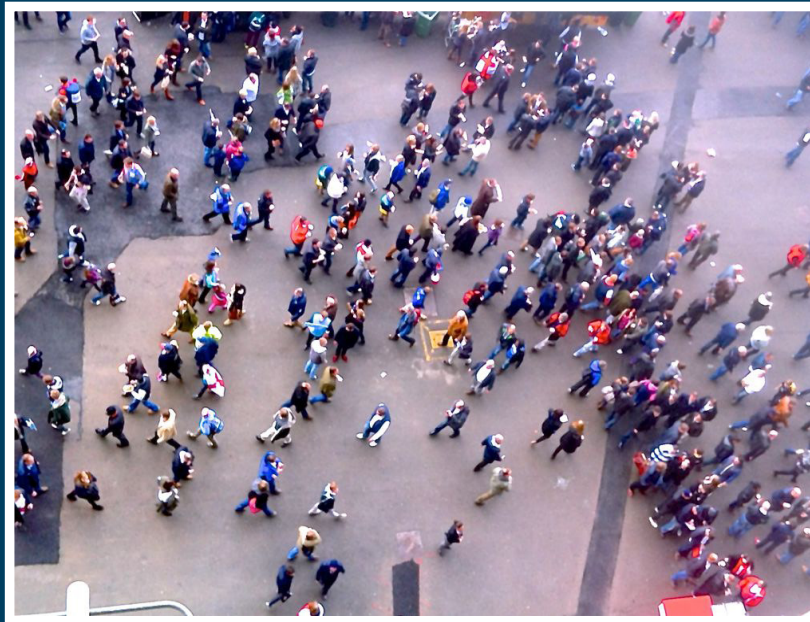


frontiers

RESEARCH TOPICS



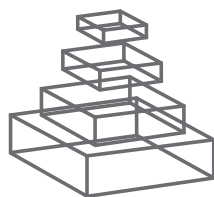
INDIVIDUAL DIFFERENCES IN ASSOCIATIVE LEARNING

Topic Editors

Robin A. Murphy and Rachel M. Msetfi



frontiers in
PSYCHOLOGY



frontiers

FRONTIERS COPYRIGHT STATEMENT

© Copyright 2007-2014
Frontiers Media SA.
All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88919-290-8

DOI 10.3389/978-2-88919-290-8

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!

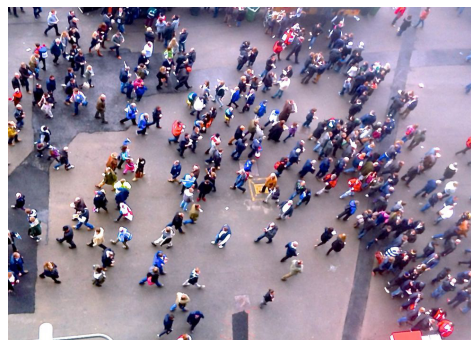
Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

INDIVIDUAL DIFFERENCES IN ASSOCIATIVE LEARNING

Topic Editors:

Robin A. Murphy, University of Oxford, United Kingdom

Rachel M. Msetfi, University of Limerick, Ireland



Picture by Robin Murphy

Theories of associative learning have a long history in advancing the psychological account of behavior via cognitive representation. There are many components and variations of associative theory but at the core is the idea that links or connections between stimuli or responses describe important aspects of our psychological experience. This Frontiers Topic considers how variations in association formation can be used to account for differences between people, elaborating the differences between males and females, differences over the life

span, understanding of psychopathologies or even across cultural contexts. A recent volume on the application of learning theory to clinical psychology is one example of this emerging application (e.g., Hazelgrove & Hogarth, 2012).

The task for students of learning has been the development, often with mathematically defined explanations, of the parameters and operators that determine the formation and strengths of associations. The ultimate goal is to explain how the acquired representations influence future behavior.

This approach has recently been influential in the field of neuroscience where one such learning operator, the error correction principle, has unified the understanding of the conditions which facilitate neuron activation with the computational goals of the brain with properties of learning algorithms (e.g., Rescorla & Wagner, 1972).

In this Frontiers Research Topic, we are interested in a similar but currently developing aspect to learning theory, which is the application of the associative model to our understanding of individual differences, including psychopathology. In general, learning theories are monolithic, the same theory applies to the rat and the human, and within people the same algorithm is applied to all individuals. If so this might be thought to suggest that there is little that learning theory can tell us about the how males and females differ, how we

change over time or why someone develops schizophrenia for instance. However, these theories have wide scope for developing our understanding of when learning occurs and when it is interfered with, along with a variety of methods of predicting these differences. We received contributions from researchers studying individual differences, including sex differences, age related changes and those using analog or clinical samples of personality and psychopathological disorders where the outcomes of the research bear directly on theories of associative learning.

This Research Topic brings together researchers studying basic learning and conditioning processes but in which the basic emotional, attentional, pathological or more general physiological differences between groups of people are modeled using associative theory. This work involves varying stimulus properties and temporal relations or modeling the differences between groups.

Table of Contents

- 05 Individual Differences in Associative Learning**
Robin A. Murphy and Rachel M. Msetfi
- 07 The Causes of Variation in Learning and Behavior: Why Individual Differences Matter**
Bruno Sauce and Louis D. Matzel
- 17 Revisiting the Learning Curve (Once Again)**
Steven Glautier
- 34 Accounting for Individual Differences in Human Associative Learning**
Nicola C. Byrom
- 42 Repeated Activation of a CS-US-Contingency Memory Results in Sustained Conditioned Responding**
Els Joos, Debora Vansteenwegen, Bram Vervliet and Dirk Hermans
- 57 Diminished Acquired Equivalence Yet Good Discrimination Performance in Older Participants**
Jasper Robinson and Emma Owens
- 65 Individual Differences in Chemotherapy-Induced Anticipatory Nausea**
Marcial Rodríguez
- 71 Pathological Gamblers are More Vulnerable to the Illusion of Control In a Standard Associative Learning Task**
Cristina Orgaz, Ana Estevez and Helena Matute
- 78 Do Personality Traits Predict Individual Differences in Excitatory and Inhibitory Learning?**
Zhimin He, Helen J. Cassaday, Charlotte Bonardi and Peter A. Bibby
- 90 Cocaine Dependent Individuals and Gamblers Present Different Associative Learning Anomalies in Feedback-Driven Decision Making: A Behavioral and ERP Study**
Ana Torres, Andrés Catena, Antonio Cándido, Antonio Maldonado, Alberto Megías and José C. Perales
- 104 Individual Differences in Discriminatory Fear Learning Under Conditions of Ambiguity: A Vulnerability Factor for Anxiety Disorders?**
Inna Arnaudova, Angelos-Miltiadis Kryptos, Marieke Effting, Yannick Boddez, Merel Kindt and Tom Beckers



Individual differences in associative learning

Robin A. Murphy^{1*} and Rachel M. Msetfi²

¹ Experimental Psychology, Behavioural Neuroscience, University of Oxford, Oxford, UK

² Department of Psychology, University of Limerick, Limerick, Ireland

*Correspondence: robin.murphy@psy.ox.ac.uk

Edited and reviewed by:

Marcel Zentner, University of Innsbruck, Austria

Keywords: learning, associationism, computational psychopathology, conditioning (psychology), individual differences

This ebook represents the scientific contribution of over 30 individuals working in laboratories in 5 countries (Belgium, Netherlands, Spain, the United Kingdom and the United States), unified by the study of associative learning and individual differences. Individual differences have a central place in the study of psychology both historically and in the present day. It was frustrating then that when we conceived the idea of developing this Frontiers Topic, we knew personally that there was considerable work being conducted in various labs that could be characterized as the “associative learning of individual differences” but we were also confounded by the observation that there was no natural scholarly home for this work to be published. By its very nature this type of interdisciplinary work is at the frontiers and the borders between disciplines. The Frontiers publishing model offered us an excellent opportunity to bring this work together. By doing so we now have a current impression of the work and ideas being grappled with in 2014. As a representative sample of this work, we think this volume makes a bold statement and would function as a useful primer for the subject.

The study of individual differences was part of the nucleus of activity for the original pioneers who settled psychology and the malthusian growth and divergence that followed during psychology’s development. The polymath Sir Francis Galton (1822–1911), for instance, a true psychological inventor, developed early psychometrics as well as the use of quantitative and statistical methods (such as the correlation coefficient and the normal distribution) for evaluating and understanding individual variation. He was well placed to make use of differences as a key psychological concept. He was a relative of Darwin and as such had a sympathetic perspective on the role that variation might play in psychological development. Despite his involvement in a range of related activities, as well as ideas that today are somewhat anti-ethical to scientific objectivity (e.g., ideas on eugenics for instance), his impact on current psychological thinking cannot be overestimated. He stood to understand psychological and individual personality variables and their relation to behavior and cognition.

This ebook owes much to Galton’s quantitative approach, both in relation to the focus on differences but also on the idea that mathematical and computational principles might provide a scientific tool for understanding these differences. For Galton his quantitative tools were used for sorting or separating the perceived groups of individuals and understanding how they might relate, for instance with the use of concept of regression toward

the mean. For associative learning, and its cousin connectionist modeling, the computational principles that have developed (for instance the Least Means Square method) provide parameters and principles for understanding the psychological differences. Models that have been developed to account for associative learning (e.g., Rescorla and Wagner, 1972; Sutton and Barto, 1981) provide explanations and descriptions for the development of acquired associations and as such these models make concrete the parameters upon which individuals might vary. Variation conceived here might be at the level of behavior as described in the papers of this volume, but the models are also used to identify neural circuits and substrates which must be the bases for the differences described.

The first three papers in this volume outline approaches to the computational analysis of individuals and provides a powerful case for how and why individual differences should be studied (Sauce and Matzel, 2013), and how the acquisition of a response or association might vary within an individual across a particular training experience (Glautier, 2013) or between individuals across the same experience (Byrom, 2013). These three contributions describe some of the elements underlying the complexity in thinking and quantitative modeling of the computational approach.

The remaining contributions highlight the variation in approaches and topics of inquiry that are amenable to this analysis. In the first, we read how differences in association and learning might be supported by memory processes and linked with differences in rumination (Joos et al., 2013) and then how differences in learning are expressed differentially over the life course (Robinson and Owens, 2013). The next set of papers describe work from the field of Computational Psychopathology asking diverse questions but with a related analytic framework. We consider the field of computational psychopathology as an important and viable method for understanding a range of human pathology. These topics are quite diverse, the study of human anxiety (Arnaudova et al., 2013) compulsive gambling (Orgaz et al., 2013), drug dependency (Torres et al., 2013) and neuroticism and personality (He et al., 2013) as well as attempts to address the physical and psychological effects and origins of nausea induced by chemotherapy treatment (Rodríguez, 2013). In all, this provides only a sample of the possible areas of human experience to which this analysis might be brought to bear and provides the reader (we hope) with a primer to excite the possibilities.

ACKNOWLEDGMENT

This work was supported by the Economic and Social Research Council (UK).

REFERENCES

- Arnaudova, I., Kryptos, A.-M., Eftting, M., Boddez, Y., Kindt, M., and Beckers, T. (2013). Individual differences in discriminatory fear learning under conditions of ambiguity: a vulnerability factor for anxiety disorders? *Front. Psychol.* 4:298. doi: 10.3389/fpsyg.2013.00298
- Byrom, N. C. (2013). Accounting for individual differences in human associative learning. *Front. Psychol.* 4:588. doi: 10.3389/fpsyg.2013.00588
- Glautier, S. (2013). Revisiting the learning curve (once again). *Front. Psychol.* 4:982. doi: 10.3389/fpsyg.2013.00982
- He, Z., Cassaday, H. J., Bonardi, C., and Bibby, P. A. (2013). Do personality traits predict individual differences in excitatory and inhibitory learning? *Front. Psychol.* 4:245. doi: 10.3389/fpsyg.2013.00245
- Joos, E., Vansteenwegen, D., Vervliet, B., and Hermans, D. (2013). Repeated activation of a CS-US-contingency memory results in sustained conditioned responding. *Front. Psychol.* 4:305. doi: 10.3389/fpsyg.2013.00305
- Orgaz, C., Estévez, A., and Matute, H. (2013). Pathological gamblers are more vulnerable to the illusion of control in a standard associative learning task. *Front. Psychol.* 4:306. doi: 10.3389/fpsyg.2013.00306
- Rescorla, R. A., and Wagner, A. R. (1972). "A theory of pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement," in *Classical Conditioning II: Current Research and Theory*, eds A. H. Black and W. F. Prokasy (New York, NY: Appleton-Century-Crofts), 64–99.
- Robinson, J., and Owens, E. (2013). Diminished acquired equivalence yet good discrimination performance in older participants. *Front. Psychol.* 4:726. doi: 10.3389/fpsyg.2013.00726
- Rodríguez, M. (2013). Individual differences in chemotherapy-induced anticipatory nausea. *Front. Psychol.* 4:502. doi: 10.3389/fpsyg.2013.00502
- Sauce, B., and Matzel, L. D. (2013). The causes of variation in learning and behavior: why individual differences matter. *Front. Psychol.* 4:395. doi: 10.3389/fpsyg.2013.00395
- Sutton, R., and Barto, A. (1981). An adaptive network that constructs and uses an internal model of its world. *Cogn. Brain Theory* 4, 217–246.
- Torres, A., Catena, A., Cándido, A., Maldonado, A., Megías, A., and Perales, J. C. (2013). Cocaine dependent individuals and gamblers present different associative learning anomalies in feedback-driven decision making: a behavioral and ERP study. *Front. Psychol.* 4:122. doi: 10.3389/fpsyg.2013.00122

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: 08 April 2014; accepted: 30 April 2014; published online: 20 May 2014.

Citation: Murphy RA and Msetfi RM (2014) Individual differences in associative learning. *Front. Psychol.* 5:466. doi: 10.3389/fpsyg.2014.00466

This article was submitted to *Personality and Social Psychology*, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Murphy and Msetfi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The causes of variation in learning and behavior: why individual differences matter

Bruno Sauce and Louis D. Matzel*

Department of Psychology, Rutgers University, Piscataway, NJ, USA

Edited by:

Robin A. Murphy, University of Oxford, United Kingdom

Reviewed by:

Robin A. Murphy, University of Oxford, United Kingdom
Nicola C. Byrom, University of Oxford, United Kingdom

***Correspondence:**

Louis D. Matzel, Department of Psychology, Rutgers University, 152 Frelinghuysen Road, Piscataway, NJ 08854, USA
e-mail: matzel@rci.rutgers.edu

In a seminal paper written five decades ago, Cronbach discussed the two highly distinct approaches to scientific psychology: experimental and correlational. Today, although these two approaches are fruitfully implemented and embraced across some fields of psychology, this synergy is largely absent from other areas, such as in the study of learning and behavior. Both Tolman and Hull, in a rare case of agreement, stated that the correlational approach held little promise for the understanding of behavior. Interestingly, this dismissal of the study of individual differences was absent in the biologically oriented branches of behavior analysis, namely, behavioral genetics and ethology. Here we propose that the distinction between “causation” and “causes of variation” (with its origins in the field of genetics) reveals the potential value of the correlational approach in understanding the full complexity of learning and behavior. Although the experimental approach can illuminate the causal variables that modulate learning, the analysis of individual differences can elucidate *how much* and *in which way* variables interact to support *variations* in learning in complex natural environments. For example, understanding that a past experience with a stimulus influences its “associability” provides little insight into how individual predispositions interact to modulate this influence on associability. In this “new” light, we discuss examples from studies of individual differences in animals’ performance in the Morris water maze and from our own work on individual differences in general intelligence in mice. These studies illustrate that, opposed to what Underwood famously suggested, studies of individual differences can do much more to psychology than merely providing preliminary indications of cause-effect relationships.

Keywords: correlational studies, learning, behaviorism, causes of variation, spatial learning, associative learning, general intelligence

In a widely influential paper, Cronbach (1957) discussed the two highly distinct approaches to scientific psychology: experimental and correlational. According to Cronbach, the experimental approach attempts to understand reality by manipulating (under simplified conditions) variables between groups/treatments. In contrast, the correlational approach attempts to understand reality by estimating the influence of variables under complex conditions between individuals. Individual differences, critical for correlational analyses, are troublesome noise for the experimental psychologist, while differences between treatments, critical to the experimental approach, are avoided among correlational psychologists. Hence, although both approaches are complementary and, as Cronbach argued, equally important to psychology, they are typically employed separately; mitigating their true explanatory potential.

As the discipline of psychology gravitated toward a more scientific framework, so too did its reliance on experimental methodologies (Ebbinghaus, 1913; Osgood, 1953). Already ubiquitous in the older sciences of physics, chemistry, and biology, the design and philosophy of controlled experiments also became part of psychologists’ mindset. Due to this wide adoption (and the advances it has prompted), the experimental approach needs little theoretical defense. If its use is still scarce in some fields

of psychology, we believe it is not due to rejection, but rather, because experimental control is often difficult to implement when studying some complex variables. However, even in very complex fields such as social psychology, the explosion of research in behavioral economics (Kahneman, 2003) illustrates the vast application (and popularity) of the experimental approach. The correlational approach, on the other hand, is not so widely embraced, and, as we will argue, needs to be better understood and more commonly implemented.

Correlational psychology has been productive for decades in fields like personality psychology, social psychology, psychometrics, clinical psychology, and developmental psychology. Although these fields focus on very distinct topics, they all try to understand what makes individuals vary according to their personality, cultural background, cognitive abilities, extreme disorders, and age. [It is worth noting that “aging” is *never* induced (i.e., experimentally manipulated). Although a comparison of two ages under controlled laboratory conditions is often described as an “experiment,” in fact, the comparison of the performance of two groups of different ages is a very narrow correlational analysis.] Acclaimed ideas like the self-determination theory (Ryan and Deci, 2000), general intelligence (Jensen, 1998), and Piaget’s theory of cognitive development (Piaget and Inhelder, 1973) are all children of the

correlational approach, and their broad impact and explanatory value is undeniable.

Even with its relative success within psychology, the correlational approach appears to provoke a disproportional distrust among psychologists. Remarkably, among those studying learning and behavior (our focus here), the correlational approach was never fully appreciated. (Even in the sub-disciplines where this approach is widely employed, it is still often admonished as being “only correlational.”) Due at least in part to the influence of behaviorism, the dominance of the experimental approach in studies of learning overwhelmed the contributions of studies of individual differences (an observation that is also true of traditional fields within behavioral neuroscience). Almost none of the principles that guide contemporary theory on learning were derived from correlational analyses. Since its origins a century ago, the behaviorists’ obsession with experimental analyses was probably a reaction to an unscientific and speculative psychology that dominated the early discipline. John Watson, the father of Behaviorism, announced in a highly influential writing (Watson, 1913) that “psychology as the behaviorist views it is a purely objective experimental branch of natural science,” and claimed that tightly controlled conditions were the answer to elucidating the basis of any behavior, from understanding his “Tortuga’s birds” to understanding the “educated European.” Later, both Tolman (1924) and Hull (1951), in a rare case of agreement, stated that correlational methods held little promise for the understanding of behavior. Tolman assumed that “individual difference variables [were] average standard values,” and that “rat-workers have always done this, perhaps unconsciously.” According to Tolman “we have tried to keep heredity normal by using large groups, age normal by using rats between 90 and 120 days old, previous training normal by using fresh rats in each new experiment, and endocrine and nutritional conditions normal by avoiding special dosages and also again by using large groups.” Tolman was, in sum, distrustful of the correlational approach, and stated that factor analyses (which epitomize the correlational method) “do not seem to suggest any simple or agreed-upon results [and, for instance, in the case of intelligence research], the controversy rages from Spearman’s one or two factors through Kelley’s and Thurstone’s three to nine factors.” Even during the revolution in learning theory in the 1960s, all critical empirical data was derived exclusively from the experimental approach (for a review of this era of rapid change, see Rescorla, 1988).

Aside from the historical reaction to non-scientific psychology, we might wonder what else led the study of behavior to become so ingrained in experiments and resistant to individual variations. The reasons for this might be the biases of psychology in relation to the role of animals. Seen sometimes as lesser organisms at a lower stage of an imaginary human scale, individual differences in animals were probably considered too simplistic in their causes to be informative (a concern that has been reinforced by the increasingly wide adoption of genetically homogeneous, inbred animals). In addition, it was easy to assume that experimental studies might elucidate an invariant framework of learning processes, mitigating any interest in individual differences. Regardless of the genesis of this bias, we believe that the correlational approach *can* provide an understanding of learning and behavior that is not attainable

through experimental studies alone, as it has done so successfully in other disciplines.

Interestingly, the waning interest in individual differences and the dismissal of the correlational approach did not occur in biological branches of behavior analysis, such as behavioral genetics and ethology. Why might these closely related fields in psychology and biology have evolved so differently? To be fair, we scientists frequently have good intuitions on how to apply the scientific method in our fields of study. However, maybe just as frequently, we fail to appreciate the full utility of those methods, and their broader implications toward our understanding of reality. For this reason, it is dangerous to simply rely on precedence and intuition to inform our methodologies. In this article, we first argue that philosophical concerns in the fields of evolution and genetics demonstrate why individual differences were so powerful in biological branches of behavior, and why we can (and should!) incorporate the same lessons in the psychological branches of behavior. It is from the distinction between “causation” and “causes of variation” (with its origins in quantitative genetics) that springs the potentially huge contribution of the correlational approach. We then use results from animal learning to illustrate how studying causes of variation can answer unique questions about the complex role that multiple psychological factors play in the expression of learning.

THREE LESSONS FROM BIOLOGY: CAUSES OF VARIATION AS THE CLUES FOR UNDERSTANDING “HOW MUCH” AND “IN WHICH WAY” PHENOTYPES EMERGE

The relevance of individual differences to scientific inquiry only became obvious after the work of Charles Darwin on evolution by natural selection (Gould, 2002). Before this, the study of life followed the same platonic idealism common in physics and chemistry. Any molecule of water, anywhere, has the same properties of an ideal molecule of water. Hence, the same was considered to be true for life. Any individual should contain more or less the same characteristics as the stereotypical (ideal) individual of that species (Bernier, 1984). And this reasoning also applied for organs, tissues, and cells. Variations, therefore, were considered imperfections around the ideal (or “true”) substrates of a system. In their early traditions, the fields of physiology and biochemistry were trying to identify the pieces that constituted a perfect engine. For digestion to occur, an organism needs the mechanisms of peristaltic movements and chloride acid in a specific order, and with a specific duration, for any given type of food. In contrast, Darwin did not focus on the ideal (or average) of a piece, but on the variation between those pieces (Gould, 2002). He was looking to what made individuals differ, and why we encounter various degrees of differences in nature. In so doing, Darwin was able to understand/discover/deduce one of the major forces of evolution (Dennett, 1995), and in fact, to grasp the adaptation of species. This was no small accomplishment, and, throughout decades of work, Darwin’s approach was primarily correlational (although he did conduct an occasional experiment, most remarkably with birds as subjects).

Almost concomitant with Darwin’s work, the focus on individual differences was critical to the discoveries of Mendel on the laws of inheritance. Although Mendel’s work was also experimental due to the manipulation of his pea plants’ attributes, this was

an “indirect” manipulation mainly intended to make the determinants of heredity (genes, chromosomes, meiotic division) simpler to observe (Griffiths et al., 2007). He did not directly manipulate the process of heredity itself, and thus could not deduce a specific cause regarding *why* a purple plant generates a purple daughter (something that can be accomplished today using transgenic techniques). What Mendel *did* deduce was a cause for the *differences* in peas (what we now know as “particular segregation”) by looking at the relevant individual differences (i.e., the ratios from breeding). Obviously, this discovery was far more important than any that could have emerged from the isolated results of experimental manipulations. This reasoning of discovering the causes underlying a system by looking at phenotypic differences gave origin to the classical approach in genetics, that later became known as “forward genetics” (Nagy et al., 2003).

The pioneering work of Darwin and Mendel reveals a very important lesson. Previously, biology mainly followed the pattern from causes (test conditions) to effects, with the attendant worries about ruling out false positives and false negatives by use of repetition and control groups. By focusing on individual differences, Darwin and Mendel made popular the opposite pattern: going from effects (the clues found in individual differences) to their causes. Now for this approach, concomitant worries arise about ruling out alternative explanations (for an in depth discussion about a similar division in scientific methods, see Cleland, 2002). The reasoning of going from causes to effects is what defines most of the experimental approach, and the reasoning behind going from effects to causes is what defines most of the correlational approach. In other words, the experimenter has to be a master puppeteer; creatively applying different treatments and proper control groups (i.e., pulling the right strings). The correlator, in turn, has to be an expert detective; creatively considering the relevant observations and variables (i.e., finding the right clues) from already existing differences in individuals (By analogy, we would rarely criticize a police investigator’s work as “only correlational.”) Absent proper control and adequate consideration, neither approach is capable of unequivocal conclusions, nor should either approach be condemned for this. This lesson, that important and historically verified conclusions have emerged from correlational research, is called here Lesson #1.

Now let us step back to think about what we can learn from analyzing the causes that underlie the emergence of individual differences. In a simple example, consider the process of combustion. We know that for combustion to occur, we need the causal factors of an oxidant (e.g., oxygen), a fuel (e.g., wood), and an external source of ignition (e.g., the strike of a match). However, oxygen (as well as fuel) is usually present in most practical situations. Thus an investigator searching for the cause of a fire in a building will most likely look for the source of external ignition (like a short-circuit of cables, an overheating of a machine, or a carelessly disposed match). On the other hand, since many other non-necessary factors could increase or reduce the intensity of a fire, a city administrator looking to reduce the incidence of fires could start reducing the most common “causes” (risk factors) for differences in fires in the past, e.g., storage of paper documents, overloaded electrical systems, and portable heat sources. Likewise, city administrators could promote those non-necessary

factors known to mitigate the damage (i.e., variation) associated with fires, e.g., fire alarms and fire sprinklers. So, although fire is caused minimally/necessarily by three factors (that have the same importance for a single fire), they can have different importance and interact with many other non-necessary factors to create different incidences of fires across buildings and cities (as well as many different responses to fire or the threat of fire). The same reasoning applies to the expression of a phenotype of a living organism. Phenotypes emerge from the interaction of genetic and environmental necessary factors. All of them are the true (and complete) causation of an individual’s phenotype, and it is meaningless to try to separate genotype and environment as distinct necessary causes for the individual’s phenotype (as oxygen, wood, and strike of a match are also inseparable as causes for lighting a fire). *All* of the causes are of the same critical importance! On the other hand, in a population, we can look for the distinct importance of (both necessary and non-necessary) causes of phenotypic variation among individuals rather than the causation of any single individual’s phenotype (Templeton, 2006). This is analogous to finding that, among all fires in a humid city, portable heating sources “caused” more fire than the storage of paper (due to the dryness created inside a room). In other words, “causation” is the inseparable causes of an idealized system, while “causes of variation” are the separated causes for *differences* in a system.

The distinction between causation and causes of variation in biology was insightfully discussed by Templeton (2006). The diseases phenylketonuria (PKU) and scurvy have closely related causes. In the case of PKU, an accumulation of phenylalanine in early life leads to mental retardation. At least two main causal factors are needed for accumulation of phenylalanine: a mutation that disrupts genes for enzymes that metabolize phenylalanine [like in the phenylalanine hydroxylase (PAH) gene], and the consumption of phenylalanine (commonly present in human diets). In scurvy, lack of vitamin C in an individual disrupts the synthesis of collagen, leading to, among other effects, open wounds and loss of teeth. Again, at least two main causal factors are needed for lack of vitamin C: the absence of vitamin C in a diet, and the incapacity to biosynthesize vitamin C (most mammals can synthesize vitamin C from simple glucose, but humans have a mutation in the gene for the L-gulonolactone oxidase (GULO) enzyme, which is required in the last step of vitamin C’s synthesis). Hence, both scurvy and PKU are (necessarily) caused by a mutant gene that leads to loss of function and by a specific diet. Yet, PKU is typically said to have a “genetic” basis, whereas scurvy is said to have an “environmental” basis. PKU is considered a genetic disease because the environmental component of the causation (i.e., phenylalanine in the diet) is nearly universal whereas the PAH mutation is rare. As a consequence, when PKU occurs in a human population, it is because the person has the mutation since virtually all of us have a diet that would promote the PKU response (given that mutation). Therefore, the phenotype of PKU is strongly associated with the PAH mutation in human populations. Scurvy is also the result of the interplay between genes and environment, but in this case the genetic component of the causation is universal in humans. However, the environmental component of the interaction of having a diet without sufficient amounts of vitamin C is rare. Therefore, the phenotype of scurvy is associated with a diet deficient in

vitamin C in human populations. In sum, while mutations and dietary habits are what cause both PKU and scurvy, genetic mutation is what causes some people to express PKU and others to not, while dietary habits cause some people to express scurvy and others to not. Different phenotypes can have the same causation, but different causes of variation!

As the above example illustrates, studying causes of variation reveals *how much* each cause influences the differences between individuals in a population. This is Lesson #2 to glean from biology. In an analogy with physiology, understanding the causational role of cholesterol in the blockage of arteries, although important to understand how the circulatory system works, provides little insight into how big the risks are of cholesterol to heart disease, or how big the role of exercise is as a mitigating factor. In other words, it tells us little about how much each cause can contribute to “real-world” variation. In this sense, the experimental analysis of the causal role of cholesterol in the blockage of arteries with no appreciation of individual differences in the causes of variations would be misleading. This quest for understanding the relative importance (i.e., “how much”) of distinct variables in the establishment of a phenotype led to a boom of new methods from founding-giants like Galton, Pearson, Wright, Fisher, and Spearman. It is not a coincidence that the complexity in trying to organize the clues that nature left in individual differences led to whole new branches of statistics, such as analysis of variance, correlations, regressions, factor analyses, and path analyses.

While the study of causes of variation is powerful, it surely has its limits, and has often been abused by scientists that treated correlations as evidence of causal relationships (for a highly critical view, see Lewontin, 2006). Maybe the best example of the confusion of this distinction between causation and causes of variation lies in the widespread misunderstanding of heritability. Like in any correlational approach, heritability estimates the causes of variation for a specific trait. Specifically, heritability measures how important the difference in genes are for the individual differences in a phenotype in a specific population and environment (Griffiths et al., 2007). A heritability of 0%, however, does not mean that genes have zero influence in the determination of the phenotype (as a matter of fact, all phenotypes have genes as causal factors); it only means that genes are not influencing the existing *individual differences* in that phenotype in that population and that environment (Visscher et al., 2008). Scurvy, as we have seen, has a heritability of 0% since all humans share the same deleterious mutation for the GULO enzyme (that synthesizes vitamin C), but that mutation certainly plays a causal role in the disease! Following the same reasoning, a heritability of 100% does not mean that genes are the sole determinants of a phenotype. Even more problematic, heritability for a specific phenotype can change drastically depending on the environment and the frequency of genes in the population (Bailey, 1997). This ephemeral and fragile aspect of heritability reveals that, although useful, it is only a gross estimation of what is an underlying complex and integrated network of causes of variation (Rockman, 2008).

Living organisms are not only complex (i.e., representing the expression of many independent factors), but are made up of many interacting (necessary and non-necessary) factors that are often shaped by selection to function as integrated units

(Pigliucci, 2003). For some complex phenotypes (like behaviors), vast networks/architectures integrate genetic, biochemical, physiological, and environmental factors across other phenotypes (Oyama, 2000). This high amount of integration of different levels makes the causes of a phenotype not only additive/subtractive, but also multiplicative, divisive, and non-linear (Templeton, 2006). Hence, experimentally modifying one component in isolation gives unpredictable, uninterpretable, or unreplicable results, and we should study multiple components simultaneously (Rockman, 2008). With the advance of genetics, the approach of forward genetics (that follow the detective’s tradition of going from individual difference to causes) is now able to reveal the details that heritability cannot (Mackay et al., 2009). Methods like quantitative trait loci (QTL) analysis and genome-wide association study (GWAS) can reveal gene effects and interactions of genes in the same locus (dominance), in different loci (epistasis) and in other phenotypes (pleiotropy; Erickson, 2005).

As seen above, the network of interacting causes in a living system is much more than the sum of the causes of its parts. This leads us to Lesson #3: studying causes of variation shows *in which way* the complicated and integrated network of causes interact. This integration of complex phenotypes is probably the main reason for the boom in the correlational approach in genetics, with remarkable advances particularly in behavioral genetics (Boake et al., 2002; and for examples of the correlational approach elucidating genetic networks in behaviors, see Ruppel et al., 2004; Edwards and Mackay, 2009; Sauce et al., 2012).

LESSONS APPLIED TO STUDIES OF LEARNING AND BEHAVIOR: THE CASE FOR GREATER FOCUS ON CAUSES OF VARIATION (AND MORE CORRELATIONAL METHODS)

It is now useful to summarize the three lessons described above in relation to Cronbach’s division of psychology according to experimental and correlational methods. While with the experimental approach we can easily determine *what* causal variables underlie learning (“causation of a behavior”), the correlational approach is better suited to determine *how much* and *in which way* variables interact in a population to produce differences in a behavior (“causes of variation of a behavior”). The difficulty inherent to correlational psychology is finding the relevant behaviors and measurements (“clues”), and discriminating between different possible causes. These difficulties are analogous (and no more or less problematic) to those encountered by the experimentalist when deciding upon the appropriate treatment/control groups to include in an experiment.

In genetics, the correlational approach is widely used to understand how much genes influence the differences in a phenotype, and in which way those genes interact to create those differences. In psychology, the same approach can be applied to the study of interacting psychological factors, like cognitive constructs and computational networks (Gallistel and Matzel, 2012). In a way, the study of psychology involves more complex considerations (and systems) than biology, since behavior is one step removed from underlying neuronal activity (Jacob, 1977). Thus it is even more imperative that we attend to causes of variation. As we already described, much work in Psychology has exploited the individual differences approach, but the study of learning and its behavioral

expression is in desperate need for insights provided by correlational methods. In other words, we need to better understand the causes of variation.

As we described above, behaviorism was by its nature an explicitly experimental approach, treating behavior as a compendium of causations, not of causes of variations. Behaviorism explained how learning happens (S-S and S-R models), what the critical variables are (e.g., CS, US, ISI, ITI, contingency, contiguity), what the properties are (e.g., extinction, inhibition, facilitation), the rates and patterns of responding (schedules of reinforcement), and general predispositions (e.g., belongingness, blocking, overshadowing; for a guide to these concepts, see Domjan, 2009). Nonetheless, it has only rarely been asked if individuals would differ in their learning capacities. For example, the acknowledgment that a simple past experience with a stimulus influences that stimulus' "associability" (as during latent inhibition) provides little insight into how other experiences interact to change it, or the relative importance of each experience in the ultimate determination of behavior.

The classic learning models of Rescorla and Wagner (1972); Pearce and Hall (1980), and others that followed were all based on the results of experimental studies, and have been varyingly successful at predicting group *average* performance (Domjan, 2009). However, those models are agnostic in relation to individual differences. In other words, they are neither informed by, nor inform about (predict) causes of variation. It is not a coincidence that most theories of learning emerged directly from the experimental data that immediately preceded them (i.e., new data often demands new theoretical frameworks). In integrated and complex networks, it becomes increasingly difficult to design experiments that produce novel or surprising results. Experimental psychology, in other words, is highly focused on observing new effects. In contrast, in correlational psychology the effects are already there, so it is more critical to make sense of the effects that have been observed.

In an example from the learning literature, the radial arm maze is a test originally designed to measure short-term ("working") memory in rats (Olton and Samuelson, 1976). During the development of the radial arm maze, many experiments were done to differentiate between variables that were needed/necessary to promote efficient performance from those that were not. Variables like algorithmic search (Roberts, 1979), auditory and olfactory guidance (Zoladek and Roberts, 1978), and marking of visited arms (Maki et al., 1984) were all "excluded" as necessary for the animals performance in the maze, suggesting that visual navigation (i.e., "spatial memory") was sufficient. In the behavioral literature, this quest for what is "necessary and/or sufficient" in learning is ubiquitously present, and reveals a mindset of the search for "causation." These experiments with the radial arm maze show what causes-effects can be, and what mice minimally need in order to find food, but not the relative importance of each variable to finding food under "normal" circumstances (either in a laboratory or in the wild). For instance, one could easily imagine a circumstance where, in the presence of degraded visual cues (for instance, in the dark spaces where rodents typically live), an animal might rely primarily on olfactory information for guidance. Thus because an animal *can* use spatial cues to guide its search, it need not necessarily (or even preferentially) do so. (It is somewhat ironic that in our quest for *precision* and *isolation* of causes, the experimental

psychologist has often lost sight of this caveat. In a recent discussion of spatial learning in one of our undergraduate classes, a perceptive student, uninitiated to the dogma of experimental psychology, asked "but in the real world, would not an animal use some combination of these strategies?" Thus what might be obvious to the uninitiated is sometimes lost on the indoctrinated.) From Lesson 3 above, we know that behaviors, like any phenotype, are notoriously complex and integrated, affording many different ways to accomplish the same goal. Therefore, the rats in the radial arm maze may differ not only in their performance, but also in the frequency with which particular strategies are recruited (i.e., how much for smell, visual tracking or algorithm) across individuals. If some rats tend to rely on one "strategy," whereas others habitually rely on alternative strategies, pooling data from both groups may be uninformative and misleading. A non-obvious cause may not be revealed if there is considerable variation within the rats in the tendency or ability to use a particular strategy. In other words, a Type II error can occur if individual variance is not taken into account (for examples and a more in depth discussion of this cases, see Kosslyn et al., 2002).

Granted from Lessons 1, 2, and 3 (above) that correlational psychology and causes of variation are critical for the study of learning and behavior, how does one proceed in actually collecting comprehensive data? As Miller (1959) suggested, multiple response variables (effects) are a problem that can be addressed with factor analysis. By substituting formal for intuitive methods, this type of analysis has been of great help in locating constructs with which to summarize observations (i.e., to organize the clues). As we have seen for genetics, individual differences result from a network of causal factors. A cause can affect multiple phenotypes, and this "pleiotropy" in genetics is what we call in psychology a "latent construct," like the *g* factor ("general intelligence") that affects many different behaviors and cognitive systems. In other situations, more than one cause is able to affect the same phenotype, and this "epistasis" in genetics is closely related to what we in psychology express by "convergent validity" (concept first appearing in Frankmann and Adams, 1962), like emotional arousal that can be defined/caused by different variables (Russell, 1978). In a factor analysis, causes that affect multiple phenotypes lead to covariance structure in a sample of individuals (Houle et al., 2002), i.e., a "latent construct." If a pair of causes affect at least one behavior in common, we see an overlap of factors (Houle et al., 2002), i.e., a "convergent validity." We will now give examples of research in learning for both cases of "latent construct" and "convergent validity" that show the importance of causes of variation and the correlational approach.

SWIMMING NAVIGATION: UNDERSTANDING THE RELATIVE IMPORTANCE OF MANY VARIABLES TO DIFFERENCES IN THE EXPRESSION OF ONE

The Morris water maze is a procedure widely used for studies of spatial learning/memory and navigation (for a review, see D'Hooge and De Deyn, 2001). In the typical paradigm, a mouse is placed into a small pool of water which contains an escape platform hidden below the water's surface. Visual cues, such as geometric patterns or colored shapes, are placed around the pool in plain sight of the animal. The platform remains in the same

position, but, on each trial, the mice are released from different starting points. Most mice learn the task (i.e., find the escape platform efficiently) surprisingly quickly, often reaching asymptotic levels of performance after three or four trials. Absent olfactory (or other) intra-maze cues or a single route that leads to the escape platform, performance on this task is presumed to strongly depend on the animal's reliance on extra-maze visual cues to guide their navigation to the invisible platform. Learning in this instance is usually calculated by the length of the path taken by the animals to find the platform.

Similar to the case of radial arm maze above, there are other (not-so-obvious) behaviors/causes that can influence the animals' performance in the Morris water maze. To assess these influences, Wolfer et al. (1998) looked for causes of variation in swimming navigation by measuring relevant variables (i.e., the right clues) inside the Morris water maze. Using a factor analysis, they found that 81% of all individual differences in performance in the Morris water maze could be largely described in terms of three statistical factors, or causes. Factor 1 explained 49% of the variability, and behaviors that loaded strongly on this factor were correlated with measures of frequent swimming near the wall, prolonged swimming times, and a low fraction of time spent in the actual target quadrant (i.e., the quadrant that contained the escape platform). Because of these clues, the authors interpreted this cause of performance as "thigmotactic behavior," and this factor was asserted to have a decidedly non-spatial origin (i.e., performance was unrelated to the animals having learned a spatial strategy). Factor 2, interpreted by the authors as "passivity," explained 19% of the variability, and correlated with reduced swimming speed and frequent floating. Finally, Factor 3, interpreted by the authors as "memory," accounted for 13% of the behavioral variability, and reflects primarily the search time spent in the former target quadrant during a probe trial (in which the escape platform was absent from the pool). This means that, although memory-guided swimming navigation in the Morris water maze is commonly regarded as being heavily dependent on spatial memory, other causes can be even more important as causes of variation in performance. All of those behaviors/causes are converging on the same behavior, i.e., "navigation," despite the relatively low contribution of spatial learning.

When using the experimental approach, we must assume that an animal behaves the way it should according to the design (parameters) of a test. In the Morris water maze, for example, a preliminary experimental comparison between a group of mutant mice carrying a disruption in the *iPA* gene (believed to play a role in the formation or modification of synaptic connections) and a group of control mice led to the conclusion that spatial memory was unaffected by the *iPA* mutation (Huang et al., 1996). However, Wolfer et al. (1998) showed that this was because the performance scores had been biased by the individual variability in the causes of thigmotaxis and passivity, which masked the subtle genotype difference in memory. With the factor analysis, the *spatial* memory impairments of the mutant mice were revealed.

The example above shows the power of the correlational approach as an aid in separating causes of variations in behavior, and in this instance, to help clean the noise from the interesting causes of variation in swimming navigation (in this case, spatial

memory). In addition, although the authors did not touch on this topic, the results from their factor analysis also showed how much each factor contributes to the differences in swimming navigation of a particular group of mice (which may be an approximation of what happens in other groups). Hence, these analyses suggest more fully how mice operate when trying to find their way across open water. The depth of this analysis could never be achieved simply through the manipulation of a single variable.

