–2779. doi: 10.1016/j.ridd.2011.05.032

Wilson, B. N., Crawford, S. G., Green, D., Roberts, G., Aylott, A., and Kaplan, B. J. (2009). Psychometric properties of the revised Developmental Coordination Disorder Questionnaire. *Phys. Occup. Ther. Pediatr.* 29, 182–202. doi: 10.1080/01942630902784761

World Health Organisation. (2013). *International Classification of Diseases-11. Proposed criteria for Pragmatic Language Impairment (IDC-11).* Geneva: WHO. Available online at: http://www.who.int/classifications/icd/revision/en/index.html [Accessed 31 January 2013].

Zwicker, J. G., Harris, S. R., and Klassen, A. F. (2012). Quality of life domains affected in children with developmental coordination disorder: a systematic review. *Child Care Health Dev.* 39, 562–580. doi: 10.1111/j.1365-2214.2012.01379.xcch

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 March 2014; accepted: 03 June 2014; published online: 20 June 2014.

Citation: Crespo-Eguílaz N, Magallón S and Narbona J (2014) Procedural skills and neurobehavioral freedom. Front. Hum. Neurosci. ***8****:449. doi: 10.3389/fnhum.2014.00449*

This article was submitted to the journal Frontiers in Human Neuroscience.

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 04 August 2014
doi: 10.3389/fnhum.2014.00587

Devaluation and sequential decisions: linking goal-directed and model-based behavior

Eva Friedel[1]*†, Stefan P. Koch[1†], Jean Wendt[1], Andreas Heinz[1], Lorenz Deserno[1,2†] and Florian Schlagenhauf[1,2†]

[1] Department of Psychiatry and Psychotherapy, Charité – Universitätsmedizin, Berlin, Germany
[2] Max Planck-Fellow Group "Cognitive and Affective Control of Behavioural Adaptation," Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Edited by:
Javier Bernacer, University of Navarra, Spain

Reviewed by:
Guillermo Horga, Columbia University Medical Center, USA
Adrian Mark Haith, Johns Hopkins University School of Medicine, USA

***Correspondence:**
Eva Friedel, Department of Psychiatry and Psychotherapy, Charité – Universitätsmedizin Berlin, Campus Charité Mitte, Charitéplatz 1, 10117 Berlin, Germany
e-mail: eva.friedel@charite.de

† These authors have contributed equally to this work.

In experimental psychology different experiments have been developed to assess goal–directed as compared to habitual control over instrumental decisions. Similar to animal studies selective devaluation procedures have been used. More recently sequential decision-making tasks have been designed to assess the degree of goal-directed vs. habitual choice behavior in terms of an influential computational theory of model-based compared to model-free behavioral control. As recently suggested, different measurements are thought to reflect the same construct. Yet, there has been no attempt to directly assess the construct validity of these different measurements. In the present study, we used a devaluation paradigm and a sequential decision-making task to address this question of construct validity in a sample of 18 healthy male human participants. Correlational analysis revealed a positive association between model-based choices during sequential decisions and goal-directed behavior after devaluation suggesting a single framework underlying both operationalizations and speaking in favor of construct validity of both measurement approaches. Up to now, this has been merely assumed but never been directly tested in humans.

Keywords: model-based and model-free learning, habitual and goal-directed behavior, 2-step decision task, devaluation task, reinforcement learning, computational modeling

INTRODUCTION

Habitual decisions arise from the retrospective, slow accumulation of rewards via iterative updating of expectations. In contrast, the goal-directed system prospectively considers future outcomes associated with an action. Thus, if outcome values change suddenly e.g., after devaluation (i.e., satiety), the goal-directed system enables quick behavioral adaptation, whereas the habitual system requires new reward experience before it can alter behavior accordingly (Balleine and Dickinson, 1998). Recently, this dual system theory has been advanced by the use of computational models of learning which either purely update reward expectations based on reward prediction errors ("model-free") or aim to map possible actions to their potential outcomes ("model-based"; Daw et al., 2005). In their comprehensive review Dolan and Dayan (2013) subsume both concepts (goal-directed/habitual and model-based/model-free) under a single framework of reflective vs. reflexive decision making. Here, model-based choices are by definition goal-directed and model-free choices rest upon habitual learning. The authors provide a historical and conceptual framework for the evolution of dual systems theories with a reflexive and a reflective control system in cognitive neuroscience. This longstanding dichotomy has been described as goal-directed vs. habitual behavior by experimental psychologists while the model-free vs. model-based theory provides a computational account of the same construct.

Dolan and Dayan (2013) rank goal-directed behavior in humans in Generation 2, evolving from animal experiments in Generation 1. Generation 3 starts with the conceptual precision of goal-directed and habitual decision making as model-based vs. model-free learning on the basis of computational accounts in a reinforcement learning context. Even though both terminologies, goal-directed and model-based behavioral control, derive from the same framework, the different operationalizations have never been directly related in a human sample.

There are two main, but experimentally distinct, approaches to test the influence of both systems: outcome devaluation and sequential decision-making. First, devaluation paradigms require participants to overcome a previously trained action after outcome devaluation. Here, the goal-directed system adapts quickly based on an explicit action-outcome association. This is in sharp contrast to the habitual system that remains initially tied to the action acquired before devaluation because it relies on a stimulus-action association without direct representation of the link between action and a now devalued outcome. These paradigms have been developed in animal research (Dickinson, 1985) and were successfully translated to human research in healthy (Valentin et al., 2007; De Wit et al., 2009; Tricomi et al., 2009) and pathological conditions (De Wit et al., 2011; Gillan et al., 2011; Sjoerds et al., 2013). Second, sequential decision-making challenges an individual with a series of subsequent decisions to finally receive a reward (Generation 3). These tasks are characterized by a state-transition structure, which probabilistically determines the entered state after a given choice. Hence, a learner that acquires and uses this task structure (e.g., using a

decision tree) by building and using an internal representation (a "model") of the task is therefore labeled as "model-based." This learner builds an internal representation of the task structure, which enables forward planning. Apparently, model-based learning is by definition goal-directed. A purely "model-free" learner neglects these transition schemes and simply repeats action sequences that were previously rewarded. Such tasks have been applied in healthy participants (Daw et al., 2011; Wunderlich et al., 2012; Smittenaar et al., 2013) and in one study in alcohol-dependent patients (Sebold et al., 2014). For both types of tasks, there is convincing evidence that human choices are influenced by both systems.

It is an on-going question whether these different measurements assess the same aspects of instrumental behavior (Doll et al., 2012; Dolan and Dayan, 2013). We assume that both measurements reflect the same construct and therefore shed light on similar mechanism from the perspectives of different experimental procedures that evolved from different fields (experimental psychology and computational theory). So far, this issue of construct validity has not been directly tested. However, the question of construct validity is important to address: in neuroscience research the two measurements have so far been treated almost equivalently and conclusions on presumably identical processes have been drawn in healthy human beings and also in severely ill individuals. Relating both measurements thus represents a coercive step to add to their conceptual precision.

To assess construct validity, we applied two tasks: a selective devaluation task (Valentin et al., 2007; Daw et al., 2011) and a sequential decision-making task (Daw et al., 2011) proven to capture the two constructs of goal-directed vs. habitual and model-based vs. model-free behavioral control separately using a within-subject design in 18 healthy male participants.

MATERIALS AND METHODS

SUBJECTS

Eighteen right-handed healthy male subjects participated in the study. All participants were assessed for Axis I or II disorder with SCID-I Interview as well as for eating disorders with the Eating Attitude Test (EAT-26) (Garner et al., 1982) indicating no psychiatric or eating disorder in any of the subjects. Participants were pre-screened to ensure that they found tomato juice, chocolate milk and fruit tea pleasant and did not show any food intolerance or were not on a diet. All participants were asked to fast for at least 6 h before their scheduled arrival time, but were permitted and motivated to drink water before the experimental procedure. Upon arrival, participants rated their hunger on a visual analog scale (VAS) and informed the instructor when they had last eaten. There were no objective measures to control if participants complied with the instruction to fast. All participants gave informed written consent and the study was approved by the Ethics Committee of the Charité Universitätsmedizin.

TASKS

Devaluation paradigm

To test goal-directed vs. habitual behavior, we used a selective devaluation paradigm with liquid food rewards (**Figures 1A,B;** Valentin et al., 2007). The two liquid food rewards were chocolate milk and tomato juice. These foods were chosen because they can be administered in liquid form, are palatable at room temperature and are distinguishable in their flavor and texture to help facilitate sensory specific satiety effects. In addition we also used a tasteless neutral water solution and fruit tea as control. The food rewards were delivered by means of separate electronic syringe pumps (one for each liquid) positioned behind a small room divider (paravent). These pumps transferred the liquids to the subjects via plastic tubes (∼6 mm diameter). The end of these tubes were held between the subject's lips like a straw and attached to the shoulder with a small adhesive tape while they were sitting in front of the computer screen performing the task.

The task consisted of three trial types: chocolate, tomato or neutral, with fully randomized order throughout the experiment (**Figure 1A**). On each trial, subjects were faced with the choice between two abstract stimuli, each of which was associated with different probabilities to receive a rewarding liquid food outcome or nothing.

The experimental procedure (**Figures 1A,B**) was divided into three steps: (1) training, (2) devaluation, and (3) test in extinction. *First,* during the *training* part, subjects learned to make choices that were associated with the subsequent delivery of these different liquid food outcomes (0.5 ml of tomato juice or chocolate milk and fruit tea). For each trial type, the overall probability of a food outcome was $p = 0.75$ for the high-probability stimulus (referring to the choice of the stimulus associated with a high-probability food outcome) with $p = 0.5$ for tomato or chocolate and $p = 0.25$ for the common outcome fruit tea. The low probability stimulus (meaning the choice of the stimulus associated with a low probability liquid food outcome) led with $p = 0.25$ to a common outcome (0.5 ml fruit tea). In the control condition, water was delivered with the same probabilities of $p = 0.75$ after a high probability stimulus choice and $p = 0.25$ after a low probability stimulus choice, respectively. The training sessions consisted of 150 trials (50 trials for each stimulus pair). To facilitate learning of the stimulus-outcome associations between the abstract stimuli and the liquid food rewards, each stimulus-outcome pair (chocolate, tomato, and neutral) was randomly assigned to one of the four spatial positions on the screen (top left, top right, bottom left, or bottom right) at the beginning of the experiment and remained constant throughout. A unique spatial location was assigned to the high-probability stimulus in all three trial-type pairs. The specific assignment of arbitrary fractal stimuli and spatial position to each particular action was fully counterbalanced across subjects. The subjects' task on each trial throughout all parts of the experimental procedure was to choose one of the two possible available stimuli on the screen which they perceived as being "more pleasant" (and thus is associated with a higher probability to receive a rewarding outcome). In a *second* step, during the *devaluation* part and after training, either tomato juice or chocolate milk was selectively devalued by feeding the subject with the food until they reported a feeling of satiety. For devaluation, participants ate either chocolate pudding or tomato soup (mean consumption in grams = 357.1, std. = 196.0) until they rated the devalued food as unpleasant and refused to consume more. *Third,* during *test in extinction* and after devaluation, participants continued with the instrumental choice paradigm in

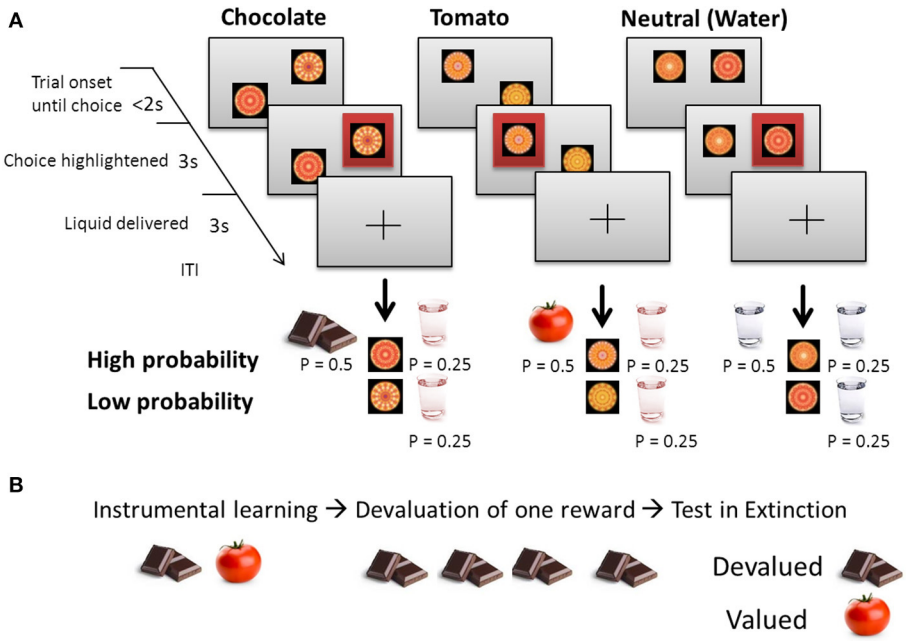

FIGURE 1 | Selective devaluation paradigm according to Valentin et al. (2007). (A) Trial structure depicted for each condition during instrumental learning. On each trial subjects had to choose between two abstract stimuli, the chosen stimulus is then highlighted. The high-probability stimulus choice leads to a food outcome (chocolate or tomato) with a probability of $p = 0.5$ and to a common outcome (fruit tea) with $p = 0.25$. The low-probability stimulus choice never leads to food outcomes and in $p = 0.25$ to the common outcome. In the neutral condition, the high-probability stimulus choice leads to water with $p = 0.75$ and the low-probability stimulus choice to water with $p = 0.25$. **(B)** After instrumental training, subjects were invited to consume either chocolate (illustrated here) or tomato soup to satiety, resulting in a selective devaluation of the consumed outcome. Subjects then underwent the same "test" procedure in extinction (chocolate and tomato were no longer presented).

extinction without delivery of the liquid food rewards tomato or chocolate (150 trials without food delivery, 50 trials for each stimulus pair). To maintain some degree of responding, the neutral fruit tea outcome continued to be available as during training with equal probability for the two available actions of $p = 0.3$ each (similar to Valentin et al., 2007). Subjects rated pleasantness of all administered foods on a visual analogous scale (VAS) before training, after training, after devaluation and after extinction. The use of an extinction procedure ensured that subjects only use information about the value of the outcome by making use of the previously learned associations between that outcome and a particular choice, as otherwise, if the tomato and chocolate outcome were presented again at test, subjects could relearn a new association, thereby confounding stimulus-response and response-outcome contributions. As reported by Valentin et al. (2007), goal-directed behavior is characterized by a significant decrease in choices of the stimulus associated with the devalued outcome, whereas habitual behavior leads to continued choosing of the stimulus associated with the devalued outcome. The number of choices was analyzed using a 2 (time: pre/post) × 2 (value: devalued/valued) repeated measures ANOVA to assess the degree of goal-directed vs. habitual choices.

For the devaluation paradigm, four participants had to be excluded from the sample (2 did not reach the learning criterion of 75% correct choices during training session and 2 refused to eat more although they did not rate the devalued food as being less pleasant after consumption and thus did not reach satiety).

Sequential decision-making task

In the sequential decision-making task (Daw et al., 2011), participants had to make two subsequent choices (each out of two options) to finally receive a monetary reward. At the first stage, each choice option led commonly (with 70% probability) to one of two pairs of stimuli and rarely (with 30% probability) to the other one. After entering the second stage, a second choice was followed by monetary reward or not, which was delivered according to slowly changing Gaussian random walks to facilitate the continuous updating of the second-stage action values. Participants performed a total of 201 trials. Crucially, a purely model-based learner uses the probabilities that underlie the transition from the first to the second stage, while a purely model-free learner neglects this task structure (**Figure 2**). Depending on the impact of previous second-stage rewards on the following first-stage choices, reinforcement learning theory predicts distinct first-stage choice patterns for model-free as opposed to model-based strategies. Model-free behavior can only generate a main effect of reward: a rewarded choice is more likely to be repeated, regardless of whether the reward followed a common or rare transition. Model-based behavior results in an interaction of the two factors, because a rare transition inverts the effect of the subsequent reward. Stay-switch behavior was analyzed as a function of reward (reward/no reward) and state (common/rare). These individual stay probabilities were subjected to a repeated-measures ANOVA with the factors reward and state.

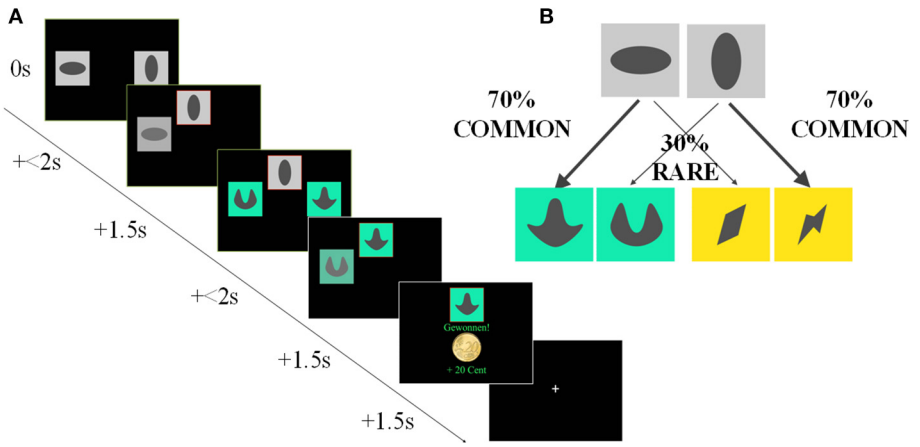

FIGURE 2 | Sequential Decision-Making Task (Two-Step), according to Daw et al. (2011). (A) Trial configuration for the Experiment. Each trial consisted of two different stages, and each stage involved a choice between two stimuli. In the first stage, subjects chose between two abstract stimuli on a gray background. The chosen stimulus was highlighted by a red frame and moved to the top of the screen, where it remained visible for 1.5 s; at the same time, the other stimulus faded away. Subjects then reached a subsequent second stage. Here subjects saw one of two further pairs of colored stimuli and again chose between these. The monetary outcome following this second stage choice (gain or no gain of 20 cent) was then presented centrally on the screen. **(B)** One pair of colored second stage stimuli occurred commonly (on 70% of trials; "common trials") after choice of one first stage stimulus, while the other pair was associated equally strongly with the other stimulus. On the remaining 30% of trials, the chosen first stage option resulted in a transition to the other second stage stimulus pair ("rare trials"). Reinforcement probabilities for each second stage stimulus changed slowly and independently according to Gaussian random walks with reflecting boundaries at 0.25 and 0.75.

With respect to the sequential decision-making task, one participant was excluded due to abortion of the experiment after half of the trials.

COMPUTATIONAL MODELING OF THE SEQUENTIAL DECISION-MAKING TASK

The aim of model-free and model-based algorithms is to learn values for each of the stimuli, which appear in the task as three pairs (s_A, s_B, s_C).s_A refers to the first-stage stimuli and s_B and s_C to the two pairs of second-stage stimuli. Here, a refers to the chosen stimuli and the indices i and t denote the stage ($i = 1$ for S_A at the first stage and $i = 2$ for S_B or S_C at the second stage) and the trial, respectively. The model-free algorithm was SARSA(λ):

$$Q_{MF_{s_{i,t+1},a_{i,t+1}}} = Q_{MF_{s_{i,t},a_{i,t}}} + \alpha_i \delta_{i,t} \quad (1)$$

$$\delta_{i,t} = r_{i,t} + Q_{MF_{s_{i+1,t},a_{i+1,t}}} - Q_{MF_{s_{i,t},a_{i,t}}} \quad (2)$$

Notably, $r(s_{1,t}) = 0$ because there are no rewards available at the first stage and $Q(s_{3,t}, a_t) = 0$ at the second stage because there are only two-stages and no third stage in this version of a sequential decision making task. All Q-values were initialized ("starting parameter") with 0. We allow different learning rates α_i for each stage i. Further, we allow for an additional stage-skipping update of first-stage values by introducing another parameter λ, which connects the two stages and allows the reward prediction error at the second stage to influence first-stage values:

$$Q_{MF_{s_{1,t+1},a_{1,t+1}}} = Q_{MF_{s_{1,t},a_{1,t}}} + \alpha_1 \lambda \delta_{2,t} \quad (3)$$

It is worth mentioning that λ additionally accounts for the main effect of reward as observed in the analysis of first-stage stay-switch behavior but not for an interaction of reward and state. Instead, the introduction of the transition matrix accounts for this interaction. Here, the model-based algorithm learns values by taking into account the transition matrix and computes first-stage values by simply multiplying the better option at the second stage with the transition probabilities:

$$Q_{MB_{s_A},a} = P(S_B|S_A, a) \times \max Q_{MF_{s_B},a} + P(S_c|S_A, a) \times \max Q_{MF_{s_C},a} \quad (4)$$

$$Q_{MB_{s_{2,t}},a_{2,t}} = Q_{MF_{s_{2,t}},a_{2,t}} \quad (5)$$

Note that this approach simplifies transition learning because transition probabilities are not learned explicitly. This approach is in line with the task instructions, and a simulation by Daw et al. (2011) verified that this approach outperforms incremental learning of the transition matrix (compare Wunderlich et al., 2012 but also see Glascher et al., 2010 or Lee et al., 2014). Finally, we connect Q_{MF} and Q_{MB} in a hybrid algorithm:

$$Q_{s_A,a} = \omega \times Q_{MB_{s_A},a} + (1 - \omega) \times Q_{MF_{s_A},a} \quad (6)$$

$$Q_{s_2,a} = Q_{MB_{s_2},a} = Q_{MF_{s_2},a} \quad (7)$$

Importantly, ω gives a weighting of the relative influence of model-free and model-based values and is therefore the model's parameter of most interest. To generate choices, we apply a softmax for Q:

$$p(i,a,t) = \frac{\exp\left(\beta_i(Q_{s_{i,t},a'_{i,t}} + \rho \times rep(a))\right)}{\sum_{a'} \exp\left(\beta_i(Q_{s_{i,t},a'_{i,t}} + \rho \times rep(a'))\right)} \quad (8)$$

Here, β controls the stochasticity of the choices and we assume this to be different between the two stages. The additional parameter ρ captures first-stage choice perseveration and rep is an indicator function that equals 1 if the previous first-stage choice was the same (Lau and Glimcher, 2005; Daw et al., 2011). In summary, the algorithm has a total of 7 parameters and can be reduced to its special cases $\omega = 1$ (4 parameters) and $\omega = 0$ (5 parameters). We fit bounded parameters by transforming them to a logistic $(\alpha, \lambda, \omega)$ or exponential (β) distribution to render normally distributed parameter estimates. To infer the maximum a posteriori estimate of each parameter for each subject, we set the prior distribution to the maximum-likelihood given the data of all participants and then use Expectation-Maximization. For an in-depth description please compare Huys et al. (2011) and Huys et al. (2012). In the computational modeling part there were no differences to Daw et al. (2011) with respect to the applied model-free and model-based algorithms as well as the softmax function.

CORRELATION ANALYSIS OF GOAL-DIRECTED AND MODEL-BASED BEHAVIOR

We assessed the degree of goal-directed behavior in the selective devaluation task by computing the interaction score from the number of choices: "*(valued stimulus pre devaluation – valued stimulus post devaluation) – (devalued stimulus pre devaluation - devalued stimulus post devaluation).*" Here, a higher score indicates more goal-directed behavior, i.e., participants more frequently preferred the valued over the devalued stimulus after devaluation. Model-based behavior in the sequential decision-making task was assessed with a similar interaction score of stay probabilities at the first stage: "*(rewarded common stimulus choice – rewarded rare stimulus choice) – (unrewarded common stimulus choice - unrewarded rares stimulus choice).*" This indicates more model-based behavior when participants more frequently stayed after having received rewards in common states and no-rewards in rare states. Further, we also use the parameter ω derived from computational modeling which balances the influences of model-free and model-based decision values. Based on a directed a priori hypothesis of a positive association between the main outcome measures of the two paradigms, we report one-tailed p-values. Due to the relatively small sample size, we apply the more conservative Spearman correlation coefficient, which is more robust against outliers.

RESULTS

DEVALUATION PARADIGM

Training

Over the course of training, participants ($n = 14$) chose the high-probability stimulus (delivering the rewarding food with a higher probability) significantly more often compared to the low probability stimulus (**Figure 3**). This was the case for all stimuli associated with the high-probability outcome in the last 10 trials of the training session [the devalued ($T = -3.50$, $p = 0.004$), valued ($T = -2.60$, $p = 0.022$), and neutral ($T = -2.73$, $p = 0.017$) condition].