GENERAL LEARNING ABILITY: UNDERSTANDING THE RELATIVE IMPORTANCE OF ONE VARIABLE TO THE DIFFERENCES IN MANY OTHERS

In our initial work on this topic, we were looking for a potential general factor that influenced learning across a variety of tasks in mice. If mice differ in their learning capacity, is there a latent factor that can influence causes of variations across disparate learning tasks? To answer this question, we tested mice in a battery of five common learning tasks (associative fear conditioning, passive avoidance, path integration, odor discrimination, and spatial navigation), each of which made unique sensory, motor, and information processing demands on the animals (Matzel et al., 2003). Unlike the more common use of genetically homogeneous animals (see above), here we used a genetically heterogeneous strain of mice in order to maximize the variability (i.e., individual differences) within the group (a useful strategy for correlational research). In our initial study, we performed a factor analysis of the performance of 56 animals across all learning tasks, and obtained a positive correlation across all tasks in which a single latent factor explained 38% of the differences between animals. In other words, animals that performed well in one task tended to perform well in other tasks of the battery. We described that latent factor (or construct) as "general learning ability" (Matzel et al., 2003).

Since the time of that initial report, similar results have been obtained with mice tested on as many as nine learning tasks (Matzel et al., 2008) and in other laboratories (Galsworthy et al., 2005; Locurto et al., 2006). All of these observations reveal one cause influencing the variation in many different learning abilities, analogous to the network of the cause of variation in human intelligence (Jensen, 1998; see Kolata et al., 2008, for a structural analysis based on observations of 250 + mice). Following this, an obvious question arose: is the latent factor that underlies performance on all tasks in our learning battery limited to an influence on *learning*? If, as has been suggested, this factor is analogous to general intelligence in humans (Blinkhorn, 2003; Kolata et al., 2008), we would expect this general cause of variation in mouse learning to interact with (i.e., cause and/or be caused by) other cognitive abilities. Does it? If yes, by how much and in which way? Breaking down the cognitive components of a general factor is similar to the case in behavioral genetics of studying the contribution of individual genes to the genetic architecture underlying the causes of variation in a behavior. Since those first observations, we have been investigating the clues behind mice's general learning differences. Among many causes of variation that we assessed, including animals' propensity for exploration or novelty seeking, working memory capacity, and attentional abilities (Matzel et al., 2006; Matzel and Kolata, 2010; Light et al., 2011), here we describe our work on reasoning capacity.

Based on what we know from the causation of learning in humans (and its analogs in artificial intelligence), we know that reasoning can create efficient heuristics that can ultimately improve learning performance. Therefore, we looked at reasoning as a potential co-variate of general learning in mice. To assess reasoning in mice, we devised a novel task based on a “decision” (or binary) tree maze (for illustration, see Matzel et al., 2011). Decision trees are commonly used in studies of decision analysis to identify strategies that are most efficient in reaching a goal. Unlike learning measures (where rate of acquisition is the critical metric), to assess reasoning we measured only animals’ *asymptotic* behavior, which can be expected to reflect the individual’s implementation of an *established* search strategy. This is important, since we were specifically interested not in *learning* ability, but rather the degree to which the animal can apply learned information in an efficient manner (thus analogous to reasoning). In this regard, two animals with the same underlying *learning* ability might express different aggregate scores in the “learning” battery due to variations in their capacity to act upon what has already been learned. We first tested the animals’ rate of acquisition on the five learning tasks that constitute our standard learning battery, and then assessed their asymptotic performance (presumed to reflect a form of reasoning) in the decision tree. When animals’ reasoning performance was compared to their factor scores for learning (representing mice’s general learning ability), we observed a strong correlation of 0.60 between these independent measures (Wass et al., 2012), i.e., aggregate learning abilities were correlated with rudimentary reasoning abilities.

The above data suggests that animals’ comprehension of the underlying structure of the decision tree, and their implementation of an efficient strategy to use this information, co-varies with their general learning abilities. This correlation is what one might expect if a latent factor influenced not just learning abilities, but rather, *general cognitive performance* (i.e., intelligence). However, performance in the decision tree maze is confounded by short-term memory duration as well as span (i.e., the animal must retain a memory of the depleted goal locations in order to operate efficiently), and so reasoning ability is not the only potential source of performance variation in this task. Thus we developed a second reasoning task (“fast mapping”), on which the animals’ performance was not subject to the same sources of noise. (Although often misunderstood to mean “replication,” “converging operations” is the method by which through independent manipulations, the effects of which have unique sets of underlying interpretations, we can “converge” on one common interpretation; Garner et al., 1956). This exemplifies the investigative work necessary when using the correlational approach. We were trying to find the right clues (reasoning instead of short-term memory) and devise adequate tests to isolate these sources of variance.

“Fast mapping” describes a process whereby a new concept or association (such as the meaning of a word) is formed based on a logical inference derived from a single exposure to limited information (Carey and Bartlett, 1978). This “inference by exclusion” is believed to play a critical role in the extraordinarily rapid and seemingly effortless acquisition of vocabulary during early human

development, and is often described as a hallmark of human reasoning. Kaminski et al. (2004), demonstrated that a Border Collie was able to accurately respond to a command to retrieve a novel object (identified by a novel term) from among set of over 200 previously learned objects. For our purposes, we designed a task to assess fast mapping in mice. Animals were familiarized with a group of objects (small plastic animals), and were then taught to associate pairs of these objects. This was accomplished by exposing the mice to one object, and then allowing them to retrieve a piece of food that was hidden under the sample object’s paired-associate. After learning a series of such object pairs (much like a word can be associated with its meaning), the animals were trained to find the relevant paired-associate within a field that contained several objects, all of which had been previously associated with different samples. This training continued for several weeks until all animals exhibited near errorless choice performance (i.e., chose the correct paired-associate from a field of familiar objects). After completing this training, animals were presented with a “fast mapping” test trial. On these trials, animals were exposed to a novel sample object, and then allowed to explore the test field which contained one novel object among a set of familiar objects (ones that had an established “meaning” based on prior training). The principle of fast mapping suggests that under these conditions, a rational animal should conclude that since the sample object was novel, the food reward should be located under the unfamiliar object in a field of otherwise familiar objects. More importantly, performance on this task makes no obvious demands of short-term memory (or at least a very minimal demand, unlike that required to perform in the decision tree described above). Hence, as any good detective would conclude, “fast mapping” allowed a better isolation of one part of the whole puzzle (analogous to a “control” in the experimental approach). We found that performance on this “cleaner” reasoning task had a correlation of 0.44 with the animals’ aggregate performance in the learning battery (Wass et al., 2012).

The results above suggest that reasoning is part of the bigger network that is also causing differences in the performance of learning tasks (i.e., the latent construct of general learning abilities, that we now call general *cognitive* abilities or GCA). However, it remains to be determined if reasoning participates in this network as a prior cause of variation in GCA, as a mediator between GCA and learning, or is simply another effect of an unspecified common antecedent. These questions could be addressed in the future with the correlational approach involving path analysis and other concepts from structural equation modeling (e.g., endogenous and exogenous variables). It is notable that these statistical techniques, maybe not coincidentally, were co-formulated by a geneticist, Sewall Wright, and a psychologist, Herbert Simons (for more on these methods, see Loehlin, 2003).

As seen in Lesson 3 above, studying individual differences within the context of theories of general mechanisms may provide insights into one of the knottier problems in psychology: understanding non-additive effects of different variables (Kosslyn et al., 2002). That is, not only may the effects of one variable alter the effects of another, but the precise degree to which the variables interact may depend on their values. These are the questions that will guide our future research.

INSIGHTS FROM CAUSES OF VARIATION: THE CASE FOR ANIMALS IN STUDYING LEARNING AND BEHAVIOR

A final case must be made from the three lessons described above: research with non-human animals can be especially powerful when studying causes of variation. In animal studies, complexity is more limited, so we are likely to find fewer (but more dominant) causes of variation even with similar levels of integration. This is because with a bigger number of causes, the potential interactions (the genetic epistasis and pleiotropy, and their environmental/psychological equivalent) are much higher. In a bigger network, the covariance of behaviors and the relative importance of causes in a species (e.g., the genes, neuronal connections, experiences, nutrition, and psychological constructs) are very difficult to understand. In other words, differences in behavior of more complex subjects like humans will reflect more influence from “other” (less dominant) causes, will be more sensitive to these other causes, and thus more difficult to predict. These extra (related or unrelated) causes and effects on individual differences can lead to an under- or over-estimation of the principal causes of behavior, and can lead us down the wrong tracks (i.e., causes for different effects). In this context, one might be compelled to ask if the intricacies of the human condition (for instance, in regard to a topic as complex as intelligence) can be adequately modeled and studied in a non-human animal such as a mouse. At some levels of analysis, the answer to this question is “no.” For instance, variations in intelligence among humans can create effects in academic/professional success, interpersonal relationships, and even prejudice (Gottfredson, 2008; Engle, 2010). These outcomes have no approximate analog in laboratory animals. However, it is exactly for this reason that the vagaries of intelligence are far simpler in animals than they are in humans. It is this simplicity along with the potential for control and invasive interventions that provide opportunities with animals that are not available to those who study intelligence in humans. Clearly, animals can never be expected to provide the complete story of any human behavior. However, much like the synergy between correlational and experimental work, the synergy between human and animal research can inform us about the human condition in ways that would be impossible with human research alone (for relevant data, see Kolata et al., 2010; for discussion and implications, see Matzel et al., in press).

The problem of complexity might explain, for example, the problem of the “missing heritability” in human intelligence. Although intelligence’s heritability is high (around 80%), it has been notoriously difficult to find its genetic causes of variation (much less its environmental influences; Deary et al., 2009). As the human brain became increasingly complex, so did the problems and tasks that humans are likely to undertake. Thus evolution probably played a bigger role in shaping human’s intelligence than it did in other animals. And because the causes of variation in human intelligence are enormously intertwined, they are necessarily harder to recognize, much less separate. Cognitive, neural, and genetic causes might be masking and/or confounding the interpretation of each’s contribution to the overall phenotype. The confusion is sufficiently great that it becomes near impossible to make sense of which the important strings (or clues) are, and which string connects to which.

Simplification by using animals is useful for experimental and correlational approaches in different ways. For experimental studies, using animals may reduce the number of necessary control conditions. For example, if Tolman (1924) employed humans as his subjects instead of rats, he would have needed more experiments to reach the same conclusions. In experimental psychology, too many extra causes (variables) complicate the experimental design, leading to many different treatments/controls. On the other hand, with correlational studies, using animals allows for more clarity to see the hidden, relevant clues, and to test for their distinct contributions and relationships (see Kolata et al., 2010, for the application to the genetics of intelligence in laboratory mice). These reasons for using animals, of course, are in addition to the better-known reasons for research with animals: convenience, cost, and number of techniques available. Of course (as noted above) animal research alone is limited in its application to the human condition. Thus both animal and human research is necessary and complimentary.

As detectives trying to understand a complex crime from professor Moriarty (here, the evolutionary process shaping a behavior across hundreds of generations), it will be extremely useful to understand smaller parts of the plan first (less complex animals, even though still considerably complex), and to later use this foundation to understand bigger parts (more complex animals, like humans and chimpanzees), and, finally, to put all the pieces together. Furthermore, many smaller parts are probably unique (with no counterpart in humans). This would ultimately inform us about how the causes of variation in other animals differ from the causes of variation in humans, and possibly provide evolutionary clues regarding *why* these differences exist. It can go beyond using animals as a generalization to humans. It can become the critical distinction between understanding *human learning* from understanding *learning* (for a similar defense of this position, see De Waal, 2009). So, as detectives, we would understand what a general Moriarty’s crime is (all designs/species for a behavior in all situations), and be more confident of when and how the next will occur.

CONCLUSION

As we have seen, biology, genetics in particular, has been extremely successful in its application of individual differences to our understanding of causes of variation. With regard to the application of correlational methods, some fields in psychology, especially in the study of learning and behavior, have been reluctant to adopt a similar strategy.

In a very influential article, Underwood (1975) argued correctly that the correlational approach can be used as a preliminary test of theories. However, Underwood argued that the use of correlations should be limited to only that, claiming that “if the correlation is substantial, the theory has a go-ahead signal, that and no more. The usual positive correlations across subjects on various skills and aptitudes allow no conclusion concerning the validity of the theory *per se*; experimental ingenuity is responsible for creating and validating a theory.” As we have discussed in this article (especially in Lessons 2 and 3), Underwood made a gross understatement about the power that comes from the study of individual differences. Correlational psychology can be much more than a mere “method of

checking viability.” It can show the importance of each cause, and in which way those variables interact in an integrated network. Furthermore, as seen in Lesson 1 above, detective/correlational work can create and validate highly ingenious and unexpected theories. Darwin and Mendel were well aware of the power of this approach, and few would dispute the magnitude (or lasting influence) of their contributions.

For all of its power, beware, though, of the irresponsible study of individual differences. Describing a multitude of correlations without considering a general mechanism or theoretical framework can be of little use, and even misleading (Kosslyn et al., 2002; Pigliucci, 2003). Darwin, one of the first to use the hypothetic deductive method, knew this rule quite well (Ayala, 2009), and this awareness might be what makes psychologists so distrustful of correlational methods. In the case of studying behavior and learning, our predecessors have successfully employed experimental methods to open the horizon for a better-guided study of causes of variation in learning and behavior. The future application of correlational methods to the study

of learning will need to use animals to simplify the questions that need to be asked in order to infer a network of causes of variation. Also, we will need to know the foundations for animal learning in order to measure it (look for clues) in a creative way so each measurement will provide its own meaningful answers.

As Cronbach (1957) urged five decades ago: “in the search for interactions we will . . . come to realize that organism and treatment are an inseparable pair and that no psychologist can dismiss one or the other as error variance.” By studying more causes of variations on individual differences, we might be able to accomplish the fruitful synergy between experimental and correlational approaches in the study of learning and behavior.

ACKNOWLEDGMENTS

Preparation of this manuscript was supported by grants from the National Institute of Aging (AG022698), the Office of Naval Research (N000141210873) and the Busch Foundation (to Louis D. Matzel).

REFERENCES

- Ayala, F. J. (2009). Darwin and the scientific method. *Proc. Natl. Acad. Sci. U.S.A.* 106(Suppl. 1), 10033–10039. doi: 10.1073/pnas.0901404106
- Bailey, R. C. (1997). Hereditarian scientific fallacies. *Genetica* 99, 125–133.
- Bernier, R. (1984). The species as an individual: facing essentialism. *Syst. Biol.* 33, 460–469. doi: 10.1093/sysbio/33.4.460
- Blinkhorn, S. (2003). Neuroscience: of mice and mentality. *Nature* 424, 1004–1005. doi: 10.1038/4241004a
- Boake, C. R. B., Arnold, S. J., Breden, F., Meffert, L. M., Ritchie, M. G., Taylor, B. J., et al. (2002). Genetic tools for studying adaptation and the evolution of behavior. *Am. Nat.* 160(Suppl. 6), S143–S159. doi: 10.1086/342902
- Carey, S., and Bartlett, E. (1978). Acquiring a single new word. *Proc. Stanford Child Lang. Conf.* 15, 17–29.
- Cleland, C. E. (2002). Methodological and epistemic differences between historical science and experimental science. *Philos. Sci.* 69, 474–496.
- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *Am. Psychol.* 12, 671–684. doi: 10.1037/h0043943
- Deary, I. J., Johnson, W., and Houlihan, L. M. (2009). Genetic foundations of human intelligence. *Hum. Genet.* 126, 215–232. doi: 10.1007/s00439-009-0655-4
- Dennett, D. C. (1995). *Darwin's Dangerous Idea*. New York: Simon & Schuster.
- De Waal, F. B. M. (2009). Darwin's last laugh. *Nature* 460, 175. doi: 10.1038/460175a
- D'Hooge, R., and De Deyn, P. P. (2001). Applications of the Morris water maze in the study of learning and memory. *Brain Res. Brain Res. Rev.* 36, 60–90.
- Domjan, M. P. (2009). *The Principles of Learning and Behavior*, 6th Edn. Belmont: Wadsworth Publishing, 696.
- Ebbinghaus, H. (1913). *Memory: A Contribution to Experimental Psychology*. New York: Teachers College, Columbia University.
- Edwards, A. C., and Mackay, T. F. C. (2009). Quantitative trait loci for aggressive behavior in *Drosophila melanogaster*. *Genetics* 182, 889–897. doi: 10.1534/genetics.109.101691
- Engle, R. W. (2010). Role of working-memory capacity in cognitive control. *Curr. Anthropol.* 51, S17–S26. doi: 10.1086/650572
- Erickson, D. (2005). Mapping the future of QTL's. *Heredity* 95, 417–418. doi: 10.1038/sj.hdy.6800699
- Frankmann, J. P., and Adams, J. A. (1962). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.* 59, 257–272.
- Gallistel, C. R., and Matzel, L. D. (2012). The neuroscience of learning: beyond the Hebbian synapse. *Annu. Rev. Psychol.* 64, 169–200. doi: 10.1146/annurev-psych-113011-143807
- Galsworthy, M. J., Paya-Cano, J. L., Liu, L., Monleón, S., Gregorayan, G., Fernandes, C., et al. (2005). Assessing reliability, heritability and general cognitive ability in a battery of cognitive tasks for laboratory mice. *Behav. Genet.* 35, 675–692. doi: 10.1007/s10519-005-3423-9
- Garner, W. R., Hake, H. W., and Eriksen, C. W. (1956). Operationism and the concept of perception. *Psychol. Rev.* 63, 149–159. doi: 10.1037/h0042992
- Gottfredson, L. S. (2008). “Of what value is intelligence?” *WISC-IV applications for clinical assessment and intervention*, 2nd Edn, eds A. Prifitera, D. Saklofske, and L. G. Weiss (Amsterdam: Elsevier), 545–563.
- Gould, S. J. (2002). *The Structure of Evolutionary Theory*. Cambridge: Harvard University Press.
- Griffiths, A. J. F., Wessler, S. R., Lewontin, R. C., and Carroll, S. B. (2007). *Introduction to Genetic Analysis*, 9th Edn. New York: W. H. Freeman and Company, 800.
- Houle, D., Mezey, J., and Galpern, P. (2002). Interpretation of the results of common principal components analyses. *Evolution* 56, 433–440.
- Huang, Y. Y., Bach, M. E., Lipp, H. P., Zhuo, M., Wolfer, D. P., Hawkins, R. D., et al. (1996). Mice lacking the gene encoding tissue-type plasminogen activator show a selective interference with late-phase long-term potentiation in both Schaffer collateral and mossy fiber pathways. *Proc. Natl. Acad. Sci. U.S.A.* 93, 8699–8704.
- Hull, C. L. (1951). *Essentials of Behavior*. New Haven: Yale University Press.
- Jacob, F. (1977). Evolution and tinkering. *Science* 196, 1161–1166. doi: 10.1126/science.860134
- Jensen, A. R. (1998). *The g Factor: The Science of Mental Ability (Human Evolution, Behavior, and Intelligence)*. New York: Praeger.
- Kahneman, D. (2003). Maps of bounded rationality: psychology for behavioral economics. *Am. Econ. Rev.* 93, 1449–1475. doi: 10.1257/000282803322655392
- 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
- Kolata, S., Light, K., and Matzel, L. D. (2008). Domain-specific and domain-general learning factors are expressed in genetically heterogeneous CD-1 mice. *Intelligence* 36, 619–629. doi: 10.1016/j.intell.2007.12.001
- Kolata, S., Light, K., Wass, C. D., Colas-Zelin, D., Roy, D., and Matzel, L. D. (2010). A dopaminergic gene cluster in the prefrontal cortex predicts performance indicative of general intelligence in genetically heterogeneous mice. *PLoS ONE* 5:e14036. doi: 10.1371/journal.pone.0014036
- Kosslyn, S. M., Cacioppo, J. T., Davidson, R. J., Hugdahl, K., Lovaglio, W. R., Spiegel, D., et al. (2002). Bridging psychology and biology: the analysis of individuals in groups. *Am. Psychol.* 57, 341–351. doi: 10.1037/0003-066X.57.5.341
- Lewontin, R. C. (2006). The analysis of variance and the analysis of causes. *Int. J. Epidemiol.* 35, 520–525. doi: 10.1093/ije/dyl062
- Light, K., Grossman, H., Kolata, S., Wass, C., and Matzel, L. D. (2011). General learning ability regulates exploration through its influence on rate of habituation. *Behav. Brain Res.* 223, 297–309. doi: 10.1016/j.bbr.2011.04.050
- Locurto, C., Benoit, A., Crowley, C., and Miele, A. (2006). The structure of individual differences in batteries of rapid acquisition tasks in mice. *J. Comp. Psychol.* 120, 378–388. doi: 10.1037/0735-7036.120.4.378

- Loehlin, J. C. (2003). *Latent Variable Models: An Introduction to Factor, Path, and Structural Equation Analysis*, 4th Edn. Abingdon: Psychology Press.
- Mackay, T. F. C., Stone, E. A., and Ayroles, J. F. (2009). The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* 10, 565–577. doi: 10.1038/nrg2612
- Maki, W. S., Beatty, W. W., Hoffman, N., Bierley, R. A., and Clouse, B. A. (1984). Spatial memory over long retention intervals: nonmemorial factors are not necessary for accurate performance on the radial-arm maze by rats. *Behav. Neural Biol.* 41, 1–6.
- Matzel, L. D., Grossman, H., Light, K., Townsend, D., and Kolata, S. (2008). Age-related declines in general cognitive abilities of Balb/C mice are associated with disparities in working memory, body weight, and general activity. *Learn. Mem.* 15, 733–746. doi: 10.1101/lm.954808
- Matzel, L. D., Han, Y. R., Grossman, H., Karnik, M. S., Patel, D., Scott, N., et al. (2003). Individual differences in the expression of a “general” learning ability in mice. *J. Neurosci.* 23, 6423–6433.
- Matzel, L. D., and Kolata, S. (2010). Selective attention, working memory, and animal intelligence. *Neurosci. Biobehav. Rev.* 34, 23–30. doi: 10.1016/j.neubiorev.2009.07.002
- Matzel, L. D., Sauce, B., and Wass, C. (in press). The Architecture of intelligence: converging evidence from studies of humans and animals. *Curr. Dir. Psychol. Sci.*
- Matzel, L. D., Townsend, D. A., Grossman, H., Han, Y. R., Hale, G., Zappulla, M., et al. (2006). Exploration in outbred mice covaries with general learning abilities irrespective of stress reactivity, emotionality, and physical attributes. *Neurobiol. Learn. Mem.* 86, 228–240. doi: 10.1016/j.nlm.2006.03.004
- Matzel, L. D., Wass, C., and Kolata, S. (2011). Individual differences in animal intelligence: learning, reasoning, selective attention and inter-species conservation of a cognitive trait. *Int. J. Comp. Psychol.* 24, 36–59.
- Miller, N. E. (1959). “Liberalization of basic S-R concepts: extensions to conflict behavior and social learning,” in *Psychology: A Study of a Science*, ed. S. Koch (New York: McGraw-Hill).
- Nagy, A., Perrimon, N., Sandmeyer, S., and Plasterk, R. (2003). Tailoring the genome: the power of genetic approaches. *Nat. Genet.* 33(Suppl.), 276–284. doi: 10.1038/ng1115
- Olton, D. S., and Samuelson, R. J. (1976). Remembrance of places passed: Spatial memory in rats. *J. Exp. Psychol. Anim. Behav. Process.* 2, 97–116. doi: 10.1037/0097-7403.2.2.97
- Osgood, C. E. (1953). *Method and Theory in Experimental Psychology*. New York: Oxford University Press.
- Oyama, S. (2000). *The Ontogeny of Information: Developmental Systems and Evolution*. Durham: Duke University Press Books.
- Pearce, J. M., and Hall, G. (1980). A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol. Rev.* 87, 532–552.
- Piaget, J., and Inhelder, B. (1973). *Memory and Intelligence*. London: Basic Books.
- Pigliucci, M. (2003). Phenotypic integration: studying the ecology and evolution of complex phenotypes. *Ecol. Lett.* 6, 265–272. doi: 10.1046/j.1461-0248.2003.00428.x
- Rescorla, R. A. (1988). Pavlovian conditioning. It's not what you think it is. *Am. Psychol.* 43, 151–160.
- Rescorla, R. A., and Wagner, A. R. (1972). “A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement,” in *Classical Conditioning II: Current Research and Theory*, eds A. H. Black and W. F. Prokasy (New York: Appleton-Century-Crofts), 64–99.
- Roberts, W. A. (1979). Spatial memory in the rat on a hierarchical maze. *Learn. Motiv.* 10, 117–140. doi: 10.1016/0023-9690(79)90040-7
- Rockman, M. V. (2008). Reverse engineering the genotype-phenotype map with natural genetic variation. *Nature* 456, 738–744. doi: 10.1038/nature07633
- Rüppell, O., Pankiw, T., and Page, R. E. (2004). Pleiotropy, epistasis and new QTL: the genetic architecture of honey bee foraging behavior. *J. Hered.* 95, 481–491. doi: 10.1093/jhered/esh072
- Russell, J. A. (1978). Evidence of convergent validity on the dimensions of affect. *J. Pers. Soc. Psychol.* 36, 1152–1168. doi: 10.1037/0022-3514.36.10.1152
- Ryan, R. M., and Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am. Psychol.* 55, 68–78.
- Sauce, B., De Brito, R. A., and Peripato, A. C. (2012). Genetic architecture of nest building in mice LG/J \times SM/J. *Front. Genet.* 3:90. doi: 10.3389/fgene.2012.00090
- Templeton, A. R. (2006). *Population Genetics and Microevolutionary Theory*. Hoboken: Wiley-Liss.
- Tolman, E. C. (1924). The inheritance of maze-learning ability in rats. *J. Comp. Psychol.* 4, 1–18. doi: 10.1037/h0071979
- Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *Am. Psychol.* 30, 128–134. doi: 10.1037/h0076759
- Visscher, P. M., Hill, W. G., and Wray, N. R. (2008). Heritability in the genomics era—concepts and misconceptions. *Nat. Rev. Genet.* 9, 255–266. doi: 10.1038/nrg2322
- Wass, C., Denman-Brice, A., Rios, C., Light, K. R., Kolata, S., Smith, A. M., et al. (2012). Covariation of learning and “reasoning” abilities in mice: evolutionary conservation of the operations of intelligence. *J. Exp. Psychol. Anim. Behav. Process.* 38, 109–124. doi: 10.1037/a0027355
- Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychol. Rev.* 20, 158–177. doi: 10.1037/h0074428
- Wolfer, D. P., Stagljar-Bozicevic, M., Errington, M. L., and Lipp, H. (1998). Spatial memory and learning in transgenic mice: fact or artifact? *News Physiol. Sci.* 13, 118–123.
- Zoladek, L., and Roberts, W. A. (1978). The sensory basis of spatial memory in the rat. *Anim. Learn. Behav.* 6, 77–81. doi: 10.3758/BF03212006

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: 08 January 2013; paper pending published: 03 April 2013; accepted: 12 June 2013; published online: 04 July 2013.
Citation: Sauce B and Matzel LD (2013) The causes of variation in learning and behavior: why individual differences matter. *Front. Psychol.* 4:395. doi: 10.3389/fpsyg.2013.00395
This article was submitted to *Frontiers in Personality Science and Individual Differences*, a specialty of *Frontiers in Psychology*.
Copyright © 2013 Sauce and Matzel. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.



Revisiting the learning curve (once again)

Steven Glautier *

School of Psychology, University of Southampton, Southampton, UK

Edited by:

Robin A. Murphy, *University of Oxford, UK*

Reviewed by:

Louis Matzel, *Rutgers University, USA*

Irina Baetu, *University of Adelaide, Australia*

***Correspondence:**

Steven Glautier, *School of Psychology, University of Southampton, Southampton, SO17 1BJ, UK*
e-mail: spg@soton.ac.uk

The vast majority of published work in the field of associative learning seeks to test the adequacy of various theoretical accounts of the learning process using average data. Of course, averaging hides important information, but individual departures from the average are usually designated “error” and largely ignored. However, from the perspective of an individual differences approach, this error is the data of interest; and when associative models are applied to individual learning curves the error is substantial. To some extent individual differences can be reasonably understood in terms of parametric variations of the underlying model. Unfortunately, in many cases, the data cannot be accommodated in this way and the applicability of the underlying model can be called into question. Indeed several authors have proposed alternatives to associative models because of the poor fits between data and associative model. In the current paper a novel associative approach to the analysis of individual learning curves is presented. The Memory Environment Cue Array Model (MECAM) is described and applied to two human predictive learning datasets. The MECAM is predicated on the assumption that participants do not parse the trial sequences to which they are exposed into independent episodes as is often assumed when learning curves are modeled. Instead, the MECAM assumes that learning and responding on a trial may also be influenced by the events of the previous trial. Incorporating non-local information the MECAM produced better approximations to individual learning curves than did the Rescorla–Wagner Model (RWM) suggesting that further exploration of the approach is warranted.

Keywords: learning curve, averaging, individual differences, mathematical model, environment structure

Objectively, associative learning theory is a thriving enterprise with a rich tradition of experimental work interpreted through the lenses of sophisticated mathematical models. However, there remains a fundamental empirical observation that is still not well captured by these models. Despite many attempts to provide an adequate account of the learning curves that are produced, even in a simple conditioning experiments, there is still considerable unexplained variation in these curves. For example, many formal models of learning lead us to expect smooth learning curves but these are seldom observed except at the level of average data. Small departures from a theoretical curve can be tolerated as measurement error but when this error is large the model must be called into question and some authors have concluded that associative models are fundamentally wrong. An alternative position, the one adopted in the current paper, is that the associative framework is essentially correct. However, it is argued that much more accurate modeling of individual learning curves is needed and can be achieved by using a more detailed representation of the stimuli provided by the learning environment. In what follows I will describe the application of a mainstream model of associative learning, the Rescorla–Wagner Model (RWM, Rescorla and Wagner, 1972), to individual learning curves. Best fitting RWM learning curves will be compared to best fitting learning curves from a modified approach which uses a more detailed representation of the stimulus environment. The modified approach, which I have named the Memory Environment Cue Array Model

(MECAM), works algorithmically in the same way as the RWM but additionally incorporates memory buffers to hold representations of the previous trial’s events. These memory representations are then processed alongside representations of the current trial. The question addressed in this paper is whether or not we can improve on the standard RWM to obtain a better model for individual learning curves by using the MECAM’s extended description of the stimulus environment. Before describing the details of the MECAM, a brief overview of the RWM and learning curve problems will be presented as a background.

The RWM is widely regarded as a highly successful and relatively simple model of associative learning (c.f. Miller et al., 1995, for an overview). In the RWM learning is described in terms of the growth of associative strength between mental representations of stimulus events. The RWM was originally developed to describe animal learning experiments, in particular experiments using Pavlovian conditioning procedures. During Pavlovian learning the RWM assumes associations are developed between mental representations of the conditioned stimulus (CS) and the unconditioned stimulus (US). For example, the experimenter may present a tone (CS) and a few seconds later an electric shock (US). After a number of CS–US pairings the experimental animal exhibits conditioned responses (CRs e.g., freezing) when the CS is presented and this is said to occur because the associations between the CS and the US representations allow excitation to spread from one representation to the other. Thus, presenting the

CS excites the representation of the US and produces the observed CRs. Informally, the presence of the CS generates an expectancy of the US. The RWM principles are sufficiently general to have been successfully imported into new domains. Since its development as a model of Pavlovian conditioning in animals the RWM has been considered a viable candidate model in a variety of human learning tasks including predictive, causal, and Pavlovian learning (e.g., Dickinson et al., 1984; Lachnit, 1988; Chapman and Robbins, 1990).

$$\Delta V = \alpha\beta(\lambda - \Sigma V) \quad (1)$$

Equation (1) is the fundamental RWM learning equation. In the equation ΔV is the change in the associative strength between the mental representation of a predictive stimulus (such as a tone CS) and the representation of the outcome (such as a shock US) that occurs on a single learning trial. ΔV is a function of two learning rate parameters, α for the CS and β for the US, and the parenthesized error term. In the error term λ is the value of the US on that trial (usually 1 or 0 for the occurrence and non-occurrence of the US, respectively) and ΣV is the summed associative strength of all the predictors that are present on the trial. The RWM is said to be error driven and competitive. It is error drive in the sense that the amount of learning depends on the difference between what occurs, λ , and what was expected, ΣV . It is competitive in the sense that the updates applied to the associative strength of a stimulus depend not just on the strength of that stimulus but also on the strength of all the other stimuli that are present on the trial— ΣV is used in the error term rather than V alone. This competitive error driven formulation is a defining feature of the RWM and has been adopted in many neural network models of learning (c.f. Sutton and Barto, 1981).

Historically, analysis of learning curves has been an important testing ground for theories of learning. Any credible theory of learning must be able to account for state transitions, as well as steady state performance. Each theory of learning makes characteristic predictions for the shape of the learning curve, the RWM is no exception. Referring to Equation (1) we can see that associative strength increases as a fixed proportion ($\alpha\beta$) of the difference between the current associative strength and the asymptote. From the RWM we therefore expect orderly negatively accelerated learning curves. Because each theory of learning makes characteristic learning curve predictions, in principle, analysis of learning curves should be theoretically decisive. Unfortunately, the utility of this approach has not been realized because of the empirically observed heterogeneity in learning curves. Smooth monotonic, S-shaped, and stepped curves have all been seen at one time or another leading Mazur and Hastie to comment “In fact, learning curves of almost every conceivable shape have been found.” (Mazur and Hastie, 1978, p. 1258). No doubt some of this variability can be accounted for by the type of task. For example, many tasks have several components, some of which might be relatively easy to learn. On this basis a task composed of simple and difficult components could produce rapid improvements in performance in the first few trials after which the rate of improvement would decline. On the other hand, a multicomponent task which involved several equally difficult components could produce a less variable rate of improvement. Thus, the shape of the

learning curve might be affected by the structure of the task that is presented and may not be straightforwardly diagnostic of the underlying process. Nevertheless, despite these interpretational problems, analyses of learning curves led to a widespread acceptance of the principle embodied in Newell and Rosenbloom’s Power Law of learning (Newell and Rosenbloom, 1981). The Power Law of learning is based on an equation of the form P (Correct Response) = $1 - \alpha t^{-\beta}$ where t is the trial number in the series, α and β are parameters of the curve. An equation of this type generates a curve in which the proportional progress toward asymptote declines with trials. In contrast, an exponential function P (Correct Response) = $1 - \alpha e^{-\alpha t}$ generates a curve in which the proportional progress toward asymptote remains constant with trials. Although there is now doubt about the status of the Power Law (Heathcote et al., 2000; Myung et al., 2000), the point to draw attention to is the critical theoretical position that has been occupied by learning curve analyses and the fact that this theoretical promise has not been realized—we cannot confidently rule in or out the RWM on the basis of its characteristic exponential form.

However, when individual learning curves are considered it is not surprising that it has proved difficult to clearly determine whether learning curves are best characterized by power functions or by exponential functions. These are relatively subtle differences occurring against a background of great variability from one participant to the next. At the level of individual learning curves there is actually little evidence of smooth learning functions, let alone clearly distinguishable exponential or power functions. One solution to this problem has been to average the individual data and then try to find the function which best describes the average curve. These average data can be well approximated by exponential or power functions. Unfortunately, this is not a viable solution because averaging of the data points generated by a function does not, in general, equal the application of that function to the average i.e., $\text{Mean}(f(i), f(j), \dots, f(n)) \neq f(\text{Mean}(i, j, \dots, n))$ (Sidman, 1952; Estes, 2002).

Although it has not been possible to adjudicate between exponential and power models of learning, analysis of the learning curve continues to stimulate important theoretical debates. The difficulty with trying to represent individual learning curves with the orderly incremental learning functions used in associative models of learning such as the RWM has led some authors to question the applicability of associative models, as a class, and to propose alternative, non-associative, mechanisms for learning. Köhler’s (1925) work on insight learning is an early example, more recent statements come from Nosofsky et al. (1994) and Gallistel and Gibbon (2002). Nosofsky et al. described the Rule-Plus-Exception (RULEX) model of classification learning in which learning is conceived of as the acquisition of simple rules for classification e.g., “if feature A is present the item belongs to category X.” In RULEX simple rules are tried first and, if these fail, exceptions and more complex rules may then be tried. The relevance of RULEX in the current context is its supposition that individual learners will test and adopt rules in idiosyncratic ways and that acquisition of a successful rule will result in step changes in learning performance. Therefore individual curves will be characterized by abrupt changes and the location of these changes in

a sequence of learning will vary randomly from participant to participant. Gallistel and Gibbon (2002) advocate an information processing model in which a response is generated when the value of a decision variable reaches a threshold value. Individuals vary in terms of the threshold value and in terms of the value of the decision variable. The result is that learning curves are expected to contain step changes varying in location from individual to individual (Gallistel et al., 2004). Neither of these models anticipate smooth individual learning curves but in both cases averaging of the individual curves produced by the models would result in smoothing. In both cases non-associative cognitive processes are proposed to explain the patterns observed in the individual data.

It is accepted that the RWM, and other modern associative models, only provide poor approximations to individual learning curves. Individual curves are highly variable from participant to participant. For example, looking ahead to the dataset to be described in more detail below, it can be seen that some participants learn quickly, apparently hitting upon a solution straight away (e.g., **Figure 7** middle panel, square symbols). Some learn quickly but might take several trials to find the solution (e.g., **Figure 7** left middle panel, square symbols). Others learn slowly with responses gradually approaching an asymptote as might be expected from the RWM (e.g., **Figure 4** left middle panel, square symbols). Furthermore, responses are often unstable showing trial-to-trial fluctuations (e.g., **Figure 2** left top panel, square symbols). Instability can occur even if an asymptote appears to have been reached (e.g., **Figure 5** right middle panel, square symbols). In these respects this human predictive learning data contains the same features described by Gallistel et al. (2004) in a variety of animal learning tasks including autoshaped pigeon key presses and eye-blink conditioning in rabbits.

The main purpose of the current paper is to explore a development in the application of the RWM with the aim of trying to obtain a better approximation to individual acquisition data within a simple associative framework. Readers familiar with associative approaches related to Stimulus Sampling Theory (Estes, 1950; Atkinson and Estes, 1963) may question the appropriateness of the RWM as the origin for this endeavor when two basic principles of Stimulus Sampling Theory appear to provide an initial step in the right direction. These principles are those of probabilistic environmental sampling and all-or-none learning (see also original paper and recent review of all-or-none learning debate Rock, 1957; Roediger and Arnold, 2012). In Stimulus Sampling Theory it is assumed that each learning trial involves a probabilistically obtained sample of stimulus elements. Given that the sampled elements may be connected to different responses there is a built in mechanism that can produce trial-by-trial response variability. Furthermore, because associations are assumed to be made in an all-or-none fashion when reinforcement occurs step-wise changes in behavior are expected. However, although Stimulus Sampling Theory is *prima-facia* a strong candidate with which to tackle the characterization of individual learning curves the RWM was chosen as a basis because of its competitive error driven formulation which has proven to be extremely useful (but not universally successful c.f. Miller et al., 1995) in accounting for a wide variety of other learning phenomena.

In developing the framework provided by the RWM the starting point was to question the assumption that participants in a learning experiment base their expectations and learning for the current trial just on the stimuli present on that trial. Actually, the learning trial is an artificial structuring of events created largely for the convenience of the experiment and there is no good reason to believe that participants actually parse their experience in this way. In fact most learning experiments have short inter-trial-intervals of just a few seconds (e.g., in Thorwart et al., 2010, ITIs of 4 s and 6 s were used in two different experiments) so that participants will still have fresh in their minds a memory of the previous trial. Evidence from several sources confirms that participants do remember previous trials and these memories can influence behavior on the current trial. For example participants remember when they have had a series of reinforced or non-reinforced trials and this affects what they expect to happen on the current trial (Perruchet et al., 2006). In the Perruchet task a long sequence of non-reinforced trials leads to an expectation that the next trial will be reinforced and vice-versa. Participants also respond to trial sequence information so that reaction time is reduced if the sequence is predictive of the response requirement, and this can occur without the participants developing a conscious expectancy for the outcome (e.g., Jones and McLaren, 2009).

In the MECA Model it is proposed that remembered stimulus elements from the previous trial are processed along with current elements and can therefore acquire associations with the outcome and contribute to the control of expectations in the same way as current elements. The MECAM works by utilizing three memory buffers in which representations of the current trial are stored alongside representations of the previous trial. The MECAM encodes the stimuli of the current trial in the *primary buffer*. Experimenter defined stimulus elements serving the CS roles are encoded along with unique configural cues (Rescorla, 1973) representing pairwise interactions between experimenter defined stimulus elements. The *secondary buffer* is a copy of the primary buffer from the previous trial plus a representation of the outcome event that served as the US on the previous trial. The *interaction buffer* contains pairwise configural cue representations for the elements from the current and previous trial. The MECAM contains a parameter ω which weights the secondary and interaction buffers. Setting these weighting parameters to zero reduces the MECAM to the RWM. The Appendix contains a detailed description of the implementations of the RWM and MECAM that were used in the simulations that will be reported below.

$$\Delta V = \omega \alpha \beta (\lambda - \Sigma V) \quad (2)$$

Equation (2) provides the learning equation used in MECAM. There is no difference between the RW and MECA models in the way associative strength updates are made except that in the MECAM the additional parameter ω is combined multiplicatively with the learning rate parameters α and β (compare Equation (1) and Equation (2)). The value of ω is allowed to vary for each cue according to the buffer in which the cue is defined. Primary buffer cues have $\omega = 1$ whereas for secondary and interaction

buffers $0 \leq \omega \leq 1$. Further details are provided in the Appendix and below there follows a short outline of MECAM's operation.

Table A2 provides an illustration of the operation of MECAM's buffers during three conditioning trials. On the first trial experimenter defined cues A and B are present along with the US outcome (an AB+ trial). The cue elements A and B appear in the primary buffer as does the configural cue ab. Cue ab is a theoretical entity used to represent the conjunction of the elements A and B. Because this is the first trial the secondary and interaction buffers are empty and only the cues A, B, and ab will have their associative strengths updated. At this point the RWM and MECAM are entirely equivalent. Differences appear on the second trial because now MECAM processes memorial representations of the events of the first trial alongside the events that occur on trial two. On trial two, three cues A, B, and C are present and there is no outcome (an ABC- trial). Configural cues ab, ac, and bc are used to represent the pairwise conjunctions of the cue elements. Thus, on trial two, there are six stimuli present in the primary buffer. There is no difference between the RWM and the MECAM in the processing of primary buffer cues. However, the MECAM additionally operates on the cue representations which now occupy the secondary and interaction buffers. There are two aspects of this operation. First, the existing associative strengths of the secondary and interaction buffer cues are combined with those in the primary buffer to produce ΣV . In this way the contents of all three buffers contribute to the outcome expectation for the trial. Second, the associative strengths of the cues present in all three buffers are updated. The cues present in the primary buffer are always just those that occur on the current trial (including configural components) whereas the secondary buffer contains a copy of all of the stimuli that occurred on the previous trial. These remembered stimuli have their own representations and associative strength. Thus, stimuli A and A_{t-1} are distinct entities, as are ab and $a_{t-1}b_{t-1}$. Because the outcome of the previous trial is just as likely, if not more likely, to be remembered than the cues, the previous trial outcome is also coded as one of the remembered stimuli in the secondary buffer (O_{t-1}). The interaction buffer encodes a subset of the configural cues that are processed by MECAM. This subset consists of pairwise configurations of the elements of the current trial and the remembered elements from the previous trial. In the Trial 2 example shown in **Table A2** the elements are A, B, and C from the current trial and elements A_{t-1} , B_{t-1} , and O_{t-1} from the previous trial. This results in nine configural cues appearing in the interaction buffer. The use of three buffers allows different ω weights to be used for different classes of stimulus entity. The third trial illustrated in **Table A2** gives a further example of how the buffer states change on the next, BC-, trial.

The MECAM is predicated on the assumption that the source of the behavioral complexity in individual learning curves is to be found in the environment to which the participants are actually exposed. A corollary is that even if the RWM is correct in its basic principles then simulations of individual participant behavior using the RWM will be inaccurate unless the input representations for the simulation match those in the individual's learning experience. The MECAM hypothesis is that during learning some of the influences on participant responding

will be due to learning of associations between trial outcomes and memories of events occurring on previous trials. If this is correct then MECAM simulations, which incorporate representations of the previous trial events as inputs to the learning and expectations for the current trial, would provide better approximations to individual learning curves than the RWM, which involves learning and expectations only for current trial events. The experiments reported below involved participants making judgements about the likelihood of an outcome in each of a series of trials. Participant responses were in the form of ratings on an 11-point scale, running from 0—event will not occur, through 5—event will/will not occur with equal likelihood, to 10—event will occur. However, these judgements are not represented directly in either the RWM or MECAM. The currency of these models is the unobserved theoretical quantity of “associative strength.” Therefore, to model the changes in these judgements during learning it was necessary to find an appropriate way to map between the theoretical quantity of associative strength and observed judgements.

Unfortunately there is little agreement on the specific mapping between association strength and behavioral response (Rescorla, 2001). This situation may seem to be a fatal flaw in any attempt to provide a testable associative theory but the problem can be circumvented in some cases by making the minimal assumption of a monotonic relationship between the strength of the CRs and association strength. This is reasonable when there are qualitatively different predictions for the effect of an experimental manipulation for the theories under consideration. For example, in a feature-negative experiment one stimulus is reinforced (A+ trials) but a compound stimulus is non-reinforced (AB- trials). The effects of adding a common feature to these trials, to give AC+ and ABC- trials, differs qualitatively for leading associative models (Thorwart et al., 2010). According to the RWM the common-cue manipulation should make the discrimination between reinforced and non-reinforced trials easier whereas according to an alternative associative model the discrimination should become more difficult (Pearce, 1994). Thus, that comparison (Thorwart et al., 2010) between two associative models only required the assumption of a monotonic mapping between association strength and response strength. However, in the current work there are no experimental manipulations with qualitatively different predictions for the RWM and MECAM. Instead, a quantitative comparison of the goodness of fit between RWM and MECAM predictions and participant responses was carried out. This needs a mapping between the model currency of association strength and behavioral response and a choice of mappings is available. Two mapping functions were selected and compared. It was assumed that strength of association could be treated as type of stimulus to which participants would respond when asked to make their predictive judgements so that a psychophysical scaling would be appropriate. Two psychophysical functions have frequently been used to relate stimulus magnitude to perceived stimulus intensity, one based on Stevens' Power Law the other based on Fechner's Law (e.g., Krueger, 1989). In the analyses below simulations were carried out using both of these mappings and comparisons between them were made.

METHODS

The simulations reported below used data from a series of six different multi-stage experiments. These experiments all used a computer-based predictive learning task with a first stage consisting of AX+, AY+, BX−, and BY− trials. Data from these trials was used in the following analyses. In this notation the letters indicate which cues are present on a trial, the plus and minus signs indicate the presence or absence of the outcome. Analysis 1 used data from Experiment 1. The data from experiments 2–6 were combined and treated as data from a single experiment, hereinafter referred to as Experiment 2, in Analysis 2.

EXPERIMENTAL METHOD

The computer-based predictive learning task was presented as a simple card game in which the participants had to learn which cards would be winning cards. Participants were presented with a series of trials each beginning with a display of a card. Participants then used the keyboard cursor keys to adjust an onscreen indicator to indicate their judgement of the likelihood that the card would win. After the participant made a judgement the trial ended with feedback on whether the card won or lost. The cards had distinctive symbols and background colors such that the symbols and colors could be used as cues to distinguish the winning and losing cards. Experiment 1 and Experiment 2 used different computer programs for implementation of the task, had different numbers of trials in the learning sequence, and used different participant populations. The five experiments that were combined for Experiment 2 were the same on all of these variables so they were analyzed together as a single experiment. Replication of the analyses on the datasets of Experiment 1 and Experiment 2 provided a test of reliability and generality of findings.

Participants

Sixty-one participants took part in Experiment 1. Their average age was 17 years and they included 18 males. They were recruited during a site visit to a sixth form (age 16–18) college in Hampshire, UK. Participation was voluntary. One hundred and forty-four participants took part in Experiment 2. Their average age was 22 years and they included 41 males. They were recruited from the student and staff at the University of Wales Swansea campus and were paid £3 for participating.

Apparatus

In Experiment 1 participants were tested in groups at three computer workstations housed in a mobile research laboratory set up in the load compartment of a specially equipped Citroen Relay van. To minimize interference between participants auditory stimuli were presented over headphones and seating was arranged so that participants could easily view only their own computer screen. The screens measured 41 cm × 26 cm (W × H) and were run in 32 bit color mode with pixel resolutions of 1440 × 900. The display was controlled by a computer program written in Microsoft Visual Studio 2008 C# language and used XNA Game Studio Version 3.1 for 3D rendering of the experimental scenario. In Experiment 2 participants were tested individually in small experimental cubicles with sounds presented over the computer speakers. The screens measured 28 cm × 21 cm (W × H)

and were run in 8 bit color mode with pixel resolutions of 640 × 480. The display was controlled by a computer program written in Borland Turbo Pascal.

Design and procedure

In all experiments participants were given a brief verbal description of the procedure before reading and signing a consent form. Next, a more detailed description of the procedure was presented on-screen for participants to read. In Experiment 1 the on-screen information was given along with a voiceover of the text, played through the headphones. The text from Experiment 1 is reproduced in full below. The text used in Experiment 2 had minor wording differences but conveyed the same information.

Thank you for agreeing to take part in this experiment. During the experiment you will be shown a series of “playing cards” on the computer screen. The cards were played in a game at Poker Faced Joe’s Casino. The experiment is divided into a series of trials, each trial representing one card game. On each trial you have to rate the likelihood that the cards on the screen will WIN or LOSE. Make your rating by adjusting the indicator using the UP and DOWN arrow keys. When you have made your rating press RETURN. When you press return the cards will be turned over and you will find out whether they win or lose. Your job is to learn what outcome to expect. At first you will not know what to expect so you will have to guess. However, as you learn, you should aim to make your predictions as accurate as possible, to reflect the true value of the cards that are in play. Review these instructions on the screen. When you are sure that you understand what is required, press the key C to continue. Please note, Poker Faced Joe’s is an imaginary casino you will not lose or gain any money by the rating you make. However, please try to make your judgements as quickly and as accurately as you can. Ask the experimenter if you have any questions or press the key C to begin.

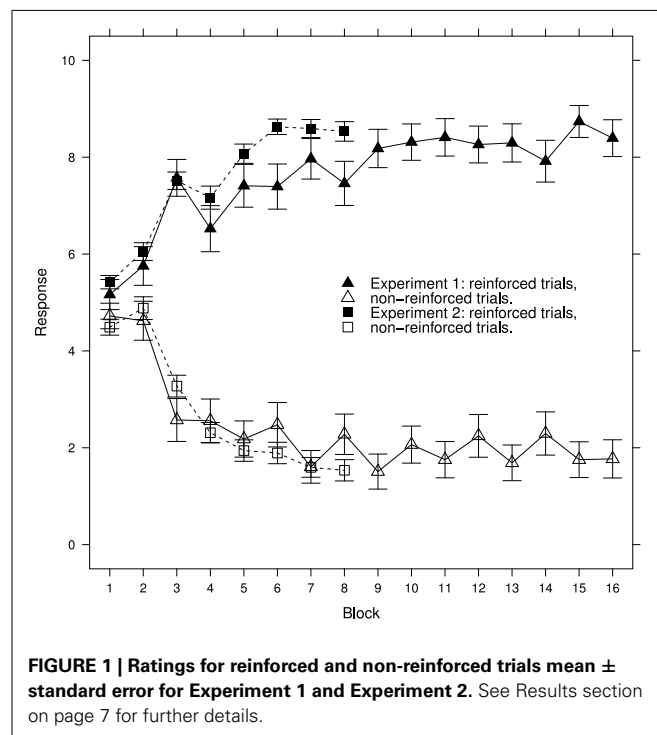


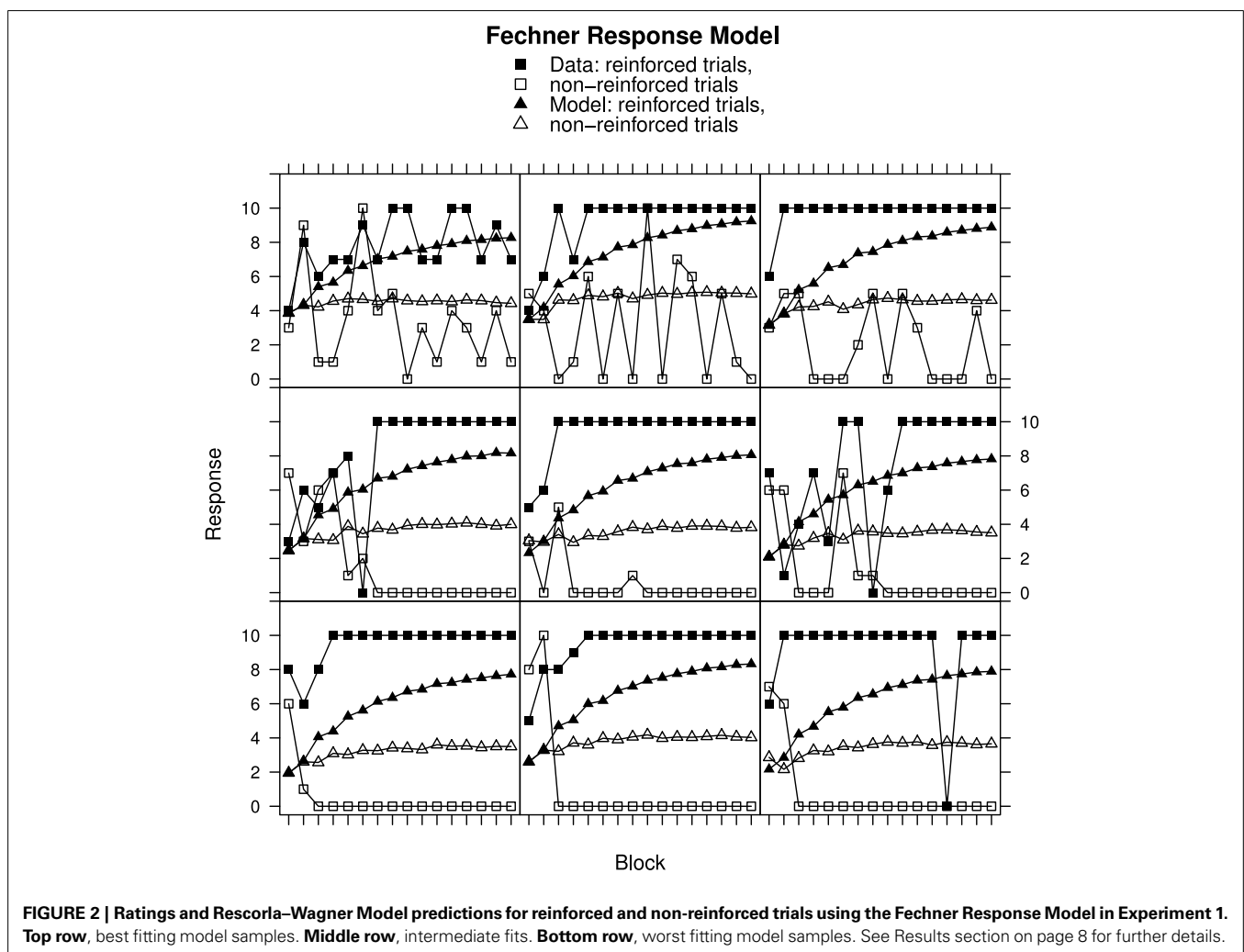
FIGURE 1 | Ratings for reinforced and non-reinforced trials mean ± standard error for Experiment 1 and Experiment 2. See Results section on page 7 for further details.

After reading the instructions participants initiated the experimental trials with a key press. There then followed a series of trials. Each trial was one of four types; AX+, AY+, BX−, or BY−. In Experiment 1 participants had eight of each trial type presented in a random order, with order randomized for each participant subject to the constraint that no more than two trials of the same type could occur in sequence. The symbols and colors serving the cue functions A, B, X, and Y were selected at random for each participant from a set of 14 symbols and a set of 13 colors (e.g., Wingdings character 94 on a pink background). The background colors were allocated to role of informative cues (A and B) and the symbols allocated to the role of redundant cues (X and Y) in an approximately counterbalanced fashion so that 30 participants had colors in the A, B roles and foreground symbols in the X, Y roles; vice-versa for the remaining 31. In Experiment 2 participants had four trials of each type presented in one of five different orders, each order randomized subject to the constraint that no more than three trials of one type could occur in sequence. Four different symbols and three different colors were used. Allocation of colors and symbols to the role of informative (A and B) and redundant (X and Y) cues was approximately counterbalanced ($n = 73$ color predictive and $n = 71$ symbol predictive). In both

experiments trials AX+ and AY+ were reinforced trials and were followed by the “win” outcome after participants made their judgements. Trials BX− and BY− were non-reinforced trials, and were followed by the “lose” outcome after participants made there judgements. Outcome feedback was in the form of onscreen text “win” and “lose” accompanied by distinctive auditory signals.

ANALYSES

Analyses 1 used data from the 61 participants who took part in Experiment 1. Analyses 2 used data from the 144 participants who took part in Experiment 2. Both analyses each involved running four simulations. Simulations of the RW and the MECA models were both run twice against the data from each participant; once with the Stevens and once with the Fechner response mappings. The simulations were carried out in order to select optimized values for model parameters i.e., the simulations involved tuning the model parameters to produce responses matched as closely as possible to those actually made by the participant. The simulations were done using a computer program written in Java and using the Apache Commons Math implementation of Hansen’s Covariance Matrix Adaptation Evolution Strategy (Hansen, 2006, 2012; Commons Math Developers, 2013). The



Covariance Matrix Adaptation Evolution Strategy (CMAES) is a derivative-free multivariate optimization algorithm which was applied to an objective function that produced the sum of squared deviations (SSD), summed over all learning trials, between the participant's response and the model. The CMAES algorithm searched for best fitting parameters for the model such that the value of the objective function was minimized. Thus, the analyses yielded, for each participant and each model, a set of parameters and an SSD value as a measure of goodness of fit. The parameters involved included the α and β learning rate parameters for the RWM and MECAM (Equation A1 and Appendix Equation A5), the buffer weights for the MECAM (ω values, Appendix Equation A5), and the parameters used to control the mapping of association to response strength in the Fechner and Stevens models (Appendix Equations A3, A4). Further details of the simulation methods are given in the Appendix. Statistical tests were performed using the R statistics package (R Core Development Team, 2012).

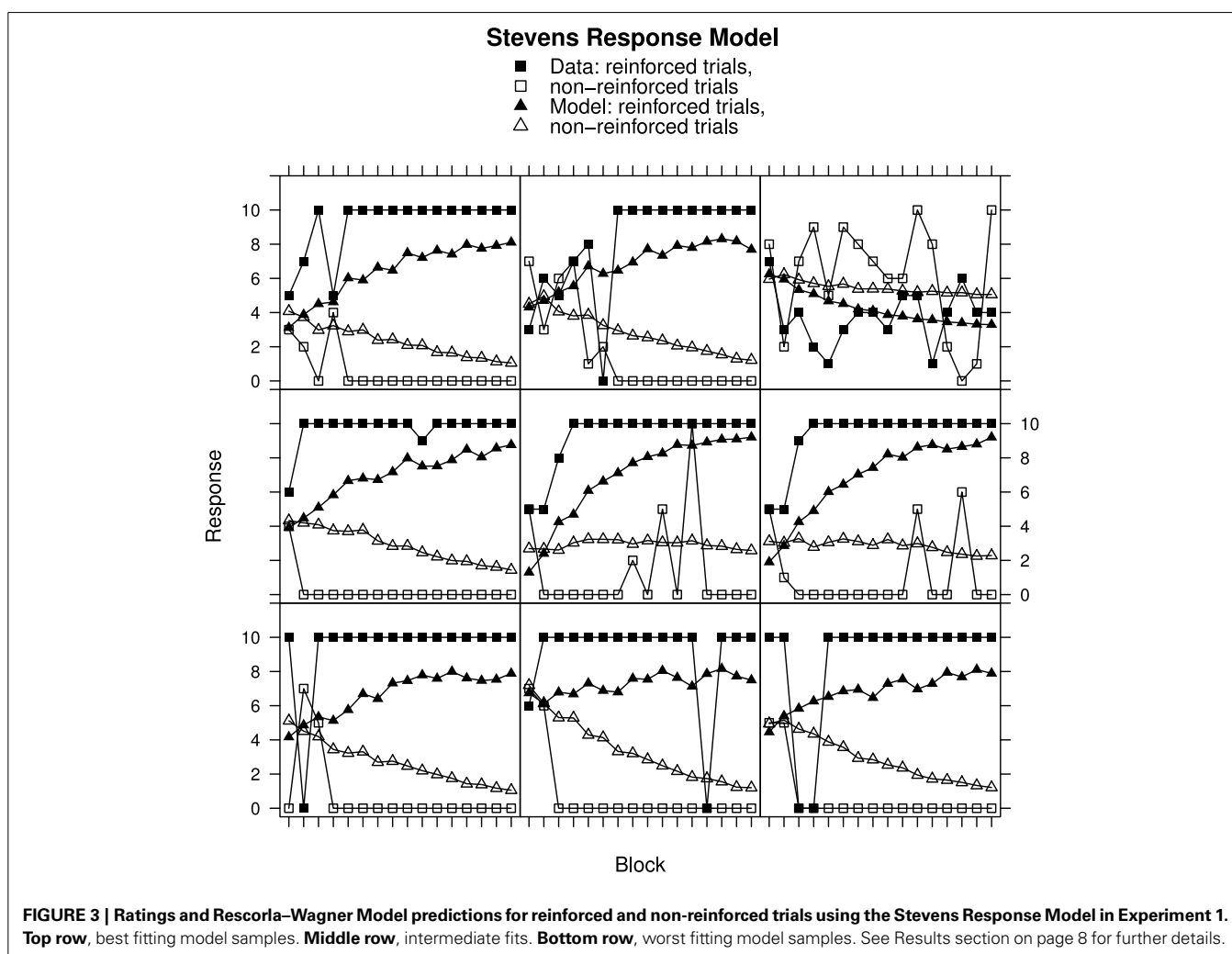
RESULTS

The results are presented in four parts. First, the average learning curves from Experiment 1 and Experiment 2 are presented.

Second, comparisons are made between the models using Stevens and Fechner response mappings. The Stevens response mapping produced better fits and, for brevity, some results are only presented graphically for the models with Stevens response mapping. Third, a comparison of the RW and MECAM models is made. Finally, a comparison of the model parameters between Experiment 1 and Experiment 2 was made to determine their stability from one dataset to another. In the results that follow the SSD values found in the optimizations were converted to Root Mean Square (RMS) measures of goodness of fit. This was done to provide comparability between Experiment 1 and Experiment 2. This was necessary because Experiment 1 had 32 learning trials whereas there were only 16 trials in Experiment 2. Thus the SSD values for Experiment 1 were larger than those in Experiment 2. Because RMS error is the average error over all data points RMS magnitude is not directly affected by the length of the trial sequence.

AVERAGE LEARNING CURVES

Figure 1 shows the average learning curves generated in Experiment 1 and 2. These curves show that learning has taken place, there are clear differences in responses to reinforced and



non-reinforced cards after the second block of trials. However, for reasons described in the introduction, the learning functions for individual participants cannot be deduced from these averages. Furthermore, these average curves hide a great deal of detail at the level of individual learning curves. In order to address both of these issues each of the following figures shows an ordered selection of individual participant data.

COMPARISON OF FECHNER AND STEVENS RESPONSE MAPPING

Figures 2, 3 show individual learning curves for samples of participants from Experiment 1 alongside model fits obtained for the Rescorla–Wagner Model equipped with the Fechner (**Figure 2**) and Stevens (**Figure 3**) response mapping models. Each figure contains nine graphics, each of which shows data for an individual participant and the associated best fitting model predictions. The data in the rows is selected to illustrate the variation in goodness of fit between model and data. The top rows represent best fits. They contain samples of participants from the lower tercile of the RMS error distributions. The middle rows contain samples of participants from the middle tercile of the RMS error distributions. The bottom rows represent worst fits. They contain samples of participants from the upper tercile of the RMS error distributions.

Figure 2 shows data from Experiment 1 plotted along with best fits from the Rescorla–Wagner Model using Fechner’s equation for mapping associative strength to response. All of the participants featured in this figure have learned to respond appropriately to the reinforced and non-reinforced cards but in several cases (e.g., top-left panel) the participants’ responses remain unstable, varying from trial-to-trial. The best fitting simulation responses mirror the overall discriminations made by the participants but do not capture the trial-to-trial variation in responding produced

by the participants, nor the downward trend in response on the non-reinforced trials. It is notable that the worst model fits, in the bottom row, occur for participants who had quickly learned the discriminations. The poor fits occur because participant responses reach asymptote within the first few trials while the model responses slowly approach their asymptotes. This results in large discrepancies between data and model on the early trials. In contrast, in the top row, the fits are better because the participant responses asymptote more slowly. Analysis of Variance on these data produced a significant 3-way interaction [$F_{(30, 870)} = 2.28, p < 0.001$] of Block (1–16) \times Reinforcement (non-reinforced “v” reinforced) \times Group (Best, intermediate, and worst RMS fit) confirming that the development of the discrimination between non-reinforced and reinforced trials differed according to the model goodness of fit.

Turning to **Figure 3**, participant data from Experiment 1 is shown alongside Rescorla–Wagner Model best fits using Stevens’ equation to map associative strength to response strength. All except one participant (top-right panel) in this sample has learned to respond appropriately. Once again the fits for the participants who learned very quickly are worse (bottom row) than for those who learned more slowly (top row) with ANOVA showing a significant interaction between Block, Reinforcement, and Group [$F_{(30, 870)} = 2.30, p < 0.001$]. In contrast to the Fechner based model, the model responses on the non-reinforced trials decline over trial blocks.

Student’s *t*-tests on the RMS error showed that the mean RMS fit was significantly better for the Stevens Response Model than for the Fechner Response Model [$t_{(60)} = 10.67, p < 0.001$]. The mean RMS error values are given in **Table 1**. A very similar picture was obtained for the analysis of Experiment 2. For brevity a

Table 1 | Parameters values obtained in Analyses 1 and 2 and model goodness of fit values (RMS).

Experiment	Model	α_{ctx}	α_{cue}	β_{rt}	β_{nrt}	k	a	c	sbw	ibw	sv	RMS
FECHNER RESPONSE MAPPING												
1	RWM	0.03 (0.008)	0.24 (0.006)	0.24 (0.005)	0.20 (0.011)	9.10 (0.236)	–	1.71 (0.111)	–	–	0.04 (0.010)	3.24 (0.078)
	MECAM	0.13 (0.013)	0.24 (0.005)	0.24 (0.005)	0.20 (0.012)	8.52 (0.323)	–	1.11 (0.118)	0.32 (0.043)	0.80 (0.039)	0.05 (0.010)	3.06 (0.071)
2	RWM	0.04 (0.006)	0.24 (0.004)	0.24 (0.002)	0.22 (0.005)	9.40 (0.107)	–	2.25 (0.087)	–	–	0.04 (0.005)	2.91 (0.057)
	MECAM	0.11 (0.009)	0.24 (0.003)	0.24 (0.003)	0.22 (0.005)	8.67 (0.168)	–	1.44 (0.088)	0.35 (0.029)	0.78 (0.024)	0.04 (0.004)	2.77 (0.057)
STEVENS RESPONSE MAPPING												
1	RWM	0.18 (0.013)	0.23 (0.008)	0.23 (0.006)	0.20 (0.012)	8.91 (0.290)	2.25 (0.107)	–	–	–	0.12 (0.007)	2.82 (0.068)
	MECAM	0.11 (0.012)	0.23 (0.006)	0.23 (0.004)	0.20 (0.011)	8.39 (0.322)	1.81 (0.118)	–	0.27 (0.036)	0.67 (0.041)	0.07 (0.009)	2.65 (0.071)
2	RWM	0.18 (0.008)	0.24 (0.004)	0.24 (0.002)	0.23 (0.005)	9.26 (0.144)	2.53 (0.067)	–	–	–	0.12 (0.004)	2.60 (0.055)
	MECAM	0.12 (0.008)	0.24 (0.003)	0.24 (0.003)	0.22 (0.006)	8.61 (0.145)	1.84 (0.081)	–	0.30 (0.025)	0.77 (0.022)	0.06 (0.004)	2.43 (0.047)

Means (standard error).

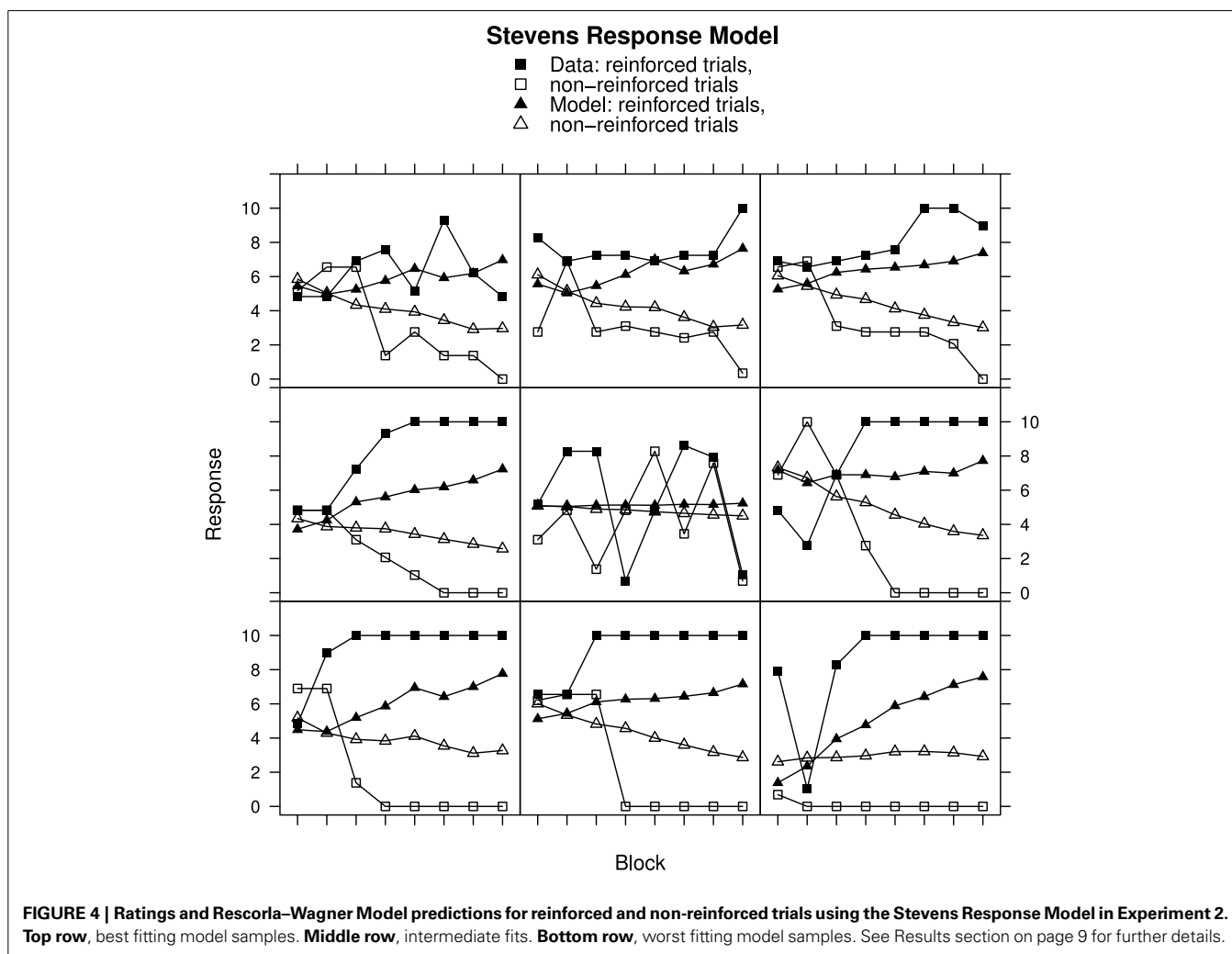
sample of participant and model data is presented for Experiment 2, only for the Stevens model, in **Figure 4**. ANOVA once again showed that the fit was related to the rate of discrimination [$F_{(14, 987)} = 3.93, p < 0.001$] and the Stevens response model also produced significantly better fits for the data of Experiment 2 than did the Fechner model [$t_{(143)} = 19.63, p < 0.001$].

COMPARISON OF RWM AND MECAM

Although the RWM captures the general trends in the data, particularly when using Stevens response mapping, consideration of the individual data in **Figures 2–4** reveals that the fitted model does not accurately reproduce the participant responses. The MECA Model was developed as an alternative application of the Rescorla–Wagner principles. The aim was to determine whether or not these shortcomings of the Rescorla–Wagner Model might be rectified by using a more elaborate model of the stimulus environment. **Figures 5, 6** show data from Experiments 1 and 2 together with best fits from the MECA Model using Stevens Response Model. In comparison with the Rescorla–Wagner Model fits (compare **Figure 3** with **5** and **Figure 4** with **6**) the MECA Model produced good fits for the participants who learn quickly the correct responses, as well as good fits for the

participants who learn more slowly. The three-way interaction of Block, Reinforcement, and Group was not significant in Experiment 1 [$F_{(30, 870)} = 1.24$] nor in Experiment 2 [$F_{(14, 987)} = 1.50$]. In addition to providing better fits overall the MECA Model also produced less stable responses from trial-to-trial and it is in that sense a better approximation to the responses produced by the participants. In many cases the trial-to-trial variation in the model predictions does not covary with the participant responses but in a number of cases there are striking correspondences (e.g., **Figure 5** middle and middle-right panels). Student's *t*-tests on the RMS error showed that the mean RMS fit was significantly better for the MECA Model than for the RWM in Experiment 1 and in Experiment 2 [$t_{(60)} = 5.68, p < 0.001$ and $t_{(143)} = 5.88, p < 0.001$, respectively]. The RMS error values are given in **Table 1**.

For Experiment 1 there was an improvement in the RMS error value for the MECA Model over the RW Model in 43 out of 61 cases—70% of participants has better fits using the MECAM, the median improvement value was 0.11. In Experiment 2 the median improvement value of MECAM over the RWM was also 0.11 with the MECAM producing smaller RMS values in 89 out of 144 participants—62% had better MECAM fits than RWM fits. **Figure 7**, gives direct comparisons of the fits of the MECAM and



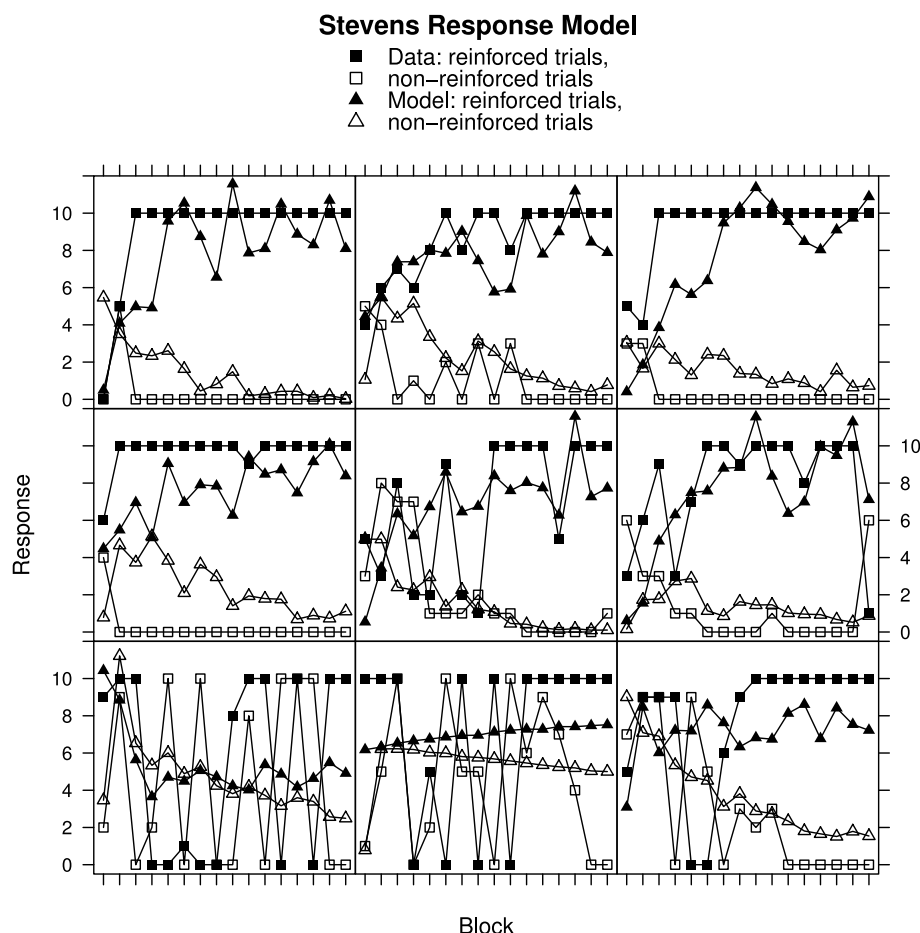


FIGURE 5 | Ratings and MECAM Model predictions for reinforced and non-reinforced trials using the Stevens Response Model in Experiment 1. Top row, best fitting model samples. Middle row, intermediate fits. Bottom row, worst fitting model samples. See Results section on page 9 for further details.

RWM to a selection of individual participants from Experiment 2. Each panel shows data from a single participant and the best fitting RWM and MECAM responses to facilitate comparison of the models. The rows in **Figure 7** are arranged to show tercile samples for participants varying according to the improvement in fit that the MECAM provided over the RWM. Participants were ranked according to the difference in RMS values between the model fits (RWM minus MECAM). A positive value on this difference score indicates that the MECAM model had a better fit than the RWM. In **Figure 7** the top row provides a sample of participants from the upper tercile of the improvement distribution (most improvement), the middle row a sample from the middle tercile, and the bottom row a sample from the lower tercile (least improvement). From left to right the RMS improvements in the top row were 0.46, 0.76, and 0.74; for the middle row they were 0.28, 0.32, and -0.01 ; and for the bottom row they were -0.26 , -0.04 , and -0.05 .

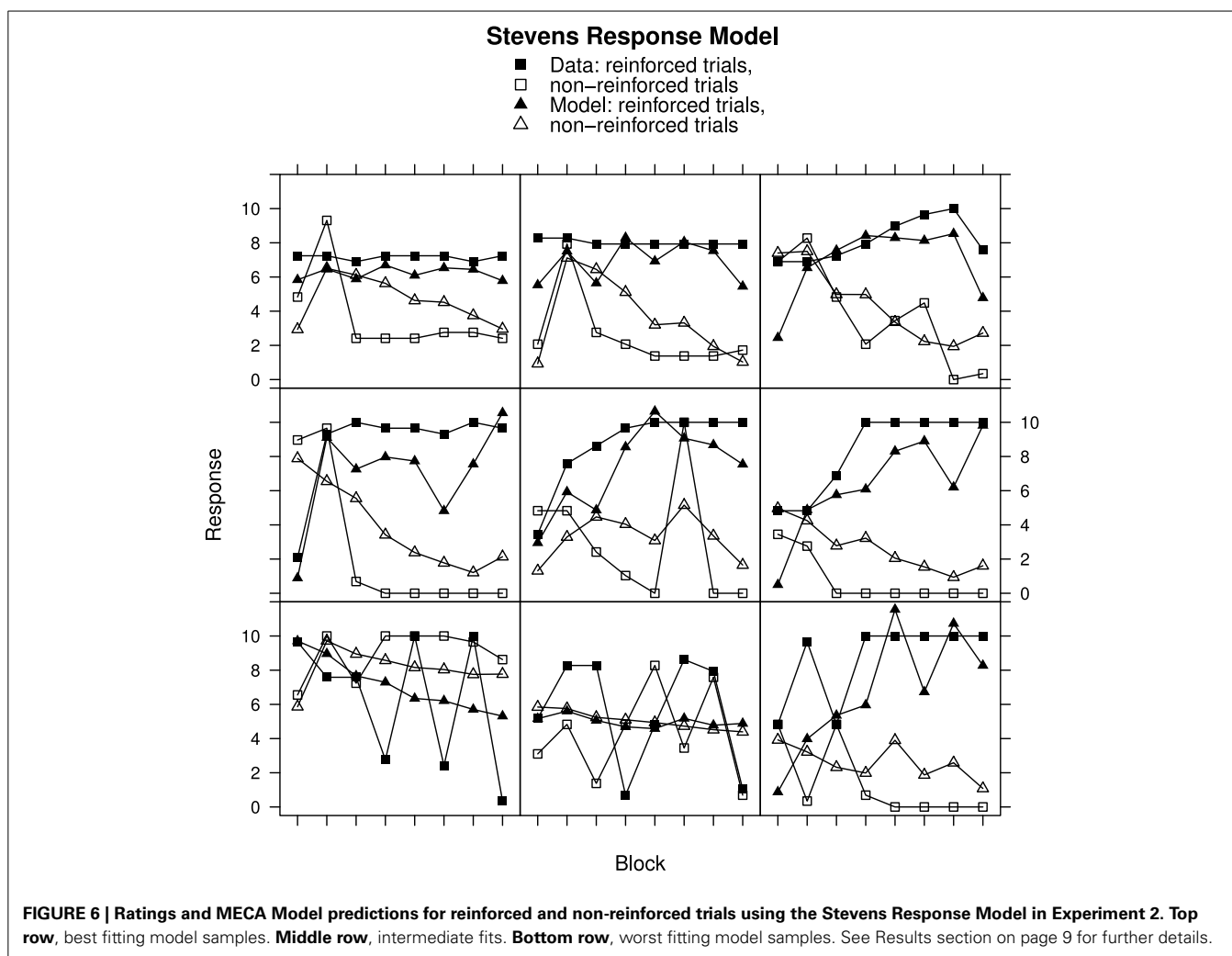
COMPARISON OF PARAMETERS FROM EXPERIMENT 1 AND EXPERIMENT 2

Multivariate Analysis of Variance (MANOVA) was used to compare Experiment 1 and Experiment 2 to assess whether or not

the fitted model parameters differed for the two datasets. The parameter values for Experiment 1 and Experiment 2 did not differ in three out of the four cases. The parameters were the same in both datasets for the MECAM Model with Fechner response mapping, and for the MECAM and RW Models with Stevens response mapping [approximate F s $F_{(9, 195)} = 1.61$, $F_{(9, 195)} = 0.84$, and $F_{(7, 197)} = 1.42$, respectively]. MANOVA did show a difference between experiments when the RWM with Fechner response mapping was considered [approximate $F_{(7, 197)} = 9.38$, $p < 0.001$]. Follow-up t -tests using Welch's correction produced significant differences only for the response mapping parameter c . Lower values of c were found in Experiment 1 than in Experiment 2 [$t_{(135)} = 3.81$, $p < 0.001$].

DISCUSSION

Two principle findings emerged. First, in this model fitting exercise, better results were obtained by using a mapping between associative strength and response strength based on Stevens' Power Law than by using a mapping based on Fechner's Law. The average model predictions using the RWM and Stevens response mapping differed from the participant data by 2.82 (Experiment 1) and 2.60 (Experiment 2) units on an 11 point response scale.