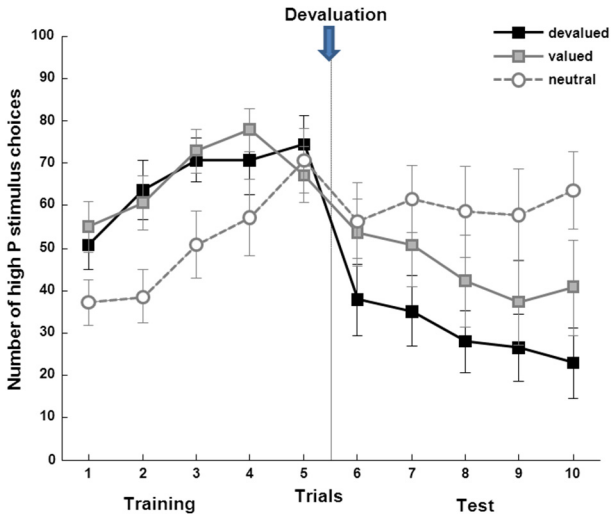

FIGURE 3 | Subjects learn to choose the high probability stimulus in all three conditions (devalued: $T = -3.50$, $p = 0.004$, valued: $T = -2.60$, $p = 0.022$; and neutral: $T = -2.73$, $p = 0.017$), with a significant decrease of instrumental choice after devaluation only for the devalued stimulus ($T = 3.15$, $p = 0.008$).

Outcome devaluation

After devaluation, participants rated the devalued food (chocolate or tomato) significantly less pleasant compared to the valued and neutral condition (**Figure 4**, $T = 2.67$, $p = 0.019$). Further, they reported significantly less hunger after the devaluation procedure (**Figure 4**, $T = 2.25$, $p = 0.042$). These results clearly indicate that the devaluation exerted its expected effect selectively for the devalued but not for the valued outcome.

Test phase in extinction

Assessing choice behavior after the devaluation procedure during the test phase in extinction, a significant time (pre/post training) × condition (devalued/valued/neutral) interaction was found ($F = 5.200$, $p = 0.040$, see **Figure 5**). This was due to a significant decrease in choice of the high-probability stimulus associated with the devalued compared to the stimulus associated with the valued and neutral outcome in the first 10 trials of the test session compared to the last 10 trials of the training session ($T = 3.15$, $p = 0.008$). Thus, participants were able to adapt their choices of stimuli as a function of the associated outcome value, providing direct behavioral evidence for goal-directed behavior as has been previously reported by Valentin et al. (2007).

SEQUENTIAL DECISION-MAKING TASK (TWO-STEP)

In line with previous studies (Daw et al., 2011; Wunderlich et al., 2012; Smittenaar et al., 2013), stay-switch behavior at the first stage revealed a significant main effect of reward ($F = 14.1$, $p =$

0.002) and a significant reward × state interaction ($F = 6.05$, $p = 0.026$, see **Figure 5**). This clearly shows that both rewards and state transitions influenced the participants' choices. Thus, a mixture of model-free and model-based strategies was observed and this was further quantified using a computational model that weights the influence of both strategies. Distribution of random-effects parameters from computational modeling is displayed in **Table 1**.

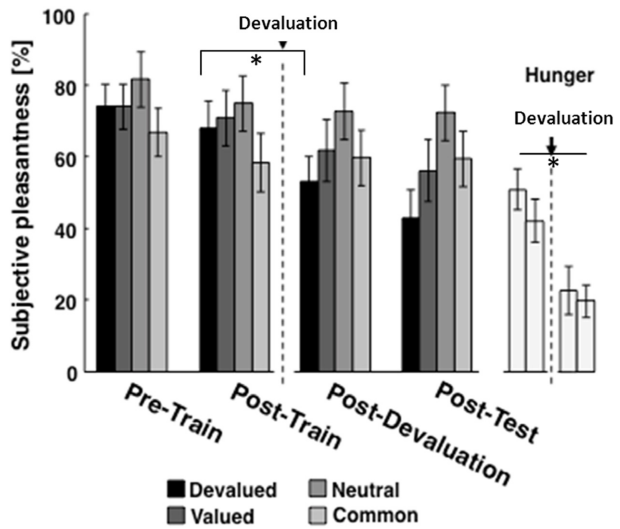

FIGURE 4 | Subjective pleasantness ratings for the devalued (chocolate or tomato), the neutral (water), and the common (fruit tea) outcome at 4 time points throughout the experimental procedure. After devaluation, participants rated the devalued food (chocolate or tomato) significantly (as indicated by *) less pleasant compared to the valued and neutral condition ($T = 2.67$, $p = 0.019$). Further, they reported significantly less hunger after the devaluation procedure (panels display subjective hunger ratings at the 4 time points, $T = 2.25$, $p = 0.042$).

CONSTRUCT VALIDITY: CORRELATION BETWEEN BOTH MEASUREMENTS

Thirteen subjects were included in the final analysis of both tasks (mean age in years = 46, std = 9). The interaction score derived from the outcome devaluation task correlated significantly with the interaction score derived from the sequential decision-making task (Spearman's rho = 0.708, $p < 0.005$, one tailed) and also with the parameter ω derived from computational modeling (Spearman's rho = 0.498, $p < 0.05$, one tailed). When removing one outlier for the model-based score ($3SD >$ mean), the correlation still remained significant in 12 participants (**Figure 6**).

Interestingly, the interaction term from the selective devaluation task did not correlate with the main effect of reward or with the parameter λ (scaling the influence of reward prediction errors on first-stage decision values) derived from computational modeling ($p > 0.75$).

DISCUSSION

In the present study, we used two reinforcement learning tasks in the same participants, selective devaluation and sequential decision-making, which are frequently used in human research.

Table 1 | Best-fitting parameter estimates shown as median plus quartiles across subjects.

	β1	β2	α1	α2	λ	ω	p
25th percentile	4.57	1.53	0.28	0.40	0.44	0.34	0.15
median	6.55	2.42	0.56	0.58	0.70	0.43	0.20
75th percentile	7.55	4.35	0.76	0.69	0.85	0.54	0.26

β, stochasticity of the choices for the first (β 1) and second (β 2) stage; α, learning rate for first (α 1) and second (α 2) stage; λ, reinforcement eligibility parameter (estimated value of the second stage should act as the same sort of model-free reinforcer for the first stage choice); ω, relative influence of model-free and model-based values; p, first-stage choice perseveration.

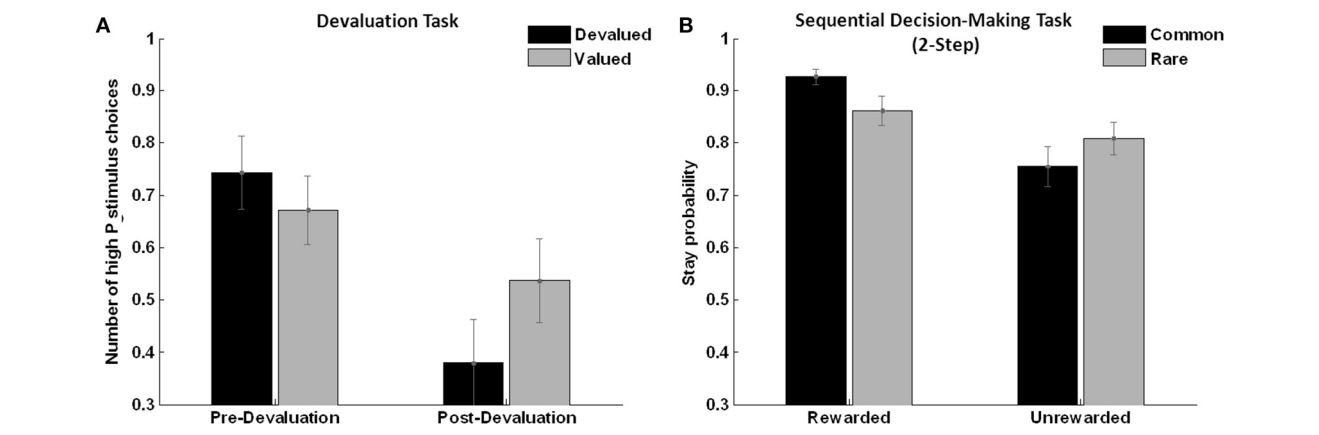

FIGURE 5 | Analysis of choice behavior. (A) Devaluation task: subjects show significantly more valued compared to devalued stimulus choices after devaluation in extinction ($n = 14$, $F = 5.20$, $p = 0.040$, error bars indicate s.e.m.), reflecting "goal-directed" behavior. **(B)** Sequential choice task: the same subjects show higher stay probabilities in the "rewarded common" as opposed to the "unrewarded common" trials (main effect reward: ($n = 17$, $F = 14.1$, $p = 0.002$), reflecting "model-based" behavior with a positive reward × frequency interaction over all subjects ($n = 17$, $F = 6.05$, $p = 0.026$).

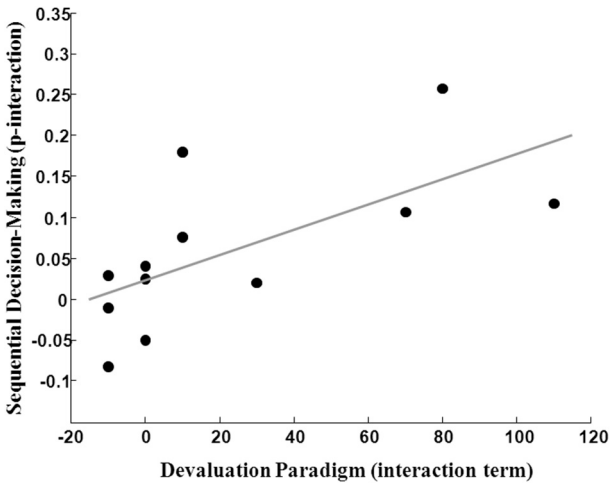

FIGURE 6 | Correlation of "model-based" [sequential decision-making (2-step) interaction term] and "goal-directed" (devaluation paradigm interaction term) behavior ($n = 12$, Spearman $R = 0.74$, $p = 0.003$).

Here, we aim to assess the construct validity of these two measurements which have both been suggested to capture the dichotomy of goal-directed or model-based vs. habitual or model-free control, respectively. In the selective devaluation task, we found evidence of goal-directed choices as subjects decreased their choice for a stimulus associated with a now devalued outcome. In the sequential decision-making task, subjects displayed model-based behavior, which is by definition goal-directed, indicating that participants used the transition structure to solve the task as it is indicated by the significant reward by state interaction and by the weighting parameter ω derived from computational modeling.

As comprehensively reviewed by Dolan and Dayan (2013) those two different operationalizations in part stem from different methodological and historical perspectives. Both selective devaluation and sequential decision-making have been used to describe similar behavioral patterns but they have never been directly related to one another in a sample of human subjects. Here we found, that measures of the individual degree of goal-directed behavior assessed with selective devaluation and model-based behavior assessed during sequential decision-making indeed correlate positively. This provides evidence for the construct validity of both measurements indicating that they measure the same concept grounded in a single common framework as suggested by Dolan and Dayan (2013).

Here, we suggest that goal-directed behavior as measured during selective devaluation reflects one of the many facets of model-based learning which is also applicable to several other tasks, in particular instrumental reversal learning (Hampton et al., 2006; Li and Daw, 2011; Schlagenhauf et al., 2014) but also Pavlovian conditioning (Huys et al., 2012; Prevost et al., 2013). This may suggest that the individual balance between the two different modes of control over instrumental choices may be relevant for a variety of tasks and reflect enduring interindividual differences that are consistent across tasks. Although this balance between goal-directed and habitual control has been considered as interindividual trait (Doll et al., 2012; Dolan and Dayan, 2013) we have to caution that the temporal stability of these measures has not been shown—as it has been the case e.g., for cognitive functions like working memory (Klein and Fiss, 1999; Waters and Caplan, 2003).

Another related question—not addressed here—concerns the notion by Daw et al. (2005) that model-free and model-based learning strategies compete with each other based on the relative certainty of their estimates (Daw et al., 2005). From this theoretical perspective, the model-based system is computationally costly: When individuals face a decision problem, the costs of opportunities of the model-based system need to rule out the benefits of the simple model-free system to govern control over a decision (also compare Niv et al., 2007). In other words, use of the model-based system should be beneficial compared to the model-free system. Lee et al. (2014) suggested that an arbitrator keeps track of the degree of reliability of the two systems and uses this information in order to proportionately allocate behavior control depending on task demands.

The sequential decision-making task used in the present study gives an individual degree of both model-free and model-based behavior. We observed that the degree of goal-directed behavior in the devaluation task was not related to measurements representing the degree of the model-free behavior during sequential decisions (as indicated by the main effect of reward or a high reinforcement eligibility parameter derived from computational modeling). This indicates the specificity of the correlation of goal-directed choices measured with the devaluation procedure and the degree of model-based behavior measured with the sequential decision-making task. One might have expected that a continued choice of the devalued option indicates habitual behavior which is then represented in a small interaction term. A correlation of the interaction term of the devaluation paradigm with measures of model-free behavior in the sequential decision-making task would have indicated that habitual behavior can be induced by the devaluation procedure. The absence of such an association is in line with the findings from Valentin et al. (2007) that on the neuronal level no activation of structures associated with habitual behavior like e.g., the putamen was observed so that the authors conclude that their selective devaluation paradigm is indeed better suited to reflect goal-directed behavior whereas habitual behavior might be observed in tasks using overtraining (Tricomi et al., 2009). To this end, associations between the balance in between model-based and model-free control determined in sequential decision-making should be related to behavioral measures of habitual responding in overtraining paradigms. In the sequential decision-making task used here the outcome probabilities driving model-free behavior during sequential decision-making were changing slowly (according to Gaussian random walks) to facilitate continuous updating of decision values. This was implemented to avoid a moment in time during the task when a purely model-free strategy becomes clearly advantageous compared to the more complex model-based strategy and might have had an effect on the development of habit-like patterns.

Thus, it is important to note that both paradigms may provide different insight into the habitual system, while goal-directed/model-based measurements are more related (and can

be better captured via the two experimental procedures). For example in another variant of devaluation (De Wit et al., 2009) alcohol-dependent patients indeed displayed an overreliance on habits at the cost of goal-directed behavior (Sjoerds et al., 2013). Using sequential decision-making in alcohol-dependent patients, another study demonstrated that model-based behavior is compromised but no difference between patients and controls was observed in terms of model-free behavior (Sebold et al., 2014). While sequential decision-making enables researchers to disentangle model-free and model-based contributions to decision-making, it may obscure enhanced habit-like patterns. To this end, paradigms are needed that are rigorously designed to capture the appropriate predominance of one or the other mode of control given a certain moment in time, also taking into account an arbitrator evaluating the performance of each of these systems (as described by Lee et al., 2014).

Limitations of our study include a relatively small sample size, thus both paradigms and the assessed measurements have been previously validated separately in larger samples (Valentin et al., 2007; Daw et al., 2011; Wunderlich et al., 2012). All results are correlational, hence inferences about causality are very limited. Nevertheless, the strong a priori hypothesis of one, single framework supports the idea of construct validity as assessed by the reported correlation.

Summing up, we suggest that the same construct of goal-directed and model-based behavior is assessed via different experimental procedures (devaluation and sequential decision-making) that validly measure this construct. This is the first study to directly compare these experiments in one sample of human participants. In conclusion, our results support the longstanding and pervasive idea of a common single framework. Therefore, we provide evidence for the construct validity, which merits the use of both experiments in assessing interindividual differences in the predominant type of behavioral control over instrumental choices.

ACKNOWLEDGMENTS

This work was supported by a grant from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) awarded to Florian Schlagenhauf: SCHL1969/2-1 (as part of FOR 1617).

REFERENCES

Balleine, B. W., and Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37, 407–419.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., and Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69, 1204–1215. doi: 10.1016/j.neuron.2011.02.027

Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711. doi: 10.1038/nn1560

De Wit, S., Barker, R. A., Dickinson, A. D., and Cools, R. (2011). Habitual versus goal-directed action control in Parkinson disease. *J. Cogn. Neurosci.* 23, 1218–1229. doi: 10.1162/jocn.2010.21514

De Wit, S., Corlett, P. R., Aitken, M. R., Dickinson, A., and Fletcher, P. C. (2009). Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. *J. Neurosci.* 29, 11330–11338. doi: 10.1523/JNEUROSCI.1639-09.2009

Dickinson, A. (1985). Actions and Habits: the development of behavioural autonomy. philosphical transactions of the royal society of london. *Ser. B Biol. Sci.* 308, 67–78.

Dolan, R. J., and Dayan, P. (2013). Goals and habits in the brain. *Neuron* 80, 312–325. doi: 10.1016/j.neuron.2013.09.007

Doll, B. B., Simon, D. A., and Daw, N. D. (2012). The ubiquity of model-based reinforcement learning. *Curr. Opin. Neurobiol.* 22, 1075–1081. doi: 10.1016/j.conb.2012.08.003

Garner, D. M., Olmsted, M. P., Bohr, Y., and Garfinkel, P. E. (1982). The eating attitudes test: psychometric features and clinical correlates. *Psychol. Med.* 12, 871–878.

Gillan, C. M., Papmeyer, M., Morein-Zamir, S., Sahakian, B. J., Fineberg, N. A., Robbins, T. W., et al. (2011). Disruption in the balance between goal-directed behavior and habit learning in obsessive-compulsive disorder. *Am. J. Psychiatry* 168, 718–726. doi: 10.1176/appi.ajp.2011.10071062

Glascher, J., Daw, N., Dayan, P., and O'doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* 66, 585–595. doi: 10.1016/j.neuron.2010.04.016

Hampton, A. N., Bossaerts, P., and O'doherty, J. P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J. Neurosci.* 26, 8360–8367. doi: 10.1523/JNEUROSCI.1010-06.2006

Huys, Q. J., Cools, R., Golzer, M., Friedel, E., Heinz, A., Dolan, R. J., et al. (2011). Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS Comput. Biol.* 7:e1002028. doi: 10.1371/journal.pcbi.1002028

Huys, Q. J., Eshel, N., O'nions, E., Sheridan, L., Dayan, P., and Roiser, J. P. (2012). Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Comput. Biol.* 8:e1002410. doi: 10.1371/journal.pcbi.1002410

Klein, K., and Fiss, W. H. (1999). The reliability and stability of the Turner and Engle working memory task. *Behav. Res. Methods Instrum. Comput.* 31, 429–432.

Lau, B., and Glimcher, P. W. (2005). Dynamic response-by-response models of matching behavior in rhesus monkeys. *J. Exp. Anal. Behav.* 84, 555–579. doi: 10.1901/jeab.2005.110-04

Lee, S. W., Shimojo, S., and O'doherty, J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. *Neuron* 81, 687–699. doi: 10.1016/j.neuron.2013.11.028

Li, J., and Daw, N. D. (2011). Signals in human striatum are appropriate for policy update rather than value prediction. *J. Neurosci.* 31, 5504–5511. doi: 10.1523/JNEUROSCI.6316-10.2011

Niv, Y., Daw, N. D., Joel, D., and Dayan, P. (2007). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology (Berl.)* 191, 507–520. doi: 10.1007/s00213-006-0502-4

Prevost, C., Mcnamee, D., Jessup, R. K., Bossaerts, P., and O'doherty, J. P. (2013). Evidence for model-based computations in the human amygdala during Pavlovian conditioning. *PLoS Comput. Biol.* 9:e1002918. doi: 10.1371/journal.pcbi.1002918

Schlagenhauf, F., Huys, Q. J., Deserno, L., Rapp, M. A., Beck, A., Heinze, H. J., et al. (2014). Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. *Neuroimage* 89, 171–180. doi: 10.1016/j.neuroimage.2013.11.034

Sebold, M., Deserno, L., Nebe, S., Schad, D., Garbusow, M., Hägele, C., et al. (2014). Model-based and model-free decisions in alcohol dependence. *Neuropsychobiology.* (in press).

Sjoerds, Z., De Wit, S., Van Den Brink, W., Robbins, T. W., Beekman, A. T., Penninx, B. W., et al. (2013). Behavioral and neuroimaging evidence for overreliance on habit learning in alcohol-dependent patients. *Transl. Psychiatry* 3, e337. doi: 10.1038/tp.2013.107

Smittenaar, P., Fitzgerald, T. H., Romei, V., Wright, N. D., and Dolan, R. J. (2013). Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. *Neuron* 80, 914–919. doi: 10.1016/j.neuron.2013.08.009

Tricomi, E., Balleine, B. W., and O'doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *Eur. J. Neurosci.* 29, 2225–2232. doi: 10.1111/j.1460-9568.2009.06796.x

Valentin, V. V., Dickinson, A., and O'doherty, J. P. (2007). Determining the neural substrates of goal-directed learning in the human

brain. *J. Neurosci.* 27, 4019–4026. doi: 10.1523/JNEUROSCI.0564-07.2007

Waters, G. S., and Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behav. Res. Methods Instrum. Comput.* 35, 550–564. doi: 10.3758/BF03195534

Wunderlich, K., Smittenaar, P., and Dolan, R. J. (2012). Dopamine enhances model-based over model-free choice behavior. *Neuron* 75, 418–424. doi: 10.1016/j.neuron.2012.03.042

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 April 2014; accepted: 15 July 2014; published online: 04 August 2014.
Citation: Friedel E, Koch SP, Wendt J, Heinz A, Deserno L and Schlagenhauf F (2014) Devaluation and sequential decisions: linking goal-directed and model-based behavior. Front. Hum. Neurosci. ***8****:587. doi: 10.3389/fnhum.2014.00587*
This article was submitted to the journal Frontiers in Human Neuroscience.

frontiers in
HUMAN NEUROSCIENCE

OPINION ARTICLE
published: 28 August 2014
doi: 10.3389/fnhum.2014.00664

The liberating dimension of human habit in addiction context

***Francisco Güell*[1]* and *Luis Núnez*[2]**

[1] *Mind–Brain Group, Institute for Culture and Society, Universidad de Navarra, Pamplona, Spain*
[2] *Centro Médico Pamplona, Pamplona, Spain*
**Correspondence: fguell@unav.es*

Edited by:
Jose Angel Lombo, Pontifical University of the Holy Cross, Italy

Reviewed by:
Robert Hester, University of Melbourne, Australia

Keywords: habit, motivation, stimulus-response, addiction, addict behavior, goal-directed behavior, neuropsychological impairment, liberating dimension

The notion of habit has acquired an important role within studies of drug addiction and dependence. In general, classical models of addiction conceive of learned compulsive behaviors in terms of a unidirectional stimulus-response model, for which habits are behavior patterns based on studies of animals and are considered to be purely automated—that is, inflexible, highly stimulus bound and insensitive to associated outcomes (Tiffany, 1990; Miles et al., 2003; Everitt and Robbins, 2005). For this approach, learning converts behavior into an automatism, or what some have termed an addictive habit (for example, Hogarth et al., 2013; Sjoerds et al., 2013). Some of these models have been expanded to incorporate motivational aspects of addiction. Such models regard reinforcement (positive or negative) as the initial and central drive for drug abuse (Robinson and Berridge, 1993; Baker et al., 2004) and are situated in a context of a larger, goal-directed, decision-making framework (Cox and Klinger, 1988; Siegel, 2005; Wes, 2006).

Within this overall picture, Sjoerds's team has proposed to expand the habit formation model by distinguishing between motor habits and motivational habits (Sjoerds et al., 2014). In the case of motor habits, behavior is based on a stimulus-response model, while motivational habits refer to compulsive behavior that is controlled by an emotional/motivational state and seems to be at least partially goal-directed. Sjoerds's proposal is a marked improvement over a strictly motor-habit notion of addiction, but we believe that it still falls short of the full context in which the notion of habit acquires its full significance. Let us examine this context more closely.

To be sure, all existing theoretical models have contributed to the understanding of drug consumption, abuse, and addiction. Generally, they affirm that habitual addictive behaviors are related to reinforcement and are conditioned by the presence of diverse environmental and motivational factors associated with the moment of consumption. With continuous consumption, the subject gradually consolidates a behavior associated with the results of consuming and, with time (a period which some designate as the appearance of substance dependency; Peer et al., 2013), the behavior becomes more and more compulsive and less flexible. Studies point out that the routine behavior responsible for addiction leads to the appearance of a state of allostasis wherein individuals take drugs no longer to feel "high," but just to feel "right" (Koob and Le Moal, 1997, 2001, 2005; Piazza and Deroche-Gamonet, 2013).

We find it interesting to note how habits are understood within this context. In studies of addiction, "habits" typically refer only to acquired behaviors that incite the subject to consume. That is, regardless of their flexibility and their relation to motivational states, habits acquired by drug addicts are considered to be those specific pathological behaviors that must be eliminated or counteracted. However, within a therapeutic framework, we find a much richer picture of habit.

Basically, such therapies pursue a modification of all the behaviors that are responsible for the consumption of drugs. The principal objective of many approaches is to fight addiction by means of learned techniques for avoiding stimuli associated with the substance (e.g., substance availability, conditioning social and living places, social groups, etc.; Tucker et al., 1990-1991). The problem is that techniques that focus on the elimination of addictive habits do not reinforce the essential supports for what is referred to as personal re-education. In the therapeutic-educative context, it is evident that one of the central consequences of addiction is the loss of habits that are necessary for personal and social realization and which are normally acquired over the course of a healthy life. Of course, if the only objective of the consumer is to obtain the substance so as to avoid the symptoms of withdrawal, any routine behavior that does not have this objective will be useless. But the problem—and here is the crux of the question—is not that addicts have lost or forgotten their daily routines. From a neuropsychological perspective (Robinson and Berridge, 2003; Verdejo-García and Bechara, 2009; García et al., 2011), it has been shown that continuous consumption of drugs deteriorates certain executive functions such that, even after abstinence has begun, cognitive flexibility in the motorization of strategies is reduced and, as a result, the capacity to organize, plan and supervise one's own daily behavior is diminished (Verdejo-García et al., 2004; Verdejo-García, 2005). The addict cannot effectively

confront his addiction until this capacity is restored.