In comparison the same figures for the Fechner response mapping were 3.24 and 2.91 (see RMS values in **Table 1**). Second, although the RWM captured general trends in the individual data, the fits were poor and significant improvements were obtained using the MECAM. Using Stevens response mapping the average MECAM predictions differed by 2.65 and 2.43 units from the participant data (Experiment 1 and Experiment 2, respectively). This latter result supports the main hypothesis of this work, that participant responses on trial n are influenced by the predictive value of the memorial representations of stimuli from the previous trial. Since the sequences in these experiments were generated randomly it is argued that the predictive contributions of trial $n - 1$ memory stimuli serve to add noise to the observed responses. Because these stimuli are unlikely to remain predictive for long sequences of trials they will tend to lose their influence toward the end of the trial sequence.

The introduction began with a statement of the theoretical significance of the form of the learning curve. Although analysis of learning curves appeared to offer a route for theory advance, the promise of ruling in or out one of two major classes of learning curve (power or exponential) has not been fulfilled. Several

factors have contributed to the difficulties including using multi-component tasks and problems with averaging (e.g., Mazur and Hastie, 1978; Heathcote et al., 2000). However, even if it is not possible to clearly determine whether or not learning curves are best characterized by power functions or by exponential functions, this does not exhaust the possibilities for theoretical analysis offered by a study of learning curves. Individual learning curve data are highly variable and idiosyncratic, and we do not yet have an accurate theoretical model of this variability. Some have argued for alternatives to associative models to understand these data (e.g., Nosofsky et al., 1994; Gallistel et al., 2004). Here it is argued here that an associative model of individual learning curves is worthy of further exploration but that such a model will require a more realistic approach to characterizing the environment of the learner. The current MECAM Model is one example of such a strategy and one of its core assumptions is that “non-local” features play a part in this environment. A second core assumption in the MECAM is that an adequate description of the stimulus environment will require recognition of interactions between elemental stimuli. Both of these core assumptions were examined in the current investigation and will be discussed below.

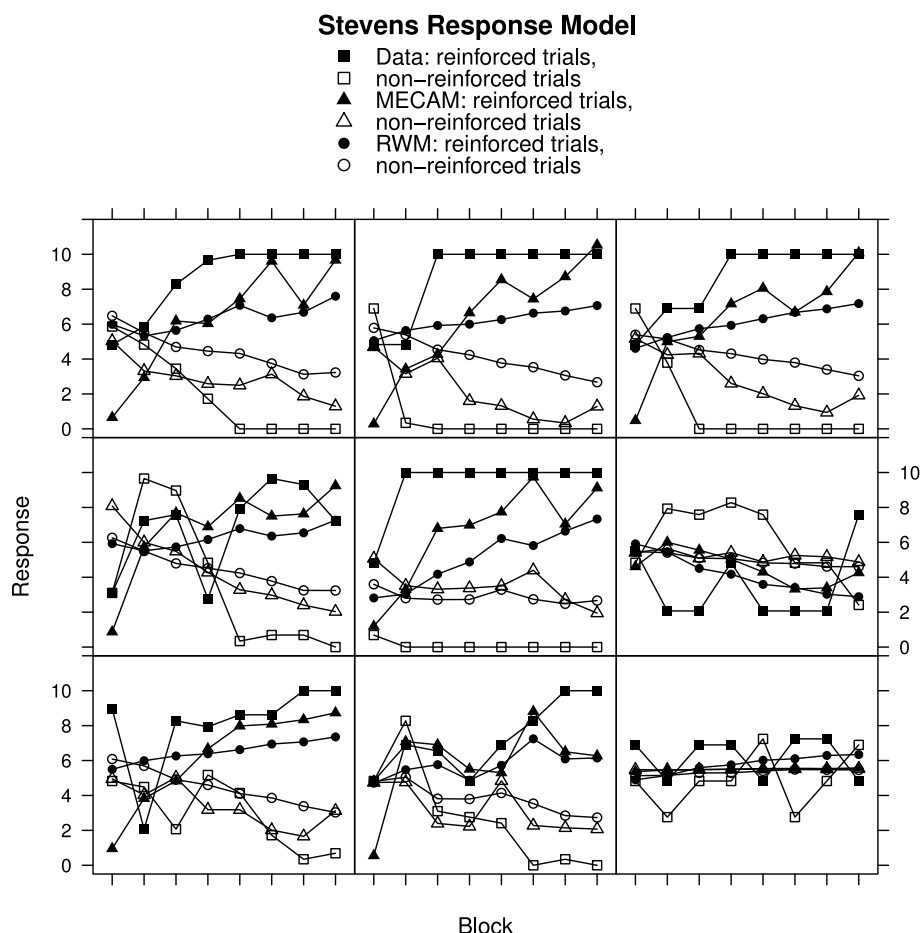


FIGURE 7 | Ratings, MECAM, and RWM predictions for reinforced and non-reinforced trials using the Stevens Response Model in Experiment 2. Top row, greatest MECAM improvement samples. Middle row, intermediate MECAM improvements. Bottom row, least MECAM improvement samples. See Results section on page 10 for further details.

This is not the first time that it has been suggested that there are non-local influences on behavior. The Perruchet effect mentioned in the introduction is another example (Perruchet et al., 2006) and there have been related suggestions in studies of sequence learning effects. Theoretical analyses of non-local influences have been explored previously in the framework of Simple Recurrent Networks (SRNs) as well as in memory buffer frameworks similar to that used in the MECAM. In the original SRN model (Elman, 1990) a three-layer neural network was used with the activations of the hidden-layer fed-back to form part of the input pattern for the current trial. This SRN was introduced as an alternative to memory buffer models of sequence learning in which the inputs of previous trials were simply repeated on the current trial. The SRN approach to sequence learning has acquired prominence but memory buffer models still appear to have some utility. Kuhn and Dienes found that a memory buffer model of learning better approximated human learning than did an SRN model (Kuhn and Dienes, 2008). Of course there are many ways in which a memory buffer model could operate and the challenge now is to develop an optimal approach. In their buffer model Kuhn and Dienes used the previous four trials and did not include

any configural cue representations. The MECA Model presented here adopted a memory buffer approach using just the previous trial and included representations of configural cues. The MECAM's implementation of both of these ideas requires further examination and development.

Use of two trials t and t_{n-1} is only an approximation to modeling the continuous time-based nature of experience. However, as argued in the introduction and as demonstrated empirically, inclusion of trial t_{n-1} results in qualitative and quantitative improvements in modeling of simple learning as compared to the same model using trial t alone. Further investigation of this approach could be carried out by using additional buffers to determine an optimal number but a more principled approach to further development of the MECAM is preferred. In MECAM the primary buffer is a focal memory store containing the events of the current trial and the secondary buffer contains a remembered version of the previous trial. The interaction buffer is a configural product of the elements in the primary and secondary buffers. MECAM currently represents time by trial-based discrete changes in the contents of these primary and secondary buffers the consequence of which is that only the current and previous trial

events can be learned about. One way to allow the possibility of events from trial t_{n-x} to play a part in MECAM's learning would be to include a model of decay and movement of the elements between the primary and secondary buffers. This would allow the buffers to contain a more heterogeneous representation of previous trials, for example the bulk of the secondary buffer could be occupied with memories of trial t_{n-1} with progressively smaller components representing trial t_{n-2} , t_{n-3} etc. Discussion of the model of buffer behavior is beyond the scope of this article but is emphasized that even a crude operationalisation of this aspect of MECAM is an improvement on modeling solely with trial t alone.

The inclusion of configural cues in the MECAM may seem questionable because there is no requirement that participants use configural cues to respond appropriately in the tasks used. Whilst some studies have shown that the weight attached to configural cues can be increased by experience (e.g., Melchers et al., 2008) there is also data to indicate that configural processes operate by default, rather than simply coming into play as necessary (e.g., Shanks et al., 1998). Thus, the simplifying assumption to exclude configural cues seems no more justified than assuming participants would only attend to the current trial. Indeed, part of the rationale for MECAM was to include aspects of the stimulus environment that are, strictly speaking, redundant for the solution of the problem at hand. The MECAM assumes that participants are responding to something when "noisy responses" occur and takes into account *measurable* components of environmental structure which previous studies have shown, in other contexts, to be important in controlling responding. It should be noted here though that the modeling exercise did not include specific comparisons of the standard RWM with and without configural cues. The primary focus was on the comparison of two models, both containing configural cues, with one model only representing the current trial (the RWM) and the other model representing the current and previous trial (the MECAM). Nevertheless we can assert that configural cues are important by looking at the optimized values of the interaction buffer weight in **Table 1**. In all fitted models this weight is substantially greater than zero and since the interaction buffer contains only configural cues this result supports their inclusion in modeling. The result for the secondary buffer is not as clear because this buffer contains a mixture of configural and elemental stimulus representations.

Thus, the MECAM principle of including an extended description of the stimulus environment, in terms of both trial history and stimulus interactions, is a reasonable way to reconcile an associative model such as the RWM with the learning curve data but the extent to which MECAM can be refined remains to be determined; MECAM as it stands is far from a complete account. The current work has provided some proof-of-concept for two major principles and future work is needed for refinement. A suggestion for a more flexible model of buffer behavior has already been mentioned and there is also a need to explore of different types of configural cue model apart from the pairwise stimulus unique-cue model used in this version of MECAM (e.g., Brandon et al., 2000).

Further developments of MECAM are justified on the basis of the statistically significant, and visible improvements, to the

modeling of individual learning curves that were obtained in the current work. However, one criticism that could be leveled at the MECAM is that the gains are small and that the model is excessively complex. Examination of the RMS error values in **Table 1** provides a metric against which to assess the size of the gains. In Experiment 1, for the Stevens response mapping, the RMS error for the MECAM was 6% less than for the RWM; in Experiment 2 the RMS error reduction was 6.5%. In these simulations MECAM was implemented with nine free parameters, a considerable increase from the RWM implied by Equation (1), which appears to include only two free parameters, α and β . It is true that the RWM is a simple model but in reality most applications of the model actually use more than these two explicitly declared free parameters. It is common practise to allow different values of α for different cue types (e.g., context cues and configural cues may have lower values) and different β values for reinforced and non-reinforced trials (e.g., Mondragon et al., 2013). If the model is intended to make quantitative rather than just qualitative predictions then inclusion of a rule to map associative strength to response strength necessarily introduces additional parameters. In the current simulations the RWM was implemented with seven free parameters so the MECAM effectively included two additional free parameters, the weights for the primary and secondary buffers sbw and ibw . It is well beyond the scope of the current paper to provide a detailed discussion of whether or not the observed gains are worth the cost of the additional parameters but two points are worthy of note. First, model complexity is not determined solely by the number of free parameters in the model (Grünwald, 2005). In fact, compared with some leading learning models (for a recent review see Wills and Pothos, 2012) the MECAM remains algorithmically simple, using the standard RWM learning rule. The aim of MECAM was to retain algorithmic simplicity and find a suitable account of the observed individual behavioral complexity in terms of the observable environmental events experienced by individual participants. Second, the model parameters were stable in two different datasets, this replication gives some assurance of the model generality.

The current test of MECAM was focussed on its ability to generate better fits to learning curve data but there are a number of other model specific predictions that would be valuable to establish the psychological validity of the concepts in MECAM. For example, because MECAM predicts an influence of the previous trial on responding to the current trial then it follows that an alternating sequence of A– and B+ trials would be learned more quickly than when A– and B+ trials were presented in a random order. Furthermore MECAM would predict considerable responding, following the alternating sequence, on the second trial of a test consisting of the sequence B– followed by T, where T is a novel test stimulus. After a randomly ordered sequence of A– and B+ trials a test consisting of B– followed by T should elicit relatively little responding. The MECAM would also give rise to the prediction that participants with better short-term memories¹ would likely have increased

¹I am grateful to a reviewer of this paper for this suggestion.

salience of events on trial t_{n-1} and thus respond differentially to a manipulation involving trial orderings. This type of test, involving model specific predictions, will ultimately be required to justify the additional complexity of the MECAM. It is clear though that we are currently in a rather uncomfortable position because models such as the RWM are unable to provide accurate quantitative approximations to the observed learning curves—a fact which is a significant shortcoming in the field of learning research.

In summary, a simple associative model such as the RWM gives only a poor approximation to individual learning curve data. It is not appropriate to rely on analysis of average curves to resolve this problem but a viable theory of learning must still be able to provide an accurate model of the individual data. The MECAM is a development of the RWM which attempts to model the complex responses that make up individual learning curves. The MECAM assumes that participant responses are subject to non-local influences (e.g., cues present on previous trial) and, because these cues are typically not predictive for long trial sequences, the influence of these cues adds noise to the observed learning curves. The improvements made by the MECA Model over the RWM suggests that this assumption is reasonable and the cue-structures defined in the current investigation are offered as an initial approximation subject to further investigation.

ACKNOWLEDGMENTS

I am grateful to Shui-I Shih and Nicola Davey for their comments on a draft of this manuscript.

REFERENCES

- Atkinson, R., and Estes, W. (1963). *Stimulus Sampling Theory*, chapter 10. Vol. 3. New York, NY: Wiley, 121–268.
- Brandon, S., Vogel, E., and Wagner, A. (2000). A componential view of configural cues in generalization and discrimination in pavlovian conditioning. *Behav. Brain Res.* 110, 67–72. doi: 10.1016/S0166-4328(99)00185-0
- Chapman, G., and Robbins, S. (1990). Cue interaction in human contingency judgement. *Mem. Cogn.* 18, 537–545. doi: 10.3758/BF03198486
- Commons Math Developers. (2013). *Apache Commons Math, Release 3.2*. Available online at: http://commons.apache.org/math/download_math.cgi
- Dickinson, A., Shanks, D., and Evenden, J. (1984). Judgment of act-outcome contingency - the role of selective attribution. *Q. J. Exp. Psychol. Sect. A Hum. Exp. Psychol.* 36, 29–50. doi: 10.1080/14640748408401502
- Elman, J. L. (1990). Finding structure in time. *Cogn. Sci.* 14, 179–211. doi: 10.1207/s15516709cog1402_1
- Estes, W. (1950). Toward a statistical theory of learning. *Psychol. Rev.* 57, 94–107. doi: 10.1037/h0058559
- Estes, W. (2002). Traps in the route to models of memory and decision. *Psychonom. Bull. Rev.* 9, 3–25. doi: 10.3758/BF03196254
- Gallistel, C., Fairhurst, S., and Balsam, P. (2004). The learning curve: implications of a quantitative analysis. *Proc. Natl. Acad. Sci. U.S.A.* 101, 13124–13131. doi: 10.1073/pnas.0404965101
- Gallistel, C., and Gibbon, J. (2002). *The Symbolic Foundations of Conditioned Behavior*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Grünwald, P. (2005). *Minimum Description Length Tutorial*. Neural Information Processing Series (MIT Press, Grünwald, Peter, Centrum voor Wiskunde en Informatica Kruislaan 413, 1098 SJ, Amsterdam, Netherlands, pdg@cw.nl), 23–79.
- Hansen, N. (2006). The CMA evolution strategy: a comparing review. *Towards New Evol. Comput.* 192, 75–102. doi: 10.1007/3-540-32494-1_4
- Hansen, N. (2012). The CMA evolution strategy. Available online at: <https://www.lri.fr/~hansen/cmaesintro.html>. Retrieved from 2 July 2013.
- Heathcote, A., Brown, S., and Mewhort, D. J. K. (2000). The power law revealed: the case for an exponential law of practice. *Psychon. Bull. Rev.* 7, 185–207. doi: 10.3758/BF03212979
- Jones, F. W., and McLaren, I. P. L. (2009). Human sequence learning under incidental and intentional conditions. *J. Exp. Psychol. Anim. Behav. Process.* 35, 538–553. doi: 10.1037/a0015661
- Köhler, W. (1925). *The Mentality of Apes*. New York, NY: Harcourt, Brace & Co.
- Krueger, L. E. (1989). Reconciling fechner and stevens: toward a unified psychophysical law. *Behav. Brain Sci.* 12, 251–267. doi: 10.1017/S0140525X0004855X
- Kuhn, G., and Dienes, Z. (2008). Learning non-local dependencies. *Cognition* 106, 184–206. doi: 10.1016/j.cognition.2007.01.003
- Lachnit, H. (1988). Convergent validation of information-processing constructs with pavlovian methodology. *J. Exp. Psychol. Hum. Percept. Perform.* 14, 143–152. doi: 10.1037/0096-1523.14.1.143
- Mazur, J. E., and Hastie, R. (1978). Learning as accumulation - re-examination of learning curve. *Psychol. Bull.* 85, 1256–1274. doi: 10.1037/0033-2909.85.6.1256
- Melchers, K., Lachnit, H., and Shanks, D. (2008). Stimulus coding in human associative learning: flexible representation of parts and wholes. *Behav. Process.* 77, 413–427. doi: 10.1016/j.beproc.2007.09.013
- Miller, R., Barnet, R., and Grahame, N. (1995). Assessment of the Rescorla–Wagner model. *Psychol. Bull.* 117, 363–386. doi: 10.1037/0033-2909.117.3.363
- Mondragon, E., Alonso, E., Fernandez, A., and Gray, J. (2013). An extension of the Rescorla and Wagner Simulator for context conditioning. *Comput. Methods Program. Biomed.* 110, 226–230. doi: 10.1016/j.cmpb.2013.01.016
- Myung, I. J., Kim, C., and Pitt, M. A. (2000). Toward an explanation of the power law artifact: insights from response surface analysis. *Mem. Cogn.* 28, 832–840. doi: 10.3758/BF03198418
- Newell, A., and Rosenbloom, P. S. (1981). *Mechanisms of Skill Acquisition and the Law of Practice*. Hillsdale, NJ: Erlbaum, 1–55.
- Nosofsky, R., Palmeri, T., and McKinley, S. (1994). Rule-plus-exception model of classification learning. *Psychol. Rev.* 101, 53–79. doi: 10.1037/0033-295X.101.1.53
- Pearce, J. (1994). Similarity and discrimination - a selective review and a connectionist model. *Psychol. Rev.* 101, 587–607. doi: 10.1037/0033-295X.101.4.587
- Perruchet, P., Cleeremans, A., and Destrebecqz, A. (2006). Dissociating the effects of automatic activation and explicit expectancy on reaction times in a simple associative learning task. *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 955–965. doi: 10.1037/0278-7393.32.5.955
- R Core Development Team (2012). R: a language and environment for statistical computing. Available online at: <http://www.R-project.org/>
- Rescorla, R. (1973). Evidence for a “unique stimulus” account of configural conditioning. *J. Comp. Physiol. Psychol.* 85, 331–338. doi: 10.1037/h0035046
- Rescorla, R. (2001). Are associative changes in acquisition and extinction negatively accelerated? *J. Exp. Psychol. Anim. Behav. Process.* 27, 307–315. doi: 10.1037/0097-7403.27.4.307
- Rescorla, R., and Wagner, A. (1972). *A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Non-reinforcement*. New York, NY: Appleton Century Crofts, 64–69.
- Rock, I. (1957). The role of repetition in associative learning. *Am. J. Psychol.* 70, 186–193. doi: 10.2307/1419320
- Roediger, and Arnold, K. M. (2012). The one-trial learning controversy and its aftermath: remembering rock (1957). *Am. J. Psychol.* 125, 127–143. doi: 10.5406/amerjpsyc.125.2.0127
- Shanks, D., Charles, D., Darby, R., and Azmi, A. (1998). Configural processes in human associative learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 1353–1378. doi: 10.1037/0278-7393.24.6.1353
- Sidman, M. (1952). A note on functional relations obtained from group data. *Psychol. Bull.* 49, 263–269. doi: 10.1037/h0063643
- Sutton, R., and Barto, A. (1981). Toward a modern theory of adaptive networks - expectation and prediction. *Psychol. Rev.* 88, 135–170. doi: 10.1037/0033-295X.88.2.135
- Thorwart, A., Glautier, S., and Lachnit, H. (2010). Convergent results in eyeblink conditioning and contingency learning in humans:

addition of a common cue does not affect feature-negative discriminations. *Biol. Psychol.* 85, 207–212. doi: 10.1016/j.biopsycho.2010.07.002

Wills, A. J., and Pothos, E. M. (2012). On the adequacy of current empirical evaluations of formal models of categorization. *Psychol. Bull.* 138, 102–125. doi: 10.1037/a0025715

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: 18 July 2013; paper pending published: 04 September 2013; accepted: 09 December 2013; published online: 26 December 2013.

Citation: Glautier S (2013) Revisiting the learning curve (once again). *Front. Psychol.* 4:982. doi: 10.3389/fpsyg.2013.00982

This article was submitted to *Personality Science and Individual Differences*, a section of the journal *Frontiers in Psychology*.

Copyright © 2013 Glautier. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX SIMULATIONS

Details of the simulations of the RWM and the MECAM are provided below. **Table A1** provides a summary of the model parameters. **Table A2** illustrates the operation of the MECAM memory buffers.

RWM simulations

The RWM simulations used the standard Rescorla–Wagner equation (Rescorla and Wagner, 1972) for updating associative strength for each stimulus present on a trial t , namely:

$$V_{i,t+1} = V_{i,t} + \alpha_i \beta (\lambda_t - \sum_{i=1}^n V_{i,t}) \quad (\text{A1})$$

In Equation (A1) the subscript t indexes the trial number and there are n stimuli present on a trial, indexed by subscript i . The update on associative strength V , is a product of α , β , and the parenthesized *error term*. λ is set to 0 for non-reinforced trials and 1 for reinforced trials. Implementation of Equation (A1) was carried out with the representation of each trial encoded to include the context, the explicit experimenter defined cues, and configural cues. For example, on an AX trial, six stimuli would be assumed to be present—C, A, X, ca, cx, and ax, where C is the experimental context (which was constant in all trials in the current simulations), A and X are the experimenter provided cues (foreground symbol and background color of the cards), and ca, cx, and ax are configural cues arising from pairwise interactions between stimulus elements C, A, and X.

The CMAES optimizing algorithm adjusted the α and β values used in Equation (A1). The α values for the configural cues were set to the average α value of the configuration elements divided by the number of elements represented (Equation A2). This scaling was chosen rather than selecting an arbitrary value on the basis that it provides a link between the salience of the elements and the configural cues, and reduces the salience of configural cues relative to element cues. Separate α values for the context and cues were selected by the optimizer, parameters α_{ctx} and α_{cue} . On reinforced trials β was set to the parameter β_{rt} and on non-reinforced trials this was scaled by multiplication with parameter β_{nrt} . The optimizer also selected the initial associative strength for all cues at the start of each simulation, parameter sv , and the parameters to control the mapping of associative strength to response strength. Optimization of sv was provided as an alternative to setting initial strength to zero or to a random value. Two models were used for response mapping, both of these use two parameters. For the mapping based on Stevens' Power Law the model response was given by Equation (A3) and for Fechner's Law the model response was given by Equation (A4). The optimizations minimized the sum of squared deviations between the model and participant responses, summing over all trials. Constraints were applied to the parameters, for RWM and MECAM simulations, as shown in **Table A1** because simulations became unstable in some cases without constraints.

$$\frac{\sum \alpha}{n^2} \quad (\text{A2})$$

$$k \left[\sum_{i=1}^n V_{i,t} \right]^a \quad (\text{A3})$$

$$k \ln \left[\sum_{i=1}^n V_{i,t} + 1 \right] + c \quad (\text{A4})$$

MECAM simulations

The MECAM simulations used a modification of Equation (A1):

$$V_{i,t+1} = V_{i,t} + \omega \alpha_i \beta \left(\lambda_n - \sum_{i=1}^n V_{i,t} \right) \quad (\text{A5})$$

In Equation (A5) an additional parameter ω is used to adjust the update to the associative strength of each cue that is present on a trial. In the MECAM the stimulus environment is assumed to consist of stimulus representations in three buffers, a *primary buffer*, a *secondary buffer*, and an *interaction buffer*. The value of ω is determined for each cue according to the buffer in which the cue is defined. The primary buffer holds representations of the stimuli present on the current trial, as specified in the implementation of the RWM described above (page 16). ω for primary buffer stimuli is set at 1. The secondary buffer holds representations of the stimuli that were present on the previous trial. ω for secondary buffer stimuli was set at the value adjusted by the optimizer, the parameter secondary buffer weight (*sbw*) is shown in **Table A1**. The primary and secondary buffers both hold elemental representations of stimuli and pairwise configural cue representations of the elemental cues as shown in **Table A2**. In **Table A2** stimuli from the previous trial are subscripted $t-1$ and configural cues are in lower case. For example A_{t-1} and $a_{t-1}b_{t-1}$ represent memories from the previous trial. A_{t-1} is the memory of element A and $a_{t-1}b_{t-1}$ is the memory of the configural cue for the co-occurrence of A and B. Note that the configural cues in the secondary buffer are remembered versions of those were created from pairwise combinations of the *stimuli* that were presented on the previous trial, they have not been created *de novo*. The interaction buffer, on the other hand, holds only configural cue representations. These representations are created from combinations of the element cues present in the primary and secondary buffers. For example aa_{t-1} is the configural cue for the co-occurrence of element A on the current trial and the memory of A from the previous trial. Note that no new configural representations that appear in the interaction buffer are created entirely from remembered elements. Thus, in the tabulated example on trial 2, we obtain configural cues such as aa_{t-1} because these consist of a current and a remembered element. However, we do not get cues such as $a_{t-1}o_{t-1}$ because this would involve two remembered elements. Configural cues involving two remembered elements only occur in the secondary buffer as remembered versions of configurations from the previous trial (e.g., $a_{t-1}b_{t-1}$).

ω for interaction buffer stimuli was set at the value adjusted by the optimizer, the parameter interaction buffer weight

(*ibw*) is shown in **Table A1**. For further illustration of how the stimulus environment is represent in the MECAM refer to **Table A2** which shows the state of the MECA buffers state in a series of three successive trials. Cues A and B

are present on the first trial, and the outcome occurs; cues A, B, and C are present on the second, non-reinforced, trial; cues B and C are present on the third, non-reinforced trial.

Table A1 | Parameters used, and optimization boundary values, in RWM and MECAM simulations.

Parameter	Description	Lower bound	Upper bound	MECAM	RWM
α_{ctx}	α value for context	0.01	0.25	✓	✓
α_{cue}	α value for cues	0.01	0.25	✓	✓
β_{rt}	β for reinforced trials	0.01	0.25	✓	✓
β_{nrt}	β scaling for non-reinforced trials	0.01	0.25	✓	✓
k	Associative strength to response mapping	0.01	10	✓	✓
a	Associative strength to response mapping	0.01	3	✓	✓
c	Associative strength to response mapping	0.01	5	✓	✓
sbw	Weight for secondary buffer stimuli	0	1	✓	×
ibw	Weight for interaction buffer stimuli	0	1	✓	×
sv	Initial associative strength	0.01	0.25	✓	✓

Table A2 | State of buffers on three successive trials; AB+, ABC−, and BC−.

	Trial 1		Trial 2		Trial 3	
	Element cues	Configural cues	Element cues	Configural cues	Element cues	Configural cues
Primary buffer	A	ab	A	ab	B	bc
	B		B	ac bc	C	
Secondary buffer			A_{t-1}	$a_{t-1}b_{t-1}$	A_{t-1}	$a_{t-1}b_{t-1}$
			B_{t-1}		B_{t-1}	$a_{t-1}c_{t-1}$
			O_{t-1}		C_{t-1}	$b_{t-1}c_{t-1}$
Interaction buffer				aa_{t-1}		ba_{t-1}
				ab_{t-1}		bb_{t-1}
				ao_{t-1}		bb_{t-1}
				ba_{t-1}		ca_{t-1}
				bb_{t-1}		cb_{t-1}
				bo_{t-1}		cc_{t-1}
				ca_{t-1} cb_{t-1} co_{t-1}		

Element cue O_{t-1} on trial 2 is the memory of the outcome that occurred on trial 1. See text above (page 16).



Accounting for individual differences in human associative learning

Nicola C. Byrom *

Department of Experimental Psychology, University of Oxford, Oxford, UK

Edited by:

Rachel M. Msetfi, University of Limerick, Ireland

Reviewed by:

Irina Baetu, University of Adelaide, Australia

Fernando Blanco, University of Deusto, Spain

***Correspondence:**

Nicola C. Byrom, Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford, OX1 3UD, UK
e-mail: nicola.byrom@psy.ox.ac.uk

Associative learning has provided fundamental insights to understanding psychopathology. However, psychopathology occurs along a continuum and as such, identification of disruptions in processes of associative learning associated with aspects of psychopathology illustrates a general flexibility in human associative learning. A handful of studies have looked specifically at individual differences in human associative learning, but while much work has concentrated on accounting for flexibility in learning caused by external factors, there has been limited work considering how to model the influence of dispositional factors. This review looks at the range of individual differences in human associative learning that have been explored and the attempts to account for, and model, this flexibility. To fully understand human associative learning, further research needs to attend to the causes of variation in human learning.

Keywords: individual differences, association learning, psychopathology, depression, perceptual processing, attention

Research into individual differences across the human population has contributed to better understanding of everything from academic achievement to crime and delinquency, from income and poverty to health (Lubinski, 2000). Studying individual difference in human learning has contributed to our understanding of the mechanisms underlying psychopathology, particularly because learning identifies a process and therefore a mechanism by which individuals might differ. As traits of psychopathology vary across the population, our understanding of the association between psychopathology and disruptions in processes of association learning, may tell us a considerable amount about the nature and extent of variation in human associative learning. While evidence that people do not all learn the same way has been used to help us understand aspects of psychopathology, this exploration of flexibility in human learning needs to be integrated into our general understanding of the mechanisms of learning so that models can accommodate the factors that produce variance in learning. To examine individual difference in all aspects of associative learning would be too board a scope for this review. To provide focus to analysis of individual differences, this paper addresses variation in learning about combinations of stimuli. Specifically, this review presents a range of examples demonstrating individual differences in the selectivity of learning and tendency to learn about individual elements or configurations and considers how models of associative learning can accommodate this variation.

Associative learning theorists understand behavior by studying how associations between stimulus representations are acquired and used. Much of this work considers which factors influence learning and how these factors exert influence. The basic model of error prediction learning, shown in Equation (1) provides us with an indication of several factors that might influence

learning. This equation was described by Rescorla and Wagner (1972).

$$\Delta V_n = \alpha_n \times \beta \times (\lambda - \Sigma V) \quad (1)$$

This equation describes change in associative strength of a stimulus (ΔV_n) as a function of prediction error; that is, the discrepancy between the outcome expected following the given stimulus and the outcome that actually occurs. Prediction error is given by the difference between the asymptote of learning (λ), the total associative strength that the unconditioned stimulus (US) can support, and the current associative strength of all stimuli present on the trial. Prediction error is multiplied by the salience or intensity of that stimulus (α) and the US (β).

To provide some examples, research has considered how stimulus representations might differ on the basis of intensity and/or salience (i.e., α) and how such differences influence learning (Perkins, 1953; Logan, 1954; Redhead and Pearce, 1995). There has also been much consideration how attention shifts between different stimuli to influence learning (Mackintosh, 1975; Pearce and Hall, 1980; Le Pelley and McLaren, 2004; de Wit and Dickinson, 2009; Harris and Livesey, 2010; Lubow, 2010; McLaren et al., 2010) and how previous experiences can modify the acquisition of new stimulus representations and their associations (Kamin, 1968; Seligman, 1972; Lubow et al., 1976). This review considers whether these factors are constant across the population, or whether the influence these factors have upon learning varies between individuals. As much of the research testing individual difference in human associative learning relates to psychopathology, this review relies heavily upon illustrations from clinically focused research. The studies discussed here demonstrate substantive individual differences in central aspects

of associative learning. The review concludes with a brief look at how models of associative learning can account for the observed individual differences.

STIMULUS SALIENCE AND SELECTIVE PREDICTION ERROR

Individual difference in terms of what is perceived to be salient may influence the acquisition of associations. The strength with which associative learning occurs tends to increase with stimulus salience (Kamin and Brimer, 1963; Kamin and Schaub, 1963). For instance, if two stimuli of different salience co-occur, stronger stimulus-outcome associations should be acquired for the more salient stimulus (Kamin, 1969; Mackintosh, 1971). Similarly, the strength of associative learning has been related to the strength of the unconditioned stimulus (US; Pavlov, 1927). For example, conditioned responding to shock in rabbits was observed to be directly related to the intensity of the shock, the US (Smith, 1968). To summarize with a relative simple example; a child playing with a toy may learn that pressing a lever on the toy causes a light to turn on. The perceived intensity or salience of the light (the outcome of the behavior) will influence the associative strength that can be supported. The perceived kinaesthetic experience of handling the lever (the intensity or salience of the stimulus) will also influence the strength of learning. Variation in terms of what individuals find salient should have a substantial impact upon the acquisition of associations and may, for example, contribute to differences in associative learning in depression and anxiety.

Depression is associated with a tendency to find certain negative information salient (Matthews et al., 1995; Mogg et al., 1995; Bradley et al., 1997; Rusting, 1998, 1999; Gotlib et al., 2004; Chan et al., 2007; Phillips et al., 2010). This should have an impact upon the associations learned. Learning with salient stimuli will occur at the expense of less salient stimuli (Mackintosh, 1971). As such, if individuals with, or at risk of developing, depression find negative information more salient, they should be more likely to learn associations with negative stimuli as opposed to positive or neutral stimuli.

When learning occurs, the strength of learning that can be supported is dependent upon the strength of the outcome, or unconditioned stimulus (i.e., Rescorla and Wagner, 1972). As in the example of the child playing with a toy, the association formed between pressing the lever and the occurrence of the outcome, the light turning on, may be influenced by how bright the light is, but also by how much lights interest the child. If the child's interest in lights is minimal, we may suggest that the perceived salience of the light, for that child, is limited. In which case, the strength of learning that the light may support should be limited. Applying this logic to individuals with depression, we may consider that the tendency to find negative information more salient may increase the perceived salience of negative outcomes. This should facilitate negative outcomes to support stronger acquisition of associative strength. This may, for instance, result in individuals with depression forming stronger associations between stimuli and negative outcomes, facilitating subsequent negative expectations. As such, the tendency to find negative information more salient may perpetuate expectation of unfavorable outcomes.

Aspects of fear conditioning associated with anxiety may be characterized by similar differences in stimulus perception.

Enhanced fear conditioning is suggested to play an important role in anxiety disorders (Craske et al., 2006; Mineka and Zinbarg, 2006). Variation in the perceived intensity of a fearful stimulus is one factor that may account for differences in the ease with which fear associations are learned or maintained (Otto et al., 2007). For instance, participants' ratings of the aversiveness of a US have been observed to correlate significantly with ability to learn to dissociate a stimulus (CS) paired with the aversive US from a CS not paired with the US (Joos et al., 2013).

The salience of a stimulus, however, is not fixed. Stimulus salience may change with experience (Mackintosh, 1975; Pearce and Hall, 1980; Le Pelley and McLaren, 2004; Le Pelley et al., 2010; Pearce and Mackintosh, 2010). Learning arguably occurs more readily with stimuli that are good predictors of an outcome while stimuli that are poor predictors of an outcome lose ability to capture attention (Mackintosh, 1975). Research into mechanisms of associative learning which may underpin symptoms of schizophrenia provide examples of individual difference in changes of stimulus salience over training.

Normally, repeated presentation of a stimulus uncorrelated with an outcome retards subsequent ability to learn about that stimulus (Lubow and Moore, 1959; Lubow et al., 1976; Lubow, 2010). This effect has been termed latent inhibition. One explanation for this effect is that repeated exposure to the stimulus reduces the salience of the stimulus, specifically affecting the attentional associability of the stimulus such that the weight of attention afforded to the stimulus is reduced relative to other stimuli (Mackintosh, 1975; Le Pelley, 2004). As attentional associability will determine which stimulus should have access to learning and which should not (Mackintosh, 1975; Le Pelley, 2004), a reduction in attentional associability should reduce learning.

This process of latent inhibition is disrupted in schizophrenia and this disruption is associated with negative symptoms of schizophrenia in particular (Lubow et al., 1976; Baruch et al., 1988; Lubow, 1989, 2010; Gray et al., 1995; Vaitl and Lipp, 1997; Raschle et al., 2001; Gal et al., 2009). In contrast, persistent latent inhibition, that is, abnormally strong processes of latent inhibition, have been observed in animal models of positive symptoms of schizophrenia (Weiner, 2003). In contrast to the wealth of research exploring disrupted latent inhibition in human participants, there has been limited work exploring the effect of persistent latent inhibition in the human population. Further research would be beneficial to help understand whether mechanisms of associative learning have relevance for understanding positive symptoms of schizophrenia. The disruption of latent inhibition associated with negative symptoms of schizophrenia, however, suggests that negative symptoms are associated with a deficit in selective attention (Solomon et al., 1981; Weiner et al., 1981, 1984) or selective prediction error (Haselgrove and Evans, 2010).

Haselgrove and Evans (2010) have used the blocking effect to further explore the relationship between selective prediction error and schizophrenia. Blocking is thought to be dependent upon selective prediction error. Kamin (1968, 1969) observed that prior training with one stimulus interferes with the acquisition of associative strength with a second stimulus when

presented in compound with the initial stimulus. For instance if a stimulus is paired with an outcome (**A+**) prior to pairing two stimuli with the same outcome (**AX+**), the associative strength acquired by the second stimulus (**X**) is reduced compared to a control. Selective prediction error is argued to underlie this effect (Haselgrove and Evans, 2010). The Rescorla and Wagner model of learning, described above in Equation 1, uses a summed error term and predicts that change in the associative strength of a stimulus depends upon the difference between the asymptote of learning supported by the outcome and the associative strength of all stimuli present on a trial. For example, on the **AX** compound trial, **A** already predicts the outcome and therefore the prediction error is minimal, preventing learning with **X**. A failure to show blocking may suggest that prediction error is non-selective, that is, on the **AX** compound trial the associative strength acquired by **A** is not considered when learning with **X** and hence learning with **X** can occur (Haselgrove and Evans, 2010).

Blocking is disrupted in schizophrenia; this disruption is associated with the negative and depressive symptoms of schizophrenia in particular (Bender et al., 2001; Moran et al., 2008). This effect has been replicated in a non-clinical sample; individuals with high levels of introverted anhedonia, the negative symptom dimension of schizotypy, show disrupted blocking (Haselgrove and Evans, 2010). Observation of this effect with the dimension of schizotypy suggests that across the general population individuals differ considerably in the selectivity of their learning.

ATTENDING TO THE CUES OR THE CONTEXT

In an associative learning paradigm participants are usually given the opportunity to learn that a stimulus predicts an outcome. Specificity is a fundamental component of this learning. That is to say, learning that a specific stimulus, and not the context in which that stimulus is presented or any other presented stimuli, predicts that the outcome of interest. To return to the original example of a child playing with a toy; pressing the lever causes a light to turn on. In playing with the toy the child has the opportunity to experience the contingency of lever pressing and the occurrence of the light. Experience of this contingency should facilitate learning that a specific cue, pressing the lever, rather than any other cue in the environment, causes the light to turn on.

One explanation for the relationship between anxiety and high levels of conditioned fear may be a deficit in specificity of learning (Baas et al., 2008; Baas, 2013). For example, if an aversive stimulus (**US**) is presented in a given context, it is likely that that context will be associated with that **US** and thus the context may begin to evoke a fear response. If the aversive **US** is always, and only, presented immediately after a specific cue, the cue can be used to predict the aversive **US**. Learning the specific association between the cue and **US** should reduce the association between the context and the aversive **US**, as the context is a less reliable predictor of the **US** than the cue. Failure to learn this specific association may be expected to result in continued general fear of the context. Studies have identified a relationship between learning a specific association between a threat cue and an aversive **US** and a reduction in general fear to the context in which the cue and aversive **US** are presented. Specifically, Baas (2013) observed that participants who failed to acquire an awareness of the relationship

between a specific threat cue and the aversive **US** rated the context in which that stimulus was presented as fearful. Fear ratings for the context were reduced in participants who acquired the specific **CS–US** association (Baas, 2013). However, this study did not observe trait anxiety to be associated with failure to learn the specific association, though it is possible that such failure to learn the specific association may relate to characteristics of anxiety such as attentional control (Derryberry and Reed, 2002; Baas, 2013).

Individual differences in specificity of learning about cues in a context may be seen in human contingency learning. Learning contingencies allows people to make judgments about how accurately events and actions predict subsequent outcomes, allowing behavior to be guided by experience (Baker et al., 2001). While positive contingencies, where the probability of an outcome occurring increases in the presence of a stimulus, are regularly encountered, we also experience zero contingencies where the outcome is no more likely to occur in the presence than the absence of a stimulus. Accuracy in identifying zero contingencies is quite poor, especially when people are asked to consider whether their actions cause an outcome (Alloy and Abramson, 1979; Baker et al., 2010). Alloy and Abramson (1979) gave participants the opportunity to press a light switch and asked them to estimate how much control they had of a light turning on and off. There was a zero contingency relationship between pressing the light switch and the light coming on; the light was just as likely to turn on during trials where the light switch was not pressed as it was during trials where the light switch was pressed. Alloy and Abramson (1979) found that depressed participants accurately judged that they had no control of the light. Non-depressed participants incorrectly estimated that they had control of the light. This effect was termed depressive realism (Alloy and Abramson, 1979). More recent experiments exploring this effect suggest that depressed participants may be less sensitive to context information (Msetfi et al., 2005). In re-running the original Alloy and Abramson experiment, Msetfi et al. (2005) varied two factors; the outcome density and the inter-trial interval (ITI). Through this experimental design the opportunity to press a light switch and the occurrence, or non-occurrence of the light is split into trials. The ITI, that is the length of time between each trial, can be varied. Outcome density, that is the proportion of trials on which outcome occurs, can also be varied while maintaining a zero contingency. For example, in a low outcome density condition the light might turn on during 25% of the trials where the light switch is pressed and 25% of the trials where the light switch is not pressed. In a high outcome density condition the light might turn on during 75% of the trials where the light switch is pressed and 75% of the trials where the light switch is not pressed.