Because in-treatment therapeutic communities (such as "Proyecto Hombre") provide a controlled environment that helps addicts to "kick the habit" and offers treatment for drug abuse, they are an ideal context for scientific research (for example, Verdejo-García, 2007). A quick look at these communities shows that one of the principal problems of drug addicts during the initial and voluntary rehabilitation process is the difficulty in acquiring daily basic routines (Daley, 1989; Verdejo-García, 2005; García et al., 2011). In the scientific literature, there are studies that suggest that rapid recovery of cognitive function during abstinence seems possible (Bates et al., 2005; Rapeli et al., 2006; Schrimsher and Parker, 2008). Anyway, in those therapeutic communities it is well know that It takes months before drug addicts are able to have a natural and reasonable daily routine and to freely assume everyday life activities—and this is only the basis for confronting addiction.

Accordingly, the following paradox arises: while most models of addiction tend to consider habits only as pathological behaviors that push the patient toward continued consumption, many therapies aim precisely at recovering *the capacity to acquire habits*, which has been damaged by continued drug use. In short, from the addiction standpoint, habit is something that needs to be eliminated, and from a therapeutic point of view, it is something that needs to be re-established. In reality, treatment consists in a mixture of both elimination and reestablishment, and both elements are considered to be of equal importance in addiction treatment for the addict, his family and his friends (Thurgood et al., 2014).

In light of this wider therapeutic context, it is evident that habits encompass much more than what is normally defined as the "habit" of drug addiction. Granted, to group all relevant behaviors under the rubric of habit does not correspond with the rigor usually demanded by science. Also, we recognize that scientific understanding of the biological basis of addiction has been advanced by simplified stimulus-response models based on tests with animals, and that the current use of terms like habit, grounded in this experimental context, has proved useful up to a point. However, as has been shown here, it is the actual scientific community that is beginning to notice the limitations of this notion of habit within more complex contexts relevant to human behavior.

Moreover, while the motivational dimension proposed by Sjoerds constitutes a significant improvement over the notion of habit as merely stimulus-response conditioning, this expanded notion of habit could still be interpreted as merely a conditioned response to a stimulus that incorporates a motivational dimension. Our suggestion is that only by taking into account the fuller, the liberating dimension of habit that is revealed in the therapeutic context can we break free from the stimulus-response model. We believe that this liberating dimension, which regulate the disposition of the subject to facilitate certain daily routines and thereby enable the subject to take on other tasks (Güell, 2014), should be acknowledged in the study of drug dependencies as the characteristic and distinctive dimension of human habits.

ACKNOWLEDGMENTS

This work was supported by Obra Social La Caixa. I am grateful to Institute for Culture and Society (Universidad de Navarra) and, specially, to Nat Barrett PhD for his suggestions.

REFERENCES

Baker, T. B., Piper, M. E., McCarthy, D. E., Majeskie, M. R., and Fiore, M. C. (2004). Addiction motivation reformulated: an affective processing model of negative reinforcement. *Psychol. Rev.* 111, 33–51. doi: 10.1037/0033-295X.111.1.33

Bates, M. E., Voelbel, G. T., Buckman, J. F., Labouvie, E. W., and Barry, D. (2005). Short-term neuropsychological recovery in clients with substance use disorders. *Alcohol. Clin. Exp. Res.* 29, 367–771. doi: 10.1097/01.ALC.0000156131.88125.2A

Cox, W. M., and Klinger, E. (1988). Amotivational model of alcohol use. *J. Abnorm. Psychol.* 97, 168–180. doi: 10.1037/0021-843X.97.2.168

Daley, D. (1989). *Relapse. Conceptual, Research and Clinical Perpsectives*. New York, NY: Haworth.

Everitt, B. J., and Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat. Neurosci.* 8, 1481–1489. doi: 10.1038/nn1579

García, G., García, O., and Secades, R. (2011). Neuropsicologíay adicción a drogas. *Papeles del Psicólogo* 32, 159–165.

Güell, F. (2014). Pre-dispositional constitution and plastic disposition: towards a more adequate descriptive framework for the notions of habits, learning and plasticity. *Front. Hum. Neurosci.* 8:341. doi: 10.3389/fnhum.2014.00341

Hogarth, L., Balleine, B. W., Corbit, L. H., and Killcross, S. (2013). Associative learning mechanisms underpinning the transition from recreational drug use to addiction. *Ann. N.Y. Acad. Sci.* 1282, 12–24. doi: 10.1111/j.1749-6632.2012.06768.x

Koob, G. F., and Le Moal, M. (1997). Drug abuse: hedonic homeostatic dysregulation. *Science* 278, 52–58. doi: 10.1126/science.278.5335.52

Koob, G. F., and Le Moal, M. (2001). Drug addiction, dysregulation of reward, and allostasis. *Neuropsychopharmacology* 24, 97–129. doi: 10.1016/S0893-133X(00)00195-0

Koob, G. F., and Le Moal, M. (2005). Plasticity of reward neurocircuitry and the "dark side" of drug addiction. *Nat. Neurosci.* 8, 1442–1444. doi: 10.1038/nn1105-1442

Miles, F. J., Everitt, B. J., and Dickinson, A. (2003). Oral cocaine seeking by rats: action or habit? *Behav. Neurosci.* 117, 927–938. doi: 10.1037/0735-7044.117.5.927

Peer, K., Rennert, L., Lynch, K. G., Farrer, L., Gelernter, J., and Kranzler, H. R. (2013). Prevalence of DSM-IV and DSM-5 alcohol, cocaine, opioid, and cannabis use disorders in a largely substance dependent sample. *Drug Alcohol Depend.* 127, 215–219. doi: 10.1016/j.drugalcdep.2012.07.009

Piazza, P. V., and Deroche-Gamonet, V. (2013). A multistep general theory of transition to addiction. *Psychopharmacology* 229, 387–413. doi: 10.1007/s00213-013-3224-4

Rapeli, P., Kivisaari, R., Autti, T., Kähkönen, S., Puuskari, V., Jokela, O., et al. (2006). Cognitive function during early abstinence from opioid dependence: a comparison to age, gender, and verbal intelligence matched controls. *BMC Psychiatry* 6:9. doi: 10.1186/1471-244X-6-9

Robinson, T. E., and Berridge, K.C. (1993). The neural basis of Drug craving: an incentive-sensitization theory of addiction. *Brain Res. Brain Res. Rev.* 18, 247–291. doi: 10.1016/0165-0173(93)90013-P

Robinson, T. E., and Berridge, K. C. (2003). Addiction. *Annu. Rev. Psychol.* 54, 25–53. doi: 10.1146/annurev.psych.54.101601.145237

Schrimsher, G. W., and Parker, J. D. (2008). Changes in cognitive function during substance use disorder treatment. *J. Psychopathol. Behav. Assess.* 30, 146–153. doi: 10.1007/s10862-007-9054-0

Siegel, P. K. (2005). *Intoxication: the Universal Drive for Mind-altering Substances*. Rochester: Park Street Press.

Sjoerds, Z., de Wit, S., van den Brink, W., Robbins, T. W., Beekman, A. T. F., Penninx, B. W. J. H., et al. (2013). Behavioral and neuroimaging evidence for overreliance on habit learning in alcohol-dependent patients. *Transl. Psychiatry* 3, e337. doi: 10.1038/tp.2013.107

Sjoerds, Z., Luigjes, J., van den Brink, W., Denys, D., and Yücel, M. (2014). The role of habits and motivation in human drug addiction: a reflection. *Front. Psychiatry* 5:8. doi: 10.3389/fpsyt.2014.00008

Thurgood, S., Crosby, H., Raistrick, D., and Tober, G. (2014). Service user, family and friends' views on the meaning of a 'good outcome' of treatment for an addiction problem. *Drugs Educ. Prev. Policy*. 21, 324–332. doi: 10.3109/09687637.2014.899987

Tiffany, S. T. (1990). A cognitive model of drug urges and drug-use behavior: role of automatic and nonautomatic processes. *Psychol. Rev.* 97, 147–168. doi: 10.1037/0033-295X.97.2.147

Tucker, J. A., Vuchinich, R. E., and Gladsjo, J. A. (1990–1991). Environmental influences on relapse in substance use disorders. *Int. J. Addict.* 25, 1017–1050.

Verdejo-García, A. (2005). Neuropsicología en el ámbito de las drogodependencias (I): evaluación de las funciones ejecutivas. *Proyecto Hombre: revista de la Asociación Proyecto Hombre* 53, 39–43.

Verdejo-García, A. (2007). Profile of executive deficits in cocaine and heroin polysubstance users: common and differential effects on separate executive components. *Psychopharmacology* 190, 517–530. doi: 10.1007/s00213-006-0632-8

Verdejo-García, A., and Bechara, A. (2009). "Neuropsicología y drogodependencias: evaluación, impacto clínico y aplicaciones parala rehabilitación," in *Manual de Neuropsicología Clínica, Madrid, Pirámide Editorial*, ed M. Perez, 179–208.

Verdejo-García, A., López-Torrecillas, F., Orozco, C., and Pérez-García, M. (2004). Clinical implications and methodological challenges in the study of the neuropsychological correlates of cannabis, stimulant and opioid abuse. *Neuropsychol. Rev.* 14, 1–41. doi: 10.1023/B:NERV.0000026647.71528.83

Wes, R. (2006). *Theory of Addiction*. Oxford: Blackwell.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 19 May 2014; paper pending published: 05 July 2014; accepted: 11 August 2014; published online: 28 August 2014.

Citation: Güell F and Núnez L (2014) The liberating dimension of human habit in addiction context. Front. Hum. Neurosci. ***8****:664. doi: 10.3389/fnhum.2014.00664 This article was submitted to the journal Frontiers in Human Neuroscience.*

frontiers in
HUMAN NEUROSCIENCE

HYPOTHESIS AND THEORY ARTICLE
published: 06 October 2014
doi: 10.3389/fnhum.2014.00764

The Wonder Approach to learning

Catherine L'Ecuyer*

Educational Consultant, Educar en el Asombro™, Barcelona, Spain

Edited by:
José Ignacio Murillo, University of Navarra, Spain

Reviewed by:
Juan Narbona, University of Navarra Clinic and School of Medicine, Spain
Carlos Alberto Blanco, Universidad de Navarra, Spain

****Correspondence:***
Catherine L'Ecuyer, Educational Consultant, Educar en el Asombro™, Moixaro 19, Sant Quirze del Vallés, 08192 Barcelona, Spain
e-mail: catherine.lecuyer.iese2004@gmail.com

Wonder, innate in the child, is an inner desire to learn that awaits reality in order to be awakened. Wonder is at the origin of *reality-based consciousness*, thus of learning. The scope of wonder, which occurs at a metaphysical level, is greater than that of curiosity. Unfortunate misinterpretations of neuroscience have led to false brain-based ideas in the field of education, all of these based on the scientifically wrong assumption that children's learning depends on an enriched environment. These beliefs have re-enforced the Behaviorist Approach to education and to parenting and have contributed to deadening our children's sense of wonder. We suggest wonder as the center of all motivation and action in the child. Wonder is what makes life genuinely personal. Beauty is what triggers wonder. Wonder attunes to beauty through sensitivity and is unfolded by secure attachment. When wonder, beauty, sensitivity and secure attachment are present, learning is meaningful. On the contrary, when there is no volitional dimension involved (no wonder), no end or meaning (no beauty) and no trusting predisposition (secure attachment), the rigid and limiting mechanical process of so-called learning through mere repetition become a deadening and alienating routine. This could be described as training, not as learning, because it does not contemplate the human being as a whole.

Keywords: Wonder Approach, learning, attachment, sensitivity, beauty, behaviorism, reality-based consciousness, reality deficit

Omnes homines natura scire desiderant.
All men by nature desire to know. (Aristotle)

INTRODUCTION

It is well documented that the organic constitution of a child's brain plays a key role in his development. But how does a child learn? Is the organic structure of the brain what drives the child to learn? Or is there any state of mind emerging from the brain that is responsible for the desire to learn? Or is the child's learning the mere result of mechanical responses to external stimulus? What is the difference between a child that seizes learning opportunities and one that does not under the same external conditions? Throughout the last decades, many neuroscientists have tried to understand the sense of self, of consciousness, in most cases recognizing that the issue escapes the scope of neuroscience. As a matter of fact, Huxley said, "how it is that any thing so remarkable as a state of consciousness comes about as the result of irritating nervous tissue, is just as unaccountable as the appearance of the Djin when Aladdin rubbed his lamp" (Huxley and Youmans, 1868).

What is the relationship between self-consciousness and learning? What is the origin of learning? Does it come from within the human being, or from without? It is organic, or intangible? Is it a by-product of the neurological makeup, or does it lie deeper than the brain?

Dan Siegel, who himself recognized that "the idea of intention is itself a philosophical puzzle" (Siegel, 2012), also said:

"When we think about psychological development, about the developing mind, it is helpful to think about what the "psyche" actually is. There is an entity called the psyche or the mind that is as real as the brain, the heart, or the lungs, although it cannot be seen directly with or without the aid of microscopes or other tools of modern technology. One definition of the psyche is: "(1) the human soul; (2) the intellect; (3) psychiatry—the mind considered as a subjectively perceived, functional entity, based ultimately upon physical processes but with complex processes of its own: it governs the total organism and its interaction with the environment" (Webster, 1996). Within this definition, we can see the central importance of understanding the psyche, the soul, the intellect, and the mind in understanding human development" (Siegel, 2001).

It is not a coincidence that world spiritual leaders took interest in Siegel's Interpersonal Neurobiology. In 1999, John Paul II invited Siegel to deliver a speech (Towards a Biology of Compassion: Relationships, the Brain and the Development of Mindsight Across the Lifespan) at the Vatican; in 2009, the Dalai Lama shared a panel with Siegel on the scientific basis of compassion.

Regardless of whether we hold religious beliefs or not, and of what they are, there is a growing sense that the motor of the human being is something intangible that cannot be seen with the eye nor can be measured with scientific instruments. Does it emerge from the brain, from interpersonal interaction (as suggested by Siegel's Interpersonal Neurobiology), is it previous to any other human process, or is it embodied within the brain?

At this point, a multidisciplinary approach is necessary in order to get a broader picture.

WONDER: A *REALITY-BASED CONSCIOUSNESS* APPROACH TO LEARNING

More than three centuries B.C., the Greek philosophers Plato and Aristotle said that the principle of philosophy was wonder (Aristotle, 2014; Plato, 2014b), the first manifestation of something intangible that moved the human being towards reality, also defined by Aquinas as "the desire to learn" and later by the English philosopher Francis Bacon as "the seed of knowledge". Chesterton talked about wonder as a principle, not a consequence: "This elementary wonder, however, is not a mere fancy derived from fairy tales; on the contrary, all the fire of fairy tales is derived from this" (Chesterton, 2004a).

More recent authors have written on the importance of wonder for the purpose of awakening ecological awareness in the child (Carson, 1965), as pedagogical proposals or tools to be used in the classroom (Legrand, 1960; Lipman and Sharp, 1986; Egan et al., 2013). But to this day, and despite the fact that it has been discussed during more than twenty-four centuries, wonder has not yet been proposed as a theory of learning.

Not only is the idea of wonder as old as Greek philosophy, it is also a universal phenomenon, well-known by any parent. *Why is it not raining upwards? Why is the moon round and not square?* Children have asked these questions since the beginning of time. When children ask these questions, they might not be demanding an answer. Rather, they might be wondering in the face of reality. They are wondering because it rains downwards and because the moon is round. When children ask these questions, they are, as Plato and Aristotle suggested, philosophizing. They are surprised at the mere fact of seeing that things "are". Babies wonder when they first see the sky, the stars, the face of their mother, when they first touch the grass, see a shadow, experience gravity and so on. As Chesterton wrote: "The most unfathomable schools and sages have never attained to the gravity which dwells in the eyes of a baby of 3 months old. It is the gravity of astonishment at the universe, and astonishment at the universe is not mysticism, but a transcendent common sense. The fascination of children lies in this: that with each of them all things are remade, and the universe is put again upon its trial. As we walk the streets and see below us those delightful bulbous heads, three times too big for the body, which mark these human mushrooms, we ought always to remember that within every one of these heads there is a new universe, as new as it was on the seventh day of creation. In each of those orbs there is a new system of stars, new grass, new cities, a new sea" (Chesterton, 2005).

THE SCOPE OF WONDER

The scope of wonder, as discussed in this present article, is greater than a mere emotional response. It is worth mentioning that many authors, a detailed analysis of which may be found in Artemenko (1972), have referred to "étonnement" (an alternative French translation for "wonder") as a spectrum of emotions ranging from a reaction to novelty, to fear, to surprise, etc. According to the Wonder Approach discussed in this article, the emotional response would be a possible consequence of wonder, not wonder as such.

Furthermore, the scope of wonder goes beyond curiosity. Curiosity is the urge to explain the unexpected (Piaget, 1969), or the urge to know more (Engel, 2011), and may be an instinctual response. Wonder is the desire to know the unknown, as well as the already known. Before the already known, a child may wonder again and again, because to wonder consists in "never taking anything for granted", even that which is already known. So regardless of whether a thing is already known, the wondering attitude is to consider this thing "as if for the first time", as well as "as if for the last time". This metaphysical manner of thinking is typical of a person that realizes that the world is, but also, that could not have been at all. We are—the world is—contingent. If we cease to exist, the world still exists... We participate in something greater than us, the world that surrounds us. Wonder is precisely what allows us to be conscious of the surrounding reality, through humility and gratitude. Wonder is a sort of *reality-based consciousness*, which perhaps could shed some light on the issue of the subjective aspect of experience that is part of what some have called the "hard problem of consciousness" (Chalmers, 1995).

THE WONDER APPROACH VS. THE BEHAVIORIST APPROACH TO EDUCATION

Contrary to the Wonder Approach would be the Behaviorist Approach to education, according to which everything is programmable and the volitional aspect is irrelevant because the child is completely dependent on the environment in order to learn. Therefore, according to this view, education would be reduced to "bombarding with information" (the more the better) and to "training in habits" (as mere mechanical repetition of actions), as reflected in John Watson's promise "Give me a dozen healthy infants, well-formed, and my own specified world to bring them up in and I'll guarantee to take any one at random and train him to become any type of specialist I might select..." (Watson, 1930). The Behaviorist Approach emphasizes the accumulation of information (knowledge), on external behaviors (skills and mechanical habits) and their emotional and physical reactions in given situations, rather than on the person's internal mental states, such as intentionality, which are much more complex.

According to the Wonder Approach, learning would start from within; it would be an inner personal "desire". The environment would be important, but the environment would not be *per se* what makes the child learn. And so it follows that "more" would not necessarily be better.

In recent years, neuroscience has come to the conclusion that more is not necessarily better and that learning is not a matter of overwhelming "enrichment" or excessive intellectual stimulation:

"There is no need to bombard infants or young children (or possibly anyone) with excessive sensory stimulation in hopes of "building better brains". This is an unfortunate misinterpretation of the neurobiological literature—that somehow "more is better". It just is not so. Parents and other caregivers can "relax" and stop worrying about providing huge amounts of sensory

bombardment for their children. This synaptic overproduction during the early years of life has been proposed to allow for a likelihood that the brain will develop properly within the "average" environment that will supply the necessary minimal amount of sensory stimulation to maintain necessary portions of this genetically created and highly dense synaptic circuitry" (Siegel, 2001).

The "unfortunate misinterpretation of the neurobiological literature" has brought on a series of "neuromyths" and false beliefs in the field of education, such as "more is better" and "earlier is better" (American Academy of Pediatrics, 1968; Goswami, 2006; Howard-Jones, 2007; Hyatt, 2007). These unfortunate misinterpretations have also encouraged false brain-based ideas in the education industry, with products such as Brain Gym®, Baby Einstein™, the use of flashcards in classrooms, attempt to repattern the child's brain through co-ordination exercises, so-called educational toys and videos, etc., all of these based on the scientifically wrong assumption that children's learning depends on an enriched environment during the period of synaptogenesis. Valuable time and money, both of which schools often lack, is being spent in obeisance to these myths (Howard-Jones, 2009). These beliefs have re-enforced the Behaviorist Approach to education and to parenting and have contributed to deadening our children's sense of wonder. The process by which this is suggested to have happened is explained below in more detail.

BEAUTY TRIGGERS WONDER IN THE CHILD

Children wonder because they realize that a thing "is", while it could "not be". What is it in the "being" of the things that surround children that trigger wonder in them? The Greek philosophers have identified some of the properties of "being", one of which is beauty. Thus, one of the properties of "being" of a thing that triggers wonder in children is beauty.

What is beauty? Does it always relate to personal taste? The beauty that philosophers refer to is not a mere esthetic beauty that depends on fashion and tastes and that usually triggers a desire for possession. The beauty to which philosophers, such as Aristotle, Plato and Aquinas refer is defined as the visible expression of truth and goodness. That is why Plato writes: "the power of the good has retired into the region of the beautiful" (Plato, 2014a). In the 21st Century, the distinction between metaphysical and cosmetic beauty might be better understood by reflecting on Dove's commercial slogan "Talk to your daughter about beauty before the beauty industry talks to your daughter".

So what would be beautiful to a child? If beauty is the visible expression of truth and goodness, beauty for a child is anything that responds to the truth and goodness of childhood. For example, children are innocent, they learn at a slower pace compared with adults, they need to trust in an attachment figure as we shall see below, they learn from within, they need silence to process information, they have a special affinity with the natural world and with mystery (a mystery is an infinite opportunity to know, which would be expected to awaken wonder, a desire to know), and so on. A beautiful environment is one that triggers wonder, which results in learning. An environment that respect a child's pace and his innocence, an educational content that goes beyond the rational and mechanical explanation of things and that leaves some space for mystery, opportunities for silence and contemplation, etc.

What is ugliness? Does it exist? Aquinas says that "beauty can be found in all existing beings" (Aquinas, 1965). This is because one of the properties of "being" is beauty, and so for the mere fact of "being", all things hold beauty in themselves, although they might do so in different proportions. Thus, ugliness may be defined as the absence of beauty, which could be partially, but never completely absent. A thing that has a small proportion of beauty in it could be defined as "empty", "vulgar", "not excellent", or "meaningless". Ugliness means less motivation for children to wonder. Children might be fascinated, their mind might be paralyzed before an ugly thing, but it does not trigger wonder in them, it does not broaden their intellectual horizons. So a relevant question would be: what would happen to children's learning if the educational system paid more attention to beauty and filtered what does not hold enough of it?

But how do we know what holds beauty and what does not? Is there an instrument that can measure the percentage of beauty in what surrounds us? Obviously, there is no such instrument. There are sensitive skins and elephant skins, so to speak. The parent's and the educator's sensitivity is what makes them able to perceive the child's needs, what is true and good for them. It is what makes them attune to beauty. In one of the most comprehensive existing studies on child care (NICHD, 2006), a mother's sensitivity (a mother's responsiveness to her children's true needs) has been considered the most consistent predictor of a child's healthy development.

SENSITIVITY IS WHAT MAKES WONDER ATTUNE TO BEAUTY

When wonder encounters reality, it attunes to its beauty. This attunement requires the child's sensitivity. Sensitivity could be defined as the capacity, not only to perceive a thing through the senses, but also to attune to the beauty that is in it. The child's attunement process is a sort of focused attention, or empathy with reality, allowing him to feel the beauty that surrounds him.

An obstacle to this attunement would be, for example, a defect in the senses, which would prevent the child from grasping the essence of a thing. This defect could be organic, or it could be the result of an environment that does not recognize his innate desire for wonder. This could be, for example, the case of a child that has been bombarded with information, strongly stimulated from without, whose senses have been crowded and overwhelmed by intensive technological multitasking and/or consuming environment. As a result, the senses' threshold of "feeling" reality goes down and wonder has less and less to expect from and to work with, until it is as though deadened. When this happens, the child becomes passive, bored and muddled and is increasingly dependant on the external environment in order to pay attention and to learn. This dependence is what would be described in the educational language as "lack of motivation".

As the threshold of "feeling" reality is dropping to dramatically low levels, the child needs more and more external stimulus in order to "feel" reality. This is when addictions could come into the picture.

This phenomenon has been considered relevant in the study of media consumption by children. Research on television viewing has established a relationship between television viewing by children under the age of three and attention problems later on in life (Christakis et al., 2004). According to the overstimulation hypothesis, "the surreal pacing and sequencing of some shows might tax the brain or parts of it, leading to short-term (or long-term) deficits" (Christakis, 2011). In Christakis' words, "prolonged exposure to rapid image change during critical period of brain development would precondition the mind to expect high levels of stimulation and that would lead to inattention in later life" (Dimitri Christakis, on TedxRainier). In other words, the child's mind gets conditioned to a reality that does not normally exist in real life. And so when the mind comes back to real ordinary life, everything seems extraordinarily boring, because it cannot see the beauty in ordinary life. As there is no beauty to attract them, children easily get distracted ("distraction" is the contrary of "attraction") and thus become completely dependent on the external environment.

In another study (Overberg et al., 2012), obese subjects could identify taste qualities less precisely than children and adolescents of normal weight. The reduced taste sensitivity makes them want to consume more. Taste sensitivity is multifactorial, so learning influence, such as exaggerated taste stimuli in early childhood, could play a role. When children's taste is over saturated, they cease to feel and so they need more food to perceive taste qualities, what could lead them to gaining more weight. Another study (Kirsh and Mounts, 2007) concluded than violent video game exposure reduces happy facial emotion recognition.

Similar conclusions have been reached in a Stanford study (Ophir et al., 2009), in which researchers looked at what heavy media multitaskers were good at, in terms of (1) capacity to filter information according to its relevancy; (2) working memory; and (3) capacity to switch efficiently from one task to the other. The study found that they were doing worst on all of these parameters. When trying to process various thoughts "at the same time", we are not attending to all of them in parallel at the same time, but rather shifting our attention back and forth among all of them, the result being that the thoughts that we are trying to attend to "at the same time" receive less of our attention, as we need to recover our train of thought every time we switch our attention from one task to the other. This is why the Nobel Prize laureate Herbert Simon said "What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it" (Simon, 1971). When the external environment overwhelms our senses, wonder is inhibited and we cease to be actively involved in paying attention. We become passive and the external input "consumes our attention", instead of us focusing on the environment. So clearly, more is not necessarily better and learning does not depend completely on the environment, but on the inner capacity to focus the attention on one thought at a time and to recognize what has meaning and what does not.

Clifford Nass, founder and director of the Communications between Humans and Interactive Media Lab, from where the study was carried out, said, "it's very troubling. And we have not yet found something that they're definitely better at than people who don't multitask (. . .) Multitaskers love irrelevancy" (Interview in Frontline, December 3th, 2009). In reality, what might be happening is that heavy media multitaskers, violent video game players and obese people who have lost taste sensitivity, like any other human being, crave beauty and meaning. But heavy external multi-source stimulus leads to the overwhelming of the senses, which could contribute to the loss of sensitivity to beauty and meaning. This makes them incapable of recognizing beauty, and so they search for beauty at random. As their craving for beauty is not easily satisfied, they then enter into an unending circle of compulsive consumption behaviors that make them feel less and less, until they can almost appear to be like philosophical zombies.

These searches for taste, for information, for images, are searches for beauty, for meaning. And a meaningful subjective experience could be described as the result of the encounter of a subject's wonder with beauty. It is meaningful because the human being is made, not only from a philosophical, but also from a neurological point of view, to be attracted by beauty, through wonder. This meaningful encounter between wonder and beauty could be what make a subject's action genuinely personal.

WONDER AND BEAUTY ARE WHAT GIVE MEANING TO THE REPETITION OF ACTIONS IN THE CHILD

According to Montessori, children's repetition is the secret of perfection (Montessori, 1986). But can *any* repetition lead to perfection? A routine is commonly defined as "a regular procedure, customary or prescribed" (Webster, 1983). In the educational context, the routine is often seen as necessary because it gives children a certain sense of security and order, as the children can anticipate what comes next. But what makes routine become an obstacle to the child's development? The routine can have an alienating effect on the child when it converts itself in a mere repetition of acts (an end in itself) that have no meaning whatsoever for the child. When this happens, the child acts in a mechanical way, is not fully conscious of what he is doing because there is no meaningful end to his action, or at least the child does not see it. As a result, the volitional, cognitive and emotional dimensions of the child are not involved, the child does not interiorize what he is doing and so there is no sustainable learning. In this context, the routine is the automation of an action. Rather than being a personal subject, the child becomes an object. This is why the result of this process would be linked to rigidity and limitation, rather than to creativity and imagination. The kind of habit involved in this situation would be the result of coercion, mere inertia, training, or perhaps addiction, but not of education. As Thomas Moore said, "Education is not the piling on of learning, information, data, facts, skills, or abilities—that's training or instruction—but is rather making visible what

is hidden as a seed" (Moore, 1997). As Aquinas (1953) points out, "before the habits of virtue are completely formed, they exist in us in certain natural inclinations, which are the beginnings of the virtues. But afterwards, through practice in their actions, they are bought to their proper completion". Virtue starts from within, not from without. In the mechanical repetition of actions, there is no real education because there is no wonder and no opportunity for beauty. Beauty is what gives the routine meaning to the child. It is what converts the routine into what Saint-Exupery called a "ritual", "what makes one day different from the other days, one hour different from the other hours" (Saint-Exupery, 2000).

So the differential element that converts the child's mere mechanical repetition of actions into a meaningful ritual is beauty. This is why Montessori had children repeating what she called "practical life exercises" (she insisted that their aim was not "practical", rather the emphasis was on the word "life") (Standing, 1998) with "motive of perfection". Montessori insisted on the importance of surrounding children with reality and beauty. As explained earlier, beauty is the visible expression of what is true and good for a child, of what the child's nature is capable of possessing. How can a child's education be the expression of truth and goodness? It is when education facilitates the child to possess that which, by his nature, he is capable of possessing. On the contrary, the education would cease to be beautiful when it does not give this opportunity to the child, or when it urges the child to possess that of which, by his nature, he is not capable of possessing. For example, a child would not be able to learn well under pressure, with high amount of external stimulus that require simultaneous thought processing, extremely fast-paced content, etc.

SECURE ATTACHMENT UNFOLDS WONDER IN THE CHILD

One of the well-known truths about children is that they need to develop a secure attachment relationship with their principal caregiver. How does the attachment process occur and how does it relate to wonder and beauty?

The attachment theory, first developed by Bowlby (1969) and Ainsworth (1967, 1969; Ainsworth et al., 1978), is now one of the most widely recognized and established theoretical approaches in the field of psychological development. Throughout the years, this theory has converted itself into "the dominant approach to understanding early social development" (Schaffer, 2007), has been confirmed by quantities of empirical research in psychology, neurobiology, pedagogy, psychiatry, etc., and is now being used as the basis of most social and childcare research and policy (NICHD, 2006).

According to Bowlby and to numerous studies, secure/insecure attachment is the function of the sensitivity the principal caretaker has towards the prompt resolution of an infant's basic needs for security, safety and protection. This is why a mother's sensitivity has been considered the most consistent predictor of a child's healthy development. This sensitivity is responsiveness, attunement to the reality of the child, with his daily life needs. So what matters is not orchestrated enrichment inputs for children, but a million small acts of responsiveness to daily life needs. Based on the responsiveness pattern, the infant will develop an "Internal Working Model", a paradigm that he has of himself and that will affect all of his future relationships.

For instance, if the infant receives the message: "Your needs cannot be attended to", he will develop the Internal Working Model "I cannot trust others", "The world is hostile", "I am not worthy", "I am not competent". The result is insecure attachment. This leads the child, teenager and adult-to-be to low self-esteem, high insecurity, low social competence and resistance to exploring the unknown. *The message that the child has interiorized is that the world is hostile, that he cannot trust what is around him*. So it would be reasonable to expect that a child with insecure attachment would have a more cynical attitude towards life, one that does not easily trust in beauty, truth and goodness. Therefore, insecure attachment would be expected to inhibit a child's capacity to perceive beauty.

On the other hand, when the infant's basic necessities are promptly addressed, he will develop the Internal Working Model: "I can trust others", "I am worthy", "I am competent". The result is secure attachment. This leads to high self-esteem and security, high social competence and interest in exploring the unknown in the child, the teenager and eventually the adult. *The message that the child has interiorized is that the world is trustworthy*. So it would be reasonable to expect that a child with secure attachment would have a greater predisposition to experience wonder, because he has a natural trusting attitude towards beauty, truth and goodness. Therefore, secure attachment would be expected to foster attunement to beauty.

Thus, one would expect the innate desire in the child for knowledge to flourish in an environment of secure attachment and to be inhibited by insecure attachment. There is a second reason for this. Once children are securely attached to their principal caregiver, they use their principal caregiver as an exploratory base to learn what is around them. What does an eight-month-old child do when introduced to a stranger? He looks at his principal caregiver, and then back and forth to the stranger, as if he were asking his caregiver for permission. What does a four-year-old child do when discovering a snail in the park? "Look mom!" This is no doubt one of the most repeated sentences in playgrounds. Children continually triangle between the reality they discover and their principal caregiver. Carson (1965) rightly points out: "If a child is to keep alive his inborn sense of wonder, he needs the companionship of at least one adult who can share it, rediscovering with him the joy, excitement, and mystery of the world we live in". In fact, securely attached children have been found to be more intellectually curious (Arend et al., 1979). And children have been found to learn better from human interaction than from an enriched environment. For example, not only is there no relation between baby videos and word or foreign language learning, but media exposure has been associated with less vocabulary and delayed language development (Kuhl et al., 2003; Chonchaiya and Pruksananonda, 2008; Richert et al., 2010; Duch et al., 2013).

That does not mean that wonder is a by-product of secure attachment, or that secure attachment precedes wonder. On the contrary, the attachment pattern develops between around 6 months and 3 years old. It would be unreasonable to say that children under 6 month-old do not experience wonder in relation to the world. Rather, it would be reasonable to say that the

attachment pattern outcome can inhibit or foster the existing potential that the child has for wonder.

If wonder is innate in the child, then it also precedes self-consciousness, which starts to appear at the age of two, when the child starts having his own biographical memory, through explicit memory (Siegel, 2012). Therefore, self-consciousness is not necessary for wonder to happen. In fact, it is notorious that infant and children have a capacity for wondering that is much greater than adults. Perhaps not having yet developed a sense of object permanence (the understanding that objects continue to exist even when they cannot be observed) has a positive effect on children's innate sense of wonder, because they literally experience what is around them, again and again, as if it were for the first time. But object permanence cannot explain wonder, because wonder is a phenomenon that occurs throughout life.

THE TRIANGLE OF WONDER: THE CHILD, THE ATTACHMENT FIGURE AND REALITY

According to the Wonder Approach, the teacher is a facilitator in the process of connecting the mind, the will and the heart of the child with what is true, good and beautiful, so that when he becomes a teenager or adult, he will eventually be able to identify and discover them "by himself".

Some interpretations of Constructivism (Piaget, 1999) suggest that the child can and should discover without any guidance. Evidence does not support educational methods such as "pure discovery without guidance" in a young learner, because if he fails to come into contact with the to-be-learned principle, discovery will not be useful in helping the learner to make sense of it (Mayer, 2004; Kirschner et al., 2006). This is because "all teaching comes from pre-existing knowledge" (Aquinas, 1953), a similar idea to what Vygotski (1978) called the *zone of proximal development*. Teaching and knowledge do not just "happen" in a magical way. The young child needs an attachment figure to mediate between him and reality, a process that some have described as *scaffolding* (Bruner, 1987; Hmelo-Silver et al., 2007).

Social Constructivism philosophy goes further by suggesting that reality is actively constructed by the child, who builds his perception through social interactions (Vygotski, 1978; Bruner, 1987). According to the Wonder Approach, neither the attachment figure nor the child can create reality ontologically speaking. Reality is prior to knowledge. As Aquinas (1953) explains, "he who teaches does not cause the truth, but knowledge of the truth, in the learner. For the propositions which are taught are true before they are known, since truth does not depend on our knowledge of it, but on the existence of things". Beyond this ontological difference, the Wonder Approach acknowledges a subjective personal dimension (the child), as well as a social dimension to learning. However, it suggests that learning is reality-based and that *reality deficit* makes learning more difficult. In fact, it has been demonstrated that infants and children learn less from 2D images than from real face-to-face situations. This is known as the *Video Deficit Effect* (Anderson and Pempek, 2005). Furthermore, a study (Diener et al., 2008) comparing infant's reactions to television and live events concluded that they look longer at, reach more to, show more interest in, and exhibit more fear to, real events. Also, when they were shown live and video events simultaneously, they had a preference for real events.

TESTABLE PREDICTIONS AND FURTHER INVESTIGATIONS

Further investigation is needed to test the Wonder construct as a valid approach to learning. Our testable prediction is that wonder, beauty, sensitivity and secure attachment provide the optimal conditions for learning in children. Wonder is innate, so it is assumed to exist in infants. Beauty is understood in our context as "what responds to the truth and goodness of childhood". Investigation is needed to define a comprehensive set of variables, although at this point in time we would expect silence, mystery, respect for a child's pace and innocence, to be optimal conditions for wonder. Sensitivity and attachment could be measured using existing tools.

It would also be relevant to investigate whether the educator's paradigm or anthropological mindset, namely the approach to learning used by the educator (wonder/behaviorist/constructivist/social) has more impact than the method used with the child. For example, the way flashcards are used by Montessori's followers is different from the way they are used by Glenn Doman's followers. We would expect the educator's paradigm to have more impact than the educational method as such.

Finally, it would be of interest to inquire into the consequences of the loss of wonder in a child. Is the educational system promoting wonder, or inhibiting it? Why? Could the loss of wonder, incurred as a result of giving overly exaggerated importance to external stimulus in learning, shed more light on the mechanisms of the increasing number of learning problems, in which environmental factor have been said to play a role? (U.S. Department of Health and Human Services, 1999).

CONCLUSION

We suggest wonder is the center of all motivation and action in the child. Wonder and beauty are what make life genuinely personal. Wonder attunes to beauty through sensitivity and is unfolded by secure attachment. When wonder, beauty, sensitivity and secure attachment are present, learning is meaningful.

On the contrary, when there is no volitional dimension involved (no wonder), no end or meaning (no beauty), no attunement between the volitional dimension and meaning (sensitivity) and no trusting predisposition (secure attachment), the rigid and limiting mechanical process of so-called learning through mere repetition becomes a deadening and alienating routine. This could be described as training, not learning, because it does not contemplate the human being as a whole.

While there is an increasing interest in an holistic and integral vision of the human being in education, there is also a tendency to conceptually fragment man into various parts and pieces, for example through theories that divide intelligence, or through the left- and right-brain balanced approach to learning, which is a consequence of an over-literal interpretation of hemisphere specialization (Goswami, 2006).

What if wonder served to bridge all of these parts and pieces in order to help make sense of them? Aristotle said, "all men by nature desire to know" (Aristotle, 2014). What if wonder were the meeting point between the volitional and the cognitive ("desire",

"to know") dimensions of the human being? This approach involves a change in paradigm because it implies a return back to reality, a switch from self-consciousness towards *reality-based consciousness* as the starting point of learning. In the midst of multidisciplinary confusion, some have been arguing in favor of the *middleman* figure of a *neuroeducator*. Before we consider experimenting this new idea on our children, perhaps it is worth opening up the multidisciplinary debate and paying some attention to the Wonder Approach. This might well be an opportunity to re-consider the classical approach to philosophy as a relevant *middleman* between neuroscience and education. Chesterton once wrote that "the world will never starve for want of wonders; but only for want of wonder" (Chesterton, 2004b). The Wonder Approach is an attempt to prove Chesterton's prophecy wrong, so that, in the midst of so many distractions, our children can wonder again before the irresistible beauty that surrounds them.

AUTHOR CONTRIBUTION

Catherine L'Ecuyer, Bachelor of Laws, MBA, European Master in Research. She is the author of "Educar en el Asombro" (L'Ecuyer, 2012), on which is based the Wonder Approach discussed in this article.

REFERENCES

Ainsworth, M. D. S. (1967). *Infancy in Uganda: Infant Care and the Growth of Attachment.* Baltimore, MD: Johns Hopkins University Press.

Ainsworth, M. D. S. (1969). Objects relations, attachment and dependency. *Child Dev.* 40, 969–1025. doi: 10.2307/1127008

Ainsworth, M. D. S., Blehar, M. C., Waters, E., and Wall, S. (1978). *Patterns of Attachment: A Psychological Study of the Strange Situation.* Hillsdale, NJ: Erlbaum.

American Academy of Pediatrics. (1968). The doman-delacato treatment of neurologically handicapped children. *Neurology* 18, 1214–1215. doi: 10.1212/wnl.18.12.1214

Anderson, D. R., and Pempek, T. A. (2005). Television and very young children. *Am. Behav. Sci.* 48, 505–522. doi: 10.1177/0002764204271506

Aquinas, T. (1953). *Questiones Disputatae de Veritate.* translated by J. V. McGlynn. Chicago: Henry Regnery Company.

Aquinas, T. (1965). *The Pocket Aquinas.* translated by V. J. Bourke. 4th Edn. New York: Washington Square Press.

Arend, R., Gove, F., and Sroufe, A. (1979). Continuity of individual adaptation from infancy to kindergarten: a predictive study of ego-resiliency and curiosity in preschoolers. *Child Dev.* 50, 950–959. doi: 10.2307/1129319

Aristotle, N. (2014). *Metaphysics.* translated by W. D. Ross. Australia: eBooks@Adelaide, The University of Adelaide.

Artemenko, P. (1972). *L'étonnement Ehez L'enfant*. Paris: J. Vrin.

Bowlby, J. (1969). *Attachment and Loss. Vol. I: Attachment.* NY: Basic Books.

Bruner, J. S. (1987). *Actual Minds, Possible Worlds.* Boston: The Jerusalem-Harvard Lectures.

Carson, R. (1965). *The Sense of Wonder.* NY: Harper & Row Publishers.

Chalmers, D. (1995). Facing up to the problem of consciousness. *J. Conscious. Stud.* 2, 200–219.

Chesterton, G. K. (2004a). *Orthodoxy.* MT: Kessinger Publishing.

Chesterton, G. K. (2004b). *Tremendous Trifles.* MT: Kessinger Publishing.

Chesterton, G. K. (2005). *The Defendant.* London: Wildside Press.

Chonchaiya, W. Y., and Pruksananonda, C. (2008). Television viewing associates with delayed language development. *Acta Paediatr.* 97, 977–982. doi: 10.1111/j.1651-2227.2008.00831.x

Christakis, D. A. (2011). The effects of fast-pace cartoons. *Pediatrics* 128, 772–774. doi: 10.1542/peds.2011-2071

Christakis, D. A., Zimmerman, F. J., DiGiuseppe, D. L., and McCarty, C. A. (2004). Early television exposure and subsequent attentional problems in children. *Pediatrics* 113, 708–713. doi: 10.1111/j.1365-2214.2004.00456_4.x

Diener, M. L., Pierroutsakos, S. L., Troseth, G. L., and Roberts, A. (2008). Video versus reality: infant's attention and affective responses to video and live presentations. *Media Psychol.* 11, 418–441. doi: 10.1080/15213260802103003

Duch, H., Fisher, E. M., Ensari, I., Font, M., Harrington, A., Taromino, C., et al. (2013). Association of screen time use and language development in hispanic toddlers: a cross-sectional and longitudinal study. *Clin. Pediatr. (Phila)* 52, 857–865. doi: 10.1177/0009922813492881

Egan, K., Cant, A. I., and Judson, G. (Eds.) (2013). *Wonder-Full Education: The Centrality of Wonder in Teaching and Learning Across the Curriculum.* Oxford, UK: Routledge.

Engel, S. (2011). Children's need to know: curiosity in schools. *Harv. Educ. Rev.* 81, 625–645.

Goswami, U. (2006). Neuroscience and education: from research to practice. *Nat. Rev. Neurosci.* 7, 406–413. doi: 10.1038/nrn1907

Howard-Jones, P. (2007). *Neuroscience and Education: Issues and Opportunities, Commentary by the Teacher and Learning Research Programme.* London: Economic and Social Research Council, TLRP.

Howard-Jones, P. (2009). Scepticism is not enough. *Cortex* 45, 550–551. doi: 10.1016/j.cortex.2008.06.002

Hmelo-Silver, C. E., Duncan, R. G., and Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: a response to kirschner, Sweller, and Clark (2006). *Educ. Psychol.* 42, 99–107. doi: 10.1080/00461520701263368

Huxley, T. H., and Youmans, W. J. (1868). *The Elements of Physiology and Hygiene: A Text-book for Educational Institutions.* NY: Appleton & Co.

Hyatt, K. J. (2007). Brain gym® building stronger brains or wishful thinking? *Remedial Spec. Educ.* 28, 117–124. doi: 10.1177/07419325070280020201

Kirschner, P. A., Sweller, J., and Clark, R. E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching . *Educ. Psychol.* 41, 75–86. doi: 10.1207/s15326985ep4102_1

Kirsh, S. J., and Mounts, J. R. W. (2007). Violent video game play impacts facial emotion recognition. *Aggress. Behav.* 33, 353–358. doi: 10.1002/ab.20191

Kuhl, P. K., Tsao, F. M., and Liu, H. M. (2003). Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proc. Natl. Acad. Sci. U S A* 100, 9096–9101. doi: 10.1073/pnas.1532872100

L'Ecuyer, C. (2012). *Educar en el Asombro.* 10th Edn. Barcelona: Plataforma.

Legrand, L. (1960). *Pour une Pégagogie de L'étonnement.* Neuchâtel: Delachaux et Niestlé.

Lipman, M., and Sharp, A. M. (1986). *Wondering at the World: Instructional Manual to Accompany KIO and GUS. Montclair.* NJ: Institute for the Advancement of Philosophy for Children (with University Press of America).

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *Am. Psychol.* 59, 14–19. doi: 10.1037/0003-066x.59.1.14

Montessori, M. (1986). *The Discovery of the Child.* NY: Ballantine Books, Azkar Books.

Moore, T. (1997). *The Education of the Heart.* NY: Thomas Moore.

National Institute of Child Health and Human Development (NICHD) (2006). *Study of Early Child Care & Youth Development.* Washington: National Institute of Child Health and Human Development.

Ophir, E., Nass, C., and Wagner, A. D. (2009). Cognitive control in media multitaskers. *Proc. Natl. Acad. Sci. U S A* 106, 15583–15587. doi: 10.1073/pnas.0903620106

Overberg, J., Hummel, T., Krude, H., and Wiegand, S. (2012). Differences in taste sensitivity between obese and non-obese children and adolescents. *Arch. Dis. Child.* 97, 1048–1052. doi: 10.1136/archdischild-2011-301189

Piaget, J. (1969). *The Psychology of Intelligence.* NY: Littlefield, Adams.

Piaget, J. (1999). *The Construction of Reality in the Child.* Oxon: Psychology Press.

Plato. (2014a). *Philebus.* translated by B. Jowett. Australia: eBooks@Adelaide, The University of Adelaide.

Plato. (2014b). *Theaetetus.* translated by B. Jowett. Australia: eBooks@Adelaide, The University of Adelaide.

Richert, R. A., Robbs, M. B., Fender, J. G., and Wartella, E. (2010). Word learning from baby videos. *Arch. Pediatr. Adolesc. Med.* 164, 432–437. doi: 10.1001/archpediatrics.2010.24

Saint-Exupery, A. (2000). *The Little Prince.* London: Mariner Books.

Schaffer, R. (2007). *Introducing Child Psychology.* Oxford: Blackwell.

Siegel, J. D. (2001). Towards an interpersonal neurobiology of the developing mind: attachment relationships, "mindsight" and neural integration. *Infant Ment. Health J.* 22, 67–94. doi: 10.1002/1097-0355(200101/04)22:1<67::aid-imhj3>3.0.co;2-g

Siegel, J. D. (2012). *The Developing Mind.* NY: Guilford.

Simon, H. A. (1971). "Designing organizations for an information-rich world," in *Computers, Communications and the Public Interest*, ed M. Greenberger (Baltimore, MD: The Johns Hopkins Press), 40–41.

Standing, E. M. (1998). *Maria Montessori: Her Life and Work.* NY: Penguin Group.

U.S. Department of Health and Human Services. (1999). *Mental Health: A report of the Surgeon General.* Rockville, M.D.: U.S. Department of Health and Services, Substance Abuse and Mental Health Services, Administration National Institute of Mental Health.

Vygotski, L. S. (1978). *Mind in Society.* London: Harvard University Press.

Watson, J. B. (1930). *Behaviorism*. Chicago: University of Chicago Press.

Webster. (1983). *New World Dictionary of the American Language*. New York: Warner Books Paperback Edition.

Webster. (1996). *Webster's Collegiat Dictionary*. New York: Random House.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 June 2014; accepted: 10 September 2014; published online: 06 October 2014.

*Citation: L'Ecuyer C (2014) The Wonder Approach to learning. Front. Hum. Neurosci. **8**:764. doi: 10.3389/fnhum.2014.00764*

This article was submitted to the journal Frontiers in Human Neuroscience.

frontiers in
HUMAN NEUROSCIENCE

OPINION ARTICLE
published: 14 November 2014
doi: 10.3389/fnhum.2014.00918

Habits as learning enhancers

Gloria Balderas*

Departamento de Filosofía, Universidad de Navarra, Pamplona, Spain
**Correspondence: gcbalderas@gmail.com*

Edited by:
Jose Angel Lombo, Pontifical University of the Holy Cross, Italy

Reviewed by:
Beat Meier, University of Bern, Switzerland

Keywords: habit, hierarchical model, action sequence, fast mapping, exclusion learning

INTRODUCTION

Habits are usually associated with both a positive and a negative consequence. The positive consequence is that habits liberate attentional resources and mechanisms (James, 1984, p. 129), thus enabling organisms to perform simultaneous or more complex actions. The negative consequence is that habits become rigid behaviors which persist despite producing harmful outcomes, as in addictions and some neurological disorders. This article proposes that habits also function as learning enhancers. The plausibility of this statement is supported by results from research on word-trained dogs. The use of an animal example has the advantage of parsimony, since it makes possible to show the capacity of habits to facilitate new learning without appealing to highly sophisticated human competences.