Varying the ITI and outcome density, Msetfi et al. (2005) observed that the original Depressive Realism effect was only present when the ITI was long and the outcome density was high. At shorter ITIs or when the outcome density was lower, non-depressed participants did not overestimate their control of the light. Interestingly, in a long ITI design participants get more exposure to the context in the absence of the outcome; that is, more experience of no-action (participants cannot press the light switch during the ITI) and no-outcome (the light never turns on during the ITI). Increasing exposure

to the no-action—no-outcome contingency increases the contingency between action and outcome. As such, under these conditions, non-depressed participants were actually correct in estimating that they had control over the outcome. The failure of the depressed participants to increase their judgments of control suggests that depressed individuals were insensitive to the no-action—no-outcome information presented during the ITI (Msetfi et al., 2005; Baker et al., 2010).

LEARNING ABOUT CONSTITUENT ELEMENTS OR CONFIGURATIONS

While linear learning refers to the acquisition and use of associations between separate stimuli and outcomes, non-linear learning refers to learning about compound stimuli as distinct configurations associated with different outcomes from those associated with the compound's constituent stimuli. The Rescorla and Wagner (1972) model of elemental learning assumes that each stimulus is processed separately so that it develops its own associative link with the outcome. When learning about, and responding to, compound stimuli, this elemental approach continues to assume that each individual stimulus develops its own associative link with the outcome. As such, the model predicts that the associative strength of a compound stimulus (i.e., V_{ab}) is the algebraic sum of the associative strength of each of the stimuli presented (i.e., $V_{ab} = V_a + V_b$). While elemental theory naturally accounts for situations where the outcome following the co-occurrence of stimuli is greater than that following the separate constituent stimuli, non-linear discrimination tasks require the opposite relationship to be learnt; where the outcome following the co-occurrence of stimuli is less than, or opposite to, that following the separate constituent stimuli. Humans and animals can successfully solve non-linear discriminations, such as negative patterning (Redhead and Pearce, 1995; Shanks and Darby, 1998; Deisig et al., 2001; Myers et al., 2001; Pearce and George, 2002; Grand and Honey, 2008; Harris et al., 2008). The traditional Rescorla and Wagner (1972) instantiation of the elemental model cannot account for this. By contrast, configural theory (Pearce, 1987) can account for non-linear discrimination learning. Configural theory (Pearce, 1987) assumes that associations form between outcomes and unitary or configural representations of the pattern of stimuli present on a given trial. As such the configuration present on a compound trial (AB) should enter into an association with an outcome independent from the associative links formed between the constituent stimuli and outcomes. Though these two classes of model make contrasting predictions about how the relationship between constituent stimuli and configurations should be learnt, there is considerable support for both models, reflecting substantial variability in non-linear learning. (Melchers et al., 2008).

It has been suggested that the perceptual properties of stimuli influence whether learning will occur with separate constituent stimuli (elemental) or configurations (configural; Lachnit, 1988; Kehoe et al., 1994; Rescorla and Coldwell, 1995; Myers et al., 2001). Others have argued that these are two separate types of learning, mediated by different neural substrates (Sutherland and Rudy, 1989; Fanselow, 1999).

Several studies have looked at whether individuals differ in their tendency to learn about constituent elements or configurations. The negative patterning discrimination ($A+$, $B+$, $AB-$) provides a useful test of configural learning, as solving the discrimination requires participants to learn that the compound stimulus is associated with a different outcome to each of its constituent stimuli. Shanks and Darby (1998) provided a suggestion that human ability to learn non-linear discriminations, such as negative patterning, might be dependent upon rule use. Shanks and Darby (1998) demonstrated that ability to learn a negative patterning discrimination was associated with later use of rule as opposed to feature based generalization (Shanks and Darby, 1998). Rule-based generalization depends on the abstraction of and generalization from a rule. Feature-based generalization depends upon the surface similarity between separate stimuli and compounds. As such, it is assumed that rule-based generalization is more complex and might require greater understanding of the discrimination (Shanks and Darby, 1998) or more working memory capacity (Wills et al., 2011).

In the Shanks and Darby (1998) experiment participants were trained on a negative patterning discrimination (i.e., $A+$, $B+$, $AB-$) intermixed with trials where separate stimuli were paired with the outcome (i.e., $I+$, $J+$) before being asked for a prediction of the outcome following the co-occurrence of the separately trained stimuli (i.e., $IJ?$). Some participants expected the outcome to occur following the IJ compound, showing feature based generalization. Others demonstrated application of a negative patterning rule, expecting no outcome to occur following the IJ compound. Rule-based generalization was associated with strong initial discrimination learning (Shanks and Darby, 1998). Wills et al. (2011) found that individuals who completed a concurrent task while learning the same initial discrimination were more likely to show feature-based generalization (Wills et al., 2011). As such, it may be that greater working memory capacity is associated with stronger non-linear discrimination learning and rule-based generalization. Recently, Baker (2013) observed performance on the Raven's Progressive Matrices (Raven, 2000) to be associated with ability to learn a negative patterning discrimination. Ravens Matrices are designed to assess reasoning ability, and as such these results may provide support for the suggestion that rule use facilitates non-linear discrimination learning, such as negative patterning.

Negative patterning, however, essentially requires learning about a configuration (that is the co-occurrence of stimuli) independently from learning about the constituent stimuli. We may thus expect that a tendency to perceive or process groups of stimuli as a unitary configuration, and not simply a cluster of co-occurring stimuli, may influence performance. Similar task requirements have been explored in other areas of psychology. For instance, face recognition is a task thought to be reliant upon configural processing (Diamond and Carey, 1986; Tanaka and Farah, 1993; Leder and Bruce, 2000; Maurer et al., 2002). Strong face recognition has been associated with a general advantage in global processing (Macrae and Lewis, 2002; Perfect, 2003); that is, tendency to process global information prior to, or with a higher priority than, the specific elements composing the global stimuli (Navon, 1977).

As individuals differ in their tendency to show a global or local processing advantage (Navon, 1977), it is possible that such variation relates to, or influences, capacity to learn about combinations of stimuli and thus learn a non-linear discrimination. Using a similar discrimination task to that developed by Shanks and Darby (1998), Byrom and Murphy (under review) found global processing to be associated with stronger ability to learn a non-linear discrimination; specifically, individuals showing a global processing advantage were better able to discriminate **BC** from **ABC** in a modified negative patterning task (**A+**, **BC+**, **ABC-**).

MODELING INDIVIDUAL DIFFERENCE IN HUMAN ASSOCIATIVE LEARNING

Use of associative learning in exploration of clinical phenomena has advanced our understanding of mechanisms underlying cognitive aspects of psychopathology. As psychopathology is widely accepted to occur along a continuum, the clinical examples presented here contribute to the demonstration of substantial individual differences in processes of associative learning. For instance, though schizophrenia is a serious mental health problem occurring with a prevalence of around 0.4% (Saha et al., 2005; McGrath et al., 2008), schizotypy, a dimension reflecting traits of schizophrenia, varies across the population (Mason et al., 2005; Mason and Claridge, 2006). Schizotypy is, like schizophrenia, associated with disruptions in latent inhibition and blocking (Moran et al., 2003; Haselgrove and Evans, 2010) as well as impaired conditional task performance (Haddon et al., 2011) and impaired visual context processing (Uhlhaas et al., 2004; Uhlhaas and Silverstein, 2005).

Models of learning may need to account for this flexibility. If the mechanisms of associative learning vary across the population, focusing on the average performance of a sample when developing models of learning may result in models which fail to accurately capture the populations' performance. Over the years there have been many modifications to simple models of learning. While these modifications allow the models to capture a broader range of experimental findings, many different factors vary during learning and as such it may not be reasonable to search for a single modification to capture all variability in learning. It is unlikely that all factors contributing to individual differences in human associative learning could be captured by one parameter.

Individual differences in many of the factors discussed above can be captured by varying the parameters present in the Rescorla and Wagner (1972) model of learning, described in Equation (1). For instance, if individuals differ in their perception of the salience of the CS or US, modifying α or β could provide flexibility to account for this variation. Varying λ allows accommodation of individual difference in the rate of learning. Further, it may be possible to account for individual difference in selectivity of learning, as observed by Haselgrove and Evans (2010) by varying the extent to which a separable (i.e., Bush and Mosteller, 1951) as opposed to a summed (i.e., Rescorla and Wagner, 1972) error term is adopted. Variation between and integration of summed and separable error terms and the relation to processes of attention have been discussed at length elsewhere (Le Pelley, 2004; Pearce and Mackintosh, 2010).

Individual difference in ability to solve a negative patterning discrimination, however, is one example of variation that cannot

be accounted for by varying existing parameters in this model. At least three different approaches have been proposed to allow for flexibility between elemental and configural models of learning; the replacement parameter, the discriminability parameter and the sampling capacity parameter. Each is discussed below.

The Replaced Elements Model (REM; Brandon et al., 2000; Wagner, 2003), conceives of stimuli as represented by multiple features or elements. The model focuses on elements that stimuli share in common and how these elements interact with elements unique to a given stimulus. In the representation of a compound there are assumed to be context independent elements which are activated whenever the stimulus is presented and context dependent elements which are activated or inhibited depending on the combinations of stimuli presented (Brandon et al., 2000). For instance, when stimulus **A** is presented alone, representations of the elements **A₁** and **A₂** may be activated. When stimulus **A** is presented in combination with stimulus **B**, the element **A₂** may be replaced by a new element, **A₃**. The model adopts the stipulation that a compound should have no more capacity to elicit associative strength than any of its constituent elements. As such, in adding and inhibiting elements, the change made to the elements represented is qualitative, with the elements represented being changed, rather than a quantitative.

The replacement parameter r allows flexibility in the proportion of context dependent elements replaced when stimuli are presented in compound (Wagner, 2003). When r is 0 no replacement occurs and as such strong generalization of associative strength between stimuli and compounds is predicted. When r is 1 there is considerable replacement of elements and as such the generalization predicted to occur between compounds and constituent stimuli should be reduced. With maximal replacement of elements, the representation of the compound should be distinct from the representation of the separate stimuli.

The discriminability parameter, suggested by Kinder and Lachnit (2003) introduces flexibility into a model of configural learning (Pearce, 1987), allowing the perceived similarity between stimuli and compounds to be altered. This also affects the extent to which generalization of associative strength is predicted. The modification assumes that as it becomes harder to identify constituent stimuli within compounds, the discriminability parameter will decrease, reducing the prediction of perceived similarity between compounds and constituent stimuli (Kinder and Lachnit, 2003).

While the replacement and discriminability parameters were developed to account for the influence of external factors such as stimulus modality (Kehoe et al., 1994), the sampling capacity parameter was developed to account for individual difference observed in human associative learning. Sampling capacity here refers to the number of stimulus features that can be sampled on a given trial. To learn about and respond to the co-occurrence of stimuli as a distinct combination, Byrom and Murphy (under review) suggest that features of each of the co-occurring stimuli must be sampled simultaneously, such that in any given sample a configuration is represented. Variation in sampling capacity should produce variation in the extent to which the features of co-occurring stimuli can be sampled and as such result in variation in ability to represent and learn about the distinct combinations of stimuli, required to

learn a non-linear discrimination. Byrom and Murphy (under review) suggest that the impact of varying sampling capacity may be modeled by incorporating a parameter, f , into a modification of Pearce's configural model of associative learning. This parameter reflects the probability of encoding a configuration, calculated from sampling capacity. For a fixed sample size, the probability of sampling a configuration of a set number of features increases as sampling capacity increases.

Pearce's (1987, 1994) configural model of learning stipulates that associative strength is acquired by the configurations of stimuli presented (i.e., **A**, **BC**, and **ABC**). However, if individuals have limited sampling capacity, they may learn about the separate stimuli and not the configurations. To allow for this flexibility, Byrom and Murphy (under review) suggest modifying Pearce's (1987, 1994) configural model of learning such that two sets of nodes may be activated by input; separate stimuli (i.e., **A**, **B**, and **C**) and presented configurations (i.e., **A**, **BC**, and **ABC**). Both sets of nodes can form associations with an unconditioned stimulus and generalization can occur between all nodes. This can be achieved by modifying Pearce's (1987, 1994) configural model of associative learning such that changes in the excitatory strength of the separate stimuli and the presented configurations is moderated by the parameter, f , reflecting sampling capacity. At a high sampling capacity, the excitatory strength of presented configurations changes across learning trials. At a low sampling capacity, the excitatory strength of the separate stimuli changes across learning trials. As Pearce's (1987, 1994) configural model is highly dependent on the influence of generalization, modification of this model must consider generalization, which, like change in excitatory strength, comes to be moderated by the parameter, f . As such, at a high sampling capacity, generalization of associative strength to separate stimuli and between

presented configurations will be high, while at a low sampling capacity generalization of associative strength to separate stimuli and between presented configurations will be low, but generalization from separate stimuli to presented configurations will be high.

The extent to which parameters can be used to make predictions about learning and behavior in novel situations is dependent upon ability to specify the parameter a-priori. Each of these modifications faces challenges in specifying parameters a-priori. The replacement parameter depends on the proportion of elements replaced when a stimulus is presented in compound. The discriminability parameter depends on ability to discriminate between stimuli. It is possible that either of these parameters may be calculated for a specific stimulus set, but many factors would be expected to interact to influence "element replacement" and stimulus discriminability, limiting the extent to which these parameters can, in general, be specified a-priori. Sampling capacity may be calculated from individual difference in tendency to show local or global processing. To do this it is necessary to have relevant data, such as participants' performance on a task such as the Navon task (Navon, 1977).

CONCLUSIONS

Individual difference in human associative learning appears to have substantial impact upon learning. To accurately understand and model human associative learning, this flexibility needs to be accounted for in terms of specific parameters. Though the introduction of new parameters to increase the flexibility of models of learning has limitations, exploring the extent to which variation in specific parameters can account for specific individual difference in human associative learning, should enhance understanding of mechanism of associative learning.

REFERENCES

- Alloy, L. B., and Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: sadder but wiser. *J. Exp. Psychol. Gen.* 108, 441–485. doi: 10.1037/0096-3445.108.4.441
- Baas, J. M. (2013). Individual differences in predicting aversive events and modulating contextual anxiety in a context and cue conditioning paradigm. *Biol. Psychiatry* 92, 17–25. doi: 10.1016/j.biopsycho.2012.02.001
- Baas, J. M., van Ooijen, L., Goudriaan, A., and Kenemans, J. L. (2008). Failure to condition to a cue is associated with sustained contextual fear. *Acta Psychol.* 127, 581–592. doi: 10.1016/j.actpsy.2007.09.009
- Baker, A. G. (2013). *Individual differences in associative learning*. Cardiff: Unpublished paper presented at The Associative Learning Symposium.
- Baker, A. G., Msetfi, R. M., Hanley, N., and Murphy, R. A. (2010). "Depressive realism? Sadly not wiser," in *Clinical Applications of Learning Theory*, eds M. Haselgrove and L. Hogarth (Hove: Psychology Press), 153–177.
- Baker, A. G., Murphy, R. A., Vallée-Tourangeau, F., and Mehta, R. (2001). "Contingency learning and causal reasoning," in *Handbook of Contemporary Learning Theories*, eds R. R. Mowrer and S. B. Klein (Mahwah, NJ: Lawrence Erlbaum Associates, Inc), 255–306.
- Baruch, I., Hemsley, D. R., and Gray, J. A. (1988). Differential performance of acute and chronic schizophrenics in a latent inhibition task. *J. Nerv. Ment. Dis.* 176, 598–606. doi: 10.1097/00005053-198810000-00004
- Bender, S., Muller, B., Oades, R. D., and Sartory, G. (2001). Conditioned blocking and schizophrenia: a replication and study of the role of symptoms, age, onset-age of psychosis and illness-duration. *Schizophr. Res.* 49, 157–170. doi: 10.1016/S0920-9964(00)00040-2
- Bradley, B. P., Mogg, K., and Lee, S. C. (1997). Attentional biases for negative information in induced and naturally occurring dysphoria. *Behav. Res. Ther.* 35, 911–927. doi: 10.1016/S0005-7967(97)00053-3
- Brandon, S. E., Vogel, E. H., and Wagner, A. R. (2000). A componential view of configural cues in generalization and discrimination in Pavlovian conditioning. *Behav. Brain Res.* 110, 67–72. doi: 10.1016/S0166-4328(99)00185-0
- Bush, R. R., and Mosteller, F. (1951). A model for stimulus generalization and discrimination. *Psychol. Rev.* 58, 413–423. doi: 10.1037/h0054576
- Chan, S. W. Y., Goodwin, G. M., and Harmer, C. J. (2007). Highly neurotic never-depressed students have negative biases in information processing. *Psychol. Med.* 37, 1281–1291. doi: 10.1017/S0033291707000669
- Craske, M. G., Hermans, D., and Vansteenwegen, D. (2006). *Fear and Learning: From Basic Processes to Clinical Implications*. Washington, DC: American Psychological Association. doi: 10.1037/11474-000
- Deisig, N., Lachnit, H., Giurfa, M., and Hellstern, F. (2001). Configural olfactory learning in honeybees: negative and positive patterning discrimination. *Lear. Mem.* 8, 70–78. doi: 10.1101/lm.8.2.70
- Derryberry, D., and Reed, M. A. (2002). Anxiety-related attentional biases and their regulation by attentional control. *J. Abnorm. Psychol.* 111, 225–236. doi: 10.1037/0021-843X.111.2.225
- de Wit, S., and Dickinson, A. (2009). Associative theories of goal-directed behaviour: a case for animal-human translational models. *Psychol. Res.* 73, 463–476. doi: 10.1007/s00426-009-0230-6
- Diamond, R., and Carey, S. (1986). Why faces are and are not special—an effect of expertise. *J. Exp. Psychol. Gen.* 115, 107–117. doi: 10.1037/0096-3445.115.2.107
- Fanselow, M. S. (1999). Learning theory and neuropsychology: configuring their disparate elements in the hippocampus. *J. Exp. Psychol. Anim. Behav. Process.* 25, 275–283. doi: 10.1037/0097-7403.25.3.275
- Gal, G., Barnea, Y., Biran, L., Mendlovic, S., Gedi, T., Halavy,

- M., et al. (2009). Enhancement of latent inhibition in patients with chronic schizophrenia. *Behav. Brain Res.* 197, 1–8. doi: 10.1016/j.bbr.2008.08.023
- Gotlib, I. H., Krasnoperova, E., Yue, D. N., and Joormann, J. (2004). Attentional biases for negative interpersonal stimuli in clinical depression. *J. Abnorm. Psychol.* 113, 127–135. doi: 10.1037/0021-843X.113.1.121
- Grand, C., and Honey, R. C. (2008). Solving, XOR. *J. Exp. Psychol. Anim. Behav. Process.* 34, 486–493. doi: 10.1037/0097-7403.34.4.486
- Gray, N. S., Pilowsky, L. S., Gray, J. A., and Kerwin, R. W. (1995). Latent inhibition in drug naive schizophrenics: relationship to duration of illness and dopamine D2 binding using SPET. *Schizophr. Res.* 17, 95–107. doi: 10.1016/0920-9964(95)00034-J
- Haddon, J. E., George, D. N., Grayson, L., McGowan, C., Honey, R. C., and Killcross, S. (2011). Impaired conditional task performance in a high schizotypy population: relation to cognitive deficits. *Q. J. Exp. Psychol.* 64, 1–9. doi: 10.1080/17470218.2010.529579
- Harris, J. A., and Livesey, E. J. (2010). An attention-modulated associative network. *Lear. Behav.* 38, 1–26. doi: 10.3758/LB.38.1.1
- Harris, J. A., Livesey, E. J., Gharaei, S., and Westbrook, R. F. (2008). Negative patterning is easier than a biconditional discrimination. *J. Exp. Psychol. Anim. Behav. Process.* 34, 494–500. doi: 10.1037/0097-7403.34.4.494
- Haselgrove, M., and Evans, L. H. (2010). Variations in selective and nonselective prediction error with the negative dimension of schizotypy. *Q. J. Exp. Psychol.* 63, 1127–1149. doi: 10.1080/17470210903229979
- Joos, E., Vansteenwegen, D., and Hermans, D. (2013). Worry as a predictor of fear acquisition in a nonclinical sample. *Behav. Modif.* 36, 723. doi: 10.1177/0145445512446477
- Kamin, L. J. (1968). “Attention-like” processes in classical conditioning,” in *Miami Symposium on the Prediction of Behaviour, Aversion Stimulation*, ed M. R. Jones (Miami: University of Miami Press), 9–33.
- Kamin, L. J. (1969). “Predictability, surprise, attention and conditioning,” in *Punishment and Aversive Behaviour*, eds B. A. Campbell and R. M. Church (New York, NY: Appleton-Century-Crofts), 279–296.
- Kamin, L. J., and Brimer, C. J. (1963). The effects of intensity of conditioned and unconditioned stimuli on a conditioned emotional response. *Can. J. Psychol.* 17, 194–200.
- Kamin, L. J., and Schaub, R. E. (1963). Effects of conditioned stimulus intensity on the conditioned emotional response. *J. Comp. Physiol. Psychol.* 56, 502–507. doi: 10.1037/h0046616
- Kehoe, E. J., Horne, A. J., Horne, P. S., and Macrae, M. (1994). Summation and configuration between and within sensory modalities in classical-conditioning of the rabbit. *Anim. Lear. Behav.* 22, 19–26. doi: 10.3758/BF03199952
- Kinder, A., and Lachnit, H. (2003). Similarity and discrimination in human Pavlovian conditioning. *Psychophysiology* 40, 226–234. doi: 10.1111/1469-8986.00024
- Lachnit, H. (1988). Convergent validation of information-processing constructs with pavlovian methodology. *J. Exp. Psychol. Hum. Percept. Perform.* 14, 143–152. doi: 10.1037/0096-1523.14.1.143
- Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A selective review and a hybrid model. *Q. J. Exp. Psychol. B.* 57, 193–243. doi: 10.1080/02724990344000141
- Le Pelley, M. E., and McLaren, I. P. L. (2004). Associative history affects the associative change undergone by both presented and absent cues in human causal learning. *J. Exp. Psychol. Anim. Behav. Process.* 30, 67–73.
- Le Pelley, M. E., Turnbull, M. N., Reimers, S. J., and Knipe, R. L. (2010). Learned predictiveness effects following single-cue training in humans. *Lern. Behav.* 38, 126–144. doi: 10.3758/LB.38.2.126
- Leder, H., and Bruce, V. (2000). When inverted faces are recognized: the role of configural information in face recognition. *Q. J. Exp. Psychol. A* 53, 513–536. doi: 10.1080/713755889
- Logan, F. A. (1954). A note on stimulus intensity dynamism. *Psychol. Rev.* 61, 77–80. doi: 10.1037/h0057534
- Lubinski, D. (2000). Scientific and social significance of assessing individual differences: “sinking shafts at a few critical points”. *Annu. Rev. Psychol.* 51, 405–444. doi: 10.1146/annurev.psych.51.1.405
- Lubow, R. E. (1989). *Latent Inhibition and Conditioned Attention Theory*. New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511529849
- Lubow, R. E. (2010). *Latent Inhibition the Corsini Encyclopedia of Psychology*: John Wiley and Sons, Inc.
- Lubow, R. E., and Moore, A. U. (1959). Latent inhibition: the effect of non-reinforced pre-exposure to the conditional stimulus. *J. Comp. Physiol.* 52, 415–419.
- Lubow, R. E., Schnur, P., and Rifkin, B. (1976). Latent inhibition and conditioned attention theory. *J. Exp. Psychol.* 2, 163–174. doi: 10.1037/0097-7403.2.2.163
- Mackintosh, N. J. (1971). Analysis of overshadowing and blocking. *Q. J. Exp. Psychol.* 23, 118–125. doi: 10.1080/0035557143000121
- Mackintosh, N. J. (1975). Theory of attention—variations in associability of stimuli with reinforcement. *Psychol. Rev.* 82, 276–298. doi: 10.1037/h0076778
- Macrae, C. N., and Lewis, H. L. (2002). Do I know you. Processing orientation and face recognition. *Psychol. Sci.* 13, 194–196. doi: 10.1111/1467-9280.00436
- Mason, O., and Claridge, G. (2006). The oxford-liverpool inventory of feelings and experiences (O-LIFE): further description and extended norms. *Schizophr. Res.* 82, 203–211. doi: 10.1016/j.schres.2005.12.845
- Mason, O., Linney, Y., and Claridge, G. (2005). Short scales for measuring schizotypy. *Schizophr. Res.* 78, 293–296. doi: 10.1016/j.schres.2005.06.020
- Matthews, G., Pitcaithly, D., and Mann, R. L. E. (1995). Mood, neuroticism, and the encoding of affective words. *Cognit. Ther. Res.* 19, 563–587. doi: 10.1007/BF02230514
- Maurer, D., Le Grand, R., and Mondloch, C. J. (2002). The many faces of configural processing. *Trends Cogn. Sci.* 6, 255–260. doi: 10.1016/S1364-6613(02)01903-4
- McGrath, J., Saha, S., Chant, D., and Welham, J. (2008). Schizophrenia: a concise overview of incidence, prevalence, and mortality. *Epidemiol. Rev.* 30, 67–76. doi: 10.1093/epirev/mxn001
- McLaren, I. P. L., Wills, A. J., and Graham, S. (2010). “Attention and perceptual learning,” in *Attention and Associative Learning: From Brain to Behaviour*, eds C. Mitchell and M. E. Le Pelley (Oxford: Oxford University Press), 133–158.
- Melchers, K. G., Shanks, D. R., and Lachnit, H. (2008). Stimulus coding in human associative learning: flexible representations of parts and wholes. *Behav. Process.* 77, 413–427. doi: 10.1016/j.beproc.2007.09.013
- Mineka, S., and Zinbarg, R. (2006). A contemporary learning theory perspective on the etiology of anxiety disorders—it’s not what you thought it was. [Review]. *Am. Psychol.* 61, 10–26. doi: 10.1037/0003-066X.61.1.10
- Mogg, K., Bradley, B. P., and Williams, R. (1995). Attentional bias in anxiety and depression—the role of awareness. *Br. J. Clin. Psychol.* 34, 17–36. doi: 10.1111/j.2044-8260.1995.tb01434.x
- Moran, P. M., Al-Uzri, M. M., Watson, J., and Reveley, M. A. (2003). Reduced Kamin blocking in non paranoid schizophrenia: associations with schizotypy. *J. Psychiatr. Res.* 37, 155–163. doi: 10.1016/S0022-3956(02)00099-7
- Moran, P. M., Owen, L., Crookes, A. E., Al-Uzri, M. M., and Reveley, M. A. (2008). Abnormal prediction error is associated with negative and depressive symptoms in schizophrenia. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 32, 116–123. doi: 10.1016/j.pnpbp.2007.07.021
- Msetfi, R. M., Murphy, R. A., Simpson, J., and Kornbrot, D. E. (2005). Depressive realism and outcome density bias in contingency judgments: the effect of the context and intertrial interval. *J. Exp. Psychol. Gen.* 134, 10–22. doi: 10.1037/0096-3445.134.1.10
- Myers, K. M., Vogel, E. H., Shin, J., and Wagner, A. R. (2001). A comparison of the Rescorla-Wagner and Pearce models in a negative patterning and a summation problem. *Anim. Learn. Behav.* 29, 36–45. doi: 10.3758/BF03192814
- Navon, D. (1977). Forest before trees—precedence of global features in visual-perception. *Cogn. Psychol.* 9, 353–383. doi: 10.1016/0010-0285(77)90012-3
- Otto, M. W., Leyro, T. M., Christian, K., Deveney, C. M., Reese, H., Pollack, M. H., et al. (2007). Prediction of “fear” acquisition in healthy control participants in a de novo fear-conditioning paradigm. [Research Support, N.I.H., Extramural]. *Behav. Modif.* 31, 32–51. doi: 10.1177/0145445506295054
- Pavlov, I. P. (1927). *Conditioned Reflexes*. Oxford: Oxford University Press.
- Pearce, J. M. (1987). A model for stimulus-generalization in pavlovian conditioning. *Psychol. Rev.* 94, 61–73. doi: 10.1037/0033-295X.94.1.61
- Pearce, J. M. (1994). Similarity and Discrimination: a Selective Review and a Connectionist Model.

- Psychol. Rev.* 101, 587–607. doi: 10.1037/0033-295X.101.4.587
- Pearce, J. M., and George, D. N. (2002). The effects of using stimuli from three different dimensions on autoshaping with a complex negative patterning discrimination. *Q. J. Exp. Psychol. B* 55, 349–364. doi: 10.1080/02724990244000061
- Pearce, J. M., and Hall, G. (1980). A model for pavlovian learning—variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol. Rev.* 87, 532–552. doi: 10.1037/0033-295X.87.6.532
- Pearce, J. M., and Mackintosh, N. J. (2010). “Two theories of attention: a review and possible integration,” in *Attention and Associative Learning, From Brain to Behaviour*, eds C. Mitchell and M. E. L. Pelley (Oxford: Oxford University Press), 11–40.
- Perfect, T. J. (2003). Local processing bias impairs lineup performance. *Psychol. Rep.* 93, 393–394.
- Perkins, C. C. (1953). The relation between conditioned stimulus intensity and response strength. *J. Exp. Psychol.* 46, 225–231. doi: 10.1037/h0063567
- Phillips, W. J., Hine, D. W., and Thorsteinsson, E. B. (2010). Implicit cognition and depression: a meta-analysis. *Clin. Psychol. Rev.* 30, 691–709. doi: 10.1016/j.cpr.2010.05.002
- Rasclé, C., Mazas, O., Vaiva, G., Tournant, M., Raybois, O., Goudemand, M., et al. (2001). Clinical features of latent inhibition in schizophrenia. *Schizophr. Res.* 51, 149–161. doi: 10.1016/S0920-9964(00)00162-6
- Raven, J. (2000). The Raven's progressive matrices: change and stability over culture and time. *Cogn. Psychol.* 41, 1–48. doi: 10.1006/cogp.1999.0735
- Redhead, E. S., and Pearce, J. M. (1995). Stimulus salience and negative patterning. *Q. J. Exp. Psychol. B* 48, 67–83.
- Rescorla, R. A., and Coldwell, S. E. (1995). Summation in autoshaping. *Anim. Learn. Behav.* 23, 314–326. doi: 10.3758/BF03198928
- Rescorla, R. A., and Wagner, A. R. (eds.). (1972). *A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement*. New York, NY: Appleton-Century-Crofts.
- Rusting, C. L. (1998). Personality, mood, and cognitive processing of emotional information: three conceptual frameworks. [Review]. *Psychol. Bull.* 124, 165–196. doi: 10.1037/0033-2909.124.2.165
- Rusting, C. L. (1999). Interactive effects of personality and mood on emotion-congruent memory and judgment. *J. Pers. Soc. Psychol.* 77, 1073–1086. doi: 10.1037/0022-3514.77.5.1073
- Saha, S., Chant, D., Welham, J., and McGrath, J. (2005). A systematic review of the prevalence of schizophrenia. *PLoS Med.* 2:e141. doi: 10.1371/journal.pmed.0020141
- Seligman, M. E. (1972). Learned helplessness. *Annu. Rev. Med.* 23, 407–412. doi: 10.1146/annurev.me.23.020172.002203
- Shanks, D. R., and Darby, R. J. (1998). Feature- and rule-based generalization in human associative learning. *J. Exp. Psychol. Anim. Behav. Process.* 24, 405–415. doi: 10.1037/0097-7403.24.4.405
- Smith, M. C. (1968). CS-US interval and US intensity in classical conditioning of the rabbit's nictitating membrane response. *J. Comp. Physiol.* 66, 679–687.
- Solomon, P. R., Crider, A., Winkelman, J. W., Turi, A., Kamer, R. M., and Kaplan, L. J. (1981). Disrupted latent inhibition in the rat with chronic amphetamine or haloperidol-induced supersensitivity—relationship to schizophrenic attention disorder. *Biol. Psychiatry* 16, 519–537.
- Sutherland, R. J., and Rudy, J. W. (1989). Configural association theory—the role of the hippocampal-formation in learning, memory, and amnesia. *Psychobiology* 17, 129–144.
- Tanaka, J. W., and Farah, M. J. (1993). Parts and wholes in face recognition. *Q. J. Exp. Psychol. A* 46, 225–245. doi: 10.1080/14640749308401045
- Uhlhaas, P. J., and Silverstein, S. M. (2005). Perceptual organization in schizophrenia spectrum disorders: empirical research and theoretical implications. *Psychol. Bull.* 131, 618–632. doi: 10.1037/0033-2909.131.4.618
- Uhlhaas, P. J., Silverstein, S. M., Phillips, W. A., and Lovell, P. G. (2004). Evidence for impaired visual context processing in schizotypy with thought disorder. *Schizophr. Res.* 68, 249–260. doi: 10.1016/S0920-9964(03)00184-1
- Vaitl, D., and Lipp, O. V. (1997). Latent inhibition and autonomic responses: a psychophysiological approach. *Behav. Brain Res.* 88, 85–93. doi: 10.1016/S0166-4328(97)02310-3
- Wagner, A. R. (2003). Context-sensitive elemental theory. *Q. J. Exp. Psychol. B* 56, 7–29. doi: 10.1080/02724990244000133
- Weiner, I., Lubow, R. E., and Feldon, J. (1981). Chronic amphetamine and latent inhibition. *Behav. Brain Res.* 2, 285–286. doi: 10.1016/0166-4328(81)90105-4
- Weiner, I., Lubow, R. E., and Feldon, J. (1984). Abolition of the expression but not the acquisition of latent inhibition by chronic amphetamine in rats. *Psychopharmacology* 83, 194–199. doi: 10.1007/BF00429734
- Weiner, I. (2003). The ‘two-headed’ latent inhibition model of schizophrenia: Modeling positive and negative symptoms and their treatment. *Psychopharmacol.* 169, 257–297. doi: 10.1007/s00213-002-1313-x
- Wills, A. J., Graham, S., Koh, Z. S., McLaren, I. P. L., and Rolland, M. D. (2011). Effects of concurrent load on feature- and rule-based generalization in human contingency learning. *J. Exp. Psychol. Anim. Behav. Process* 37, 308–316. doi: 10.1037/a0023120

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: 18 January 2013; accepted: 14 August 2013; published online: 04 September 2013.

Citation: Byrom NC (2013) Accounting for individual differences in human associative learning. *Front. Psychol.* 4:588. doi: 10.3389/fpsyg.2013.00588

This article was submitted to *Personality Science and Individual Differences*, a section of the journal *Frontiers in Psychology*.

Copyright © 2013 Byrom. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Repeated activation of a CS-US-contingency memory results in sustained conditioned responding

Els Joos*, Debora Vansteenwegen, Bram Vervliet and Dirk Hermans*

Faculty of Psychology and Educational Sciences, University of Leuven, Leuven, Belgium

Edited by:

Rachel M. Msetfi, University of Limerick, Ireland

Reviewed by:

John Zelenski, Carleton University, Canada

Miguel A. Vadillo, University College London, UK

*Correspondence:

Els Joos, Faculty of Psychology and Educational Sciences, University of Leuven, Tiensestraat 102, 3000 Leuven, Belgium

e-mail: els.joos@ppw.kuleuven.be;

Dirk Hermans, Faculty of Psychology and Educational Sciences, University of Leuven, Tiensestraat 102, box 3712, 3000 Leuven, Belgium

e-mail: dirk.hermans@ppw.kuleuven.be

Individuals seem to differ in conditionability, i.e., the ease by which the contingent presentation of two stimuli will lead to a conditioned response. In contemporary learning theory, individual differences in the etiology and maintenance of anxiety disorders are, among others, explained by individual differences in temperamental variables (Mineka and Zinbarg, 2006). One such individual difference variable is how people process a learning experience when the conditioning stimuli are no longer present. Repeatedly thinking about the conditioning experience, as in worry or rumination, might prolong the initial (fear) reactions and as such, might leave certain individuals more vulnerable to developing an anxiety disorder. However, in human conditioning research, relatively little attention has been devoted to the processing of a memory trace after its initial acquisition, despite its potential influences on subsequent performance. Post-acquisition processing can be induced by mental reiteration of a conditioned stimulus-unconditioned stimulus (CS-US)-contingency. Using a human conditioned suppression paradigm, we investigated the effect of repeated activations of a CS-US-contingency memory on the level of conditioned responding at a later test. Results of three experiments showed more sustained responding to a “rehearsed” CS+ as compared to a “non-rehearsed” CS+. Moreover, the second experiment showed no effect of rehearsal when only the CS was rehearsed instead of the CS-US-contingency. The third experiment demonstrated that mental CS-US-rehearsal has the same effect regardless of whether it was cued by the CS and a verbal reference to the US or by a neutral signal, making the rehearsal “purely mental.” In sum, it was demonstrated that post-acquisition activation of a CS-US-contingency memory can impact conditioned responding, underlining the importance of post-acquisition processes in conditioning. This might indicate that individuals who are more prone to mentally rehearse information condition more easily.

Keywords: conditioning, human learning, CS-US-contingency, rehearsal, post-acquisition processing, conditioned suppression

INTRODUCTION

In classical conditioning, a learning experience is often considered to end when the conditioning stimuli are no longer present. This is based on the fact that conditioning refers to the contingent presentation of an originally neutral stimulus (conditioned stimulus, CS) together with a biologically relevant unconditioned stimulus (US), resulting in the CS becoming a signal for US-onset and thus evoking a conditioned response (CR) during subsequent presentations (Bouton, 2007). This CR can be decreased or eliminated by non-reinforced presentations of the CS, a procedure called “extinction.” It is generally assumed that conditioning comprises both learning and memory: the learning of a CS-US-contingency builds up a memory (“encoding”), which is stored (“consolidation”) and can be reactivated upon future confrontations with that CS (“retrieval”). The strength of the CR is a function of these three processes.

Most conditioning research is focused on the encoding phase (which comprises the actual learning) and much less on consolidation and retrieval. However, the latter phases may have major effects on long-term conditioning. For instance, in the case of

Pavlovian conditioning, human participants may mentally reflect upon the conditioning experience by repeatedly reactivating either the CS-representation, the US-representation, or the entire experience (CS-US-contingency memory). This repeated thinking about a negative experience might be akin to repetitive thought processes such as worry and rumination, as will be discussed in more detail shortly. The current research aimed to investigate the role of individual differences in such repetitive thought on the strength of conditioned responding.

Current evidence suggests that repeated reactivation of the conditioning memory results in higher CRs at a later test compared to conditions that do not include such active post-acquisition processing. First, the impact of US-rehearsal has been investigated by Davey and colleagues (Jones and Davey, 1990; Davey and Matchett, 1994). After conditioning, participants in the experimental group were asked to rehearse the US whenever the word “think” was presented on the screen, while control participants rehearsed either a non-aversive event or an unrelated aversive event. It was demonstrated that participants who rehearsed the US after acquisition retained a skin conductance response (SCR)

during subsequent CS-presentations while this was not true for controls. Arntz et al. (1997) replicated this finding using SCR's, but not when relying on anxiety ratings. Arguably, mental repetition of the US leads to stronger conditioned responding upon subsequent CS-presentations.

A second active post-acquisition procedure that could result in stronger conditioned responding is mentally reiterating the CS-US-contingency. As it is well-known that the contingency between the CS and the US is important in determining a CR, the procedure of repeated post-acquisition activation of the CS-US-contingency memory merits investigation as well. A preliminary indication of this effect can be found in a study by Yaremko and Werner (1974) who showed that repeatedly imagining a previously presented tone-shock-contingency elicited more pronounced electrodermal responses during subsequent extinction than imagining the same stimuli in an unpaired way. Imagining was cued by auditory presentations of the words "tone" and "shock." Although not set up as a study about post-acquisition processing in conditioning (thus lacking appropriate control for acquisition strength), this study at least suggests that repeated mental activation of a CS-US-contingency impacts subsequent CRs. A second line of studies, performed in rabbits, provides only indirect support for a role of rehearsal in conditioning. Wagner and colleagues (e.g., Wagner et al., 1973; Terry and Wagner, 1975) investigated rehearsal as an explanatory mechanism for the fact that a US needs to be unexpected for CS-US-learning to occur. They suggested that rehearsal of conditioning events was crucial in conditioning. Furthermore, given that the rehearsal capacity of an organism is limited, they predicted (and showed) that a surprising event that would command rehearsal could interfere with the necessary rehearsal and thus with the learning of other CS-US-pairings. Based on these findings, Wagner (1981) developed the model of Standard Operating Procedures (SOP) which states that conditioned associations require the joint rehearsal of the representations of the CS and the US in memory.

It is surprising that these post-acquisition processes have received only little attention in human conditioning research, while they play a central role in the memory literature. For instance, rehearsal, a type of post-acquisition processing defined as the covert or overt repetition of information (Atkinson and Shiffrin, 1971) is studied extensively and is implicated as an important factor in most models of memory functioning (Anderson, 1999). It is well established that more rehearsal, both in frequency as in length of the rehearsal period, typically results in enhanced memory for the rehearsed information (Ebbinghaus, 1885/1913; Johnson, 1980). As conditioning relies on both learning and memory (Bouton and Moody, 2004), it seems obvious to study post-acquisition rehearsal processes in conditioning. In a first attempt to address this issue, we (Joos et al., 2012b) investigated the role of post-acquisition processing in fear learning, by examining the impact of rehearsing an aversive conditioned association (CS = picture; US = human scream) on subsequent fear responding. Fear responding to the picture-CS which was previously paired with the scream persisted in participants who rehearsed this contingency, but decreased in participants who had been asked to rehearse a different contingency. In the current manuscript, the role of post-acquisition processing is studied in

a conditioned suppression paradigm that allows us to investigate whether the earlier findings in fear conditioning also apply in a more neutral contingency learning task. More importantly, giving it's less time-consuming nature, this paradigm allows a more in depth analysis of the influence of rehearsal on conditioned responding using different rehearsal procedures (see below).

Besides the theoretical relevance of post-acquisition processes in conditioning, studying these processes is ecologically valid as well. In general, it is believed that some aspects of anxiety disorders are explained by conditioning processes (Rachman, 1991; Craske et al., 2006; Field, 2006), i.e., undergoing a conditioning experience (such as a car accident) can install subsequent fear reactions (such as driving phobia). However, large differences exist in whether or not an individual develops an anxiety disorder after such an (aversive) learning experience.