Evidence has been found that dogs are able to *fast map* (Kaminski et al., 2004). In studies of language acquisition, the ability to make accurate assumptions about the referent of an unfamiliar word is called *fast mapping*, a phenomenon that has been observed especially in toddlers (Carey and Bartlett, 1978; Swingley, 2010). This article argues that the training in words forms habits that predispose dogs to establish a new word-object association.

The definition of learning as ontogenetic adaptation (De Houwer et al., 2013, p. 633) and the hierarchical view of habit (Dezfouli and Balleine, 2012, 2013) are expounded in Section Learning and Habits. Taking into account these notions, the results of experiments on *fast mapping* in dogs are presented in Section Fast Mapping in Dogs and Learned Habits, to show that habits work as learning enhancers. Finally, there is a brief section of Concluding Remarks.

LEARNING AND HABITS

This section presents a functional definition of learning and the hierarchical view of habits. These two notions serve as a framework to present the results on fast mapping in dogs. Learnings defined as ontogenetic adaptation are "changes in the behavior of an organism that are the result of regularities in the environment of that organism" (De Houwer et al., 2013, p. 633). This definition applies to any observable behavior of living organisms, provided that this behavior is a response to stimuli in their (past or present) environment.

The relevant regularities for learning are stimuli or behaviors that are repeated over time or that are present more than one at a time. Importantly, since this definition includes the behavior of the organism itself as a regularity, it can favor the claim that habits are learning enhancers.

Causality between regularities and behavior is functional: the regularity in the environment can be described as an independent variable whose properties determine the behavior (the dependent variable). In this sense, to say that an organism has learned something is equivalent to a hypothesis about how a (past or present) regularity has caused a change in behavior. Moreover, the definition means that learning is only an adaptation—because it occurs due to a regularity—but not that it must be adaptive (advantageous for the organism).

Dezfouli and Balleine (2012, 2013) have presented evidence for the hierarchical view of habits. In their proposal, habits are more complex than goal-directed actions. Habits are action sequences—macro actions composed of primitive actions—under the control of a global goal-directed system that also governs goal-directed actions. In virtue of this system, the organism can opt for simple (goal-directed) actions or launch a sequence to achieve its goals in efficient manner.

To explain how these sequences are consolidated, Dezfouli and Balleine distinguish between *closed-loop* and *open-loop* execution. At the beginning of learning, feedback is crucial. The organism needs a reward or some clues in the environment to identify and perform the proper behavior (*closed-loop execution*). In advanced stages of training, a step in the sequence is conditioned by the previous step, regardless of feedback stimuli or reward (*open-loop execution*). This independence accounts for the insensitivity to the outcome shown in experiments of reward devaluation and contingency degradation that are standard measures to determine if a habit has been acquired (Dickinson et al., 1983; Dickinson, 1985).

When a sequence is consolidated, reaction times decrease. This occurs because the organism is not evaluating the reward for each primitive action but is acting based on an average of the reward received by previous executions of the macro action. If the environment changes and habitual behavior becomes maladaptive, the sequence can disintegrate after some time when the average reward is diminished.

Since the sequence constitutes a unit and the steps are interdependent, the organism tends to complete it. Each primitive action performed functions as a signal to execute the next action. However, the goal-directed system can regain control to

facilitate the learning of new sequences (Dezfouli and Balleine, 2012, pp. 1047–1048; 2013, pp. 10–11).

In addition, the hierarchical view predicts that if the organism must make a decision in the initial state of a sequence, it exhibits habitual behavior; but if the decision point coincides with a mid-state sequence, the behavior will be goal-directed.

FAST MAPPING IN DOGS AND LEARNED HABITS

A purpose of training dogs with words is to elucidate whether other species share some of the mechanisms involved in human language. Typically, the association between a label and an item is done by presenting simultaneously the object and its name; then the dog is allowed to explore or play with the object; finally, the animal is requested to deliver the item and is rewarded if its behavior is correct.

Several studies have confirmed that some dogs are able to relate an unfamiliar word with a new object, an ability similar to human *fast mapping*.

Kaminski et al. (2004) have examined Rico, a dog that has learned over 200 label-item associations. First, the performance of Rico was tested with a simple version of the fetching-game: the dog was asked to bring a familiar item from another room. Rico correctly brought 37 items during 40 trials. Second, *fast mapping* was tested in sessions in which an unknown item was placed among 7 familiar items: after requesting for one or two familiar items, a new word was used to ask Rico to bring an item. Rico brought the new item in 7 of 10 sessions.

The researchers assumed that Rico's performance could include, among others, a general mechanism for exclusion learning. Markman and Abelev (2004) have suggested that Rico could choose the correct item due to a bias toward novelty; but this objection has been refuted by showing that dogs are able to ignore new items (Fischer et al., 2004; Aust et al., 2008; Pilley and Reid, 2011; Grassmann et al., 2012). In what follows it is assumed that dogs are capable of learning by exclusion. I will attempt to show that this learning is supported by the acquisition of two habits, described according to the hierarchical view: the item-label association and the fetching-game.

At first glance, it seems that the label-item association does not constitute a sequence; however it is a complex behavior. Even in the absence of distracting items, it is possible to distinguish three primitive actions: (i) *search*; (ii) *match*; and (iii) *approximation*. The dog has to look for the object (i); match the item with its label (ii) when the correct item is in view; and show some other behavior (iii) indicating that recognition (i.e., take, paw). It could be insisted that the association is a simple behavior because is identified with the matching (a cognitive response); but in the experimental context, this response is accessible to the observer only because is preceded by the search and followed by another action. Therefore, the execution of the association task must also include these steps. When the macro action has been acquired, animals run it fast.

The fetching-game is a sequence separable from the association task. There is evidence that dogs are able to combine different types of orders with various label-item pairs, therefore dogs can learn different games with the same objects (i.e., pointing-game) (Pilley and Reid, 2011; Ramos and Ades, 2012).

The *fetching-game* is the main macro action which includes *selecting the correct item* as a subordinate macro action. The fetching-game consists in going for and delivering the requested item. Primitive actions begin when Rico receives the order to bring the object. Rico executes three main tasks: (a) go for; (b) select; and (c) deliver; *b* is in turn divided in *look-for*, *match*, and *take*. It should be added that since each matching is different, the animal acquires new sequences every time it learns a word; so that each label-item pair increases its resources for an efficient performance.

When a familiar item is requested *a*, *b*, and *c* function as a unit and the steps are executed without interruption. In the experiments of fast mapping, this behavior changes, and this change can reveal the role of habits in both the detection and the solution of a problem.

DETECTING THE PROBLEM

The dog immediately executes *a* (go for) in response to the request with the unknown word, but the execution stops at *b* (select). This suggests that Rico detects a problem. Rico executes *look-for* but is stuck in *match*, therefore it can not *take* (see also Pilley and Reid, 2011, p. 193). This behavior fits the hierarchical view, in which each primitive action within a sequence is a signal to execute the next action. In this case, *take* can not start because *match* was not executed. Since Rico does not dispose of a name-object association that enables it to complete the task, it is in a situation where it has to make a decision in the middle of the selection task, so the goal-directed system regains control. After solving the problem, the fetching-game sequence follows its tendency to completion and Rico returns to the sequence: it goes to *take* and to *c* (deliver). This description also follows the hierarchical view because at the starting point the behavior begins as a habit, when a decision is required it becomes goal-directed and ends again as a habit after overcoming the difficulty.

SOLVING THE PROBLEM

Successful selection of new items is explained (in part) because dogs use the exclusion mechanism to eliminate options. This process involves the goal-directed system. Nevertheless, exclusion requires a criterion to determine which items must be excluded. The key point is that this criterion is provided by the association sequences consolidated during training. The learned label-item pairs prevent the animal from matching previously labeled items with new labels, and therefore, they guide it to match the new sound with the unnamed item. In addition, the context of the fetching-game also models the behavior of the animal since in this main macro action the reward depends on delivering a specific (correct) item which forces the animal to choose one item rather than perform any other action.

CONCLUDING REMARKS

Exclusion learning involved in fast mapping can be described according to the hierarchical view of habits. To the extent that habits are consolidated sequences, they can be considered as a type of behavioral regularity. According to the

definition of learning as ontogenetic adaptation, behavioral regularities can lead to learning. The performance of Rico and other dogs manifest how habits modulate behavior and guide the animal to detect and solve a problem.

The problem is the absence of a label-item pair that allows completion of the sequences of association and fetching-game. The typical response of word-trained dogs is to stop acting: they do not choose any of the known items. In this situation the role of the habit as learning enhancer resides in the dynamism of the sequence that tends to be completed. If this dynamism is interrupted, the goal-directed system starts the exclusion process.

In addition, overcoming the problem requires habits in two ways. First, the overall context of game-fetching constrains the behavior of the animal to choose only one item. Moreover, the set of available association sequences provides the criteria that eliminate all familiar items that already have a name.

Thus, in this example of behavioral research on animals, acquired habits can be seen as regularities that lead to new learning. This case shows the plausibility of habits as learning enhancers in a parsimonious way. If this claim is accepted, the possibility of a new line of research is open.

ACKNOWLEDGMENTS

I am very grateful to the extremely helpful comments of the reviewer of this journal and to the members of the Mind-Brain Project of the Institute for Culture and Society (ICS) of the University of Navarra.

REFERENCES

Aust, U., Range, F., Steurer, M., and Huber, L. (2008). Inferential reasoning by exclusion in pigeons, dogs, and humans. *Anim. Cogn.* 11, 587–597. doi: 10.1007/s10071-008-0149-0

Carey, S., and Bartlett, E. (1978). "Acquiring a single new word," in *Proceedings of the Stanford Child Language Conference/Papers and Reports on Child Language Development* (Stanford University), 17–29.

De Houwer, J., Barnes-Holmes, D., and Moors, A. (2013). What is learning? On the nature and merits of a functional definition of learning. *Psychon. Bull. Rev.* 20, 631–642. doi: 10.3758/s13423-013-0386-3

Dezfouli, A., and Balleine, B. W. (2012). Habits, action sequences and reinforcement learning. *Eur. J. Neurosci.* 35, 1036–1051. doi: 10.1111/j.1460-9568.2012.08050.x

Dezfouli, A., and Balleine, B. W. (2013). Actions, action sequences and habits: evidence that goal-directed and habitual action control are hierarchically organized. *PLoS Comput. Biol.* 9:e1003364. doi: 10.1371/journal.pcbi.1003364

Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Philos Trans. R. Soc. Lond. B Biol. Sci.* 308, 67–78. doi: 10.1098/rstb.1985.0010

Dickinson, A., Nicholas, D. J., and Adams, C. D. (1983). The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. *Q. J. Exp. Psychol.* 35, 35–51. doi: 10.1080/14640748308400912

Fischer, J., Call, J., and Kaminski, J. (2004). A pluralistic account of word learning. *Trends Cogn. Sci.* 8, 481. doi: 10.1016/j.tics.2004.09

Grassmann, S., Kaminski, J., and Tomasello, M. (2012). How two word-trained dogs integrate pointing and naming. *Anim. Cogn.* 15, 657–665. doi: 10.1007/s10071-012-0494-x

James, W. (1984). *Psychology, Briefer Course*. Vol. 12. Cambridge, MA: Harvard University Press.

Kaminski, J., Call, J., and Fischer, J. (2004). Word learning in a domestic dog: evidence for "fast mapping." *Science* 304, 1682–1683. doi: 10.1126/science.1097859

Markman, E. M., and Abelev, M. (2004). Word learning in dogs? *Trends Cogn. Sci.* 8, 479–481. doi: 10.1016/j.tics.2004.09.007

Pilley, J. W., and Reid, A. K. (2011). Border collie comprehends object names as verbal referents. *Behav. Process.* 86, 184–195. doi: 10.1016/j.beproc.2010.11.007

Ramos, D., and Ades, C. (2012). Two-item sentence comprehension by a dog (Canis familiaris). *PLoS ONE* 7:e29689. doi: 10.1371/journal.pone.0029689

Swingley, D. (2010). Fast mapping and slow mapping in children's word learning. *Lang. Learn. Dev.* 6, 179–183. doi: 10.1080/15475441.2010.484412

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 14 June 2014; accepted: 28 October 2014; published online: 14 November 2014.

Citation: Balderas G (2014) Habits as learning enhancers. Front. Hum. Neurosci. ***8****:918. doi: 10.3389/fnhum.2014.00918*

This article was submitted to the journal Frontiers in Human Neuroscience.

frontiers in
HUMAN NEUROSCIENCE

OPINION ARTICLE
published: 25 July 2014
doi: 10.3389/fnhum.2014.00525

Toward a new conception of habit and self-control in adolescent maturation

Jose Víctor Orón Semper*

Mind-Brain Group, Institute Culture and Society, Universidad de Navarra, Pamplona, Spain
**Correspondence: josevictororon@gmail.com*

Edited by:
Jose Angel Lombo, Pontifical University of the Holy Cross, Italy

Reviewed by:
Jose Angel Lombo, Pontifical University of the Holy Cross, Italy
Tomas Alonso Ortiz, UCM, Spain
Francisco Ceric, Universidad del Desarrollo, Chile

Keywords: habit, self-control, adolescent, grit, emotion regulation

Neuropsychology shows us that adolescent maturation involves three areas: ejective functions, personal identity, and socialization and this maturation is not reached without emotion regulation. If we look at what this emotion regulation is made up of, psychology will tell us motivation, stress, resilience, emotional cognition, self-control, and habits are fields in which emotion regulation is useful. All of them are looking at the same thing but from different points of view. We can consider forming good habits as the outcome of reached emotional regulation by continued effort of self-control. Currently, neuroscience has seen habit like motor routine and for that reason links habits with corticostriatal pathways, but this is a narrow view of habit. In this opinion article we propose others cerebral process that fit better with a more general conception of the habit. This is developed during adolescence through emotion regulation, so education could be crucial to reach healthy or unhealthy habit.

THE FRAME OF ADOLESCENCE

Lately, certain singularities of adolescence have been presented. Lag between cortical and subcortical maturation could explain adolescence's behavior (Ernst et al., 2009; Somerville and Casey, 2010), but we also think this should be present with other transformations typical of the age they are related with self-control and habits.

Nowadays we can confidently say neuropsychological maturation of human beings, far from being closed in the early years of life, extends until the end of the second decade or more. The specific challenge of adolescence is split in three fields: executive functions, identity, and socialization (Crone and Dahl, 2012). Mental processes of executive functions are mainly supported by the prefrontal cortex (García et al., 2009; Delgado-Mejía and Etchepareborda, 2013). Identity and socialization interact with each other and mainly rest in default mode (Dennis and Thompson, 2013; Teicher et al., 2013). These systems work together (Smallwood et al., 2012; Chen et al., 2013), but it is not only the maturation of these systems but also, as we will see, a global maturation and change of the whole brain. Singularity of adolescence is that from that age, their maturation needs are not only a convenient environment and time, but also the youth need to make good decisions and have healthy life experiences. So at the end of adolescence, around middles twenties, we can find young adults or eternal adolescents (Blakemore, 2008; Choudhury et al., 2012; Crone and Dahl, 2012; Giedd, 2012) and emotion regulation is a key component for successful adolescence (Zins et al., 2005; Crone and Dahl, 2012). Knowing that being a teenager does not mean committing inevitably, risky actions. That is because it is not the same sensation seeking or risky actions. Belonging to a given age group neither forces us to commit risky actions, nor guarantee us to be sensible. Only self-control education guarantees us to be sensible (Romer et al., 2010). As we are going to see, all the cerebral systems which support personal maturation mature through adolescence. Nevertheless some systems, like default mode, continue to change throughout life (Campbell et al., 2013).

SELF-CONTROL AND HABITS FROM PSYCHOLOGY

Self-control makes reference to knowing how to deal with our impulses in relation to our long-term goals. On the one hand, this long-term orientation has to do with motivation aspects, and on the other hand, self-control is developed in a stressful or temptation environment. So, we can understand self-control is like the daily way to develop self-regulation (Duckworth et al., 2013b).

Habit can be understood more generally than neuroscience. Neuroscience usually understands habit as a repetition of a given behavior. This is a mechanistic vision. Habit makes reference to an internal state that we can reach through voluntary repetition, and favor to behave in a given way, if we want it to (Bernacer and Giménez-Amaya, 2013; Bernacer et al., 2014). This frees us to pay attention to all the processes and allows us to focus on other processes. So acquiring good habits allows adolescents to successfully transit to adulthood. During childhood and adolescence the named habit is "grit" (Duckworth, 2013; Tough, 2013), what reminds us the philosophical term of perseveration. Grit is a better predictor for success than quotient intelligence (Duckworth et al., 2010, 2013a). Another process that comes from psychology is self-concept. This makes us orientated to behave in one way (Dweck, 2000).

The reason to present self-control and habit together is because maintained self-control creates perseverance, or grit, which is a habit and favors self-control. Sometimes they are presented independently. For instance the experience of the sweet with children aged 4-years-old (Duckworth et al., 2013b) is seen like self-control, but it is evident that parents who bring up their children until 4-years-old, are the same who bring them up for the rest of their lives, where they create habits.

A NEW PROPOSAL FROM SELF-CONTROL AND HABIT IN NEUROSCIENCE

ABOUT SELF-CONTROL

We have to consider several elements

1. Amygdala and accumbens activation. Amygdala by its relations with hippocampus and prefrontal cortex (Kobera et al., 2008) is part of the process of knowing how to wait and not to be hasty, and also for taking on disadvantages because there is a later reward (Pesoa, 2010). Accumbens by its relation with hypothalamus has resources which help to not fall into addiction (Hoebel et al., 2007). Moreover, accumbens by its relations with cortical and sub cortical regions is part of a process of knowing how to delay reward or give up a present good for a future greater good (Cardinal et al., 2002).
2. Traditionally, the reactive character of both nucleuses has been exaggerated, when indeed it is an "educate" reactivity. Glutamatergic projections from prefrontal cortex affect accumbens' dopaminergic receptors fixing one way to react when accumbens receive dopamine from ventral tegmental area and substantia nigra (Picciotto, 2013).
3. We need not forget orbitofrontal cortex, which makes a biological brake over received impulse subcortical. It allows the "fast way" more affective to integrate with the "slowly way" more rational—then the decision-making system works well (Cardinal et al., 2002; Roech et al., 2007; Sladky et al., 2013).
4. The decision making system uses frontoparietal net to make the decision and other operculocingular to keep the action (Fair et al., 2007).

ABOUT HABIT

We can think of all changes in activation which free prefrontal cortex to be in charge of the given process and then work in other aspects of the same process or even others. These changes create tendencies to act.

5. Changing the component of each net and gaining specificity in a given activity (Fair et al., 2007; Dennis and Thompson, 2013).
6. One important area is medial prefrontal cortex, in where we store long-term assessments of our lived experiences. Moreover medial prefrontal cortex sends directly projections to premotor and motor areas. It is useful to not imitate who we are looking at and also to keep our initiative to decide when to act. So this area is highly related with our personalization (Isoda and Noritake, 2013). Hippocampus is more active for short-term, medial prefrontal cortex for long-term (Bonnici et al., 2012) and lateral and medial parietal for supporting our believes and self-concept because they are part of default mode. This system is active in the process of self-reference and consciousness (Mason et al., 2007; Fransson and Marrelec, 2008).
7. There is one event well-known as "switch backward" and it happens at the end of adolescence. This process frees prefrontal cortex from having to do everything. So it is free for working on other things. It reminds us the concept of habit of the present topic. We are going to number several of them:

   (a) Ventromedial of prefrontal cortex changes its activation to entorhinal and temporal cortex for leading attention and then affects to episodic codification (Schott et al., 2011);
   (b) Medial prefrontal cortex changes its activation to temporoparietal junction for mentalization and perspective taken (Crone and Dahl, 2012);
   (c) From anterior cingulate cortex to parietal and occipital for filtering what is irrelevant (Velanova et al., 2008);
   (d) From dorsomedial prefrontal cortex to superior and posterior temporal for distinguishing between physical cause and intentional cause (Pfeifer and Blakemore, 2012);
   (e) From medial prefrontal cortex to temporal cortex for self-concept (Sebastian et al., 2008);
   (f) From dorsolateral prefrontal cortex to anterior cingulate cortex for impulse control (Fair et al., 2007).

CONCLUSION

We have hypothesized several cerebral changes than could support a widely idea of habit and self-control. And as the period when these processes are formed is during adolescence, we highlighted adolescence education. The issue is not whether they reach habits, they will get it, however the issue is what kind of habits they are.

In this opinion article, we have marked only some points to offer a broad view of habit and self-control. These assertions need to be contextualized therefore in a more general frame. It is also needed to make differences between emotion, cognition, decision making, and so on in order to integrate them into a singular action. So we need to think about how to relate functional levels to neuroanatomical ones. And we need to consider the differences of importance among neurotransmitters because their influence has multilevel explanation. All of this shows the complexity of habit and self-control

ACKNOWLEDGMENT

Supported by Fundacion La Caixa.

REFERENCES

Bernacer, J., Balderas, G., Martinez-Valbuena, I., Pastor, M. A., and Murillo, J. I. (2014). The problem of consciousness in habitual decision making. *Behav. Brain Sci.* 37, 21–22. doi: 10.1017/S0140525X13000642

Bernacer, J., and Giménez-Amaya, J. M. (2013). "On habit learning in neuroscience and free will," in *Is Science Compayible with Free Will?*, eds A. Suarez and P. Adams (New York, NY: Springer), 177–193.

Blakemore, S. J. (2008). The social brain in adolescence. *Nat. Rev.* 9, 267–277. doi: 10.1038/nrn2353

Bonnici, H. M., Chadwick, M. J., Lutti, A., Hassabis, D., Weiskopf, N., and Maguire, E. A. (2012). Detecting representations of recent and remote autobiographical memories in vmPFC and Hippocampus. *J. Neurosci.* 32, 16982–16991. doi: 10.1523/JNEUROSCI.2475-12.2012

Campbell, K. L., Grigg, O., and Saverino, C. (2013). Age differences in the intrinsic functional

connectivity of default network subsystems. *Front. Aging Neurosci.* 5:73. doi: 10.3389/fnagi.2013.00073

Cardinal, R. N., Parkinson, J. A., Hall, J., and Everitt, B. J. (2002). Emotion and motivation: the role of the amygdale, ventral striatum, and prefrontal cortex. *Neurosci. Biobehav. Rev.* 26, 321–352. doi: 10.1016/S0149-7634(02)00007-6

Chen, A. C., Oathes, D. J., Chang, C., Bradley, T., Zhou, Z. W., Williams, L. M., et al. (2013). Causal interactions between fronto-parietal central executive and default-mode networks in humans. *Proc. Natl. Acad. Sci. U.S.A.* 110, 19944–19949. doi: 10.1073/pnas.1311772110

Choudhury, S., McKinney, K. A., and Merten, M. (2012). Rebelling against the brain: public engagement with the 'neurological adolescent'. *Soc. Sci. Med.* 74, 565–573. doi: 10.1016/j.socscimed.2011.10.029

Crone, E. A., and Dahl, R. E. (2012). Understanding adolescence as a period of social–affective engagement and goal flexibility. *Nat. Rev.* 16, 636–650. doi: 10.1038/nrn3313

Delgado-Mejía, I., and Etchepareborda, M. C. (2013). Trastornos de las funciones ejecutivas. diagnóstico y tratamiento. *Rev. Neurol.* 57(Suppl. 1), S94–S103.

Dennis, E. L., and Thompson, P. M. (2013). Mapping connectivity in the developing brain. *Int. J. Dev. Neurosci.* 31, 525–542. doi: 10.1016/j.ijdevneu.2013.05.007

Duckworth, A. L. (2013). What sets high achievers apart? *Monit. Psychol.* 44:11.

Duckworth, A. L., Kimand, B., and Tsukayama, E. (2013a). Life stress impair self-control in early adolescence. *Front. Psychol.* 3:608. doi: 10.3389/fpsyg.2012.00608

Duckworth, A. L., Tsukayamaa, E., and Geier, A. B. (2010). Self-controlled children stay leaner in the transition to adolescence. *Appetite* 54, 304–308. doi: 10.1016/j.appet.2009.11.01

Duckworth, A. L., Tsukayama, E., and Kirby, T. A. (2013b). Is it really self-control? Examining the predictive power of the delay of gratification task. *Pers. Soc. Psychol. Bull.* 39, 843–855. doi: 10.1177/0146167213482589

Dweck, C. (2000). *Self-Theories: Their Role in Motivation, Personality and Development*. New York; London: Psychology Press.

Ernst, M., Romeo, R. D., and Andersen, S. L. (2009). Neurobiology of the development of motivated behaviors in adolescence: a window into a neural systems model. *Pharmacol. Biochem. Behav.* 93, 199–211. doi: 10.1016/j.pbb.2008.12.013

Fair, D. A., Dosenbach, N. U. F., Church, J. A., Cohen, A. L., Brahmbhatt, S., Miezin, F. M., et al. (2007). Development of distinct control networks through segregation and integration. *Proc. Natl. Acad. Sci. U.S.A.* 104, 13507–13512. doi: 10.1073/pnas.0705843104

Fransson, P., and Marrelec, G. (2008). The precuneus/posterior cingulate cortex plays a pivotal role in the default mode network: evidence from a partial correlation network analysis. *Neuroimage* 42, 1178–1184. doi: 10.1016/j.neuroimage.2008.05.059

García, A., Enseñat, A., Tirapu, J., and Roig-Roviraa, T. (2009). Maduración de la corteza prefrontal y desarrollo de las funciones ejecutivas durante los primeros cinco años de vida. *Rev. Neurol.* 48, 435–440.

Giedd, J. N. (2012). The digital revolution and adolescent brain evolution. *J. Adolesc. Health* 51, 101–105. doi: 10.1016/j.jadohealth.2012.06.002

Hoebel, B. G., Avena, N. M., and Rada, P. (2007). Accumbens dopamine-acetylcholine balance in approach and avoidance. *Curr. Opin. Pharmacol.* 7, 617—627. doi: 10.1016/j.coph.2007.10.014

Isoda, M., and Noritake, A. (2013). What makes the dorsomedial frontal cortex active during reading the mental states of others? *Front. Neurosci.* 7:232. doi: 10.3389/fnins.2013.00232

Kobera, H., Feldman, L., Barrett, B. C., Bliss-Moreau, E., Lindquist, K., and Wager, T. D. (2008). Functional grouping and cortical–subcortical interactions in emotion: a meta-analysis of neuroimaging studies. *Neuroimage* 42, 998–1031. doi: 10.1016/j.neuroimage.2008.03.059

Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., and Macrae, C. E. (2007). Wandering minds: the default network and stimulus-independent thought. *Science* 315, 393–395. doi: 10.1126/science.1131295

Pesoa, L. (2010). Emotion and cognition and the amygdala: from "what is it?" to "what's to be done?" *Neuropsychologia* 48, 3416–3429. doi: 10.1016/j.neuropsychologia.2010.06.038

Pfeifer, J. H., and Blakemore, S.-J. (2012). Adolescent social cognitive and affective neuroscience: past, present, and future. *Soc. Cogn. Affect. Neurosci.* 7, 1–10. doi: 10.1093/scan/nsr099

Picciotto, M. R. (2013). An indirect resilence to adiction. *Nat. Neurosci.* 16, 521–523. doi: 10.1038/nn.3375

Roech, M. R., Calu, D. J., Burke, K. A., and Schoenbaum, G. (2007). Should I stay or should I go?: transformation of time-discounted rewards in orbitofrontal cortex and associated brain circuits. *Ann. N.Y. Acad. Sci.* 1104, 21–34. doi: 10.1196/annals.1390.001

Romer, D., Duckworth, A. L., Sznitman, S., and Park, S. (2010). Can adolescents learn self-control? Delay of gratification in the development of control over risk taking. *Prev. Sci.* 11, 319–330. doi: 10.1007/s11121-010-0171-8

Schott, B. H., Niklas, C., Kaufmann, J., Bodammer, N. C., Machts, J., Schütze, H., et al. (2011). Fiber density between rhinal cortex and activated ventrolateral prefrontal regions predicts episodic memory performance in humans. *Proc. Natl. Acad. Sci. U.S.A.* 108, 5408–5413. doi: 10.1073/pnas.1013287108

Sebastian, C., Burnett, S., and Blakemore, S. J. (2008). Development of the self-concept during adolescence. *Trends Cogn. Sci.* 12, 441–446. doi: 10.1016/j.tics.2008.07.008

Sladky, R., Höflich, A., and Küblböck, M. (2013). Disrupted effective connectivity between the amygdala and orbitofrontal cortex in social anxiety disorder during emotion discrimination revealed by dynamic causal modeling for fMRI. *Cereb. Cortex*. doi: 10.1093/cercor/bht279. [Epub ahead of print].

Smallwood, J., Brown, K., Baird, B., and Schooler, J. W. (2012). Cooperation between the default mode network and the frontal–parietal network in the production of an internal train of thought. *Brain Res.* 1428, 60–70. doi: 10.1016/j.brainres.2011.03.072

Somerville, L. H., and Casey, B. J. (2010). Developmental neurobiology of cognitive control and motivational systems. *Curr. Opin. Neurobiol.* 20, 236–241. doi: 10.1016/j.conb.2010.01.006

Teicher, M. H., Anderson, C. M., Ohashi, K., and Polcari, A. (2013). Childhood maltreatment: altered network centrality of cingulate, precuneus, temporal pole and insula. *Biol. Psychiatry*. doi: 10.1016/j.biopsych.2013.09.016. [Epub ahead of print].

Tough, P. (2013). *How Children Succeed: Grit, Curiosity and the Hidden Power of Character*. London: Random House Books.

Velanova, K., Wheeler, M. E., and Luna, B. (2008). Maturational changes in anterior cingulate and frontoparietal recruitment support the development of error processing and inhibitory control. *Cereb. Cortex* 18, 2505–2522. doi: 10.1093/cercor/bhn012

Zins, J. E., Weissberg, R. P., Wang, M. C., and Wlberg, H. J. (eds.). (2005). *Building Academic Success on Social and Emotional Learning*. New York, NY: Colombia University.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 31 March 2014; accepted: 28 June 2014; published online: 25 July 2014.

Citation: Orón Semper JV (2014) Toward a new conception of habit and self-control in adolescent maturation. Front. Hum. Neurosci. ***8****:525. doi: 10.3389/fnhum.2014.00525*

This article was submitted to the journal Frontiers in Human Neuroscience.

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 02 July 2014
doi: 10.3389/fnhum.2014.00489

From episodic to habitual prospective memory: ERP-evidence for a linear transition

Beat Meier[1,2] *, Sibylle Matter[1], Brigitta Baumann[1], Stefan Walter[1,2] and Thomas Koenig[2,3]

[1] Institute of Psychology, Experimental Psychology and Neuropsychology, University of Bern, Bern, Switzerland
[2] Center for Cognition, Learning and Memory, University of Bern, Bern, Switzerland
[3] Department of Psychiatric Neurophysiology, University Hospital of Psychiatry, Bern, Switzerland

Edited by:
Javier Bernacer, University of Navarra, Spain

Reviewed by:
Robert West, Iowa State University, USA
Grit Herzmann, The College of Wooster, USA

***Correspondence:**
Beat Meier, Institute of Psychology, Experimental Psychology and Neuropsychology, University of Bern, Fabrikstrasse 8, 3012 Bern, Switzerland
e-mail: beat.meier@psy.unibe.ch

Performing a prospective memory task repeatedly changes the nature of the task from episodic to habitual. The goal of the present study was to investigate the neural basis of this transition. In two experiments, we contrasted event-related potentials (ERPs) evoked by correct responses to prospective memory targets in the first, more episodic part of the experiment with those of the second, more habitual part of the experiment. Specifically, we tested whether the early, middle, or late ERP-components, which are thought to reflect cue detection, retrieval of the intention, and post-retrieval processes, respectively, would be changed by routinely performing the prospective memory task. The results showed a differential ERP effect in the middle time window (450–650 ms post-stimulus). Source localization using low resolution brain electromagnetic tomography analysis suggests that the transition was accompanied by an increase of activation in the posterior parietal and occipital cortex. These findings indicate that habitual prospective memory involves retrieval processes guided more strongly by parietal brain structures. In brief, the study demonstrates that episodic and habitual prospective memory tasks recruit different brain areas.

Keywords: intention, habit, recognition, covariance mapping, N300, P3b, parietal old/new effect, prospective positivity

INTRODUCTION

Typically, habits are formed without intention. However, there are situations in which we intentionally and deliberately want to form a habit, for example, when we must remember to take medication on a regular basis. This situation is referred to as habitual prospective memory and the neural basis of its formation is the goal of the present study. Prospective memory can be defined as the ability to remember to perform a previously formed intention at the appropriate occasion. It is highly relevant in everyday life and is involved in tasks such as remembering to buy groceries on the way home from work, to keep an appointment or to comply with a medication prescription regiment. Prospective memory tasks can be classified as episodic when they are concerned with one-time events and they can be classified as habitual when they need to be executed repeatedly (cf., Meacham and Singer, 1977; Meacham and Leiman, 1982; Kvavilashvili and Ellis, 1996; Einstein et al., 1998; Graf, 2005). Although there has been a considerable interest in prospective memory in the last two decades, the main focus was on episodic prospective memory tasks. In particular, the question whether remembering an episodic prospective memory task can occur spontaneously or whether strategic monitoring for the retrieval occasion is necessary is at the core of the current theoretical debate (see McDaniel and Einstein, 2007; Kliegel et al., 2008 for overviews).Only a few studies were concerned with habitual prospective memory (Einstein et al., 1998; Elvevag et al., 2003; Vedhara et al., 2004; Matter and Meier, 2008; Cuttler and Graf, 2009) and these studies were mainly concerned with habitual prospective memory performance deficits in older adults and in patient populations, or with questions related to medication adherence. However, none of these studies has examined the neural correlates of habitual prospective memory. The goal of the current study was to fill this gap and to identify the electrophysiological signature of the transition from episodic to habitual prospective memory using event-related potentials (ERPs).

In previous ERP studies different characteristic modulations of prospective memory have been identified. The *N300* represents an occipital–parietal negativity in an early time window about 300 ms after stimulus-onset and is elicited when prospective targets are compared to ongoing task trials, or when remembered targets are compared to missed prospective memory target trials (West and Covell, 2001; West et al., 2001; West and Ross-Munroe, 2002; West, 2005, 2008). Moreover, the N300 is sensitive to the amount of available attentional resources, that is, increased attentional demands of the ongoing task disrupted the efficiency of prospective memory target detection and led to an attenuation of the N300. Therefore, this component is associated with processes related to the detection of the prospective memory targets and can be considered as the prospective component of a prospective memory task (i.e., remembering that something must be done).

The prospective positivity occurs between 400 and 1200 ms after stimulus-onset which is distributed across central, parietal, and occipital brain areas (see West, 2005, 2008). This positivity is elicited when prospective memory target trials are compared to prospective lures and also, when prospective memory target trials are compared to ongoing task trials (West et al., 2001; West

and Ross-Munroe, 2002; West and Krompinger, 2005). This component can be further subdivided into three components, Pb3, parietal old/new effect, and sustained parietal positivity (West, 2011). The *P3b* is a relatively large positivity over parietal regions and it typically peaks between 300 and 800 ms post-stimulus. It is elicited when infrequent targets are detected, for example during the oddball task (e.g., Kok, 2001). A further component is the *parietal old/new effect*, an effect typically found in studies of recognition memory (Rugg and Curran, 2007). Therefore, it is thought to be associated with processes related to the retrieval of the intention and can be considered as the retrospective component of a prospective memory task (i.e., remembering what has to be done). Both the *P3b* and the *parietal old/new effect* occur in about the same time window, but can be distinguished by their functional relevance.

In addition, a further component which occurs in the later part of this time window and which is expressed mainly on parietal electrodes has been identified. This *sustained parietal positivity* is thought to be related to post-retrieval processes which may support the realization of the intention once it is retrieved (West and Krompinger, 2005; West, 2007; West et al., 2007). Thus, the prospective component (remember that) and the retrospective component (remember what), which are inherent in a prospective memory task, are supported by different ERP-components (Zimmermann and Meier, 2006, 2010; West et al., 2007).

So far, it is not known whether the ERP-components are differentially associated with episodic and habitual prospective memory. It has been proposed that as a task becomes habitual, it requires less attention and its execution becomes more automatic (Einstein et al., 1998; Dismukes, 2008). Therefore it is possible that the detection of prospective memory targets requires less attention and as a consequence, the N300 which has been shown to depend on attentional processes may be attenuated as a task becomes habitual. Moreover, studies with the oddball paradigm have shown that with habituation the P3b is reduced and thus a reduction of the P3b might also be expected when a prospective memory task becomes habitual (Ravden and Polich, 1998).

However, when the dual-task nature of a prospective memory task is considered the opposite result is also possible (cf., Smith, 2003; Bisiacchi et al., 2009). In dual-task paradigms the P3b produced by a secondary task typically decreases in amplitude when the difficulty of a primary task is increased (Strayer and Kramer, 1990; Kramer et al., 1991; Watter et al., 2001). In a habitual prospective memory task the difficulty of the primary (i.e., ongoing) task typically decreases with practice. Thus, the amplitude of the secondary (i.e., the prospective memory task) may increase when the task changes from episodic to habitual. Moreover, with increasing practice of responding to the prospective memory target events this particular stimulus category may be integrated into the task-set of the ongoing task, thus leading to a change of the dual-task structure, and as a consequence to a change in resource allocation (i.e., an increase in spontaneous retrieval).

Further, it is also possible, that by repeatedly retrieving a particular task the parietal old/new effect is affected. From studies of recognition memory it is known that this component is enhanced when confidence in recollection is increased (Curran, 2004). Moreover, as a task becomes habitual the execution of the intention may also become more automatized and accordingly a change in the sustained parietal positivity may occur.

To test these possibilities we conducted two experiments. In Experiment 1, we used verbal materials. The ongoing task was a lexical decision task and the prospective memory task was to respond to a specific target word. In Experiment 2, we used non-verbal materials. The ongoing task was a perceptual discrimination task using abstract shapes and the prospective memory task was to respond to the category of white shapes. In both experiments, a total of forty prospective memory targets occurred and the main question was whether we would find differences between ERP components of the first versus the second half of the experiment.

In order to investigate the neural signature of the episodic to habitual transition, we defined three time windows which were derived from previous ERP studies: an early time window lasting from 250 to 450 ms after stimulus-onset to assess the N300, a middle time window lasting from 450 to 650 ms after stimulus-onset to assess the P3b and the parietal old/new effect and a late time window lasting from 650 to 850 ms to assess the sustained parietal positivity.

EXPERIMENT 1

METHOD

Participants

Twenty-two right-handed psychology students (mean age = 26.5 years, SD = 8.1; 16 female) participated in the study. They were recruited from the departmental subject pool and received course credits for participation. All of them had normal or corrected-to-normal visual acuity and reported no evidence of neurological compromise. The study was approved by the Institutional Review Board and informed written consent was obtained from each participant.

Materials

For a lexical decision task, a total of 610 high frequency nouns (no proper names, no animals) with a length of 4–9 letters were selected from the CELEX database (Baayen et al., 1995). For each word a non-word was generated by keeping the position of the first and the last letter while randomly changing the position of the middle letters, resulting in a pool of 1220 stimuli for the lexical decision task.

From these materials five different blocks were composed. A baseline block contained 40 pseudo-randomly selected trials (20 words and 20 non-words). Four experimental blocks contained 305 trials composed of 295 letter-strings from the stimulus pool and 10 additional prospective memory targets. The word "Hund" (German for "dog") which is the most typical member of the category animal served as the prospective memory target (Hager and Hasselhorn, 1994). A specific rather than a categorical prospective memory target was used because in everyday life habitual prospective memory tasks are generally cued by specific target events (e.g., taking medication after breakfast; cf., Meacham and Leiman, 1982; Dismukes, 2008). In each block, the first prospective memory target was presented at the 30th

position and the last (i.e., the 10th) was presented at the 300th position, respectively. The remaining prospective memory targets were presented at pseudo-randomized intervals of 20, 30, or 40 trials between the first and the last prospective memory target. Across experimental blocks, a total of 100 four-letter-words were randomly distributed. These were used as control items.

Stimuli were presented in uppercase and in 36-point black Arial font against a white background in the center of the screen. The letter-strings were surrounded by a black rectangle with a border width of 2 mm which remained on the screen during the whole experiment. The purpose of the rectangle was to facilitate the fixation of the center of the screen and to minimize eye-movements. The experiment was controlled by E-Prime 1.1 software (Psychology Software Tools, www.pstnet.com) running on an IBM-compatible computer with a 17" VGA monitor.

Procedure

After obtaining consent, the electroencephalography (EEG) recording equipment was set up and the participants were instructed for the lexical decision task. Specifically, they were informed that they would see letter-strings on the computer screen and that for each one they had to decide whether it was a word or a non-word by pressing the "B"-key with the right index finger for a word and the "N"-key with the right middle finger for a non-word. Next, EEG recording started and the baseline block was administered. Then, participants received the prospective memory task instruction. They were informed that an additional goal of the study was to investigate how well they would remember to carry out an intended activity in the future. Participants were asked to press the "H"-key on the computer keyboard with a finger of their right hand whenever the word "Hund" was presented. The test phase consisted of four experimental blocks, separated by short breaks during which participant were told to relax.

Each lexical decision trial lasted 2000 ms. First, a letter-string surrounded by a rectangle was presented for a fixed duration of 1000 ms. Then the letter-string was removed and the empty rectangle stayed for another 1000 ms resulting in a 2000 ms response window. When a participant forgot to press the "H"-key for a prospective memory target trial, a message appeared in the center of the rectangle to remind the participant of the prospective memory task. To continue, participants were instructed to press the "H"-key. This procedure was used to make sure that the task became habitual and that a large number of prospective memory target trials was available for the ERP-analysis. The whole experiment lasted ∼50 min.

EEG recording and analysis

The EEG was digitized (500 Hz, 0.015 to 250 Hz bandpass) and stored from 62 electrodes located according to an extended version of the International 10–20 System using a Brainproducts EEG system. Inter-electrode impedances were kept below 5 kΩ. All electrodes were recorded against Fz. Eye-movements were monitored with two additional electrooculogram (EOG) channels.

For offline data analysis, first, an independent component analysis (ICA) based eye-movement correction was applied (Delorme et al., 2007). Across subjects, between one and three ICA components were considered as related to horizontal and vertical eye-movements and were thus removed. Further periods with remaining artifacts were identified and removed according to visual inspection. Electrodes F1 and F2 had to be excluded in all datasets due to technical problems. The data were filtered offline with a bandpass filter from 0.5 to 20 Hz, the reference channel Fz was reinstated and the data were recomputed against average reference. Artifact-free EEG epochs were extracted from 100 ms before stimulus presentation to 1000 ms after stimulus presentation for correct responses. No pre-stimulus baseline correction was applied to avoid confounding effect of an eventual pre-stimulus CNV.

Identification of the prospective memory modulations

A first analysis was conducted to identify the three prospective memory components N300, parietal old/new effect, and sustained parietal positivity (West, 2005, 2008). Separate ERPs for prospective memory target trials and for four-letter control words from the ongoing task were computed and averaged across subjects. The differences between ongoing task ERPs and prospective memory target ERPs were calculated in three post-stimulus time windows derived from the literature: from 250 to 450 ms (N300), from 450 to 650 ms (parietal old/new effect), and from 650 to 850 ms (sustained parietal positivity).

All ERP comparisons were based on paired topographic analyses of variances (TANOVAs), normalized across electrodes, as a global test for topographic differences (Strik et al., 1998). TANOVAs have been shown to yield similar conclusions as previously used statistical analysis (Wirth et al., 2007), but minimize problems of redundant testing across electrodes or pre-selection of sites for testing. Differences that were significant in the TANOVA were further explored using paired *t*-maps, informing about the scalp distribution of the signal-to-noise ratio of an effect and allowing comparisons with previous studies. Furthermore, since topographic differences assessed by a TANOVA must have resulted from differences in active brain regions, significant TANOVA differences were also investigated using voxel-wise *t* tests of low resolution brain electromagnetic tomography analysis (LORETA, Pascual-Marqui et al., 1994). For source localisation, software from the Cuban Neuroscience Center, Havanna was used, employing an average brain model of the Montreal Neurological Institute (Collins et al., 1994). A forward model consisting of three spheres was used to model piecewise homogenous compartments of the brain, skull and scalp, with radios of 95, 99, and 103 mm respectively. As conductivity ratios 1, 0.0125, 1 for the brain, skull and scalp, respectively, were used (cf., Oostendorp et al., 2000; Zhang et al., 2006). A grid of 3244 points constrained to the gray matter modeled the intracerebral electrical sources. The grid has a resolution of 7, 7, and 8 mm for X, Y, Z axes, respectively. With this information the physical term (electric lead field) that relates the intra-cerebral activity to scalp electric fields was computed. Inverse solutions of the individual mean maps in the significant analysis window were computed for each condition using the LORETA method, normalized for variance across voxels, and a paired *t* test was computed in each voxel. The following contrasts were investigated:

1. Event-related potential differences between the prospective memory and the ongoing task. This is a replication of previous findings and can be considered as *prospective memory effect*.
2. Event-related potential differences between the first and second half of the experiment for prospective memory trials. As in the first half of the experiment the prospective memory task is considered as more episodic and in the second half it is considered as more habitual this difference can be considered as the episodic to habitual *transition effect*.
3. Event-related potential differences between the first and second half of the experiment for ongoing task trials. In order to control for a more general effect that is rather related to repeatedly performing the task and which is not related to the prospective memory task *per se*, this difference can be considered as a *practice effect*.

Covariance mapping

Additionally, a more fine-grained analysis was conducted to investigate the trajectory of the transition effect. Rather than just comparing the first and the second half of the experiment, we tested a linear model using a covariance mapping approach. Covariance mapping allows to identify scalp fields (i.e., covariance maps) that correlate linearly with an external, continuous measure (Koenig et al., 2008). In the present case, this external measure was time, and covariance maps were computed for each subject and separately using each of the artifact-free prospective memory trials (representing the transition effect) and using each of the artifact-free trials of the ongoing task (representing the practice effect). These individual covariance maps were averaged within the early (250–450 ms), middle (450–650 ms), and late (650–850 ms) time window. Topographic consistency tests were applied to test whether the individual mean covariance maps were similar across subjects, which, if significant, would indicate that across subjects, there was a common set of brain regions that showed a linear relation of activation strength with time (Koenig and Melie-García, 2010). Furthermore, to distinguish the transition effect from the practice effect in the covariance maps, these were again compared using paired TANOVAs. The comparison of the covariance maps is mathematically identical to compute differences of prospective memory and ongoing task trials at different time points and then assess the change of this difference as a function of time.

Finally, we estimated the actual trajectory of the transition effect across the 40 trials. This analysis was based on the covariance maps obtained from the prospective memory trials averaged across subjects. The fit of all valid individual single ERP trials with these mean covariance maps across subjects was computed, separately for each time point of the analysis window, and separately for each trial (excluding wrong responses and those with artifacts). These fits were then averaged both across all time points of the analysis window and across all subjects, and plotted against the trial number.

RESULTS

Behavioral data

Prospective memory task. Prospective memory performance was measured as proportion of correct responses to the target word. Performance was 0.91 (SE = 0.02) for the first half and 0.93 (SE = 0.01) for the second half of the experiment, respectively. A paired-samples t test revealed no significant difference $t(21) = -1.1$, $p > 0.05$. Mean reaction time for correct prospective memory targets was 889 ms (SE = 24) and 854 ms (SE = 22) for the first and second part of the experiment, respectively. A paired-samples t test revealed a significant difference, $t(21) = 2.3$, $p < 0.05$, indicating shorter reaction times for the second compared to the first part of the experiment.

Ongoing task. Proportion of correct ongoing task responses was 0.953 (SE = 0.006) for the first and 0.950 (SE = 0.008) for the second half of the experiment. A paired-samples t test revealed no significant difference, $t(21) = 1.1$, $p > 0.05$. Mean reaction time of the ongoing task trials (correct responses) was 690 ms (SE = 15) and 683 ms (SE = 15) for the first and the second part of the experiment, respectively. A paired-samples t test revealed no significant difference $t(21) = 1.4$, $p > 0.05$.

To test whether performing the ongoing lexical decision task was affected by the additional requirement of the prospective memory task, the difference between lexical task performance in the ongoing task and in the baseline trials was calculated. Mean reaction time difference was −37 ms (SE = 16) and −46 ms (SE = 18) for the first and the second part of the experiment, respectively. Accuracy difference was −0.01 (SE = 0.01) both between the practice and the first part as well as between the practice and the second part of the experiment. The results of t tests showed no costs associated with performing the prospective memory task (all ts ≤ 1.5; all ps > 0.05).

Electrophysiological data

Identification of the prospective memory modulations. The TANOVAs comparing prospective memory target trials and control words revealed significant effects in the early time window (250–450 ms, $p < 0.001$), in the middle time window (450–650 ms, $p < 0.05$) and in the late time window (650–850 ms, $p < 0.001$), respectively.

On top of **Figure 1** the grand-mean traces of prospective memory target trials and control words at electrodes where their differences were most pronounced are presented (left), and the t-maps of differences between the two conditions (right). In the early time window, the t-maps revealed a bilateral negativity at electrodes over occipital, parietal, and temporal regions of the scalp indicating the N300 (largest t-value at electrode P4, $t = 7.7$). This was accompanied by positivity at frontal electrodes (cf., West et al., 2001; West and Krompinger, 2005). For the middle time window, a positivity was found at frontal, central, and parietal regions of the scalp, indicating both a P3b and a parietal old/new effect (largest t-value at electrode F4, $t = 4.4$). For the late time window, a positivity was found at parietal electrodes only, indicating the sustained parietal positivity (largest t-value at electrode P3, $t = 6.4$). Therefore, the typical ERP modulations for prospective memory were identified which is a pre-condition for further analyses.