In contemporary learning theory, it is suggested that such individual differences in the etiology and maintenance of anxiety disorders could, among others, be explained by individual differences in temperamental variables, such as trait anxiety and behavioral inhibition (Levey and Martin, 1981; Mineka and Zinbarg, 2006; Mineka and Oehlberg, 2008; Joos et al., under review). However, we hypothesize that the differential tendency to engage in repetitive thought, such as worry or rumination, might also be an important factor in explaining differences in conditionability. Not only is trait anxiety highly associated with repetitive thought (e.g., Meyer et al., 1990), but it is also demonstrated that trait worry predicts the strength of fear acquisition (Otto et al., 2007; Joos et al., 2012c). Participants scoring higher on the Penn State Worry Questionnaire (Meyer et al., 1990) demonstrated enhanced fear learning.

As such, differences in the strength of a CR, could in part be explained by differences in how people process a learning experience when the conditioning stimuli are no longer present. Individuals might repeatedly reflect upon an aversive conditioning experience, which might be akin to repetitive thought processes such as worry and rumination (Watkins, 2008). The potential role of repetitive thought in anxious responding is supported by the fact that individual differences in worry and rumination correlate with anxious symptoms (e.g., Meyer et al., 1990; Segerstrom et al., 2000; Fresco et al., 2002; Muris et al., 2004; Ehring et al., 2011). Moreover, repetitive thought (worry and/or rumination) has even been shown to predict the level of anxiety or anxiety symptoms in prospective designs (Nolen-Hoeksema, 2000; Segerstrom et al., 2000; Calmes and Roberts, 2007; Hong, 2007; McLaughlin et al., 2007; Watkins, 2008). In sum, the more one engages in repetitive thought, the more anxiety symptoms are experienced. Hence, given the role of conditioning processes in anxiety, we believe that repeatedly thinking about a conditioning experience, as in worry or rumination, might prolong the initial (fear) reactions and as such, might leave certain individuals more vulnerable to developing an anxiety disorder.

Given these correlational findings, we wanted to investigate experimentally the influence of differential post-acquisition processing (i.e., mental rehearsal) on subsequent conditioned responding. To this aim, we modeled differences in repetitive thought by experimentally inducing repeated activation of a conditioning experience. In the present studies, we targeted

post-acquisition processing of the CS-US-contingency, rather than activation of the CS- or the US-representation. In Experiments 1 and 2, participants were primed to rehearse the CS-US-contingency by presenting the CS and a verbal label referring to the US, in line with Yaremko and Werner (1974). This procedure allowed us to control the content of participant's thoughts. Moreover, this cued rehearsal procedure might resemble repetitive thought in real-life. A cue activates both the mental representations of the CS and the US and the association between them, but the US is never directly experienced. This resembles cued recall of fear memory by real-life confrontations with the phobic stimulus (e.g., driving a car). Given that this rehearsal procedure might entail additional acquisition trials (due to visual presentation of the CS), the procedure is contrasted with a purely mental rehearsal procedure, as used by Davey and colleagues (Jones and Davey, 1990). As repetitive thought is often purely mental as well, this procedure might more closely resemble ruminative thinking, i.e., repetitive thought that is cued by intrusions or memories of the conditioning event. In general, we hypothesize that mental reiteration of a CS-US-contingency, in the absence of real US-presentations, results in more conditioned responding compared to when no repeated activation occurs. This repetitive mental evocation of the CS-US-contingency memory is further referred to as "rehearsal."

We used a conditioned suppression paradigm, known as the Martians preparation, which has proven to be sensitive to a wide range of CS-US-contingency manipulations (for a review, see Franssen et al., 2010). This preparation was developed by Arcediano et al. (1996) to create a human analog for the conditioned suppression task used in animal conditioning. In such a task, the amount of suppression of an operant response serves as a behavioral measure of the strength of Pavlovian conditioning. As the Martians preparation is developed for use in humans, an instructed US is employed, rather than a biologically significant US. The task is set up as a computer game in which participants utilize their laser-gun (space bar) to shoot Martians. However, shooting during activation of the anti-laser shield (US) results in an inescapable invasion of Martians. Space scenes (CSs) predict the occurrence of the anti-laser shield. Learning is evident when participants refrain from bar pressing during a CS+. Using this preparation it was examined whether rehearsing a CS-US-contingency results in more conditioned suppression to a rehearsed CS+ than to a non-rehearsed CS+.

EXPERIMENT 1

MATERIALS AND METHODS

Participants

Participants were 42 volunteers aged between 18 and 53 years ($M = 22.17$, $SD = 6.01$). They participated in partial fulfillment of course requirements or were paid for their participation. The study was conducted in accordance with the requirements of the ethical committee of the faculty. All gave informed consent and were instructed that they could decline further participation at any time. They were uninformed about the purpose of the experiment and had no previous experience with the Martians preparation, apart from one participant who was excluded specifically for this reason. Three other participants were excluded due to problems during the procedure (talking and being distracted during bar

pressing). The results for 38 participants (ages 18–27, $M = 21.05$, $SD = 2.20$; 31 women) were included in the data-analysis.

Stimuli and apparatus

All participants were tested individually. Participants responded using the space bar of the keyboard. The Martians preparation was implemented into a flexible Windows95™ environment by Baeyens and Clarysse (1998), using Microsoft Visual C++ 5.0 and was recently adapted into MartiansV2 by Franssen et al. (2010).

Background pictures of four multi-colored space scenes served as CSs and were counterbalanced across individuals. CS-duration was 1.5 s, but was extended to 3 s during crucial test trials. The US consisted of a 0.5 s white flashing screen (5 flashes at a rate of 10 flashes/s; interflash time = 50 ms) accompanied by a metallic sound played in continuous looping (73 dBa). All sounds were presented binaurally through headphones (Philips SHP 2000). The images of the Martians and the explosions that appeared after "shooting" a Martian were multi-colored stimuli measuring 50×50 pixels. A screenshot of the Martians preparation is presented in Figure 1.

Procedure

In the Martians computer game, participants have to shoot incoming Martians by pressing the space bar (operant behavior). A Pavlovian CS-US-contingency is superimposed on this operant task. The US is described as an anti-laser shield. Participants have to refrain from bar pressing during activation of this anti-laser shield because otherwise, an inescapable invasion of Martians follows. The Martians procedure typically consists of various phases. In the two experiments presented here, these were: pre-training phase, US-only phase, acquisition, and acquisition test phase, rehearsal phase, and rehearsal test phase.

During the *Pre-training phase*, participants learned to emit a regular pattern of operant responding (bar pressing). Martians landed on the screen in rows from left to right and from top to bottom at a rate of 4/s. A full screen consisted of 7 rows and 10 columns (inter-row distance = 20 pixels, inter-column distance = 20 pixels). If full, the screen rolled up in a continuous fashion to make room for new Martians. Participants learned to press the space bar at the same rate as the appearance of Martians (4/s). In that case, only explosions rather than Martians appeared. However, if participants barpressed at a higher rate, not all Martians were shot and explosions appeared occasionally. The task for the participant was to make as many explosions appear as possible. Neither CSs nor USs were presented during this phase.

During the entire experiment, instructions were given both orally and visually on the computer screen (for an overview of instructions, see Baeyens et al., 2001). Participants could practice the bar pressing behavior for 25 s (100 Martians) during which the experimenter gave oral guidance if needed. After this phase (and after the US-only phase, the acquisition/acquisition test phase and the post-rehearsal phase) visual feedback was provided in the form of hit percentage.

The purpose of the next phase, the *US-only phase* was to introduce the instructed US, represented by the so-called "anti-laser shield" (combination of a flashing screen and a metallic sound). During the US, Martians appeared at the same rate as before. In

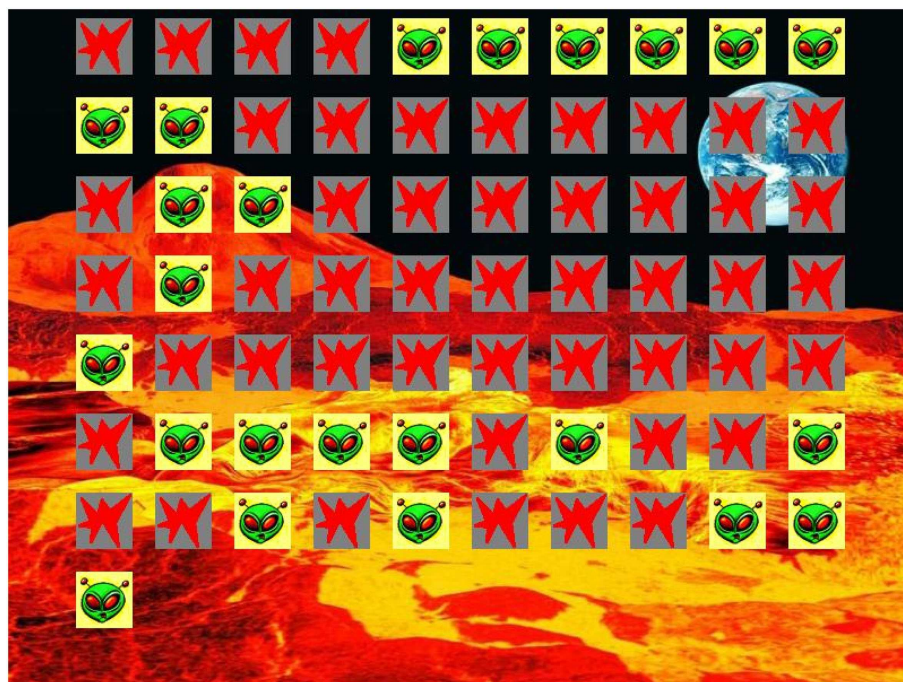


FIGURE 1 | Screenshot from the Martians preparation. Depicted are Martians and explosions (when Martian is shot by pressing the space bar) against a CS-background picture of a space scene.

this phase no CSs were presented. Participants learned to refrain from bar pressing when the anti-laser shield was activated, since pressing the space bar during this period evoked an inescapable invasion of Martians. An invasion lasted for 5 s and consisted of the landing of “thousands” of Martians (at a rate of 20/s) accompanied by a new sound played in continuous looping (79 dBa). During this invasion, bar pressing was ineffective (no explosions appeared contingent upon bar pressing). The US-only phase entailed four trials. On average, the inter-trial interval lasted for 7.5 s ($SD = 2.5$ s). This was the case throughout the whole experiment. The first two trials were used by the experimenter to explain (a) what an anti-laser shield looks like and (b) what happens if one presses the space bar during the anti-laser shield. Throughout the following two trials, participants could practice avoiding bar presses during the US which was virtually impossible as the USs appeared unannounced.

The *Acquisition phase* entailed the introduction of Pavlovian CS-US-contingencies which were superimposed on the operant baseline task. Participants were instructed that indicators would appear (background pictures) that might predict the occurrence of the US. They had to learn to distinguish good (CS+) from bad (CS−) predictors. In case of a good predictor, participants had to refrain from bar pressing to avoid pressing during the anti-laser shield. In case of a bad predictor, this suppression behavior was undesirable as not pressing the space bar would result in the successful landing of numerous Martians.

The Acquisition phase included training with two different CS+s and two different CS−s. A CS+ was immediately followed by the US. A CS− was never followed by the US. Space scenes were

counterbalanced between participants, serving either as the CS+ that would be rehearsed (CS+_R) or that would not be rehearsed (CS+_{NR}) or as one of both CS−s (CS−_A or CS−_B). This resulted in 12 counterbalancing conditions. All CSs lasted for 1.5 s and were presented five times each, generating 20 (randomized) trials. For the CS+ trials, a 80% reinforcement schedule was used in order to obtain suboptimal conditioning.

The Acquisition phase was immediately followed by the *Acquisition Test phase*. This transition was not noticeable to participants. The Acquisition Test phase comprised one non-reinforced presentation of every CS. Trial order was again randomized. During this test phase, every CS was presented for 3 s instead of 1.5 s to allow a more accurate measurement of the suppression behavior. Throughout the Acquisition and the Acquisition Test phase, trials lasted for 7 s. During the ITT's, the background screen was black. The light in the room was dimmed during these phases.

In the *Rehearsal phase*, the crucial manipulation was implemented. The goal was to prompt participants to mentally rehearse one of the CS-US-contingencies they had acquired in the previous phase. The CS+ that was part of the rehearsed CS-US-contingency is referred to as the rehearsed CS+. The other CS+ is called the non-rehearsed CS+. As a background for this differential mental rehearsal participants were asked to engage in a so-called attention training task that “could affect their future task performances.” More precisely, they were requested to focus their attention on one of the background pictures (CS+_R) that was previously presented and to think about this background and how it co-occurred with the flashing anti-laser shield. When they noticed being distracted, participants had to gently refocus

their attention on the background – anti-laser shield-compound. Participants were prompted to keep refocusing their attention whenever necessary. The training task was set up as a cover story to ensure rehearsal of both the CS and the US. The stimuli of the other CS-US-contingency were never presented during this phase. The background picture (CS+_R) and the word “anti-laser shield” (in Dutch) were presented six times for 15 s, alternated with a 10-s black screen. This was done to repeatedly draw the attention of the participants to the screen. The attention training task lasted for 2 min 20 s and was conducted twice. After each training, participants rated how easy/difficult it was for them to focus (and keep focused) their attention on the background and the flashing anti-laser shield on a 21-point scale ranging from –100 (*very difficult*) to +100 (*very easy*) in steps of 10. In between both training tasks, a filler task consisting of two questionnaires was administered¹.

After the Rehearsal phase the effects of mental rehearsal were monitored. Participants were redirected to the Martians computer task for the *Rehearsal Test phase*. Once more, the light was dimmed. Participants were instructed that the task was identical as before and that they were again expected to shoot Martians to stop them from invading Earth. They had to avoid bar pressing during the anti-laser shield and pay attention to the signals (CSs) to infer US-occurrence. No further instructions were given. However, when a participant asked whether the anti-laser shield would occur again, he/she was told that this possibility existed. The phase consisted of three blocks of one unreinforced presentation of the four CSs (CS+_R, CS+_{NR}, CS–_A, and CS–_B) to ensure a reliable assessment of conditioned responding after rehearsal. Within each block of four trials, trial order was randomized. Since testing occurred under extinction, the first test trial was the most crucial one as non-reinforced presentations might have reduced conditioned responding in the subsequent test trials. After the test phase, participants were thanked for their participation.

RESULTS

Manipulation check

After each training task, participants rated the difficulty of focusing their attention on a scale ranging from –100 (*very difficult*) to +100 (*very easy*). If participants reported being distracted during rehearsal, this might have influenced the quality of CS-US-rehearsal. The mean attention scores for the first and second attention training as well as the overall mean score for both tasks are presented in **Table 1**. As negative scores indicate difficulties to focus attention, we excluded for further analyses three participants who obtained a negative score averaged over both training tasks.

¹In Experiment 1, the Penn State Worry Questionnaire – Past Day (PSWQ-PD; Joos et al., 2012a) and the Sensitivity to Punishment-Sensitivity for Reward-questionnaire (SPSRQ; Torrubia et al., 2001) were administered at this stage. In Experiment 2, the Penn State Worry Questionnaire (PSWQ; Meyer et al., 1990), the Mindfulness Attention Awareness Scale (MAAS; Brown and Ryan, 2003), and the Questionnaire upon Mental Imagery (QMI; Sheehan, 1967) were administered. In Experiment 3, the MAAS and the PSWQ-PD were filled out. There were no significant associations between the questionnaire scores and the effect of rehearsal on conditioning performance. These null findings might be attributed to the fact that the rehearsal manipulation in the current studies overruled the effect of individual difference variables on the strength of the conditioning response.

Dependent variable

In conditioned suppression tasks like the Martians task, suppression of the operant response (bar pressing) serves as a measure of the strength of classical conditioning. Participants' behavior is expressed in terms of suppression ratios (SRs) of the form $a/(a + b)$, where a is the number of responses during the CS, and b the number of bar presses in an equal period of time immediately preceding CS-onset. This implies that a SR equaling 0.5 indicates no suppression at all, while a SR equaling 0 designates complete suppression of the operant response.

Acquisition test

Figure 2 displays SRs as a function of Phase and CS-type. Data from the Acquisition Test phase were analyzed using a repeated measures ANOVA with CS-type (CS–_{Average}/CS+_R/CS+_{NR}) as within-subjects variable². The main effect of CS-type was significant, $F(2, 68) = 23.64$, $p < 0.0001$, $MSE = 0.007$, $\eta_p^2 = 0.41$, indicating successful differential acquisition. Planned comparisons indicated that the SRs for the CS+_R and the CS+_{NR} were significantly lower than the SR for CS–_{Average}, $F_{CS+R}(1, 34) = 22.01$, $p < 0.0001$, $MSE = 0.009$; $F_{CS+NR}(1, 34) = 40.10$, $p < 0.0001$, $MSE = 0.008$. Given the interest in a differential effect between both CS+s after rehearsal, a non-differential level of conditioning for both CS+s is a prerequisite, which was fulfilled, $F(1, 34) = 1.89$, $p = 0.18$.

Rehearsal test

Figure 2 suggests that the rehearsal manipulation had an effect on the level of suppression to the CS+s at test. This was supported by a 2 (Phase: Acquisition test/Rehearsal test average) \times 2 (CS-type: CS+_R/CS+_{NR})-repeated measures ANOVA, including average SRs over three test trials (Rehearsal test average), which revealed a significant Phase \times CS-type interaction, $F(1, 34) = 5.53$, $p < 0.05$, $MSE = 0.005$, $\eta_p^2 = 0.14$. This indicates a different course of suppression over time to the rehearsed than to the non-rehearsed CS+. Planned comparisons at test demonstrated that the CS+_R evoked a stronger CR than the CS+_{NR}, $F(1, 34) = 4.26$, $p < 0.05$, $MSE = 0.005$. A decrease in conditioned suppression is present for both CS+s, but is stronger for the non-rehearsed CS+, $F(1, 34) = 32.67$, $p < 0.0001$, $MSE = 0.006$, than for the rehearsed CS+, $F(1, 34) = 5.63$, $p < 0.05$, $MSE = 0.008$.

Thus, the hypothesis about the effect of rehearsal on conditioned responding is supported. As the test phase comprised three trials which were conducted under extinction, the effect of rehearsal was also investigated for responding on the first test trial only. Using a 2 (Phase: Acquisition test/Rehearsal test 1) \times 2 (CS-type: CS+_R/CS+_{NR})-repeated measures ANOVA, it was shown that the crucial Phase \times CS-type interaction was significant, $F(1, 34) = 6.77$, $p < 0.05$, $MSE = 0.009$, $\eta_p^2 = 0.17$. Again, planned comparisons demonstrated that after rehearsal, the CS+_R evoked a stronger CR than the CS+_{NR}, $F(1, 34) = 5.69$, $p < 0.05$, $MSE = 0.01$. Moreover, conditioned responding to the

²In all 3 three experiments, data were first analyzed using an ANOVA with CS-type (CS–_A, CS–_B, CS+_R, and CS+_{NR}) as within-subjects variable (and Group as between-subjects variable in Experiments 2 and 3). Given that in every group of participants the SRs did not differ between both CS–s, SRs for both stimuli were always averaged for use in subsequent analyses.

Table 1 | Mean attention score (and standard deviations) on the first and the second attention training task and average for both tasks, as a function of experimental group for Experiments 1, 2, and 3.

		Attention training 1		Attention training 2		Average	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Experiment 1		58.05	28.77	40.34	37.72	49.20	30.47
Experiment 2	CS-US-rehearsal	56.25	33.73	32.50	39.92	44.38	31.84
	CS-rehearsal	66.25	18.61	29.58	40.59	47.92	24.54
	Visual rehearsal	46.30	40.11	34.44	35.99	40.37	36.90
Experiment 3	Mental rehearsal	23.08	50.89	21.53	45.67	22.31	45.59
	Control	49.60	37.47	27.60	49.69	38.60	41.72

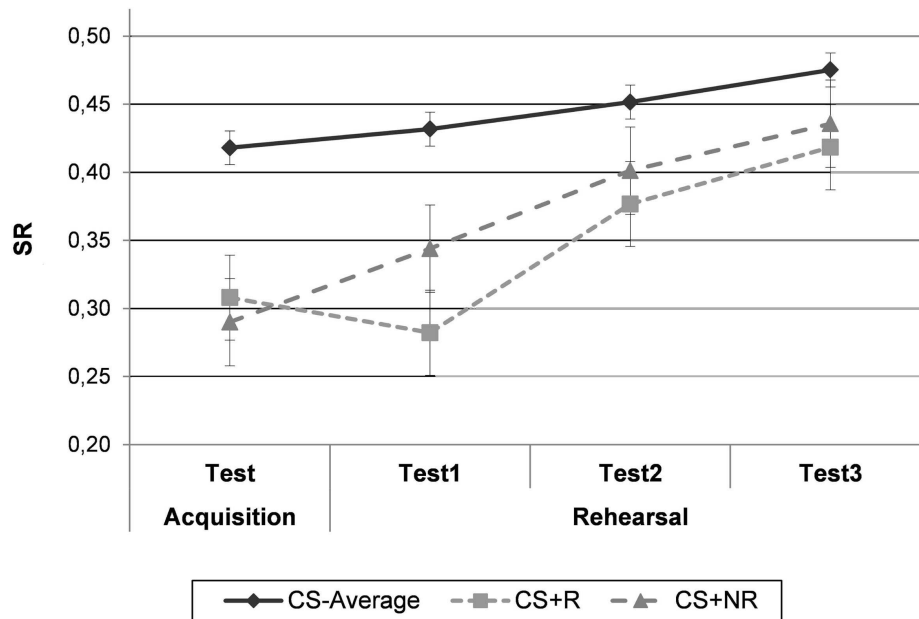


FIGURE 2 | Experiment 1. Mean suppression ratio's (SR) as a function of phase (acquisition test, rehearsal test 1, rehearsal test 2, and rehearsal test 3) and CS-type (CS_{Average}, CS_R, and CS_{NR}). CS_{Average} = the average for both CS₊'s. CS_R = rehearsed CS₊, CS_{NR} = non-rehearsed CS₊. Error bars denote standard error.

non-rehearsed CS₊ attenuated from acquisition test to rehearsal test 1, $F(1, 34) = 6.03$, $p < 0.05$, $MSE = 0.009$, while this was not the case for the rehearsed CS₊, $F(1, 34) = 1.48$, $p = 0.23$.

DISCUSSION

Experiment 1 was set up to test whether repeated activation of a previously acquired CS-US-contingency memory impacts conditioning effects to that CS in the long-term. The results clearly show stronger conditioned responding to the rehearsed CS₊ as compared to the non-rehearsed CS₊, indicating that mental reiteration of a CS-US-experience strengthens subsequent conditioned responding. An important observation is however that, rather than causing an increment in responding, rehearsal seems to sustain responding, while the absence of rehearsal results in decreased CRs. This pattern is further investigated in Experiment 3.

After demonstrating an effect of rehearsal, we wanted to explore the boundary conditions of this effect. As both CS₊s were paired with the same US, the observed rehearsal effect cannot be

attributed to rehearsal of the US alone. Indeed, US-rehearsal would elicit the same level of responding to both CS₊s at test. However, at this point it is unclear whether repeated activation of the CS alone would result in increased conditioned responding as well. Indeed, it might be the case that the observed effect should be attributed to CS-rehearsal rather than to rehearsal of the CS-US-contingency. Experiment 2 was set up to investigate this possibility.

EXPERIMENT 2

In addition to investigating whether the findings of Experiment 1 could be replicated, this experiment was conducted with the aim to extend these promising findings. More precisely, it was investigated whether mental reiteration of a CS alone results in increased or sustained responding as well, in which case the observed rehearsal effect in Experiment 1 could be attributed to CS-rehearsal.

That CS-rehearsal could increment responding has been shown in studies regarding sensitization or incubation. First, rehearsal of

the CS might result in a sensitization of conditioned responding, i.e., an increase in responsiveness caused by the (covert) repetition of a stimulus (Groves and Thompson, 1970). As such, this might underlie a strengthening of the CR after CS-US-rehearsal. Second, in fear conditioning, it is proposed that short unreinforced presentations of the CS might result in an *increment*, rather than a decrement, in responding. Since Eysenck (1968) termed this phenomenon “incubation,” some studies have provided tentative support for this hypothesis (e.g., Rohrbaugh and Riccio, 1970; Rohrbaugh et al., 1972). Until now, little evidence exists that repeated CS-only presentations promote a *progressive increase* in CR strength (Nicholaichuk et al., 1982; Kaloupek, 1983). Most studies demonstrate that short duration CS-presentations evoke resistance to extinction, rather than incubation (Stone and Borkovec, 1975; Sandin and Chorot, 1989). A process of incubation, either defined as an increment in conditioned responding or a resistance to extinction, would result in more conditioned suppression to the rehearsed CS+ (as compared to the non-rehearsed CS+) after the rehearsal phase as well.

To test the impact of rehearsing the CS without reference to the US, two conditions were included. Besides the “CS-US-Rehearsal”-group, a replication of Experiment 1, this study comprised a control condition, “CS-Rehearsal.” Participants in this condition were requested to rehearse the CS, instead of the CS-US-contingency. Conversely, it is important to note that most learning theories would predict extinction, characterized by a *decrement* rather than an increment in CR, after unreinforced CS-presentations (Hermans et al., 2006). Given these conflicting predictions, it is important to include this control group. If conditioned responding is only sustained after CS-US-rehearsal and not after CS-rehearsal, the data from Experiment 1 cannot be ascribed to mental reactivations of the memory of the CS alone.

MATERIALS AND METHODS

Participants

Fifty psychology students participated in return for course credits. All participants provided informed consent and were instructed that they could decline further participation at any time during the experiment. They were all uninformed about the purpose of the experiment. Two participants were excluded due to technical problems. The remaining 40 females and 8 males, aged 17–21 years ($M = 18.13$, $SD = 0.74$), were randomly assigned either to the condition “CS-US-Rehearsal” or the condition “CS-Rehearsal,” resulting in 24 participants in each condition.

Procedure

The same apparatus, software, and stimuli were used as in Experiment 1. For both conditions, the Pre-training phase, the US-only phase and the Acquisition and Acquisition Test phase were identical to those in Experiment 1. Only the Rehearsal phase differed between both experiments. Similarly, the aim of this phase was to evoke mental rehearsal of previously presented stimuli. However, while participants in one condition rehearsed a CS-US-contingency as in Experiment 1, participants in the other condition had to mentally rehearse a CS+, without reference to the US. Hence, participants in this condition were merely asked to focus their attention on one of the background pictures that

was previously presented. As always, they were requested to gently refocus their attention when they noticed being distracted. As in Experiment 1, the attention training cover story was applied to obtain rehearsal. During the Rehearsal phase, the same parameters were used. More precisely, in the “CS-US-Rehearsal”-group, the background picture, and the word “anti-laser shield” (in Dutch) appeared on the screen six times. In the “CS-Rehearsal”-condition, only the background picture (CS+_R) was presented, again for six times alternated with black screens.

The attention training was again executed twice and each training phase was followed by a short rating of the difficulty to focus their attention. The training phases were separated by the administration of three filler questionnaires. Upon completion of the Rehearsal phase, an unrelated computer task (causal learning task) was administered. Subsequently, participants were redirected to the first computer for the Rehearsal Test phase, which consisted of three blocks, each containing one unreinforced presentation of every CS.

RESULTS

Manipulation check

The mean attention scores for the first and second training task (see Table 1) did not differ between both conditions, $t_1(46) = 1.27$, $p = 0.21$; $t_2(46) = 0.25$, $p = 0.80$, nor did the attention score when averaged over both tasks, $t(46) = 0.43$, $p = 0.67$. As in Experiment 1, we excluded for further analyses the (two) participants who obtained a negative score when averaged over both training phases.

Acquisition test

Suppression ratio's for each condition are presented in Figure 3, as a function of Phase and CS-type. Acquisition data were analyzed using a 2×3 -repeated measures ANOVA with Group (“CS-US-Rehearsal”/“CS-Rehearsal”) as a between-subjects variable and CS-type (CS–_{Average}/CS+_R/CS+_{NR}) as a within-subjects variable. Participants displayed differential conditioned responding to the CS+s as compared to the CSs. Overall, there was a main effect of CS-type, $F(2, 88) = 73.83$, $p < 0.0001$, $MSE = 0.007$, $\eta_p^2 = 0.63$, but no significant Group \times CS-type interaction, $F(2, 88) = 0.04$, $p = 0.96$, indicating no group differences in differential responding.

Planned comparisons confirmed that in the “CS-US-Rehearsal”-condition, the CS+_R and the CS+_{NR} generated significantly more conditioned responding than the CS–_{Average}, $F_{CS+R}(1, 44) = 44.53$, $p < 0.0001$, $MSE = 0.008$; $F_{CS+NR}(1, 44) = 53.21$, $p < 0.0001$, $MSE = 0.008$. Participants demonstrated an equal amount of suppression for both CS+s, $F(1, 44) = 0.55$, $p = 0.46$. This indicates that no differences in responding existed between both CS+s before the onset of the rehearsal phase. Similarly, participants in the “CS-Rehearsal”-condition demonstrated more conditioned responding to the CS+_R, $F(1, 44) = 44.70$, $p < 0.001$, $MSE = 0.008$, and the CS+_{NR}, $F(1, 44) = 52.57$, $p < 0.001$, $MSE = 0.008$, than to the CS–_{Average}. Again, no significant differences emerged between both CS+s, $F(1, 44) = 0.45$, $p = 0.51$.

Rehearsal test

The left panel of Figure 3 (“CS-US-Rehearsal”-condition) shows a data pattern that generally replicates Experiment 1. After rehearsal,

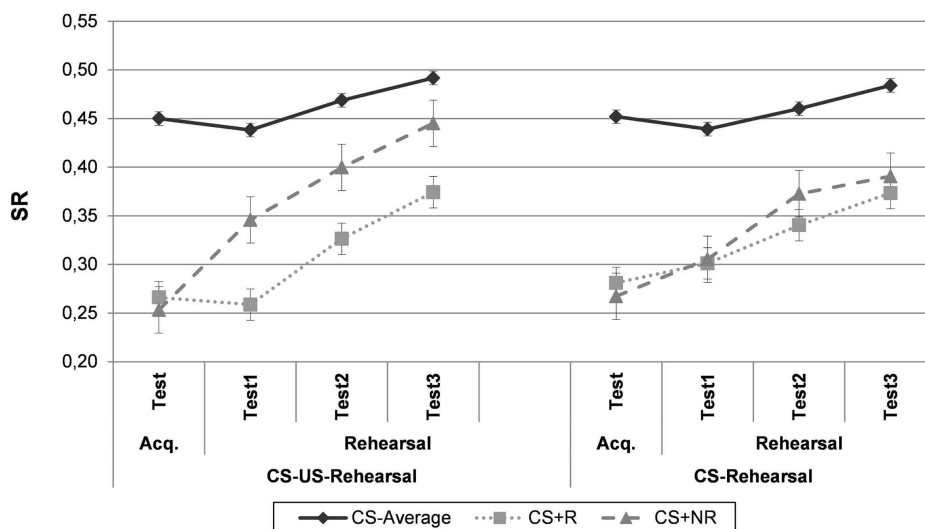


FIGURE 3 | Experiment 2. Mean suppression ratio's for the "CS-US-Rehearsal"-group, who rehearsed the CS-US-contingency, and the "CS-Rehearsal"-group, who rehearsed only the CS, as a function of phase (acquisition test, rehearsal test 2,

and rehearsal test 3) and CS-type (CS-Average, CS+R, and CS+NR). CS-Average = the average for both CS-'s. CS+R = rehearsed CS+, CS+NR = non-rehearsed CS+, Acq. Test = acquisition test. Error bars denote standard error.

the CS+R evokes more conditioned responding than the CS+NR. In line with predictions, this data pattern is absent for the "CS-Rehearsal"-group. To investigate whether mentally rehearsing the CS-US-contingency or the CS alone differentially impacts subsequent CRs, a 2 (Group: "CS-US-Rehearsal"/"CS-Rehearsal") \times 2 (Phase: Acquisition test/Rehearsal test average) \times 2 (CS-type: CS+R/CS+NR)-repeated measures ANOVA was conducted. As in the previous experiment, data of the three test trials were averaged to obtain a more reliable assessment. The Group \times Phase \times CS-type interaction failed to reach significance, $F(1, 44) = 2.44$, $p = 0.13$. However, after exclusion of the two participants who experienced difficulties focusing their attention during the first attention task rather than exclusion of those participants with a negative score when averaged over both tasks, the three-way interaction was marginally significant, $F(1, 44) = 3.82$, $p = 0.057$, $MSE = 0.004$. Although only marginally significant, the partial eta squared (η_p^2) of .08 suggests that this interaction can be interpreted as a medium to large effect (Stevens, 2002).

Follow-up analyses using simple interactions showed that for the "CS-US-Rehearsal"-condition, the Phase (Acquisition test/Rehearsal test average) \times CS-type (CS+R/ CS+NR) interaction was significant, $F(1, 44) = 10.62$, $p < 0.005$, $MSE = 0.005$, $\eta_p^2 = 0.19$. Planned comparisons confirmed that the CS+R produced significantly more conditioned suppression than the CS+NR, $F(1, 44) = 15.22$, $p < 0.001$, $MSE = 0.004$, at rehearsal test. Moreover, the decrease in suppression was significant for the CS+NR, $F(1, 44) = 45.74$, $p < 0.0001$, $MSE = 0.005$, but non-significant for the CS+R, $F(1, 44) = 3.95$, $p = 0.05$. For the "CS-Rehearsal"-condition, the overall Phase (Acquisition test/Rehearsal test average) \times CS-type (CS+R/ CS+NR) interaction failed to reach significance, $F(1, 44) = 1.32$, $p = 0.26$. Rehearsing a CS+ alone does not seem to impact subsequent

CRs to this CS. This conclusion is corroborated by planned comparisons showing no difference in SR between both CS+s at test, $F(1, 44) = 0.87$, $p = 0.36$. Responding to both CS+s decreased significantly from acquisition to rehearsal test, $F_{CS+R}(1, 44) = 5.84$, $p < 0.05$, $MSE = 0.007$; $F_{CS+NR}(1, 44) = 19.23$, $p < 0.0001$, $MSE = 0.005$.

Given that the Group \times Phase \times CS-type interaction failed to reach significance for the average SR over three test trials and given that rehearsal test 1 is probably the most valid test trial, the impact of rehearsal was also assessed for the first test trial only. A 2 (Group) \times 2 (Phase: Acquisition test/Rehearsal test 1) \times 2 (CS-type) ANOVA was conducted, which revealed a marginally significant three-way interaction, $F(1, 44) = 3.98$, $p = 0.052$, $MSE = 0.007$, $\eta_p^2 = 0.08$ (medium to large effect). Moreover, after exclusion of the two participants with difficulties to focus during the first attention task (instead of averaged over both tasks), this interaction was significant, $F(1, 44) = 6.07$, $p < 0.05$, $MSE = 0.006$, $\eta_p^2 = 0.12$ (medium to large effect), providing support for the differential impact on conditioned responding of rehearsing a CS-US-contingency versus a CS alone.

Follow-up analyses targeting the data for the "CS-US-Rehearsal"-condition, yielded a significant Phase \times CS-type interaction, $F(1, 44) = 10.75$, $p < 0.005$, $MSE = 0.007$, $\eta_p^2 = 0.20$. Planned comparisons showed that the CS+R produced significantly more suppression than the CS+NR at rehearsal test 1, $F(1, 44) = 13.48$, $p < 0.001$, $MSE = 0.008$. Comparable to Experiment 1, the SR for the CS+R remained intact after rehearsal, $F(1, 44) = 0.33$, $p = 0.57$, while responding to the CS+NR decreased from acquisition to rehearsal test 1, $F(1, 44) = 13.24$, $p < 0.001$, $MSE = 0.008$. Taken together, these data replicate the finding of a strengthened CR to a CS after mental CS-US-rehearsal. Data for the "CS-Rehearsal"-condition showed that the Phase

(Acquisition test/Rehearsal test) \times CS-type interaction was non-significant, $F(1, 44) = 0.29$, $p = 0.59$, indicating that CS-rehearsal did not impact responding. Planned comparisons provided additional support for this conclusion by showing no difference in SR between both CS+s at rehearsal test 1, $F(1, 44) = 0.03$, $p = 0.87$. Unexpectedly, responding did not show a significant decrement between acquisition test and rehearsal test 1 for both CS+s, $F_{CS+R}(1, 44) = 0.53$, $p = 0.47$; $F_{CS+NR}(1, 44) = 2.15$, $p = 0.15$, as was the case for rehearsal test average.

In sum, it seems that while rehearsal of the CS-US-compound sustains conditioned responding to the rehearsed CS+ as compared to the non-rehearsed CS+, this is not the case when only the CS+ is rehearsed.

DISCUSSION

Experiment 2 was set up to test whether the effects of Experiment 1 were due to reactivations of the CS-memory or of the CS-US-memory. First, the results of the “CS-US-Rehearsal”-group in Experiment 2 replicated the findings of Experiment 1. As such, this study provides additional evidence that mental reiteration of a CS-US-contingency strengthens conditioned responding to that CS+ relative to responding to a non-rehearsed CS+. In addition, the results from the “CS-Rehearsal”-group showed no difference between the rehearsed and the non-rehearsed CS+ and as such, no evidence for sensitization or incubation of responding after CS-rehearsal was provided. This points to the conclusion that mental repetition of only the CS is not sufficient to produce the effect of Experiment 1 (sustained CRs at test). Because Experiments 1 and 2 also showed no effect on the CS+ that was conditioned to the same US but not rehearsed (CS+_{NR}) in the CS-US-rehearsal groups, the observed effect should probably not be attributed to rehearsal of the US either. The main conclusion is that mental reiteration of a CS-US-contingency causes sustained conditioned responding, while reactivation of the CS or the US alone does not have this effect.

It is important to note that we cannot rule out the possibility that participants in the “CS-Rehearsal”-group might have also thought about the US during instructed CS-rehearsal. However, given the clear difference in instructions (and cueing on the screen) between both experimental groups, we believe that participants in the “CS-US-rehearsal”-condition at least thought more about the CS-US-compound than participants in the “CS-rehearsal”-condition.

An important question is to what extent our rehearsal manipulation might simply constitute additional acquisition trials. During the rehearsal phase, we presented the CS-picture and a verbal reference to the US. Although this procedure does not comprise experience with the actual US (sensory characteristics), it may produce additional learning of the mere CS-US-contingency. During “pure” rehearsal, these additional contingency experiences would be internally generated (thinking back of the co-occurrences of the CS and the US), whereas they were externally generated in Experiments 1 and 2. Moreover, repetitive thought does not only occur during presentations of the phobic stimuli (CS). Indeed, individuals often think about past conditioning experiences *in the absence* of the CS or US, so a purely mental rehearsal procedure would seem more akin to repetitive thought. Therefore, we

conducted a third experiment, which included a new condition in which participants were prompted by a neutral signal to rehearse the CS-US-contingency in a purely mental way, i.e., without being primed by the visual presence of the CS and a verbal reference to the US. This procedure was in line with the paradigm used by Davey and colleagues (Jones and Davey, 1990; Davey and Matchett, 1994). A non-differential rehearsal effect in the visually aided rehearsal condition as in the purely mental rehearsal condition would indicate that the observed rehearsal effect should not be attributed to additional training during rehearsal.

A second important note is that both in Experiments 1 and 2 the rehearsal effect seems partly driven by a *decrease* in CR to the non-rehearsed CS+, rather than by an *increase* in CR to the rehearsed CS+. This decrement may reflect the natural course of conditioned responding over time; rehearsal would then prevent this spontaneous decrease in responding (see General Discussion). However, it is also possible that the CS-US-rehearsal trials primarily produce their effect by reducing responding to the non-rehearsed CS+, such that rehearsal of one CS-US-contingency interferes with responding to the other CS+, which was presented during the same learning phase. Such interference has been shown before, for instance in studies by Pineno and colleagues (Matute and Pineno, 1998; Pineño et al., 2000) demonstrating impaired responding to X when X+ training was followed by A+ training. Similarly, in our studies the decrease in CR to the non-rehearsed CS+ can be considered as the result of stimulus competition between elementally trained CSs evoked by mental rehearsal trials pairing the rehearsed CS+ and the US. A related memory phenomenon is *retrieval-induced forgetting*, which refers to the situation where retrieval of a subset of formerly studied material (e.g., CS+_R – US) causes subsequent forgetting of the non-retrieved material (e.g., CS+_{NR} – US; Bäuml et al., 2010, p. 1048). In a recent study by Ortega-Castro and Vadillo (2013), retrieval-induced forgetting was demonstrated using word pairs, where several cues predicted a common outcome. As such, rehearsal/retrieval of one CS-US-contingency might induce forgetting of the non-rehearsed contingency.

In order to evaluate this possibility, we included an extra control group in Experiment 3 who “rehearsed” an irrelevant picture-word pair after acquisition. This group will show the natural course of responding to a CS+ from acquisition to test. A significant decline in CRs for both CS+s in this group would indicate that the observed decrease in CRs to the non-rehearsed CS+ should not be attributed to interference.

EXPERIMENT 3

Experiment 3 comprised three experimental groups. The first group, “Visual Rehearsal,” was largely a replication of the CS-US-rehearsal groups in the previous experiments, including visually aided CS-US-rehearsal. The second group, “Mental Rehearsal,” entailed a purely mental rehearsal procedure, without any visual guidance (except during instructions). Finally, participants in the “Control”-condition rehearsed an unrelated picture-word pair. We expected sustained suppression to the rehearsed CS+, but a decrease in responding to the non-rehearsed CS+ in both rehearsal groups. Moreover, a decline in CRs to both CS+s was expected in the “Control”-condition.

MATERIALS AND METHODS

Participants

Eighty students participated either in return for course credits or as paid volunteer. They provided informed consent and were uninformed about the purpose of the study. Participants were randomly allocated to either the “Visual Rehearsal”-condition ($n = 27$), the “Mental Rehearsal”-condition ($n = 27$) or the “Control”-condition ($n = 26$). The data of two participants (from “Mental Rehearsal”-group and “Control”-condition) had to be excluded due to apparatus failure. The remaining 78 participants (63 women) had a mean age of 19.63 (SD = 2.46; range 17–34).