Analysis of the episodic to habitual transition effect. The analysis of the transition effect was based on an average of 17.8 (range = 15–20) valid prospective memory target trials per

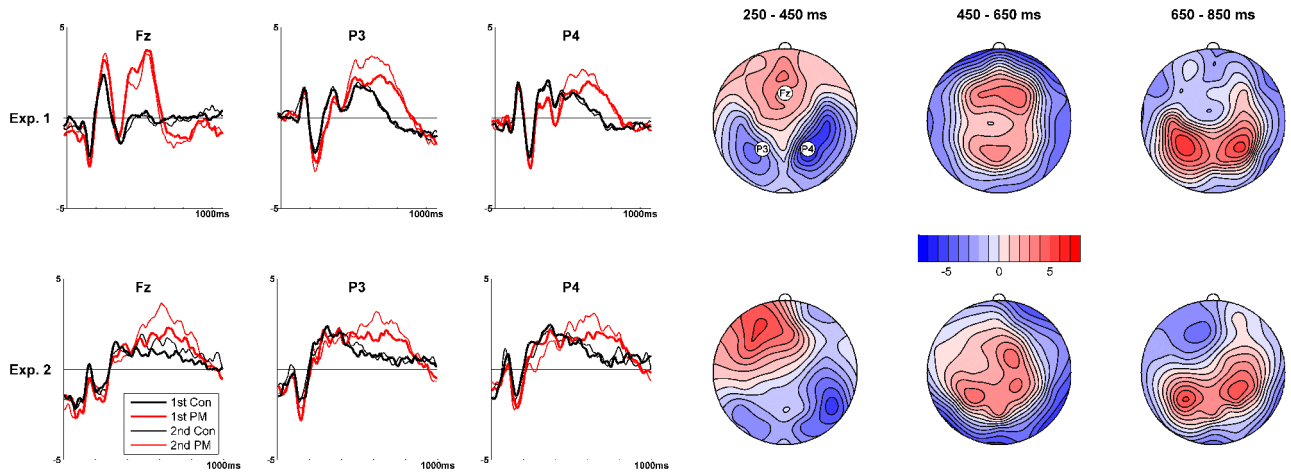

FIGURE 1 | Prospective memory components for Experiment 1 and Experiment 2. Left: ERPs for prospective memory targets (PM; red lines) and control trials (Con; black lines) at selected electrodes, for first (1st) and second half (2nd), respectively. Right: *t*-maps comparing the ERPs from the prospective task with the ERPs from the ongoing task for the early (250–450 ms), middle (450–650 ms), and late time windows (650–850 ms), reflecting the N300, the parietal old/new effect and the late parietal positivity, respectively. The colorbar indicates *t*-values.

subject from the first half, and 18 valid prospective memory target trials (range = 15–20) from the second half. TANOVAs comparing the first and second half ERPs in the three time windows yielded a significant difference in the middle time window ($p < 0.001$), but neither in the early nor in the late time windows ($p = 0.23$ and $p = 0.14$, respectively). **Figure 2A** (top) shows the *t*-map and selected traces of the transition effect in the significant time window. The largest differences were observed at parietal electrodes (largest *t*-value at electrode PO1, *t*-value = 5.0). The traces at the selected electrode Pz show higher amplitudes in the second half (printed in red color) compared to the first half (printed in black color) of the experiment in the middle time window. Our data therefore suggest that the transition of the prospective memory task from episodic to habitual affected either the P3b which would indicate a reallocation of processing capacity or the parietal old/new effect, which is thought to represent the retrospective component of prospective memory.

Low resolution brain electromagnetic tomography analysis voxel-wise statistics are shown in **Figure 3** and indicate that the transition from episodic to habitual prospective memory was associated with an increase in activity in parieto-occipital areas and a decrease in frontal activity. Statistically, regions with significantly higher current density in the second half were identified in occipital and superior parietal brain areas and regions with significantly lower current density in the second half were spread across superior, medial and inferior frontal brain areas ($p < 0.05$).

Analysis of the practice effect. The analysis of the practice effect was based on four-letter control words. From the first half of the experiment, on average 38.9 valid trials per subject (range = 36–41) were available and from the second half, 48.1 trials (range = 41–52) we available. None of the TANOVAs comparing first and second half ERPs was significant ($p = 0.23$, $p = 0.29$, and $p = 0.95$, for the early, middle, and late time windows, respectively). In **Figure 2B** (top), the shapes of the ERPs elicited by the control words from the ongoing task from the first (printed black color) and the second half (printed in red color) are presented. The two waves did not show any apparent differences. Thus, when comparing ERPs recorded in the second against the first half of the experiment, the presence of a significant effect obtained in the prospective memory trials, and the absence of an effect in the control task supports the notion that the transition effect is specific for prospective memory.

Covariance mapping. For the prospective memory trials, the covariance analysis with time as linear predictor yielded covariance maps that were consistent across subjects in the middle ($p < 0.001$) and late ($p < 0.001$) time window, but not in the early time window ($p > 0.99$). These covariance maps revealed a central posterior positivity. Covariance analysis of the practice effect was significant in the early ($p < 0.001$) and in the middle time window ($p < 0.05$), but not in the late window ($p > 0.99$). To compare the differences between the covariance maps of the prospective memory and the control trials, TANOVAs were computed with normalized data. The results showed no difference in the early time window ($p = 0.89$), but the covariance maps differed significantly in the middle and the late time window (both $p < 0.05$).

The estimated trajectory of the transition effect across the 40 trials is shown in **Figure 4** (left side). As expected, there was an overall negative fit with the covariance maps in the trials of the first (more episodic) half of the experiment, and a positive fit in the second (more habitual) half. The fit with the covariance maps can be interpreted as an index for the transition from a more episodic to a more habitual mode of processing. It appears that this transition is rather linear. Indeed, a linear regression of the points of the trajectory explained 74% of the variance ($r = 0.86$).

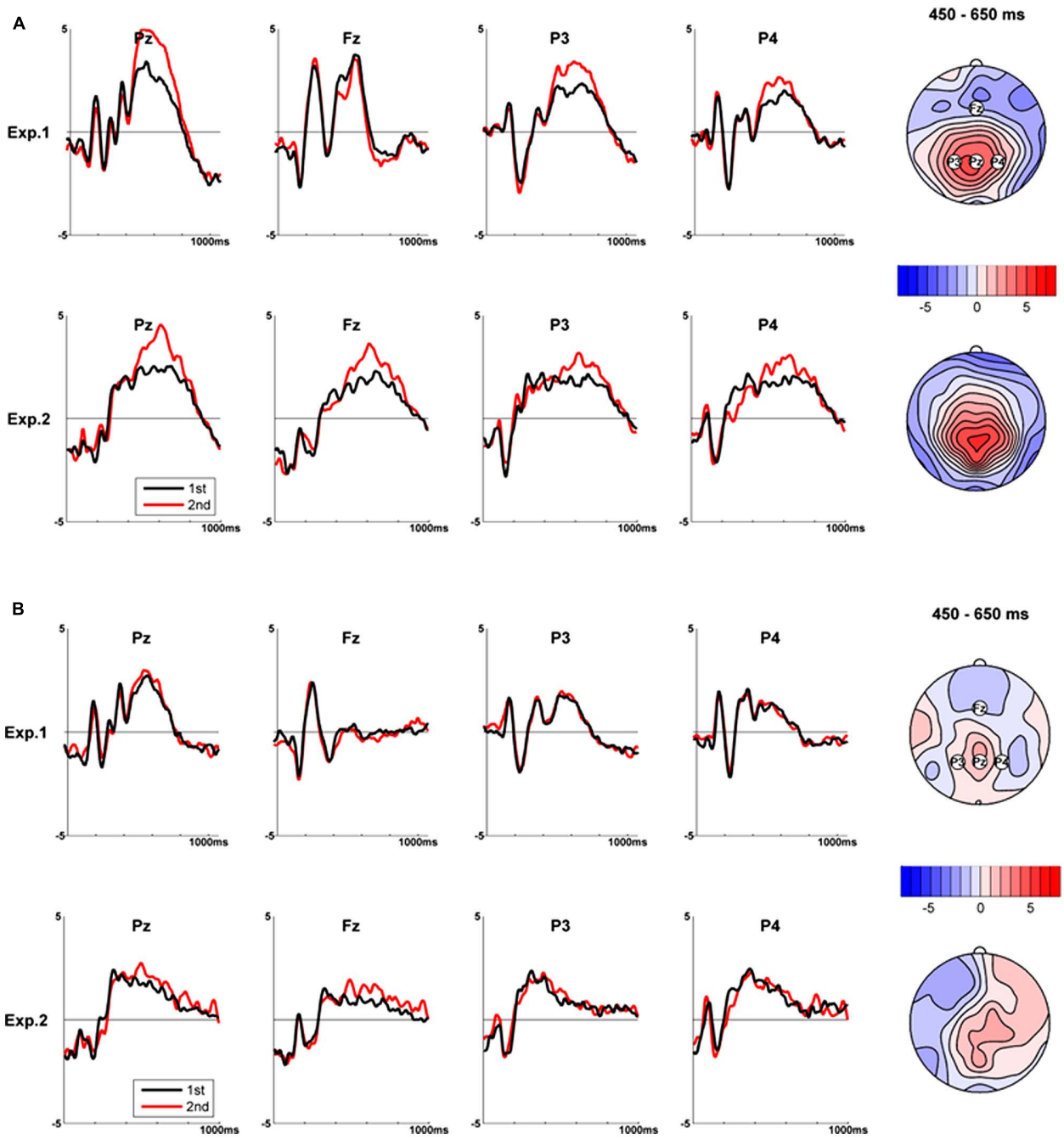

FIGURE 2 | Comparison of first versus second test halves for Experiment 1 and Experiment 2. (A) Transition effect. **(B)** Practice effect. Left: ERPs for prospective memory targets **(A)** and control items **(B)** from the first (black lines) and second half (red lines) at selected electrodes. Right: corresponding *t*-maps in the middle time window (450–650 ms).

DISCUSSION

The primary goal of Experiment 1 was to test whether the neural signature changes when a prospective memory task changes from episodic to habitual. In order to accomplish this goal we first tested whether we would find the three neural components which are typically associated with a prospective memory task (i.e., the N300, the P3b, the parietal old/new effect, and the sustained parietal positivity). As expected, these components were identified in an early, middle, and late time window, respectively and therefore the precondition for testing the transition from episodic to habitual prospective memory was met. Next, we compared these components in the first and the second half of the experiment. We found a difference in the middle time window only. That is, the parietal old/new effect was stronger in the second compared to the

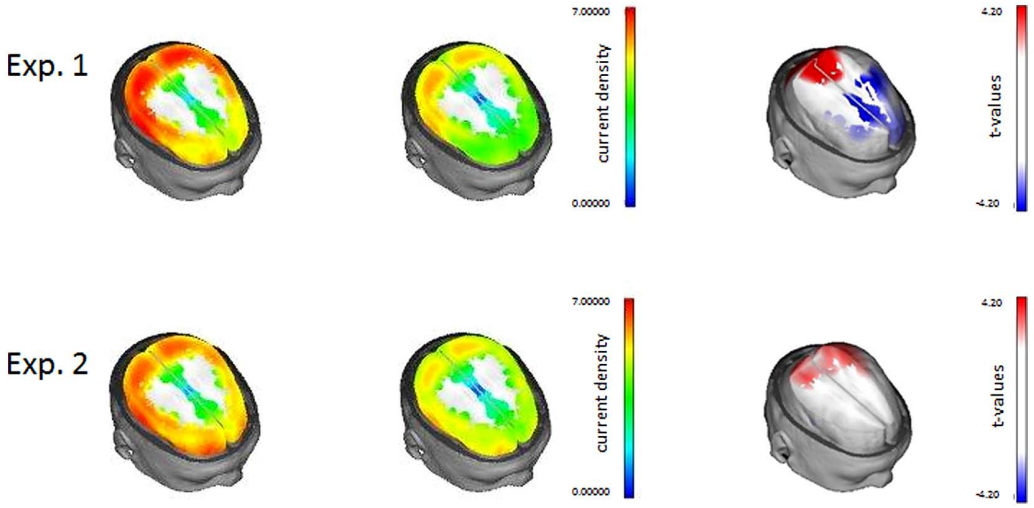

FIGURE 3 | Results from low resolution brain electromagnetic tomography analysis (LORETA) for Experiment 1 (top) and Experiment 2 (bottom). On the left side the 1st and 2nd half data are presented, the colorbar indicates current density. On the right side the transition effect is presented; the colorbar indicates *t*-values.

first half of the experiment as expressed by an increased activation at centro-parietal electrodes and a decrease of activation at frontal electrodes. Using LORETA, this transition effect was localized in parieto-occipital and frontal brain regions. Specifically, when the task changed from episodic to habitual there was an increase in brain activity in parieto-occipital areas and a decrease in brain activity in frontal areas. Covariance mapping further revealed that the differences between the ERP activation patterns in the first half and the second half of the experiment for prospective memory targets compared to control words was significant mainly in the middle and the late time windows, which are both associated with parietal activations. Finally, a plot of the fit across each single ERP trial for each individual with the covariance map across all participants revealed a linear relationship, indicating that the transition from episodic to habitual is rather continuous than categorical.

The results confirm that episodic and habitual prospective memory can be differentiated on a neural level. However, the distinction is rather quantitative than qualitative, as in both halves of the experiment the three components of prospective memory were found. As the only difference was found in the middle time window, in which the P3b and the parietal old/new effect occurred, the critical difference seems to be related either to a reallocation of processing capacity or to a facilitation of retrieval processes. The first interpretation is consistent with findings from dual-task studies of the oddball paradigm which showed that with more difficult primary tasks fewer resources are available for the secondary task which is expressed in decrease of P3b amplitude (e.g., Kramer et al., 1991; Watter et al., 2001). As the resource demands for the ongoing task decrease with practice a reallocation of processing capacity to the prospective memory task in the second half of the experiment is likely. However, the findings would also be consistent with results from research in recognition memory where, similarly, an increase of the parietal old/new effect occurs with high confidence in recollection (e.g., Curran, 2004). The result of a decrease in frontal activation might be related to the fact that the more habitual the prospective memory task becomes, fewer resources must be recruited for monitoring the prospective memory targets. This interpretation would be consistent with results from functional magnetic resonance imaging (fMRI) studies showing the involvement of frontal areas when monitoring for the prospective memory targets is required (e.g., Gilbert et al., 2006; Simons et al., 2006).

On a behavioral level, the accuracy of performing the prospective memory task was high from the beginning and close to ceiling. This was intended by design in order to include as many valid EEG-trials as possible in the ERP-analyses. The results further showed that performing the prospective memory task became faster in the second compared to the first half while performance of the lexical decision task remained constant in terms of both accuracy and response times. Together the behavioral results suggest that when the prospective memory task changed from episodic to habitual, performance became faster and this was not accompanied by a cost in ongoing task performance. This result indicates that, in fact, performing the prospective memory task required less attention and its execution became more automatic (cf., Einstein et al., 1998; Dismukes, 2008). Thus, the combination of the behavioral results and also the inspection of the *t*-maps are in line with the interpretation of resource allocation and the P3b as the source of the difference between the first and the second part of the experiment.

In Experiment 1 we used one specific prospective memory target which was presented repeatedly across the experiment. In contrast, the control word condition was composed of different four-letter words that were not repeated across the experiment. Therefore, it is possible that our results have been influenced by the fact that compared to control words the prospective memory targets became more familiar across the experiment. As a

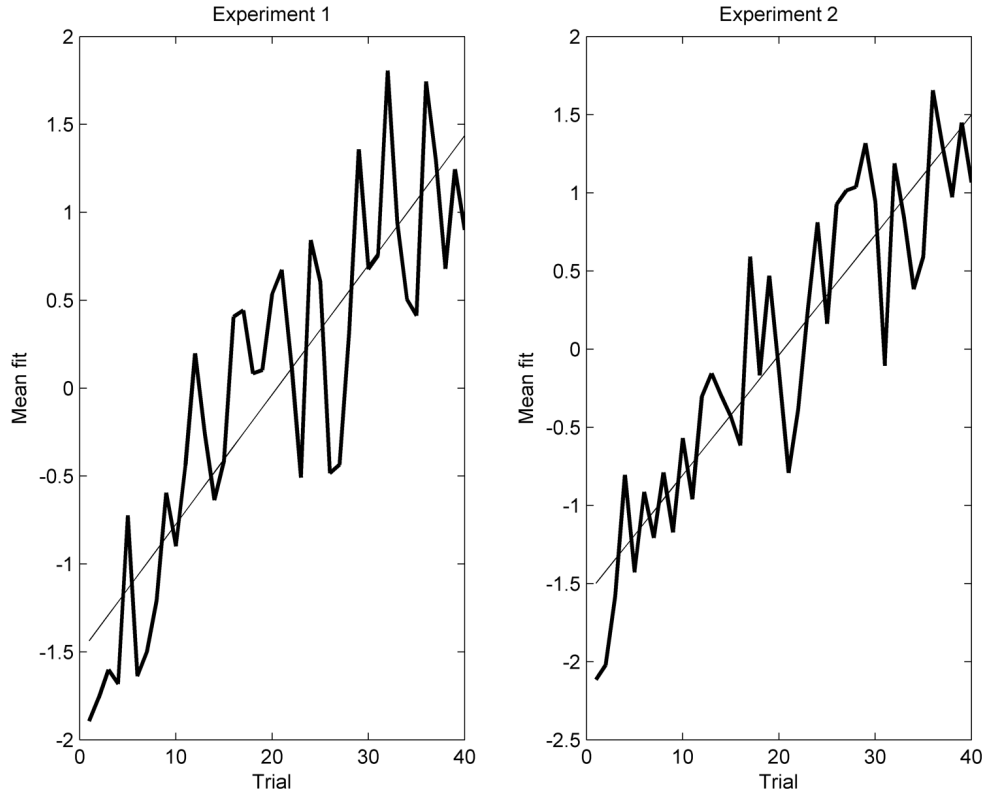

FIGURE 4 | Mean fit across all subjects of the single trial ERP (vertical axis) with the mean covariance map plotted against trial number (horizontal axis). The more positive the value, the more the single trial ERP data resembled the obtained covariance maps. Experiment 1(left), Experiment 2 (right). Both experiments showed a rather continuous increase of the fit that was well approximated with a linear regression line.

consequence an alternative explanation would be that the differential effects are simply related to prospective memory target familiarity. In order to exclude this alternative explanation we designed a second experiment. In Experiment 2 we used categorical rather than specific prospective memory targets. In addition, the control items were closely matched to the prospective memory targets. Moreover, in Experiment 2, we used a non-verbal perceptual discrimination task rather than a (verbal) lexical decision task. We reasoned that if the results from Experiment 1 are in fact specific to the transition between an episodic and a habitual prospective memory task, then the same pattern of results should be found in Experiment 2, independent of the particular ongoing task and independent of the particular kind of prospective memory targets.

EXPERIMENT 2

METHOD

Participants

A total of 20 new right-handed volunteers who had not participated in Experiment 1 were recruited. One participant had to be excluded from the analyses due to evidence of neurological compromise and two participants had to be excluded due to perspiration artifacts in the EEG-data. The data of the remaining 17 participants were used (mean age = 28 years, SD = 4; 10 female). All of them had normal or corrected-to-normal visual acuity and no evidence of neurological compromise. The study was approved by the Institutional Review Board and informed written consent was obtained.

Materials

For a perceptual discrimination task, a total of 1240 abstract shapes were used. From the materials of Slotnick and Schacter (2004) we selected 1200 shapes, subdivided into 30 differently colored sets each with 40 different shapes. A set of white shapes was used for the prospective memory task. In order to form a control condition, the set of white shapes used for the prospective memory task was dyed with a light yellow color. Accordingly, prospective memory items and control items consisted of exactly the same shapes and differed only by their color. For each color set five identical and five non-identical shape-pairs were created. The materials were equally divided into four experimental blocks. Each block consisted of 310 different shapes in 31 different colors. A particular shape appeared twice in each block, once in an identical shape-pair and once in a non-identical shape-pair.

For practice 36 trials with 18 identical and 18 non-identical shape-pairs with six different colors were used. These were different from the shapes of the experimental blocks.

As in Experiment 1 the first and the last (10th) prospective memory targets were positioned at the 30th and the 300th trial

of every experimental block. The remaining prospective memory targets were distributed between them at pseudo-randomized intervals of 20, 30, or 40 trials. To counterbalance the relative order of prospective memory and control trials, the first and the last (10th) control targets were presented as 15th and 285th trial in two of the blocks, and as 40th and 310th trial in the other two blocks, respectively, the remaining control targets were presented at pseudo-randomized intervals of 15, 20, 25, 30, 35, 40, or 45 trials. Block order was randomized for each participant.

Procedure

The procedure of Experiment 2 was similar to Experiment 1. For the perceptual discrimination task participants were informed that they will see a pair of shapes on the computer screen and that they have to decide whether the two shapes are identical or not by pressing the "B" or the "N"-key with the index finger and the middle finger of the right hand. For the prospective memory task, participants were instructed to press the "M"-key with the ring finger of the right hand whenever a white shape-pair was presented.

Each perceptual discrimination trial lasted 2000 ms. Stimuli were presented in pairs, side by side horizontally, against a black background, in the center of the screen. A gray fixation-cross was presented between the two shapes and the shape-pairs were surrounded by a gray rectangle. A colored shape-pair was presented for 1000 ms, then the shape-pair was removed and the fixation-cross and the rectangle stayed for another 1000 ms, resulting in a 2000 ms response window. The whole experiment lasted ∼50 min.

EEG recording and analysis

The EEG was digitized (500 Hz, 0.015 to 250 Hz bandpass) and stored from 64 electrodes located according to an extended version of the international 10–20 system using an Easycap EEG system. Inter-electrode impedances were kept below 5 kΩ. All electrodes were recorded against Fz. Eye-movements were monitored with two additional EOG channels.

First, the sampling rate was reduced to 200 Hz. Across subjects, between 1 and 3 ICA components (Delorme et al., 2007) were recognized as eye-movement related and offline removed from the data. Based on visual inspection further periods with artifacts were removed from further analysis. The data were filtered offline using a bandpass filter from 0.5 to 20 Hz and recomputed against average reference. Artifact-free EEG epochs were extracted from stimulus presentation to 1000 ms after stimulus presentation for correct responses.

In order to identify the prospective modulations the same analyses were conducted as described in Experiment 1. The early time window lasted from 250 to 450 ms after stimulus (representing the N300), the middle time window lasted from 450 to 650 ms after stimulus-onset and the late time window lasted from 650 to 850 ms (representing the P3b, the parietal old/new effect, and the sustained parietal positivity). Further, the same analyses as in Experiment 1 were conducted with the same time windows defined above to replicate the episodic to habitual prospective memory transition effect. For a better understanding, again covariance analyses and LORETA were used.

RESULTS

Behavioral data

Prospective memory task. Prospective memory performance was measured as proportion of correct responses to the white target shapes. Performance was 0.75 (SE = 0.04) for the first half and 0.93 (SE = 0.03) for the second half of the experiment, respectively. A paired-samples t test revealed a significant difference, $t(16) = -5.97$, $p < 0.001$. Mean reaction time for correct prospective memory targets was 848 ms (SE = 24) for the first half and 834 ms (SE = 21) for the second half of the experiment. A paired-samples t test revealed no significant difference, $t(16) = 0.92$, $p > 0.05$.

Ongoing task. Proportion of correct ongoing task responses was 0.85 (SE = 0.03) for the first half and 0.88 (SE = 0.03) for the second half of the experiment, respectively. A paired-samples t test revealed no significant difference, $t(16) = -1.38$, $p > 0.05$. Mean reaction time of correct ongoing task responses was 993 ms (SE = 37) for the first half and 956 ms (SE = 40) for the second half, respectively. A paired-samples t test revealed no significant difference, $t(16) = 2.11$, $p > 0.05$.

To test whether performing the ongoing perceptual discrimination task was affected by the additional requirement of the prospective memory task, the difference between performance in the ongoing task and the baseline trials was calculated. Mean reaction time difference was 22 ms (SE = 37) and −20 ms (SE = 37) for the first and the second half, respectively. Accuracy difference was 0.01 (SE = 0.01) between baseline and first half and 0.02 (SE = 0.02) between baseline and second half. The results of t tests showed no cost across the experiment (all ts ≤ 1.5; all ps > 0.05).

Electrophysiological data

Identification of the prospective memory modulation. The comparison of prospective memory target trials and control shapes using TANOVAs and t-maps revealed significances in the early time window (250–450 ms), $p < 0.001$ (largest t-value at electrode P8: $t = 6.6$), in the second time window (450–650 ms), $p < 0.001$ (largest t-value at electrode FC2: $t = 4.9$), and in the third time window (650–850 ms), $p < 0.005$ (largest t-value at electrode P3: $t = 5.8$). t-maps (**Figure 1**, bottom) confirmed that the prospective memory components had a similar topography as Experiment 1; these topographies corresponded largely to those interpreted as N300, P3b, parietal old/new effect, and the sustained parietal positivity.