Procedure

The apparatus, software, and stimuli were again identical as in Experiment 1. Moreover, the same procedure was used, with only the Rehearsal phase differing from the previous studies. During this phase, all participants received instructions to rehearse the co-occurrence of two related stimuli, using the attention training cover story (with same parameters). In both rehearsal conditions (Visual/Mental Rehearsal), participants received the same instructions as in the CS-US-rehearsal conditions from the previous experiments asking them to focus their attention on one of the background pictures (CS+_R) and how it co-occurred with the anti-laser shield. A slide with the CS and a verbal reference to the US was additionally presented on the computer screen to ensure that all participants were aware of the stimuli on which to “focus their attention.” While participants in the “Visual Rehearsal”-condition were presented with this CS-picture and a verbal reference to the anti-laser shield during the rehearsal phase, participants in the “Mental Rehearsal”-group saw only an exclamation mark, prompting them to a purely mental repetition of the conditioning stimuli. Participants in the “Control”-condition were also requested to focus their attention on a picture, a word and how they co-occurred. To ensure that they had equal visual experience with the CS-picture and the anti-laser shield as participants in the “Mental Rehearsal”-group, control participants were presented with a CS+ -picture and a verbal reference to the anti-laser shield as an example of a possible picture-word pair they could encounter in the following phase. Subsequently, it was further clarified that they had to focus on clouds and how these co-occurred with rain. A visual display of a picture of clouds and the word “rain” was presented during these instructions and during the following Rehearsal phase.

After each training phase, participants rated how difficult it was to focus their attention. In between both training phases, three questionnaires were administered. The Rehearsal phase and the Rehearsal Test phase were separated by an unrelated computer task (causal learning task).

RESULTS

Manipulation check

The attention score during the first and second attention training task and the score when averaged over both tasks (see **Table 1**) did not significantly differ between groups, as evidenced by the absence of an effect of group in several one-way ANOVAs with Group

as between-subjects variable, $F_{\text{training 1}}(2, 75) = 2.89$, $p = 0.06$, $F_{\text{training 2}} < 1$, $p = 0.57$, $F_{\text{training 1+2}}(2, 75) = 1.50$, $p = 0.23$. As in Experiments 1 and 2, participants who obtained a negative attention score averaged over both training phases were excluded. This was the case for five participants (18.52%) in the “Visual Rehearsal”-condition, seven participants (26.92%) in the “Mental Rehearsal”-condition and four participants (16.00%) in the “Control”-condition.

Acquisition test

Figure 4 depicts SRs for each condition as a function of Phase and CS-type. As can be seen in the graph, participants show successful differential acquisition with higher SRs to the CS–_{Average} than to the CS+_s. However, the level of responding to the CS+_s after acquisition seems to differ according to the experimental group. This is corroborated using a 3×3 -repeated measures ANOVA with Group (“Visual Rehearsal”/“Mental Rehearsal”/“Control”) as between-subjects variable and CS-type (CS–_{Average}/ CS+_R/ CS+_{NR}) as within-subjects variable. This ANOVA yielded a main effect of CS-type, $F(2, 118) = 125.25$, $p < 0.0001$, $\text{MSE} = 0.005$, $\eta_p^2 = 0.68$, which was qualified by a significant Group \times CS-type interaction, $F(4, 118) = 2.76$, $p < 0.05$, $\text{MSE} = 0.005$, $\eta_p^2 = 0.09$.

Further analyses showed that participants in the “Visual Rehearsal”-condition demonstrated less suppression to CS–_{Average} than to CS+_R, $F(1, 59) = 58.40$, $p < 0.0001$, $\text{MSE} = 0.006$, and CS+_{NR}, $F(1, 59) = 84.89$, $p < 0.0001$, $\text{MSE} = 0.006$. Both CS+_s evoked a non-differential amount of suppression, $F(1, 59) = 2.44$, $p = 0.12$. The same pattern was evident for participants in the “Control”-condition, with higher SRs to both CS+_s than to CS–_{Average}, $F_{\text{CS+R}}(1, 59) = 80.53$, $p < 0.0001$, $\text{MSE} = 0.006$; $F_{\text{CS+NR}}(1, 59) = 66.25$, $p < 0.0001$, $\text{MSE} = 0.006$, and a non-differential level of responding to both CS+_s, $F(1, 59) = 1.61$, $p = 0.21$. However, while participants in the “Mental Rehearsal”-condition again demonstrated more suppression to the CS+_s than to the CS–_{Average}, $F_{\text{CS+R}}(1, 59) = 21.80$, $p < 0.0001$, $\text{MSE} = 0.006$; $F_{\text{CS+NR}}(1, 59) = 45.40$, $p < 0.0001$, $\text{MSE} = 0.006$, they showed significantly more suppression to the CS+_{NR} than to the CS+_R, $F(1, 59) = 5.01$, $p < 0.05$, $\text{MSE} = 0.004$, which is unexpected given that the procedure was identical for both CS+_s until then.

Rehearsal test

Figure 4 suggests a general decrease in CRs to both CS+_s in the “Control”-condition and a smaller decrease in conditioned responding to the CS+_R than to the CS+_{NR} after CS-US-rehearsal in both rehearsal groups, which is in line with our hypotheses. This was supported by a 3 (Group: “Visual Rehearsal”/“Mental Rehearsal”/“Control”) $\times 2$ (Phase: Acquisition test/Rehearsal test average) $\times 2$ (CS-type: CS+_R/ CS+_{NR})-repeated measures ANOVA. As before, data of the three test trials were combined. This analysis revealed a significant Phase \times CS-type interaction, $F(1, 59) = 7.46$, $p < 0.01$, $\text{MSE} = 0.003$, $\eta_p^2 = 0.11$, that subsumed under a significant Group \times Phase \times CS-type interaction, $F(1, 59) = 3.27$, $p < 0.05$, $\text{MSE} = 0.003$, $\eta_p^2 = 0.10$, indicating that the course of responding to both CS+_s was differentially influenced by the rehearsal manipulation depending on the experimental group.

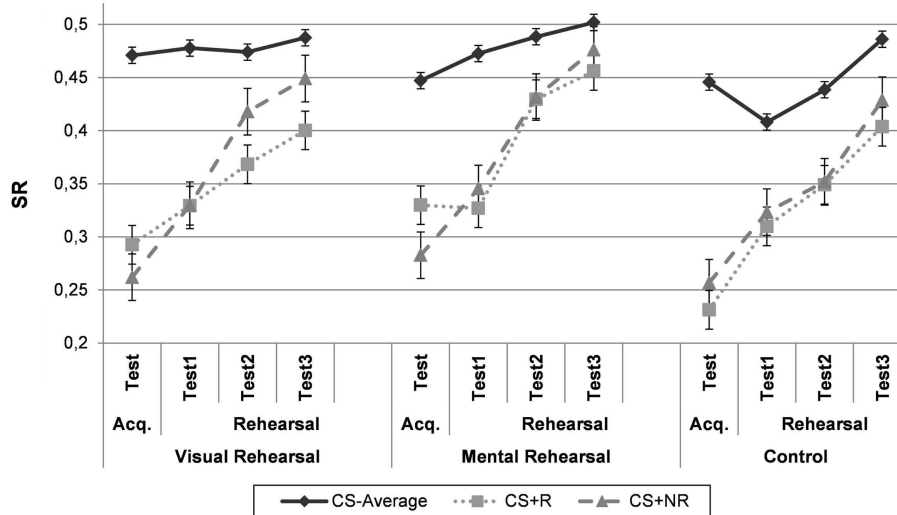


FIGURE 4 | Experiment 3. Mean suppression ratio's for the "Visual Rehearsal"-group, who rehearsed the CS-US-association with visual guidance, the "Mental Rehearsal"-group, who rehearsed the CS-US-contingency without visual guidance, and the "Control"-condition, who rehearsed an irrelevant picture-word pair, as

a function of phase (acquisition test, rehearsal test 1, rehearsal test 2, and rehearsal test 3) and CS-type (CS-Average, CS+R, and CS+NR). CS-Average = the average for both CS-'s. CS+R = rehearsed CS+, CS+NR = non-rehearsed CS+, Acq. Test = acquisition test. Error bars denote standard error.

The three-way interaction was further explored using simple interactions and planned comparisons. In the "Control"-condition, conditioned suppression significantly decreased at the same rate for both CS+s during rehearsal of an unrelated picture-word pair, $F_{CS+R}(1, 59) = 36.74$, $p < 0.0001$, $MSE = 0.004$; $F_{CS+NR}(1, 59) = 24.72$, $p < 0.0001$, $MSE = 0.005$, as evidenced by a non-significant Phase \times CS-type interaction, $F(1, 59) = 0.25$, $p = 0.62$. In contrast, in the "Visual Rehearsal"-condition, the Phase \times CS-type interaction was significant, $F(1, 59) = 7.62$, $p < 0.01$, $MSE = 0.003$, indicating a stronger decrease in suppression for the CS+NR than for the CS+R. Responding to both CS+s decreased from acquisition to test, but this decrease was larger for the CS+NR, $F(1, 59) = 39.28$, $p < 0.0001$, $MSE = 0.005$, than for the CS+R, $F(1, 59) = 13.73$, $p < 0.001$, $MSE = 0.004$. These results point toward an effect of visually aided CS-US-rehearsal on subsequent CRs. Likewise, the Phase \times CS-type interaction in the "Mental Rehearsal"-condition also reached significance, $F(1, 59) = 6.03$, $p < 0.05$, $MSE = 0.003$, with a smaller decrease in CRs for the CS+R, $F(1, 59) = 12.17$, $p < 0.001$, $MSE = 0.004$, than for the CS+NR, $F(1, 59) = 33.04$, $p < 0.0001$, $MSE = 0.005$.

To investigate whether visually aided CS-US-rehearsal had a different impact on responding than purely mental CS-US-rehearsal, the Group ("Visual Rehearsal"/"Mental Rehearsal") \times Phase \times CS-type interaction was assessed. This interaction failed to reach significance, $F(1, 59) = 0.007$, $p = 0.94$, suggesting that both forms of rehearsal influenced responding in the same way. Furthermore, given our hypothesis that the decline in responding to the CS+NR after CS-US-rehearsal reflects a natural course of responding, the 3 (Group) \times 2 (Phase) interaction was assessed for the CS+NR. This interaction was non-significant, $F(2, 59) = 0.41$, $p = 0.66$, indicating that responding to

the CS+NR decreases to the same extent after CS-US-rehearsal as after rehearsal of an irrelevant picture-word pair.

As the first test trial is considered the most valid one, the data were also analyzed when taking into account only the change in responding from post-acquisition to the first rehearsal test trial. A 3 (Group: "Visual Rehearsal"/"Mental Rehearsal"/"Control") \times 2 (Phase: Acquisition test/Rehearsal test 1) \times 2 (CS-type: CS+R/CS+NR)-repeated measures ANOVA revealed a non-significant Group \times Phase \times CS-type interaction, $F(1, 59) = 1.25$, $p = 0.29$. Further analyses for the three conditions separately revealed that for participants in the "Control"-condition, the Phase \times CS-type interaction was not significant, $F(1, 59) = 0.13$, $p = 0.72$. Both CS+s evoked significantly less suppression at rehearsal test 1 than at acquisition test, $F_{CS+R}(1, 59) = 12.07$, $p < 0.001$, $MSE = 0.005$, $F_{CS+NR}(1, 59) = 7.57$, $p < 0.01$, $MSE = 0.006$, indicating that rehearsal did not impact responding. In the "Mental Rehearsal"-group, responding to the CS+NR significantly decreased between acquisition and rehearsal test 1, $F(1, 59) = 6.13$, $p < 0.05$, $MSE = 0.006$, while no such decrease was present for the CS+R, $F(1, 59) = 0.02$, $p = 0.90$, suggesting sustained responding to the CS+R. The Phase \times CS-type interaction was only marginally significant, $F(1, 59) = 3.37$, $p = 0.07$, $MSE = 0.006$. The same pattern emerged for participants in the "Visual Rehearsal"-group. Again, CRs significantly decreased for the CS+NR during the rehearsal phase, $F(1, 59) = 8.25$, $p < 0.01$, $MSE = 0.006$, but not for the CS+R, $F(1, 59) = 2.78$, $p = 0.10$. The Phase \times CS-type interaction was however not significant, $F(1, 59) = 0.86$, $p = 0.36$.

Based on visual inspection of Figure 4, we tested whether the Phase \times CS-type interaction differed between both rehearsal groups, given that the rehearsal effect on the first test trial seems more pronounced in the "Mental Rehearsal"-condition.

The Group (“Visual/Mental Rehearsal”) \times Phase \times CS-type interaction failed to reach significance, $F(1, 59) = 0.51$, $p = 0.48$, indicating a non-differential effect of visually aided and purely mental rehearsal.

Given that the pattern of results in **Figure 4** suggests a delayed effect of rehearsal in the “Visual Rehearsal”-group, we examined the change in responding between acquisition test and the final test trial for both CS+s using a 3 (Group) \times 2 (Phase: Acquisition test/Rehearsal test 3) \times 2 (CS-type) ANOVA. In line with the previous findings, the same rate of CR decrease was observed for both CS+s in the “Control”-condition, $F_{\text{CS+R}}(1, 59) = 47.89$, $p < 0.0001$, $\text{MSE} = 0.007$, $F_{\text{CS+NR}}(1, 59) = 41.75$, $p < 0.0001$, $\text{MSE} = 0.007$, with a non-significant Phase \times CS-type interaction, $F(1, 59) = 0.0003$, $p = 0.99$. In contrast, in both rehearsal groups, suppression declined significantly stronger for the CS+NR, $F_{\text{Mental Rehearsal}}(1, 59) = 47.84$, $p < 0.0001$, $\text{MSE} = 0.007$; $F_{\text{Visual Rehearsal}}(1, 59) = 51.78$, $p < 0.0001$, $\text{MSE} = 0.007$, than for the CS+R, $F_{\text{Mental Rehearsal}}(1, 59) = 23.25$, $p < 0.0001$, $\text{MSE} = 0.007$; $F_{\text{Visual Rehearsal}}(1, 59) = 19.57$, $p < 0.0001$, $\text{MSE} = 0.007$, as evidenced by a significant Phase \times CS-type interaction in both the “Mental Rehearsal,” $F(1, 59) = 5.73$, $p < 0.05$, $\text{MSE} = 0.004$, and the “Visual Rehearsal”-condition, $F(1, 59) = 9.25$, $p < 0.005$, $\text{MSE} = 0.004$. The overall Group \times Phase \times CS-type interaction was however only marginally significant, $F(1, 59) = 2.60$, $p = 0.08$, $\text{MSE} = 0.004$. In sum, although suggested by the data pattern, no strong evidence exists that visually aided CS-US-rehearsal has a more delayed effect on responding than purely mental rehearsal.

DISCUSSION

The data pattern in the “Visual Rehearsal”-group replicated the findings of Experiments 1 and 2. After rehearsal of the CS+R, suppression to the CS+NR decreased significantly stronger than suppression to the CS+R. Moreover, responding to the CS+R seemed to persist longer as evidenced by a non-significant decline from post-acquisition to the first rehearsal test, while this decrease was significant for the CS+NR. Importantly, the same pattern emerged for the “Mental Rehearsal”-condition. Again, the decrease in responding was significantly stronger for the CS+NR than for the CS+R and responding to CS+R sustained on the first rehearsal test trial. This suggests that the rehearsal effects of Experiments 1 and 2 should not be attributed to the fact that the rehearsal trials simply constitute additional acquisition trials. Rehearsal impacts responding both when the to-be-rehearsed information is externally or internally generated.

Importantly, the crucial Group \times Phase \times CS-type interaction was significant, demonstrating an effect of rehearsal in both CS-US-rehearsal groups, but not in the “Control”-group. However, three important notes should be made. First, in the “Mental Rehearsal”-condition, responding to both CS+s already differed at the end of the acquisition phase, which is unexpected given that both CS+s underwent the exact same procedure until then. Although responding to the CS+R shows a slower decrease than responding to the CS+NR, the CS+R does not evoke significantly more suppression after rehearsal than the CS+NR, which might be attributed to this unexpected post-acquisition difference. Second, because of the difference between both rehearsal conditions in baseline responding to the CS+s the non-significance of the

Group (Visual Rehearsal/Mental Rehearsal) \times CS-type interaction should be interpreted with caution. Third, in contrast to Experiments 1 and 2, the rehearsal effect is most strongly present for the overall rehearsal test and to a lesser extent for the first test trial only.

The significant decline for both CS+s in the “Control”-condition seems to demonstrate that the natural course of conditioned suppression is to decrease over time, rather than to persist at the same level. This suggests that the decrease in responding to the CS+NR during CS-US-rehearsal in the current and the previous experiments, should not be attributed to some kind of interference from the CS+R.

A final important remark is that more participants had to be excluded because of a negative attention score than in the previous experiments. Although the number of participants with a negative score was not significantly associated with the experimental group, $\chi^2(2) = 1.033$, $p = 0.60$, this exclusion was especially remarkable in the “Mental Rehearsal”-condition, where 26.92% of the participants were omitted. Presumably, this should be attributed to the less concrete nature of the rehearsal task in this group, where only an exclamation mark was presented.

GENERAL DISCUSSION

In human conditioning studies, relatively little attention has been devoted to the processing of a memory trace after its initial acquisition. In an attempt to explore the role of active post-acquisition processing in conditioning, we experimentally induced repeated activation of a CS-US-contingency memory and tested whether this impacted conditioned responding at test. Experiment 1 showed that rehearsing a previously acquired CS-US-contingency leads to stronger CRs to the rehearsed than to a conditioned, but non-rehearsed CS+. In Experiment 2 this effect was replicated and, in addition, it was shown that no such difference occurred when rehearsal was focused on the CS alone (rather than the CS-US-contingency). Experiment 3 demonstrated that the same rehearsal effect is found regardless of how mental rehearsal is induced. Priming rehearsal through visual presentation of the CS-picture and a verbal reference to the US (as in Experiments 1 and 2) has the same effect as a purely mental procedure. Moreover, results of the “Control”-condition suggest that the natural course of responding to CS+s after acquisition (and during an irrelevant rehearsal task) is a decrement in suppression. An important limitation of Experiment 3 is formed by the post-acquisition differences in responding between the conditions. While the rehearsed and the non-rehearsed CS+ evoked the same amount of responding in the “Visual Rehearsal” and the “Control”-condition, this was not the case in the “Mental Rehearsal”-condition. The reason for this unexpected difference is unclear, but it might complicate interpretation of our findings. However, overall, the data show that repeatedly activating the memory trace of a CS-US co-occurrence impacts subsequent conditioned responding, even in the absence of direct experience with the phenomenological aspects of CS-US-pairings. Interestingly, a recurrent finding is that rather than increasing the level of responding, CS-US-rehearsal results in persistent CRs, while the absence of mental activation causes responding to decline. We will discuss this issue in more detail below.

Based on Experiments 1 and 2 an important question was to what extent the rehearsal manipulation entailed learning processes rather than memory processes. Indeed, the procedure of repeatedly activating the CS-US-contingency might have constituted additional *training* trials, given that the CS-US-contingency was partially presented. However, the results of Experiment 3 showed that purely mental rehearsal, which was not cued by the CS and a verbal reference to the US and was thus internally generated, affected conditioned responding in a similar way.

This issue demonstrates that research on rehearsal effects in conditioning is located at the interface of learning and memory. Two alternative positions exist regarding the interrelation of these two processes. One perspective is that learning occurs only when external input is present, while memory processes pertain to internally generated input. In that case, learning occurred in Experiments 1 and 2, while the rehearsal manipulation in Experiment 3 elicited a memory process, rather than a learning process. An alternative viewpoint is to define learning as a change in behavior due to experience (e.g., Bower and Hilgard, 1981). In this perspective, learning occurred in all three experiments, as the results showed a change in conditioned behavior compared to when no repeated activation was induced. In sum, both learning and memory seem crucial in understanding conditioning effects (cf. Bouton and Moody, 2004) and their interplay is an important but generally ignored topic.

The current research ties up with an increasing body of research demonstrating that mental representations of a conditioning stimulus can influence conditioned responding in the absence of the physical stimulus. An overview of this literature is provided by Dadds et al. (1997), Holland (1990), and Pickens and Holland (2004). In short, there are two important lines of evidence for the fact that activation of the mental representation of a stimulus can induce learning to that stimulus. First, in US-revaluation studies it is found that conditioned responding decreases/increases after devaluation/inflation of the US without directly experiencing the CS-US-contingency (e.g., Rescorla, 1974; White and Davey, 1989; de Jong et al., 1996). Second, in studies on representation-mediated learning (Holland, 1990; Pickens and Holland, 2004) it is typically shown that an associatively activated stimulus representation can substitute for actually presented stimuli. Besides demonstrating that an association may be formed with a stimulus when the stimulus is not presented, it is also demonstrated that associations may be formed between two stimuli even when *both* of the stimuli are absent, rather than only one of them (e.g., in animals: Holland and Sherwood, 2008; in humans: Le Pelley and McLaren, 2001).

Besides the theoretical importance of our results in bringing research traditions on memory and learning closer together, the idea of mental rehearsal is clinically relevant as well. Overall, the data indicate that not only the acquisition experiences themselves, but also the way in which one cognitively engages in the memories of these events, has an impact on conditioned responding. Clinical observations suggest that individuals differ in their tendency to engage in repetitive thought such as worry and rumination. Repetitive thought is defined as “the process of thinking attentively, repetitively, or frequently about one’s self and one’s world” (Segerstrom et al., 2003, p. 909). Hence, this

variable can be considered as a form of active rehearsal. After experiencing a traumatic event, rehearsal of this negative event together with associated stimuli might strengthen the acquired CS-US-association. Previous work by Otto et al. (2007), as well as a more recent study in our laboratory (Joos et al., 2012c), points in that direction. Otto et al. (2007) found trait worry to be a good predictor of the strength of fear acquisition. We replicated the finding that individuals with a higher level of worry demonstrated more pronounced fear acquisition. Moreover, this association could not be explained by trait anxiety (Joos et al., 2012b). One way to explain this relation between worry and conditioning strength is that the high trait-worriers mentally repeat the CS-US-contingency during acquisition and therefore show stronger conditioned responding in the fear conditioning task. Of course, post-acquisition mental repetition of the fear memory is only one route that might play a role. As such, these studies do not provide direct evidence of the impact of rehearsal after acquisition and differ in this respect from the experimental studies presented in this paper.

Given the conclusion that post-acquisition rehearsal impacts conditioned responding, an important question pertains to the exact processes that are responsible for this effect. A first candidate in explaining the results is *consolidation*, which refers to the progressive post-acquisition strengthening of memory traces in long-term memory (Dudai, 2004). Repeated activation of the CS-US-contingency might strengthen the association between the mental representations of the CS and the US in memory, resulting in a cognitive consolidation of this memory trace.

A second mechanism focuses on the decrease in conditioned responding to the non-rehearsed CS+. As suggested before, the CS-US-rehearsal trials might *interfere* with responding to the non-rehearsed CS+ at test. However, the results of the “Control”-condition of Experiment 3 indicate that the recurrently observed decrease in responding to the non-rehearsed CS+ is the natural course of responding and should therefore not be attributed to interference by a related stimulus (CS+_R), as the same reduction in CRs is evident when no related stimulus was rehearsed.

That the natural course of conditioned suppression after rehearsal is to decrease, rather than to persist at the same level supports the likelihood of a third possible underlying mechanism, i.e., *prevention-of-forgetting*. Indeed, this decrease might be considered as a type of forgetting. Hence, mentally rehearsing a CS-US-contingency might prevent forgetting of this memory trace through repeated activation. This idea is in line with the notion that rehearsal prevents the loss of information in short-term memory (e.g., Atkinson and Shiffrin, 1971; Portrat et al., 2008). More specifically, as forgetting in conditioning probably relates to a lack of accessibility of the memory trace rather than a loss of information (Anderson, 2000; Bouton, 2004), rehearsal might counteract a spontaneous decrease in accessibility of the memory trace. In all three reported experiments, we see a decrease in CR strength to the non-rehearsed CS+ between acquisition and test, supporting the claim that participants “forget” this association to some extent. For the “CS-Rehearsal”-group in Experiment 2, conditioned responding to both CS+s also decreases between acquisition and test, but this decrease fails to reach statistical significance on the first test trial. However, the non-significant

Group \times Phase (Acquisition test/Rehearsal test 1) interaction for the non-rehearsed CS+ suggests that both groups show the same decreasing pattern of responding to the non-rehearsed CS+. Overall, our data seem to support the notion that rehearsal renders the CS-US-memory more accessible, resulting in stronger CRs upon subsequent CS-presentations compared to presentations of the non-rehearsed CS.

In conclusion, the present studies show that repeated post-acquisition activation of a CS-US-contingency sustains conditioned responding. Through experimental induction of post-acquisition CS-US-repetition, it was shown that active post-encoding processes such as rehearsal, which are frequently studied

in the memory literature, might play an important role in conditioning as well. In particular, long-term conditioning effects may be largely influenced by such memory processes. Additionally, our results indicate that individual differences in the tendency to reflect upon past experiences, as in worry or rumination, might create differences in conditioned responding, due to varying levels of post-acquisition activation of the CS-US-memory.

ACKNOWLEDGMENTS

Els Joos is research assistant for the Research Foundation – Flanders (FWO-Vlaanderen). The research was supported by GOA funding (GOA/2007/03; 3H051018).

REFERENCES

- Anderson, J. R. (1999). *Learning and Memory: An Integrated Approach*, 2nd Edn. New York, NY: Wiley.
- Anderson, J. R. (2000). *Cognitive Psychology and Its Implications*, 5th Edn. New York, NY: Worth Publishing.
- Arceidiano, F., Ortega, N., and Matute, H. (1996). A behavioural preparation for the study of human Pavlovian conditioning. *Q. J. Exp. Psychol.* 49B, 270–283. doi:10.1080/713932633
- Arntz, A., Spit, S., and Merckelbach, H. (1997). Cued UCS rehearsal and the impact of painful conditioned stimuli: UCS rehearsal increases SCR's but reduces experienced pain. *Behav. Res. Ther.* 35, 591–605. doi:10.1016/S0005-7967(97)00020-X
- Atkinson, R. C., and Shiffrin, R. M. (1971). The control processes of short-term memory. *Sci. Am.* 224, 82–90. doi:10.1038/scientificamerican0871-82
- Baeyens, F., and Clarysse, J. (1998). *Martians for Windows™95 [Computer Program]*. Leuven: University of Leuven.
- Baeyens, F., Vansteenwegen, D., Hermans, D., Vervliet, B., and Eelen, P. (2001). Sequential and simultaneous feature positive discriminations: occasion setting and configural learning in human Pavlovian conditioning. *J. Exp. Psychol. Anim. Behav. Process.* 27, 279–295. doi:10.1016/S0023-9690(03)00058-4
- Bäumli, K.-H., Pastötter, B., and Hanslmayer, S. (2010). Binding and inhibition in episodic memory – cognitive, emotional, and neural processes. *Neurosci. Biobehav. Rev.* 34, 1047–1054. doi:10.1016/j.neubiorev.2009.04.005
- Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learn. Mem.* 11, 485–494. doi:10.1101/lm.78804
- Bouton, M. E. (2007). *Learning and Behavior: A Contemporary Synthesis*. Sunderland, MA: Sinauer Associates.
- Bouton, M. E., and Moody, E. W. (2004). Memory processes in classical conditioning. *Neurosci. Biobehav. Rev.* 28, 663–674. doi:10.1016/j.neubiorev.2004.09.001
- Bower, G. H., and Hilgard, E. R. (1981). *Theories of Learning*, 5th Edn. Englewood Cliffs, NJ: Prentice Hall.
- Brown, K. W., and Ryan, R. M. (2003). The benefits of being present: mindfulness and its role in psychological well-being. *J. Pers. Soc. Psychol.* 84, 822–848. doi:10.1037/0022-3514.84.4.822
- Calmes, C. A., and Roberts, J. E. (2007). Repetitive thought and emotional distress: rumination and worry as prospective predictors of depressive and anxious symptomatology. *Cognit. Ther. Res.* 30, 343–356. doi:10.1007/s10608-006-9026-9
- Craske, M. G., Hermans, D., and Vansteenwegen, D. (eds). (2006). *Fear and Learning: From Basic Processes to Clinical Implications*. Washington, DC: American Psychological Association Books.
- Dadds, M. R., Bovbjerg, D. H., Redd, W. H., and Cutmore, T. R. H. (1997). Imagery in human classical conditioning. *Psychol. Bull.* 122, 89–103. doi:10.1037/0033-2909.122.1.89
- Davey, G. C. L., and Matchett, G. (1994). Unconditioned stimulus rehearsal and the retention and enhancement of differential “fear” conditioning: effects of trait and state anxiety. *J. Abnorm. Psychol.* 103, 708–718. doi:10.1016/S0005-7967(97)00020-X
- de Jong, P., Muris, P., and Merckelbach, H. (1996). UCS inflation and human aversive autonomic conditioning. *Biol. Psychol.* 43, 227–239. doi:10.1016/0301-0511(96)05190-3
- Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annu. Rev. Psychol.* 55, 51–86. doi:10.1146/annurev.psych.55.090.902.142050
- Ebbinghaus, H. (1885/1913). *Memory: A Contribution to Experimental Psychology*. Available at: <http://psychclassics.yorku.ca/Ebbinghaus>
- Ehring, T., Zetsche, U., Weidacker, K., Wahl, K., Schönfeld, S., and Ehlers, A. (2011). The Perseverative Thinking Questionnaire (PTQ): validation of a content-independent measure of repetitive negative thinking. *J. Behav. Ther. Exp. Psychiatry* 42, 225–232. doi:10.1016/j.jbtep.2010.12.003
- Eysenck, H. J. (1968). A theory of the incubation of anxiety/fear responses. *Behav. Res. Ther.* 6, 309–321. doi:10.1016/0005-7967(68)90064-8
- Field, A. P. (2006). Is conditioning a useful framework for understanding the development and treatment of phobias? *Clin. Psychol. Rev.* 26, 857–875. doi:10.1016/j.cpr.2005.05.010
- Franssen, M., Clarysse, J., Beckers, T., van Vooren, P., and Baeyens, F. (2010). A free software package for a human online conditioned suppression preparation. *Behav. Res. Methods* 42, 311–317. doi:10.3758/BRM.42.1.311
- Fresco, D. M., Frankel, A. N., Menin, D. S., Turk, C. L., and Heimberg, R. G. (2002). Distinct and overlapping features of rumination and worry: the relationship of cognitive production to negative affective states. *Cognit. Ther. Res.* 26, 179–188. doi:10.1023/A:1014517718949
- Groves, P. M., and Thompson, R. F. (1970). Habituation: a dual-process theory. *Psychol. Rev.* 77, 419–450. doi:10.1037/h0029810
- Hermans, D., Craske, M. G., Mineka, S., and Lovibond, P. F. (2006). Extinction in human fear conditioning. *Biol. Psychiatry* 60, 361–368. doi:10.1016/j.biopsych.2005.10.006
- Holland, P. C. (1990). Event representations in pavlovian conditioning: image and action. *Cognition* 37, 105–131. doi:10.1016/0010-0277(90)90020-K
- Holland, P. C., and Sherwood, A. (2008). Formation of excitatory and inhibitory associations between absent events. *J. Exp. Psychol. Anim. Behav. Process.* 34, 324–335. doi:10.1037/0097-7403.34.3.324
- Hong, R. Y. (2007). Worry and rumination: differential associations with anxious and depressive symptoms and coping behavior. *Behav. Res. Ther.* 45, 277–290. doi:10.1016/j.brat.2006.03.006
- Johnson, R. E. (1980). “Memory-based rehearsal,” in *Psychology of Learning and Motivation*, Vol. 14, ed. G. H. Bower (New York: Academic Press), 263–307. doi:10.1016/S0079-7421(08)60163-2
- Jones, T., and Davey, G. C. (1990). The effects of cued UCS rehearsal on the retention of differential fear conditioning: an experimental analogue of the worry process. *Behav. Res. Ther.* 28, 159–164. doi:10.1016/0005-7967(90)90028-H
- Joos, E., Vansteenwegen, D., Brunfaut, E., Bastiaens, T., Demyttenaere, K., Pieters, G., et al. (2012a). The Penn State Worry Questionnaire – past day: development and validation of a measure assessing daily levels of worry. *J. Psychopathol. Behav. Assess.* 34, 35–47. doi:10.1007/s10862-011-9265-2
- Joos, E., Vansteenwegen, D., and Hermans, D. (2012b). Post-acquisition repetitive thought in fear conditioning: an experimental investigation of the effect of CS-US-rehearsal. *J. Behav. Ther. Exp. Psychiatry* 43, 737–744. doi:10.1016/j.jbtep.2011.10.011
- Joos, E., Vansteenwegen, D., and Hermans, D. (2012c). Worry

- as a predictor of fear acquisition in a non-clinical sample. *Behav. Modif.* 36, 723–750. doi:10.1177/0145445512446477
- Kaloupek, D. G. (1983). The effects of compound in vivo and imaginal exposure: a test of fear enhancement models. *Behav. Ther.* 14, 345–356. doi:10.1016/S0005-7894(83)80098-7
- Le Pelley, M. E., and McLaren, I. P. L. (2001). Retrospective revaluation in humans: learning or memory? *Q. J. Exp. Psychol.* 54B, 311–352. doi:10.1080/713932762
- Levey, A., and Martin, I. (1981). “Personality and conditioning,” in *A Model for Personality*, ed. H. J. Eysenck (Berlin: Springer), 123–168.
- Matute, H., and Pineno, O. (1998). Stimulus competition in the absence of compound conditioning. *Anim. Learn. Behav.* 26, 3–14. doi:10.3758/BF03199157
- McLaughlin, K. A., Borkovec, T. D., and Sibrava, N. J. (2007). The effects of worry and rumination on affect states and cognitive activity. *Behav. Ther.* 38, 23–38. doi:10.1016/j.beth.2006.03.003
- Meyer, T. J., Miller, M. L., Metzger, R. L., and Borkovec, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behav. Res. Ther.* 28, 487–495. doi:10.1016/0005-7967(90)90135-6
- Mineka, S., and Oehlberg, K. (2008). The relevance of recent developments in classical conditioning to understanding the etiology and maintenance of anxiety disorders. *Acta Psychol. (Amst.)* 127, 567–580. doi:10.1016/j.actpsy.2007.11.007
- Mineka, S., and Zinbarg, R. (2006). A contemporary learning theory perspective on the etiology of anxiety disorders: it's not what you thought it was. *Am. Psychol.* 61, 10–26. doi:10.1037/0003-066X.61.1.10
- Muris, P., Roelofs, J., Meesters, C., and Boomsma, P. (2004). Rumination and worry in nonclinical adolescents. *Cognit. Ther. Res.* 28, 539–554. doi:10.1023/B:COTR.0000045563.66060.3e
- Nicholaichuk, T. P., Quesnel, L. J., and Tait, R. W. (1982). Eysenck's theory of incubation: an empirical test. *Behav. Res. Ther.* 20, 329–338. doi:10.1016/0005-7967(82)90092-4
- Nolen-Hoeksema, S. (2000). The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. *J. Abnorm. Psychol.* 109, 504–511. doi:10.1037/10021-843X.109.3.504
- Ortega-Castro, N., and Vadillo, I. A. (2013). Retrieval-induced forgetting and interference between cues: training a cue–outcome association attenuates retrieval by alternative cues. *Behav. Processes* 94, 19–25. doi:10.1016/j.beproc.2012.11.010
- Otto, M. W., Leyro, T. M., Christian, K., Deveney, C. M., Reese, H., Pollack, M. H., et al. (2007). Prediction of “fear” acquisition in healthy control participants in a de novo fear-conditioning paradigm. *Behav. Modif.* 31, 32–51. doi:10.1177/0145445506295054
- Pickens, C. L., and Holland, P. C. (2004). Conditioning and cognition. *Neurosci. Biobehav. Rev.* 28, 651–661. doi:10.1016/j.neubiorev.2004.09.003
- Pineño, O., Ortega, N., and Matute, H. (2000). The relative activation of associations modulates interference between elementally trained cues. *Learn. Motiv.* 31, 128–152. doi:10.1006/lmot.1999.1047
- Portrat, S., Barrouillet, P., and Camos, V. (2008). Time-related decay or interference-based forgetting in working memory? *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 1561–1564. doi:10.1037/a0013356
- Rachman, S. (1991). Neo-conditioning and the classical theory of fear acquisition. *Clin. Psychol. Rev.* 11, 155–173. doi:10.1016/0272-7358(91)90093-A
- Rescorla, R. (1974). Effect of inflation of the unconditioned stimulus value following conditioning. *J. Comp. Physiol. Psychol.* 86, 101–106. doi:10.1037/h0035964
- Rohrbaugh, M., and Riccio, D. C. (1970). Paradoxical enhancement of learned fear. *J. Abnorm. Psychol.* 75, 210–216. doi:10.1037/h0028974
- Rohrbaugh, M., Riccio, D. C., and Arthur, A. (1972). Paradoxical enhancement of conditioned suppression. *Behav. Res. Ther.* 10, 125–130. doi:10.1016/S0005-7967(72)80005-6
- Sandin, B., and Chorot, P. (1989). The incubation theory of fear/anxiety: experimental investigation in a human laboratory model of Pavlovian conditioning. *Behav. Res. Ther.* 27, 9–18. doi:10.1016/0005-7967(89)90114-9
- Segerstrom, S. C., Stanton, A. L., Alden, L. E., and Shortridge, B. E. (2003). A multidimensional structure for repetitive thought: what's on your mind, and how, and how much? *J. Pers. Soc. Psychol.* 85, 909–921. doi:10.1037/0022-3514.85.5.909
- Segerstrom, S. C., Tsao, J. C. I., Alden, L. E., and Craske, M. G. (2000). Worry and rumination: repetitive thought as a concomitant and predictor of negative mood. *Cognit. Ther. Res.* 24, 671–688. doi:10.1023/A:1005587311498
- Sheehan, P. W. (1967). A shortened form of Betts' questionnaire upon mental imagery. *J. Clin. Psychol.* 23, 386–389. doi:10.1002/1097-4679(196707)23:3<386::AID-JCLP2270230328>3.0.CO;2-S
- Stevens, J. P. (2002). *Applied Multivariate Statistics for the Social Sciences*, 4th Edn. Hillsdale: Lawrence Erlbaum Associates.
- Stone, N. M., and Borkovec, T. D. (1975). The paradoxical effect of brief CS exposure on analogue phobic subjects. *Behav. Res. Ther.* 13, 51–54. doi:10.1016/0005-7967(75)90051-0
- Terry, W. S., and Wagner, A. R. (1975). Short-term memory for “surprising” versus “expected” unconditioned stimuli in Pavlovian conditioning. *J. Exp. Psychol. Anim. Behav. Process.* 104, 122–133. doi:10.1037/0097-7403.1.2.122
- Torrubia, R., Avila, C., Molto, J., and Caseras, X. (2001). The sensitivity to punishment and sensitivity to reward questionnaire (SPSRQ) as a measure of Gray's anxiety and impulsivity dimensions. *Pers. Individ. Dif.* 31, 837–862. doi:10.1016/S0191-8869(00)00183-5
- Wagner, A. R. (1981). “SOP: a model of automatic memory processing in animal behavior,” in *Information Processing in Animals, Memory Mechanisms*, eds N. E. Spear and R. R. Miller (Hillsdale, NJ: Erlbaum), 5–47.
- Wagner, A. R., Rudy, J. W., and Whitlow, J. W. (1973). Rehearsal in animal conditioning. *J. Exp. Psychol.* 97, 407–426. doi:10.1037/h0034136
- Watkins, E. R. (2008). Constructive and unconstructive repetitive thought. *Psychol. Bull.* 134, 163–206. doi:10.1037/0033-2909.134.2.163
- White, K., and Davey, G. C. L. (1989). Sensory preconditioning and UCS inflation in human fear conditioning. *Behav. Res. Ther.* 27, 161–166. doi:10.1016/0005-7967(89)90074-0
- Yaremko, R. M., and Werner, M. (1974). Cognitive conditioning: imagined stimulus contiguity and the third interval conditional GSR. *Integr. Psychol. Behav. Sci.* 9, 215–221. doi:10.1007/BF03001505

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 December 2012; accepted: 13 May 2013; published online: 30 May 2013.

Citation: Joos E, Vansteenwegen D, Vervliet B and Hermans D (2013) Repeated activation of a CS-US-contingency memory results in sustained conditioned responding. *Front. Psychol.* 4:305. doi: 10.3389/fpsyg.2013.00305
This article was submitted to *Frontiers in Personality Science and Individual Differences*, a specialty of *Frontiers in Psychology*.

Copyright © 2013 Joos, Vansteenwegen, Vervliet and Hermans. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.



Diminished acquired equivalence yet good discrimination performance in older participants

Jasper Robinson * and Emma Owens

School of Psychology, The University of Nottingham, Nottingham, UK

Edited by:

Robin A. Murphy, University of Oxford, UK

Reviewed by:

Irina Baetu, University of Adelaide, Australia

Sam Hannah, University of Saskatchewan, Canada

*Correspondence:

Jasper Robinson, School of Psychology, The University of Nottingham, Room B33, University Park, NG7 2RD Nottingham, UK
e-mail: jasper.robinson@nottingham.ac.uk

We asked younger and older human participants to perform computer-based configural discriminations that were designed to detect acquired equivalence. Both groups solved the discriminations but only the younger participants demonstrated acquired equivalence. The discriminations involved learning the preferences [“like” (+) or “dislike” (–)] for sports [e.g., tennis (t) and hockey (h)] of four fictitious people [e.g., Alice (A), Beth (B), Charlotte (C), and Dorothy (D)]. In one experiment, the discrimination had the form: At+, Bt–, Ct+, Dt–, Ah–, Bh+, Ch–, Dh+. Notice that, e.g., Alice and Charlotte are “equivalent” in liking tennis but disliking hockey. Acquired equivalence was assessed in ancillary components of the discrimination (e.g., by looking at the subsequent rate of “whole” versus “partial” reversal learning). Acquired equivalence is anticipated by a network whose hidden units are shared when inputs (e.g., A and C) signal the same outcome (e.g., +) when accompanied by the same input (t). One interpretation of these results is that there are age-related differences in the mechanisms of configural acquired equivalence.