Analysis of the episodic to habitual transition effect. For episodic prospective memory ERPs, the mean of artifact free valid trials per subject was 15.18 (range = 6–19), for individual habitual prospective memory ERPs, it was 18.35 (range = 10–20). As in Experiment 1, TANOVAs revealed a significant ERP difference of more episodic compared to more habitual trials in the middle time window, $p < 0.005$ (largest t-value at electrode Pz: $t = 6.0$), but not in the early or late time windows (ps > 0.05). The traces of the two conditions at Pz, as well as t-maps computed from the middle time window are shown in **Figure 2A** (bottom). This pattern is very similar to that of Experiment 1.

The results from LORETA source localisation are presented in **Figure 3** (bottom). They suggest that prospective memory ERPs of the second part compared to the first part of the experiment are associated with increased activity in the parietal lobes. This also replicates the findings of Experiment 1. Further, increased activity was also found in both occipital lobes. Brain regions with significantly higher current density in prospective memory ERPs of second compared to first half of the experiment were the precuneus, cuneus, occipital pol, superior occipital gyrus, superior parietal lobus, and inferior parietal lobus. In contrast to Experiment 1, however, ERPs of the second half of the experiment did not show a decrease in current density in the frontal lobes.

Analysis of the practice effect. For computing ERPs of control stimuli, the average number of trials was 15.53, (range = 8–20) for the first half, and 16.70 (range = 9–20) for the second half of the experiment. None of the TANOVAs revealed a significant difference for the practice effect (first time window: $p = 0.47$, second time window: $p = 0.54$, and third time window: $p = 0.33$). These results are presented in **Figure 2B** (bottom).

Covariance mapping. For the prospective memory trials, covariance maps were consistent across subjects in all three time windows, all ps < 0.001. As in Experiment 1, the covariance maps revealed positivity over frontal, central and parietal regions and negativity over frontal and temporal regions in the middle and late time window. For the control shapes, there was no evidence for consistent covariance maps across subjects in the early and late time windows ($p = 0.58$ and $p = 0.59$, respectively). However, there was a consistent topography in the middle time window ($p < 0.001$). Comparisons of covariance maps of the prospective memory trials and the control figures in the three time windows showed no significant differences in the early and late time windows (with $p = 0.26$ and $p = 0.62$, respectively), but they differed in the middle time window ($p = 0.047$).

As in Experiment 1, the estimated trajectory of the transition effect across trials indicated that the change of the covariance maps of prospective trials followed a linear gradient across the 40 trials (see **Figure 4**, on the right). Similar to Experiment 1, a regression analysis was calculated. The linear trajectory explained 81% of the variance ($r = 0.89$).

DISCUSSION

The primary goal of Experiment 2 was to test whether the neural signature changes observed in Experiment 1 can be generalized to a non-verbal ongoing task and with categorical prospective memory intention. As in Experiment 1, we first tested the presence of the three neural components which are typically associated with a prospective memory task (i.e., the N300, the P3b, the parietal old/new effect, and the sustained parietal positivity). These components were found and therefore the precondition for testing the transition from episodic to habitual prospective memory was met. When we tested these components for changes from the first to the second half of the experiment, we found a difference in the middle time window only as in Experiment 1. This was expressed by an increased activation at centro-parietal electrodes and a decrease of activation at frontal electrodes. Using LORETA, this transition effect was accompanied by an increase in brain activity in parieto-occipital areas. In Experiment 2, there was no activation difference in frontal regions when comparing the first and second half of the experiment. The differences in stimulus material and ongoing task demands may be responsible for this result.

However, as in Experiment 1, covariance mapping also revealed differences between the ERP activation patterns in the first half and the second half of the experiment in the middle time window. A plot of the fit indices across each single ERP trial for each individual with the covariance map across all participants revealed a similar linear relationship as in Experiment 1, further supporting the assumption that on a neural level, the transition from episodic to habitual follows a continuous linear function.

On a behavioral level, prospective memory performance was lower in the first part of the experiment and increased with routine in the second, more habitual part of the experiment. As a consequence fewer data-points were available for calculating ERPs. This may have been one source for the lack of finding a reduction in frontal activations compared to Experiment 1. No accuracy differences were found for the ongoing task. Moreover, no differences were found in the reaction times, neither for the prospective memory task nor for the ongoing task. As in Experiment 1 no performance costs were associated with adding a prospective memory task, indicating the automatic nature of habitual prospective memory. To summarize, Experiment 2 replicated the main results of Experiment 1, in particular the transition effect as based on differences in parietal activations in the middle time window (between 450 and 650 ms) and thus indicates that these are independent from the particular ongoing task and the particular kind of prospective memory targets. Moreover, together the present experiments also show the generality of the transition effect across different degrees of processing overlap between the ongoing task and the prospective memory task (Meier and Graf, 2000).

GENERAL DISCUSSION

This is the first study that addressed the transition from episodic to habitual prospective memory. In two separate experiments with different ongoing tasks and different kinds of intention specificity, we showed that with routine, the ERP-component in the time window between 450 and 650 ms post-stimulus became consistently larger. This result indicates that compared to episodic prospective memory, in habitual prospective memory resource allocation changes and intention retrieval is probably facilitated. The results confirm that episodic and habitual prospective memory can be differentiated on a neural level.

Moreover, the results indicate that the changes are rather quantitative than qualitative. This is reflected in the fact that the predicted ERP components that are typically involved in the realization of delayed intentions were present for earlier and later trials of both experiments. The N300 which is associated with prospective memory target detection was not changed when a task became more habitual. This indicates that target detection is a robust process that alerts the cognitive system that a significant event has occurred. The invariance of the N300 to habitualization is in line with result from West et al. (2003) who found a

similar N300 for prospective memory targets and for prospective memory lures that were perceptually distinct. In contrast, West et al. (2006) found that the amplitude for prospective memory targets that were embedded in a 1-back task was reduced compared to when they were embedded in a 2- or 3-back task, suggesting that the neural correlates of cue detection were sensitive to the availability of attentional resources. According to this latter finding it may be surprising that with the increasing availability of processing resources associated with a task becoming habitual the N300 remained constant. However, it seems that the amount of resource changes was not sufficient to affect the N300 as it was when West et al. compared 2- and 3-back trials.

The most important result of the present study is the ERP difference that was consistently found in the middle time window, in which the P3b and the parietal old/new effect occurred. Thus, as a task becomes habitual a reallocation of processing capacity, a facilitation of retrieval processes, or a combination thereof seems to occur. It is possible that the transition effect reflects memory retrieval in the earlier trials and a reallocation of processing capacity in the later trials once ongoing task processing is proceduralized and the representation of the prospective memory target has stabilized. This interpretation is consistent with findings from dual-task studies of the oddball paradigm which showed that with more difficult primary tasks fewer resources are available for the secondary task which is expressed in decrease of P3b amplitude (e.g., Kok, 2001; Watter et al., 2001). By the same logic, with increasing practice the ongoing task gets easier, thus freeing resources for the prospective memory task as expressed by an increase of P3b amplitude. It is also consistent with results from recognition memory in which high confidence in recollection is associated with an increased parietal old/new effect and with the results of a recent prospective memory study that showed enhanced prospective positivity for easier compared to more difficult prospective memory targets (Cona et al., 2013). Notably the present study was not designed to distinguish between these two possibilities. We were motivated by a more modest goal, namely to test whether we would find any differences in the neural signature of the transition between episodic and habitual prospective memory. Future studies are necessary to test the relative contribution of the P3b and the parietal old/new effect for this transition.

Finally, we also found a robust sustained parietal positivity which was not changed in the course of the experiment. Although the functional role of this late component is not clear yet it seems to be related to post-retrieval processes or as suggested more recently by processes related to task-set reconfiguration (Bisiacchi et al., 2009; West, 2011). This idea is consistent with the result that the sustained parietal positivity was larger for prospective memory targets than for prospective memory lures in a study by West et al. (2001). In the context of the present study it is reasonable to assume that these processes and task requirements remain stable.

On a behavioral level, the accuracy of performing the prospective memory task was high from the beginning and close to ceiling. This was intended by design in order to include as many valid trials for the ERP-analyses. In Experiment 1, the results further showed that performing the prospective memory task became faster in the second compared to the first half while performance of the lexical decision task remained constant. In Experiment 2, prospective memory performance increased while ongoing task performance remained constant. Together the behavioral results suggest that when the prospective memory task changed from episodic to habitual, performance became faster and this was not accompanied by a cost in ongoing task performance. This result indicates that, in fact, performing the prospective memory task required less attention and its execution became more automatic (cf., Einstein et al., 1998; Dismukes, 2008). In line with this interpretation, responding to prospective memory targets became statistically (Experiment 1) and numerically (Experiment 2) faster, however, due to the bivalent nature of prospective memory targets, performance did not reach the level of the ongoing task (cf., Meier and Rey-Mermet, 2012).

On a neuroanatomical level, further analyses involving source localisation revealed that the neural changes associated with the transition effect are related to an increase in activity in parietal brain areas and, at least in Experiment 1, it was also related to a decrease in frontal brain activity. Frontal cortex activation has generally been discussed as reflecting "retrieval effort" (Schacter et al., 1996) and as being involved in strategic and intentional retrieval of stored representations (Fletcher and Henson, 2001). For example, Simons et al. (2006) found higher activation in frontal regions, specifically in lateral BA 10 (and deactivation in medial BA 10) associated with cue identification and also intention retrieval. These effects were more pronounced when the prospective memory task required higher demands on the retrieval of the intention. Gilbert et al. (2006) found that reaction times to tasks that had provoked lateral BA 10 activations were slower than reaction times in their control tasks. Lateral BA 10 is presumably activated whenever additional attention resources are spent to an external stimulus and when resources are invested to handle this stimulus (Burgess et al., 2008). In the present study episodic to habitual prospective memory transition was accompanied by a frontal deactivation on the neuronal level and shorter reaction time on the behavioral level indicating that retrieval was more automatic and less attention resources had to be spent in more habitual compared to more episodic prospective memory task trials.

Parietal cortex activation is often involved in episodic memory retrieval (Cabeza and Nyberg, 2000; Wagner et al., 2005). According to the mnemonic-accumulator hypothesis (Wagner et al., 2005), memory strength is expressed by activity in the parietal cortex which is assumed to temporally integrate memory-strength signal. Thus, the higher parietal activation identified in habitual prospective memory with ERPs and LORETA is consistent with fMRI findings of higher confidence in the recollection of the intended action.

Overall, the pattern of decreased frontal activation and increased parieto-occipital activation that accompanies the transition from episodic to habitual prospective memory is compatible with the multi-process framework of prospective memory (cf., McDaniel and Einstein, 2000; Meier et al., 2006, 2011). According to this framework, one route toward successful prospective memory is via reflexive associative processes. It is assumed that retrieval cues interact with memory traces previously associated with the

cues and deliver the intention reflexively to awareness (McDaniel et al., 2004). After repeated performance of the prospective memory task memory traces may be stronger and the association between cue and intention more pronounced supporting reflexive associative processes.

ACKNOWLEDGMENT

This research was supported by a grant from the Swiss National Science Foundation (Grant Nr. 100013-109734) to Beat Meier.

REFERENCES

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX Lexical Database (Release 2) [CD-ROM]*. Philadelphia, PA: Linguistic Data Consortium.

Bisiacchi, P. S., Schiff, S., Ciccola, A., and Kliegel, M. (2009). The role of dual-task and task-switch in prospective memory: behavioural data and neural correlates. *Neuropsychologia* 47, 1362–1373. doi: 10.1016/j.neuropsychologia.2009.01.034

Burgess, P. W., Dumontheil, I., Gilbert, S. J., Okuda, J., Schölvinck, M. L., and Simons, J. S. (2008). "On the role of rostral prefrontal cortex (area 10) in prospective memory," in *Prospective Memory: Cognitive, Neuroscience, Developmental, and Applied Perspectives*, eds M. Kliegel, M. A. McDaniel, and G. O. Einstein (Mahwah, NJ: Erlbaum), 235–360.

Cabeza, R., and Nyberg, L. (2000). Imaging cognition II: an empirical review of 275 PET and fMRI studies. *J. Cogn. Neurosci.* 12, 1–47. doi: 10.1162/08989290051137585

Collins, D. L., Neelin, P., Peters, T. M., and Evans, A. C. (1994). Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J. Comput. Assist. Tomogr.* 18, 192–205. doi: 10.1097/00004728-199403000-00005

Cona, G., Bisiacchi, P. S., and Moscovitch, M. (2013). The effects of focal and nonfocal cues on the neural correlates of prospective memory: insights from ERPs. *Cereb. Cortex* doi: 10.1093/cercor/bht116 [Epub ahead of print].

Curran, T. (2004). Effects of attention and confidence on the hypothesized ERP correlates of recollection and familiarity. *Neuropsychologia* 42, 1088–1106. doi: 10.1016/j.neuropsychologia.2003.12.011

Cuttler, C., and Graf, P. (2009). Sub-clinical compulsive checkers show impaired performance on habitual, event- and time-cued episodic prospective memory tasks. *J. Anxiety Disord.* 23, 813–823. doi: 10.1016/j.janxdis.2009.03.006

Delorme, A., Sejnowski, T., and Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage* 34, 1443–1449. doi: 10.1016/j.neuroimage.2006.11.004

Dismukes, R. K. (2008). "Prospective memory in aviation and everyday settings," in *Prospective Memory: Cognitive, Neuroscience, Developmental, and Applied Perspectives*, eds M. Kliegel, M. A. McDaniel, and G. O. Einstein (New York: Erlbaum), 411–431.

Einstein, G. O., McDaniel, M. A., Smith, R. E., and Shaw, P. (1998). Habitual prospective memory and aging: remembering intentions and forgetting actions. *Psychol. Sci.* 9, 284–288. doi: 10.1111/1467-9280.00056

Elvevag, B., Maylor, E. A., and Gilbert, A. L. (2003). Habitual prospective memory in schizophrenia. *BMC Psychiatry* 3:9. doi: 10.1186/1471-244X-3-9

Fletcher, P. C., and Henson, R. N. (2001). Frontal lobes and human memory: insights from functional neuroimaging. *Brain* 124, 849–881. doi: 10.1093/brain/124.5.849

Gilbert, S. J., Spengler, S., Simons, J. S., Frith, C. D., and Burgess, P. W. (2006). Differential functions of lateral and medial rostral prefrontal cortex (area 10) revealed by brain-behavior associations. *Cereb. Cortex* 16, 1783–1789. doi: 10.1093/cercor/bhj113

Graf, P. (2005). "Prospective memory retrieval revisited," in *Dynamic Cognitive Processes*, eds N. Ohta, C. M. Mac Leod, and B. Uttl (Tokyo: Springer), 305–332.

Hager, W., and Hasselhorn, M. (1994). *Handbuch deutschsprachiger Wortnormen*. Göttingen: Hogrefe.

Kliegel, M., McDaniel, M. A., and Einstein, G. O. (eds). (2008). *Prospective Memory: Cognitive, Neuroscience, Developmental, and Applied Perspectives*. Mahwah, NJ: Erlbaum.

Koenig, T., and Melie-García, L. (2010). A method to determine the presence of averaged event-related fields using randomization tests. *Brain Topogr.* 23, 233–242. doi: 10.1007/s10548-010-0142-1

Koenig, T., Melie-García, L., Stein, M., Strik, W., and Lehmann, C. (2008). Establishing correlations of scalp field maps with other experimental variables using covariance analysis and resampling methods. *Clin. Neurophysiol.* 119, 1262–1270. doi: 10.1016/j.clinph.2007.12.023

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology* 38, 557–577. doi: 10.1017/S0048577201990559

Kramer, A. F., Strayer, D. L., and Buckley, J. (1991). Task versus component consistency in the development of automatic processing: a psychophysiological assessment. *Psychophysiology* 28, 425–437. doi: 10.1111/j.1469-8986.1991.tb00726.x

Kvavilashvili, L., and Ellis, J. (1996). "Varieties of intentions: some distinctions and classifications," in *Prospective Memory: Theory and Application*, eds M. Brandimonte, G. O. Einstein, and M. A. McDaniel (Mahwah, NJ: Erlbaum), 23–51.

Matter, S., and Meier, B. (2008). Prospective memory affects satisfaction with the contraceptive pill. *Contraception* 78, 120–124. doi: 10.1016/j.contraception.2008.04.007

McDaniel, M. A., and Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieval: a multiprocess framework. *Appl. Cogn. Psychol.* 14, 127–144. doi: 10.1002/acp.775

McDaniel, M. A., and Einstein, G. O. (2007). *Prospective Memory: An Overview and Synthesis of an Emerging Field*. Thousand Oaks, CA: Sage.

McDaniel, M. A., Guynn, M. J., Einstein, G. O., and Breneiser, J. (2004). Cue-focused and reflexive-associative processes in prospective memory retrieval. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 605–614. doi: 10.1037/0278-7393.30.3.605

Meacham, J. A., and Leiman, B. (1982). "Remembering to perform future actions," in *Memory Observed: Remembering in Natural Contexts*, ed. U. Neisser (San Francisco: W. H. Freeman), 327–342.

Meacham, J. A., and Singer, J. (1977). Incentive effects in prospective remembering. *J. Psychol.* 97, 191–197. doi: 10.1080/00223980.1977.9923962

Meier, B., and Graf, P. (2000). Transfer appropriate processing for prospective memory tests. *Appl. Cogn. Psychol.* 14, 11–27. doi: 10.1002/acp.768

Meier, B., and Rey-Mermet, A. (2012). Beyond monitoring: after-effects of responding to prospective memory targets. *Conscious. Cogn.* 21, 1644–1653. doi: 10.1016/j.concog.2012.09.003

Meier, B., von Wartburg, P., Matter, S., Rothen, N., and Reber, R. (2011). Performance predictions improve prospective memory and influence retrieval experience. *Can. J. Exp. Psychol.* 65, 12–18. doi: 10.1037/a0022784

Meier, B., Zimmermann, T. D., and Perrig, W. J. (2006). Retrieval experience in prospective memory: strategic monitoring and spontaneous retrieval. *Memory* 14, 872–889. doi: 10.1080/09658210600783774

Oostendorp, T. F., Delbeke, J., and Stegeman, D. F. (2000). The conductivity of the human skull: results of in vivo and in vitro measurements. *IEEE Trans. Biomed. Eng.* 47, 1487–1492. doi: 10.1109/TBME.2000.880100

Pascual-Marqui, R. D., Michel, C. M., and Lehmann, D. (1994). Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *Int. J. Psychophysiol.* 18, 49–65. doi: 10.1016/0167-8760(84)90014-X

Ravden, D., and Polich, J. (1998). Habituation of P300 from visual stimuli. *Int. J. Psychophysiol.* 30, 359–365. doi: 10.1016/S0167-8760(98)00039-7

Rugg, M. D., and Curran, T. (2007). Event-related potentials and recognition memory. *Trends Cogn. Sci.* 11, 251–257. doi: 10.1016/j.tics.2007.04.004

Schacter, D. L., Alpert, N. M., Savage, C. R., Rauch, S. L., and Albert, M. S. (1996). Conscious recollection and the human hippocampal formation: evidence from positron emission tomography. *Proc. Natl. Acad. Sci. U.S.A.* 93, 321–325. doi: 10.1073/pnas.93.1.321

Simons, J. S., Scholvinck, M. L., Gilbert, S. J., Frith, C. D., and Burgess, P. W. (2006). Differential components of prospective memory? Evidence from fMRI. *Neuropsychologia* 44, 1388–1397. doi: 10.1016/j.neuropsychologia.2006.01.005

Slotnick, S. D., and Schacter, D. L. (2004). A sensory signature that distinguishes true from false memories. *Nat. Neurosci.* 7, 664–672. doi: 10.1038/nn1252

Smith, R. E. (2003). The cost of remembering to remember in event-based prospective memory: investigating the capacity demands of delayed intention performance. *J. Exp. Psychol. Learn.* 29, 347–361. doi: 10.1037/0278-7393.29.3.347

Strayer, D. L., and Kramer, A. F. (1990). An analysis of memory-based theories of automaticity. *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 291–304. doi: 10.1037/0278-7393.16.2.291

Strik, W. K., Fallgatter, A. J., Brandeis, D., and Pascual-Marqui, R. D. (1998). Three-dimensional tomography of event-related potentials during response inhibition: evidence for phasic frontal lobe activation. *Electroencephalogr. Clin. Neurophysiol.* 108, 406–413. doi: 10.1016/S0168-5597(98)00021-5

Vedhara, K., Wadsworth, E., Norman, P., Searle, A., Mitchell, J., Macrae, N., et al. (2004). Habitual prospective memory in elderly patients with Typ 2 diabetes: implications for medication adherence. *Psychol. Health Med.* 9, 17–27. doi: 10.1080/1354850031000163724

Wagner, A. D., Shannon, B. J., Kahn, I., and Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends Cogn. Sci.* 9, 445–453. doi: 10.1016/j.tics.2005.07.001

Watter, S., Geffen, G. M., and Geffen, L. B. (2001). The n-back as a dual-task: P300 morphology under divided attention. *Psychophysiology* 38, 998–1003. doi: 10.1111/1469-8986.3860998

West, R. (2005). "The neural basis of age-related decline in prospective memory," in *Cognitive Neuroscience of Aging*, eds R. Cabeza, L. Nyberg, and D. Park (New York, NY: Oxford University Press), 246–264.

West, R. (2007). The influence of strategic monitoring on the neural correlates of prospective memory. *Mem. Cognit.* 35, 1034–1046. doi: 10.3758/BF03193476

West, R. (2008). "The cognitive neuroscience of prospective memory," in *Prospective Memory: Cognitive, Neuroscience, Developmental, and Applied Perspectives*, eds M. Kliegel, M. A. McDaniel, and G. O. Einstein (New York: Erlbaum), 261–282.

West, R. (2011). The temporal dynamics of prospective memory: a review of the ERP and prospective memory literature. *Neuropsychologia* 49, 2233–2245. doi: 10.1016/j.neuropsychologia.2010.12.028

West, R., Bowry, R., and Krompinger, J. (2006). The effects of working memory demands on the neural correlates of prospective memory. *Neuropsychologia* 44, 197–207. doi: 10.1016/j.neuropsychologia.2005.05.003

West, R., and Covell, E. (2001). Effects of aging on event-related neural activity related to prospective memory. *Neuroreport* 12, 2855–2858. doi: 10.1097/00001756-200109170-00020

West, R., Herndon, R. W., and Crewdson, S. J. (2001). Neural activity associated with the realization of a delayed intention. *Cogn. Brain Res.* 12, 1–9. doi: 10.1016/S0926-6410(01)00014-3

West, R., and Krompinger, J. (2005). Neural correlates of prospective and retrospective memory. *Neuropsychologia* 43, 418–433. doi: 10.1016/j.neuropsychologia.2004.06.012

West, R., McNerney, M. W., and Travers, S. (2007). Gone but not forgotten: the effects of cancelled intentions on the neural correlates of prospective memory. *Int. J. Psychophysiol.* 64, 215–225. doi: 10.1016/j.ijpsycho.2006.09.004

West, R., and Ross-Munroe, K. (2002). Neural correlates of the formation and realization of delayed intentions. *Cogn. Affect. Behav. Neurosci.* 2, 162–173. doi: 10.3758/CABN.2.2.162

West, R., Wymbs, N., Jakubek, K., and Herndon, R. W. (2003). Effects of intention load and background context on prospective remembering: an event-related brain potential study. *Psychophysiology* 40, 260–276. doi: 10.1111/1469-8986.00028

Wirth, M., Horn, H., Koenig, T., Stein, M., Federspiel, A., Meier, B., et al. (2007). Sex differences in semantic processing: event-related brain potentials distinguish between lower and higher order semantic analysis during word reading. *Cereb. Cortex* 17, 1987–1997. doi: 10.1093/cercor/bhl121

Zhang, Y. C., van Drongelen, W., and He, B. (2006). Estimation of in vivo brain-to-skull conductivity ratio in humans. *Appl. Phys. Lett.* 89, 223903–2239033. doi: 10.1063/1.2398883

Zimmermann, T. D., and Meier, B. (2006). The rise and decline of prospective memory performance across the lifespan. *Q. J. Exp. Psychol.* 59, 2040–2046. doi: 10.1080/17470210600917835

Zimmermann, T. D., and Meier, B. (2010). The effect of implementation intentions on prospective memory performance across the lifespan. *Appl. Cogn. Psychol.* 24, 645–658. doi: 10.1002/acp.1576

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 March 2014; accepted: 17 June 2014; published online: 02 July 2014.
Citation: Meier B, Matter S, Baumann B, Walter S and Koenig T (2014) From episodic to habitual prospective memory: ERP-evidence for a linear transition. Front. Hum. Neurosci. ***8****:489. doi: 10.3389/fnhum.2014.00489*
This article was submitted to the journal Frontiers in Human Neuroscience.