Keywords: acquired equivalence, attentional set, ageing, discrimination learning, connectionism, associative learning, healthy aging, configural processing

INTRODUCTION

Experiments on “acquired equivalence” have revealed important information about the way in which animals encode stimulus representations. For example, Honey and Ward-Robinson (2001) gave rats acquired equivalence training in which a tone would signal food delivery (t+) and a clicker would not (c–) in two distinctly decorated Skinner boxes (A and C). But in two other Skinner boxes (B and D), the tone and click signalled the alternative outcome (i.e., t– and c+). The complete discrimination can be represented as: At+, Ac–, Bt–, Bc+, Ct+, Cc–, Dt–, Dc+. It was evident that rats had learned the discrimination because they anticipated the delivery of food on reinforced (+) trials by approaching the site of delivery and refrained from this on the non-reinforced (–) trials. Notice that no single stimulus uniquely predicts either outcome: all stimuli are equally often reinforced and non-reinforced and it is necessary for rats to learn about specific configurations of stimuli. Influential theoretical accounts of such learning (e.g., Rescorla, 1976; Pearce, 2002) provide accounts of solution of the At+, Ac–, Bt–, Bc+, Ct+, Cc–, Dt–, Dc+ discrimination in which the eight trial types are represented by eight “configural” stimuli, each being associated with the appropriate outcome. However, Honey and Ward-Robinson gave an additional stage of training that produced results not anticipated by these models. In the subsequent stage, rats were split into two groups to receive different types of “reversal training,” in which at least some of the trial outcomes were switched. For group Whole, all trial types were reversed (i.e., At–, Ac+, Bt+, Bc–, Ct–, Cc+, Dt+, Dc–) but for group Part only half of the trial types were reversed (At+, Ac–, Bt–, Bc+, Ct+, Cc–, Dt+,

Dc–). Both groups’ performances were reduced by the reversal from the original stages and both recovered; however, group Whole’s performance recovered more quickly than group Part’s did. It is this feature of the data that challenges alternative configural learning theories (e.g., Rescorla, 1976; Pearce, 2002). Notice that in the pre-reversed discrimination these Skinner boxes indicate the equivalent reinforcement arrangements for the tone and click. Informally expressed, it is as though rats’ representations of Skinner boxes A and C (and B and D) had “acquired equivalence” during pre-reversal training. Thus, new learning during the reversal may transfer between A and C (and between B and D). For group Part, the acquired equivalence between A and C (and between B and D) will lead to conflicting information because A and C no longer indicated equivalent tone/click reinforcement relationships. But for group Whole, although the tone/click reinforcement relationships have all reversed, A and C (and B and D) remain equivalent. It is notable that non-configural forms of acquired equivalence are possible (e.g., Honey and Hall, 1989) but they are interpretable in simpler terms than those considered here (e.g., Ward-Robinson and Hall, 1999).

Honey et al. (2010) describe this finding, and others like them (e.g., Ward-Robinson and Honey, 2000; Hodder et al., 2003), in terms of a three-layer connectionist network, which will be described in detail in the Discussion. Those authors also note that their model will adequately explain the finding that discriminations involving “intra-dimensional” shifts are mastered more quickly than those involving “extra-dimensional” shifts (e.g., Owen et al., 1991; Barense et al., 2002). This suggestion is theoretically significant because, if substantiated, it would

undermine claims that non-human animals' demonstrations of intra-dimensional transfer are not actually demonstrations of a genuine attentional process (cf., Mackintosh, 1974). It is also clinically significant because intra-/extra-dimensional shift experiments in rats demonstrate the role of the prefrontal cortex in "attentional set" in frontal lobe disorders (Owen et al., 1991; Dias et al., 1997; Birrel and Brown, 2000; Hampshire and Owen, 2010). Deficits in performance in intra-/extra-dimensional shift have been reported in apparently healthy, older human volunteers (Owen et al., 1991; see also, Barense et al., 2002). That observation and Honey et al.'s assertion that the same psychological processes outlined in their model, govern not only acquired equivalence and intra-/extra-dimensional set shifting, make several predictions. In particular, manipulations that affect intra-/extra-dimensional set shifting, should also affect acquired equivalence. We report here results of two experiments that support that prediction by demonstrating acquired equivalence performance to be diminished in (healthy) older participants relative to younger participants.

EXPERIMENT 1

Honey and Ward-Robinson (2001) demonstrated acquired equivalence in rats using an appetitive conditioning procedure. We adapted their procedure for use with older and younger participants in Experiment 1 whose design is summarized in Figure 1. Older and younger participants were required to learn about four fictitious characters' like or dislike of two sports. Two of the characters liked the same two sports and disliked the two alternative sports (Stage 1). Acquired equivalence could be demonstrated over a series of "reversals" (Stage 2 and Stage 3) in which some or all of the previously liked sports became disliked and vice versa. If the two pairs of characters had acquired equivalence, participants' performance should recover more rapidly from the whole reversal than from the part reversal. The new question we asked here was: would this acquired equivalence effect be different in a group of older participants?

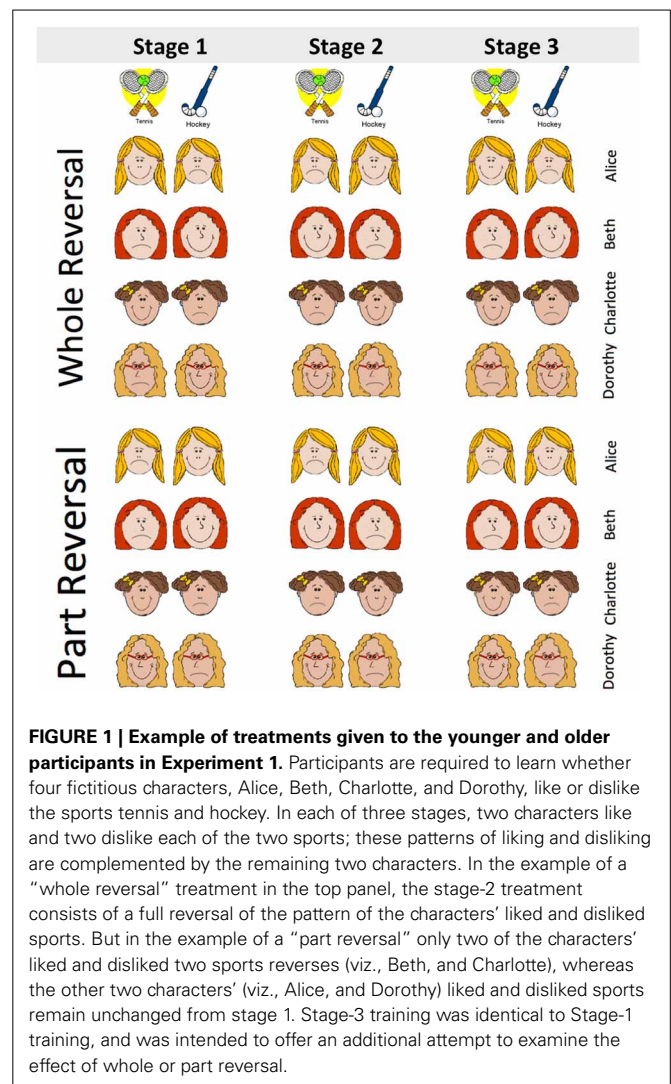
MATERIALS AND METHODS

PARTICIPANTS

Group Y comprised two men and fourteen women with a mean age of 20.8 years (range: 20–24 years); Group O comprised seven men and eight women with a mean age of 64.4 years (range: 55–77 years). Participants were a self-selected sample of respondents to recruitment posters in public places (cafés, Post Offices, etc., Group O) or were University of Nottingham students who gained course credit for participation (Group Y). All participants were naive with respect to the stimuli used in the experiment.

APPARATUS AND STIMULI

Experiments were run in a small quiet room in the School of Psychology, University of Nottingham. Stimuli were presented and responses were recorded on a laptop (Toshiba Portégé A200). From opposite corners its screen measured 31 cm. Participants used a separate keyboard that was connected to the laptop and positioned such that the participant could use the keyboard while looking at the laptop screen. The keyboard consisted of a standard QWERTY keyboard with a number pad to the right hand



side. With the exception of the keys numbered 1, 2, 3, 4, and 5 that ran along the top of the QWERTY part of the keyboard and the space key, black stickers covered the letter/number of each key. Our intention here was to direct participants' responses to the keys 1, 2, 3, 4, and 5 during the experiment. A pair of headphones (Panasonic RP-HT225) was plugged into the laptop and was used to present the auditory stimuli described below.

The following cartoon depictions were used as stimuli: (a) a pair of crossed tennis rackets and ball with "Tennis" written below them; (b) a hockey stick and ball with "Hockey" written below them; (c) four characters' faces with "neutral" facial expressions, "happy" facial expressions and "sad" facial expressions (i.e., twelve images of the characters). Each character's neutral image had her name (Alice, Beth, Charlotte or Dorothy) written below it. Each character's happy image had "X Likes this Sport" written above it where X is the character's name; sad images were similarly accompanied by text that read "X Dislikes this Sport." These stimuli occupied a screen area of around 400 mm². Auditory data files had been created on a computer (iMac, Apple Computers) using a synthetic voice ("Victoria"). These files read aloud the

text “Correct,” “Incorrect” and “Have a guess next time” and had durations of between 0.5 and 1.5 s. During the experiment, the character and sport stimuli were presented side by side and vertically central. The character appeared only on the left-hand side; the sport appeared only on the right-hand side. A numbered scale comprising the numbers 1, 2, 3, 4, and 5 could be presented below the neutral images of the characters’ faces with 1 on the left and 5 on the right. The word “Dislike” appeared to the left of “1” and the word “Like” appeared to the right of “5.”

PROCEDURE

Participants from Groups Y and O were randomly assigned to whole-reversal (W) or part-reversal (P) groups, to create Group YW, Group YP, Group OW, and Group OP, see **Figure 1**. Participants’ mean ages in groups YW, YP, OW, and OP were, respectively 20.1, 21.5, 64.6, and 64.3 years. There were seven women and one man in both Group YW and YO; there were five women and two men in Group OW; and there were three women and five men in Group OP.

All participants were given training in which they were asked to learn which sports (Tennis and Hockey) four fictitious characters (Alice, Beth, Charlotte, and Dorothy) liked. Participants keyed “5” for liked sports and “1” for disliked sports. Keys in-between could be used for less confident responses. For the purposes of feedback (see below), keying 4 or 5 were “correct” on like trials and “incorrect” on dislike trials; and keying 1 or 2 were “correct” on dislike trials and “incorrect” on like trials. Keying 3 was neither correct nor incorrect. For all participants, each sport was liked by two of the characters and disliked by the other two characters; each character liked one sport and disliked the other. Thus, each character agreed in her opinion of the two sports with one other character and disagreed with the two remaining characters. We counterbalanced stimulus arrangements such that for some participants Alice and Charlotte (and therefore, Beth and Dorothy) had equivalent sports opinions and for others Alice and Beth (and therefore, Charlotte and Dorothy) has equivalent sports opinions. Neither Alice and Dorothy nor Beth and Charlotte shared sports opinions for any participants. For some participants the shared opinions were based on liking Tennis (and, therefore, disliking Hockey); for others participants the shared opinions were based on liking Hockey (and therefore, disliking Tennis). The orthogonal arrangement of this counterbalancing created four different discriminations, which were given to similar numbers of participants.

Participants read a standard instruction sheet that gave an indication of the rationale of the experiment and emphasized participants’ entitlement to leave the experiment. Instructions were then presented on the laptop. A scenario was described involving the participant learning which of two sports the four characters liked. Instructions described making 1–5 key responses to indicate each character’s like/dislike of the sports. At the end of the instruction phase the experimenter went through an example of how to use the keyboard to register responses. The experimenter checked that the participant understood and was comfortable with the task, and then left the room. The instruction phase repeated then the participants pressed the spacebar to initiate the trials.

The sequence of one type of trial is exemplified in **Figure 2**. Each trial consisted of: (1) The 1.5-s, centrally located presentation of the keyboard character “+,” (2) The presentation of a character and a sport, during which the participant had unlimited time to select a response from 1 (dislike) to 5 (like) of the scale that was presented below them, (3) Information about the characters like/dislike of the sport was given for 2.9 s. This comprised presentation of text (e.g., Alice likes this Sport) with the accompanying “happy” or “sad” version of the character and the Tennis or Hockey picture, (4) Auditory feedback was given (“Correct,” “Incorrect” or, where 3 was keyed, “Have a guess next time”).

During stage-1 training, each of the eight trial types was given 24 times in an irregular sequence (i.e., 96 liked and 96 disliked trials). During stage-2 training, participants received an identical treatment but on a reversed version of the task: for participants in Groups YW and OW, all four of the characters now liked the sports they had previously disliked and now disliked the sports they had previously liked. For participants in Groups YP and OP, only two of the characters’ opinions of the sports reversed. Stage 3 was the final stage of training and was identical to the first except that only sixteen trials of each of the eight trial types was given. Group YP and OPs’ partial reversals were systematically varied and were arranged so that each particular discrimination was matched by a subgroup of YW and OW. Thus, performance differences between whole/part reversals could not be attributed to differences in difficulty of the specific trials types in their discrimination.

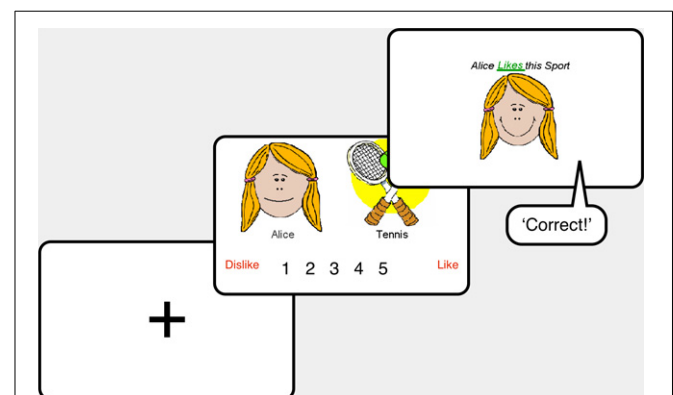


FIGURE 2 | Illustration of the the sequence of events on each trial of training in Experiments 1 and 2. (1) The leftmost panel represents the presentation on the computer’s screen of the fixation cross (“+”), which occurred at the beginning of each trial for 1.5 s. There was no requirement of the participant; (2) The central panel represents the presentation on the computer screen of the character, the sport and the rating scale. This example represents a trial in which a participant was asked to rate Alice’s like/dislike of tennis, however, other trial types occurred (see, e.g., **Figures 1, 3**).

This slide remained until the participant had made their rating; (3) The rightmost panel represents the feedback given to the participant following their previous rating. In this example, the participant correctly rated Alice as liking tennis (i.e., the participant gave Alice a rating of ≥ 4 for tennis), which was accompanied by the spoken word “Correct” and by the statement “Alice likes this sport.” Full details of the feedback given on incorrect trials and on the other types of trial are given above.

RESULTS

Initial examination of raw trial-by-trial data revealed that all participants mastered the task rapidly, for example reaching asymptotic discrimination after about six blocks of the eight trial types. Our analysis focuses, therefore, on the terminal six trials of established training and the initial six trials of the reversal. Here, group differences were not masked by rapid discrimination learning. Data on “dislike” trials were transformed to match the scale of the “like” trials. That is, 1 (the correct response) was recoded as 5, 2 as 4, 3 remained as 3, 4 as 2 and 5 as 1. This obtained a like/dislike-independent response measure in which 5s indicate the correct response and 1s indicate the incorrect response. These data are summarized in **Figure 3**. We see that all four groups’ performance before both reversals was good (around the asymptote of 5) and that it declined on both of the reversals, recovering quickly. Inspection of the two Y groups, indicates that group YW recovered from the disruption of the reversal more quickly than group YP. However, no such pattern can be seen in the O groups: groups OW and OP show no obvious difference in recovery. This description of the data was supported by an analysis of variance (ANOVA) with within-subject variables of: (1) cycle (i.e., the first and second twelve-trial cycles of established discrimination and subsequent reversal), (2) established training (end of stages 1 and 2) versus reversal stage (beginnings of stages 2 and 3), and (3) trial; and between-subject variables of: (1) age (i.e., Y versus O), and (2) reversal group (i.e., W versus P). The analysis revealed main effects of trial, $F_{(5, 135)} = 4.9$, $p < 0.001$, $\eta_p^2 = 0.156$, reversal stage, $F_{(1, 27)} = 40.7$, $p < 0.001$, $\eta_p^2 = 0.602$ and cycle, $F_{(1, 27)} = 5.0$, $p < 0.034$, $\eta_p^2 = 0.158$. The Cycle \times Trial \times Age interaction, $F_{(5, 135)} = 2.9$, $p < 0.017$, $\eta_p^2 = 0.097$, and the Cycle \times Trial \times Reversal Stage

interaction, $F_{(5, 135)} = 2.3$, $p < 0.046$, $\eta_p^2 = 0.079$ were significant. No other main effect was significant. The source of the interaction involving the age variable was examined using similar analyses separated for young and older participants. Analysis of older participants’ data yielded a main effect of reversal stage only, $F_{(1, 13)} = 15.0$, $p < 0.003$, $\eta_p^2 = 0.537$. No other statistic was significant and, of most importance, none was significant that involved the reversal-group variable, smallest $p > 0.121$. However, the corresponding analysis of the younger participants’ data yielded reliable main effects of cycle, reversal, and trial, and reliable Reversal \times Trial, and Reversal \times Trial \times Reversal Group interactions, largest $p < 0.024$, $F_{(5, 70)} = 2.7$, $\eta_p^2 = 0.167$. The source of the younger participants’ Reversal \times Trial \times Reversal Group interaction was examined using a pair of ANOVAs with data split across the reversal stage variable (i.e., on established discrimination data and reversed data) with only cycle and reversal group as variable. No significant statistics were obtained for the established discrimination data, smallest $p > 0.134$. The corresponding ANOVA for the reversed data yielded a significant main effects of cycle and trial and a significant Trial \times Reversal Group interaction, largest $p < 0.012$, $F_{(1, 29)} = 7.3$, $\eta_p^2 = 0.203$. Simple main-effects analysis on this interaction using separate error terms for each trial, revealed younger participants’ whole reversal performance to be superior to partial reversal performance on the fifth trial, $F_{(1, 14)} = 13.7$, $p < 0.003$, $\eta_p^2 = 0.495$.

EXPERIMENT 2

The results of Experiment 1 join those of Honey and Ward-Robinson (2001) and Hodder et al. (2003) in showing an acquired equivalence effect by an improved rate of “whole” reversal learning relative to “part” reversal learning in younger participants.

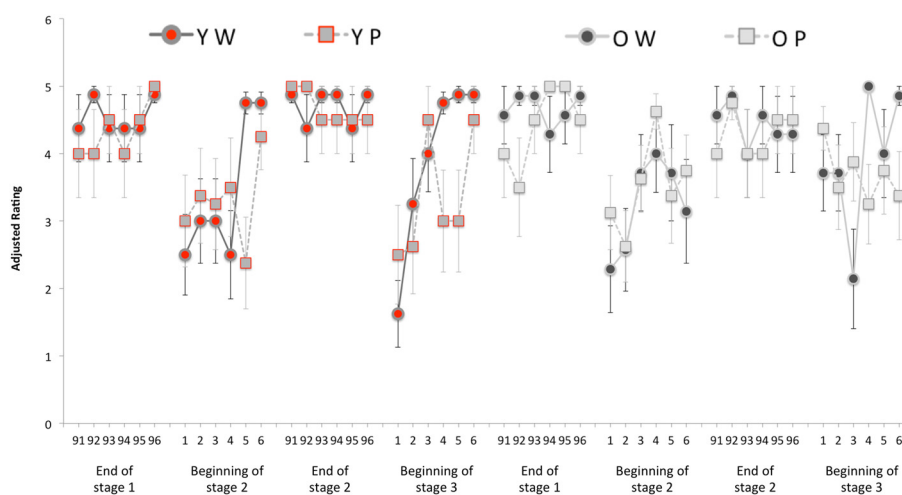


FIGURE 3 | Means, and one standard error of each mean, of data from Experiment 1. The leftmost four sets of six trials of data are from the two younger participant groups (groups YW and YP); the rightmost data are the corresponding data from two older participant groups (groups OW and OP). For all four groups, and running from left to right, the four sets of data represent: (1) The final six liked and final six disliked trials of stage 1 (i.e., trials 91 through to 96 of stage 1); (2) the first six liked and first six disliked

trials of stage 2 (i.e., trials 1 through to 6 of stage 2); (3) the final six liked and final six disliked trials of stage 2 (i.e., trials 91 through to 96 of stage 2); (4) the first six liked and first six disliked trials of stage 3 (i.e., trials 1 through to 6 of stage 3). The ratings are expressed in a like/dislike-independent form such that data from dislike trials were transformed to match the scale of the like trials. Thus, here data are pooled over the like and dislike trial types and scores of 5 is (maximally) correct and a score of 1 is (maximally) incorrect.

Our new finding is that this difference in whole/part learning rate was absent in older participants. Before considering fully the implications of this finding, we sought to replicate it using similar logic to that of Experiment 1. For Experiment 1 to reveal acquired equivalence, it is necessary for the benefit of acquired equivalence to more than offset the cost of relearning new character-sport relationships. In Experiment 2, which is summarized in **Figure 4**, we followed Honey and Ward-Robinson (2001) in the use of a design that avoids this compromise. Older and younger participants were required to learn the four characters like/dislike of four sports. For the Congruent treatment, each of the characters' like/dislike of the the four sports was matched with one other character. For the Incongruent treatment, no one character's sport like/dislike was matched with any other character. Acquired equivalence could be demonstrated by the finding that the

discrimination was mastered more rapidly in the congruent than the incongruent condition. Again, we asked whether the extent of acquired equivalence would be different in the two age groups.

MATERIALS AND METHODS

PARTICIPANTS, APPARATUS, AND STIMULI

Group Y comprised six men and ten women with a mean age of 21.2 years (range: 18–24 years); Group O comprised five men and eleven women with a mean age of 64.8 years (range: 53–77 years). The apparatus and stimuli were those used in Experiment 1. Experiment 2 used an additional two sport stimuli, bowling and netball to make a total of four sports for the four characters. All unspecified details of participants, apparatus and stimuli were identical to those of Experiment 1.

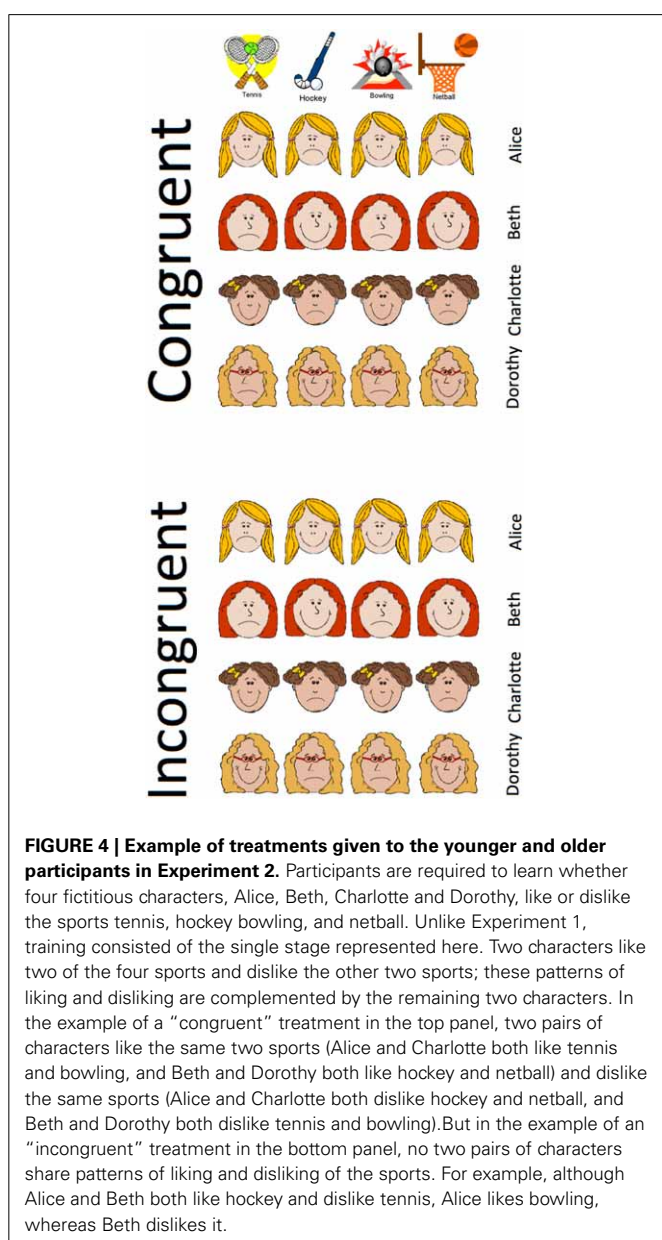
PROCEDURE

Participants from Groups Y and O were randomly assigned to congruent (C) or incongruent (I) groups, to create Group YC, Group YI, Group OC, and Group OI. The mean ages and numbers of women and men in these groups was, respectively: 21.0, 22.5, 64.5, and 65.1 years; and 6:2, 4:4, 5:3, and 6:2. All participants were given training in which they were asked to learn which of the four sports (Tennis, Hockey, Bowling, and Netball) the four fictitious characters (Alice, Beth, Charlotte, and Dorothy) liked. For all participants, each of the four sports was liked by two of the characters and disliked by the other two characters; each character liked two sports and disliked the other two sports. For the congruent groups, each character shared her pattern of sport liking and disliking with one other character and the two remaining characters had the complementary pattern of liking and disliking of sports. For all participants in the congruent groups Alice was equivalent to Charlotte and Beth was equivalent to Dorothy. For approximately half of the participants in the two congruent groups this was based upon Alice and Charlotte's shared liking of Bowling and Tennis (and shared disliking of Netball and Hockey); for the remainder of the participants in the two congruent groups, equivalence was based upon Alice and Charlotte's shared disliking of Bowling and Tennis (and their shared liking of Netball and Hockey). The arrangements of Alice and Charlotte's liking and disliking of Bowling and Netball was the same for the two incongruent groups as for the two congruent groups. The incongruent groups' treatment differed from the congruent groups' treatment in the four characters' liking and disliking of Tennis and Hockey: for approximately half of the participants in the two incongruent groups, Alice and Beth liked Tennis (and disliked Hockey); but for the remainder Alice and Beth disliked Tennis (and liked Hockey). Notice that for the incongruent groups no two characters were exactly alike in their pattern of sports liking.

All participants received 256 trials in random sequence with the constraint that each of the sixteen trial types created by the combinations of the four characters and four sports occurred once in each block of sixteen trials. Unspecified procedural details were identical to those of Experiment 1.

RESULTS

The results of Experiment 2 are summarized in **Figure 5**. As in Experiment 1, the scale for dislike trials was reversed to match that



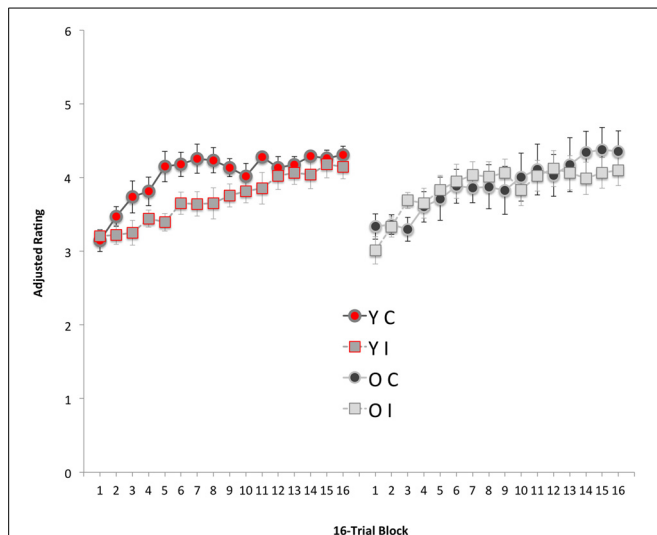


FIGURE 5 | Means, and one standard error of each mean, of data from Experiment 2. The leftmost 16 8-trial blocks of data are from the two younger participant groups (groups YC and YI); the rightmost data are the corresponding data from two older participant groups (groups OC and OI). The ratings are expressed in a like/dislike-independent form such that data from dislike trials were transformed to match the scale of the like trials. Thus, here data are pooled over the like and dislike trial types and scores of 5 is (maximally) correct and a score of 1 is (maximally) incorrect.

of the like trials and data were collapsed over like and dislike trials. Initial inspection and analysis revealed that the older participants' discrimination performance was robust, though the response buttons were often not the most extreme (i.e., responses of 1 and 5). This feature of the data indicates that older participants may have differed from younger participants in their response bias (i.e., tending to make more accurate, but more modest, responses). For example, on the 16th block of training, only three of the sixteen younger participants gave mean responses that were not 5 s or 1 s, however, at that point, fifteen of the sixteen older participants gave scores that were not 5 s or 1 s ($\chi^2 = 15.4, p < 0.001$). To correct for this bias each datum was normalized by multiplying it by a normalization ratio (cf., Ringo, 1988; Baxter and Murray, 2001). The normalization ratio was computed for each block by dividing the arithmetic mean of all data for that block (i.e., ignoring age and congruency designation) by the mean for the age group (i.e. ignoring only congruency designation) on that block. This process acted to moderate younger participants' responses and boost older participants' responses, which was irrespective of congruency designation.

We see that participants in all groups learned the relationships between the characters and the sports and, it seems, less rapidly than the discrimination in Experiment 1, which could be the result of the additional number of trial types. Group YC appeared to master the discrimination more rapidly than Group YI. The question of key interest is whether the older participants would also show acquired equivalence. In particular, would Group OC's performance show superiority over Group OI's? As in Experiment 1, the older participants appear to have satisfactorily learned the discrimination but do not demonstrate acquired equivalence.

This description of the data was supported by an ANOVA with block as a within-subject variable and age and congruency as between-subject variables, which revealed a main effect of block, $F_{(15, 420)} = 16.6, p < 0.001, \eta_p^2 = 0.373$ and a Block \times Age \times Congruency interaction, $F_{(15, 420)} = 1.8, p < 0.035, \eta_p^2 = 0.060$. No other statistics were significant, smallest $p > 0.216, F_{(1, 28)} = 1.6, \eta_p^2 = 0.054$.

The source of the Block \times Age \times Congruency interaction was located by performing a pair of separate, 2×16 ANOVAs on younger and older participants' data. The ANOVA on the younger participants' data yielded a main effect of block and a Block \times Congruency interaction, smaller $p < 0.014, F_{(15, 210)} = 2.1, \eta_p^2 = 0.128$. The congruency main effect was not significant, $F_{(1, 14)} = 3.9, p < 0.066, \eta_p^2 = 0.222$. The source of the Block \times Congruency interaction in younger participants' data was located using simple main-effects analysis with separate error-terms for each level of block. This showed responding of Group YC to be superior to that of Group YI on blocks, 5, 6, 7, and 8, largest $p < 0.049, F_{(1, 14)} = 4.6, \eta_p^2 = 0.248$.

The corresponding 2×16 ANOVA on older participants' data yielded only a main effect of block, $F_{(15, 210)} = 5.7, p < 0.001, \eta_p^2 = 0.288$. Neither the congruency main effect nor its interaction with block was significant, $F_s < 1$.

DISCUSSION

We sought to test Honey et al.'s (2010) claim that acquired equivalence of configural learning and intra-dimensional/extra-dimensional set-shifting experiments may employ a common mechanism. We reasoned that because performance at attentional set-shifting is reduced in healthy, relatively aged subjects (Owen et al., 1991; Barense et al., 2002), if Honey et al.'s assertion is correct, performance at acquired equivalence should also be reduced. Our new findings supported that suggestion. They do not unambiguously confirm that there is a relationship between configural learning and intra-dimensional/extra-dimensional set-shifting (e.g., one brought about by their reliance on a common psychological process). For example, configural learning and set shifting could be governed by independent psychological processes, each being affected by some aspect of ageing. Nonetheless, our new results represent a first and necessary step in the conclusion that configural learning and set shifting are governed by a common process.

The force of that argument relies on the specificity of the reduction in performance. That is, older participants' reduced performance at an acquired equivalence task is not theoretically decisive if it is part of a more general pattern of reduction. This could be obtained by some general disadvantage, perhaps a reduction in working memory performance, inhibition or simply less familiarity with computer-based tasks than the younger participants. Participants may have differed in their motivation to participate (younger participants gained course credit, older participants did not) or in the level or style of their educations (younger participants were current university students, older participants were not). Examination of performance that is not part of the acquired equivalence task is key to evaluating these possibilities. Older participants' performance in Experiment 2 did indicate some general deficit in discrimination relative to younger

participants (which was evident before data were normalized). Of course, older people may present a general change in performance that is especially pronounced in acquired equivalence tasks. Such an interaction between tasks and the effects of age on performance could generate the results obtained. We cannot eliminate such an account but we noted above that the general performance deficiency exhibited by older participants appeared to be a response bias and rather than a discrimination deficiency. Further evidence against the suggestion that the age-associated change in acquired equivalence is merely part of a general decline comes from Experiment 1. Here, the acquired equivalence deficit was not accompanied by a general change in performance (e.g., the initial ANOVA did not generate any significant statistics that involved the variable age). We have no ready explanation for the inconsistency across experiments of the age-related response bias; but because it is uncorrelated with effects on acquired equivalence it is, without some additional elaboration, an inadequate explanation of our findings.

Leaving to one side for a moment the age-related effects of performance, the findings of the acquired equivalence of configural learning may be accommodated by a connectionist model (Honey et al., 2010), whose main features are summarized in **Figure 6**. Individual elements of the discrimination, here the characters and the sports, are represented at the input layer of the network. Presentation of two items (e.g., Alice and tennis) will tend to generate activity in the network's hidden layer. Hidden-unit activity is subject to a "winner-take-all" process in which the single most active unit will suppress activity in less active units. At first, hidden unit selection will be stochastic: one lucky unit (e.g., "w")

will be active when activity is generated by the outcome (here, the information being that the character likes or dislikes the sport). The development of hidden-unit → output unit connection strength will be supported by the co-occurrence of activity sustained by "feedback" from the output unit back to the hidden unit. It is this feedback process that give this model its capacity to accommodate acquired equivalence. On a correctly answered trial (e.g., one corresponding to "Alice likes tennis"), after some training, "Alice" and "tennis" will generate activity in the hidden unit, "w," which will generate activity in the "like" hidden unit. Here "like" is also the outcome of the trial (i.e., the participant is informed that "Alice likes tennis"). This "correct" outcome will tend to stimulate further activity, via a feedback connection, to hidden unit "w." If we ignore any intervening trials and consider next what will happen when the participant receives a trial in which they are asked if Charlotte likes tennis. The presence of tennis in the input layer will tend to provoke activity in the hidden unit "w," which codes for liked character-sport combinations; w's activity now provokes activity in the like output unit. On this occasion the participant is likely to correctly indicate that "Charlotte likes tennis," which will again provoke like → "w" feedback and will improve connection strength between Charlotte and "w" and between tennis and "w." Of course, some intervening trials will involve tennis also being disliked by some characters (viz., Beth and Dorothy). Thus, at intermediate points of training there is no reason to suppose that the presence of tennis on an "Alice likes tennis" trial will correctly activate hidden unit "w": it could equally well activate hidden unit "y" (which codes for Beth and Dorothy's dislike of tennis). On such trials in which "y" is incorrectly selected, the dislike output-unit will be activated by "y" but the actual trial outcome will activate the like output-unit. This means that the output-unit → hidden-unit feedback signal will not sustain activity in the hidden unit "y," and the capacity of Alice to activate it will diminish. Over multiple trials these processes will tend to encourage sharing of hidden units, thus generating acquired equivalence.

It follows from the analysis above that the disruption of the conjoint hidden- and input-unit activity on correct trials will lead to a reduction in the sharing of hidden units. Understanding such a process may be key to understanding age-related changes seen in acquired equivalence here, and, by extension, those seen in attentional set experiments (e.g., Owen et al., 1991; Barense et al., 2002). One way that this could occur has already been proposed to explain similar effects of neural manipulations on configural acquired equivalence (e.g., Coutureau et al., 2002; Iordanova et al., 2007). Here, older people's networks function as described above but with the single exception that the feedback signal from the output-layer to the hidden-layer is weakened or is absent. As outlined above, the feedback signal is a necessary step in the sole means by which input units come to share hidden units; the absence of this signal will, therefore, prevent sharing of hidden units and, therefore, prevent acquired equivalence. The absence of shared hidden units will not prevent the engagement of (trial-unique) hidden units in learning. The hidden unit that is most active on a particular trial will still tend to become associated with the output unit and it will not be activated by any other trial type. The mechanism of learning in older people, then, becomes

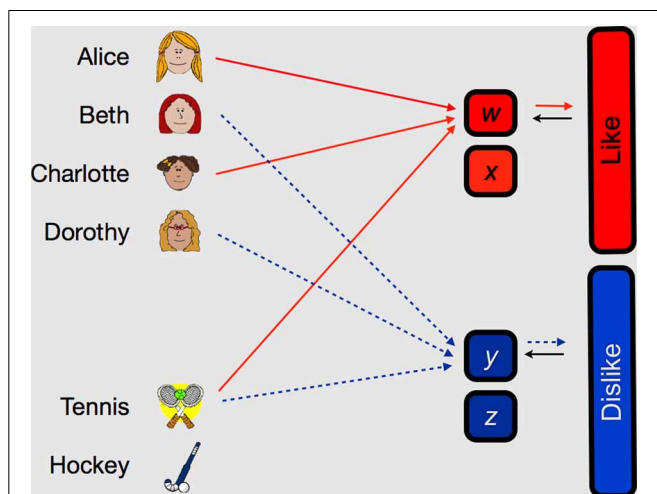


FIGURE 6 | Depiction of theoretical analysis of acquired equivalence.

The three-layered network is composed of: input units, that represent the individual components of each stimulus of each trials (e.g., both Alice and tennis); hidden units, that represent the combination of stimuli on that trial (e.g., Alice with tennis); and the output units, that represent the trial's correct outcome (e.g., that Alice likes tennis). Input unit → hidden unit and hidden unit → output unit connections begin with weights of random strength that approximate zero. Weight changes occur as learning progresses. An output unit → hidden unit connection gives feedback to the hidden unit about the trial's outcome.

like that described by models such as those of Rescorla (1976) and Pearce (2002): each unique combination of character and sport will require its own, unique hidden-unit. The translation from this model to ageing people is unclear but it seems possible that they reflect developmental changes in cortical regions in rhinal (e.g., Coutureau et al., 2002 or prefrontal brain-regions e.g., Iordanova et al., 2007)

Whatever the precise detail of the deficit in performance, our current results demonstrate the generality of demonstrations of acquired equivalence reported by others (e.g., Ward-Robinson and Honey, 2000; Honey and Ward-Robinson, 2001; Coutureau et al., 2002; Hodder et al., 2003; Iordanova et al., 2007) and its absence in older participants. We noted also that the parallel

between these facts and age-related deficits in performance on intra-dimensional/extra-dimensional set tasks (e.g., Owen et al., 1991; Barense et al., 2002) could be the result of their being underpinned by a shared mechanism and that this does not require an attentional component (cf., Honey et al., 2010).

ACKNOWLEDGMENTS

This research was supported by an Experimental Psychology Society, University Bursary Scheme grant that supported Emma Owens' collection experimental work. We gratefully acknowledge Allie Broome, Lone Hørlyck, and Claire Petelczyc for their help. Some results were reported at A Festschrift for Geoffrey Hall, Meeting of the Experimental Psychology Society, Hull, UK, 2012.

REFERENCES

- Barense, M., Fox, M., and Baxter, M. (2002). Aged rats are impaired on an attentional set-shifting task sensitive to medial frontal cortex damage in young rats. *Learn. Mem.* 9, 191–201. doi: 10.1101/lm.48602
- Baxter, M. G., and Murray, E. A. (2001). Opposite relationship of hippocampal and rhinal cortex damage to delayed nonmatching-to-sample deficits in monkeys. *Hippocampus* 11, 61–71. doi:10.1002/1098-1063(2001)11:1<61::AID-HIPO1021>3.0.CO;2-Z
- Birrel, J. M., and Brown, V. J. (2000). Medial frontal cortex mediates perceptual attentional set shifting in the rat. *J. Neurosci.* 20, 4320–4324.
- Coutureau, E., Killcross, A. S., Good, M., Marshall, V. J., Ward-Robinson, J., and Honey, R. C. (2002). Acquired equivalence and distinctiveness of cues: II. Neural manipulations and their implications. *J. Exp. Psychol. Anim. Behav. Process.* 28, 388–396. doi: 10.1037/0097-7403.28.4.388
- Dias, R., Robbins, T. W., and Roberts, A. C. (1997). Dissociable forms of inhibitory control within prefrontal cortex with an analog of the Wisconsin card sort test: restriction to novel situations and independence from “on-line” processing. *J. Neurosci.* 17, 9285–9297.
- Hampshire, A., and Owen, A. M. (2010). “Clinical studies of attention and learning,” in *Attention and Associative Learning From Brain to Behaviour*, eds C. J. Mitchel and M. Le Pelley (Oxford: Oxford University Press), 385–406.
- Hodder, K. I., George, D. N., Killcross, A. S., and Honey, R. C. (2003). Representational blending in human conditional learning: implications for associative theory. *Q. J. Exp. Psychol. B* 56, 223–238. doi: 10.1080/02724990244000269
- Honey, R. C., Close, J., and Lin, T. C. E. (2010). “Acquired distinctiveness and equivalence: a synthesis,” in *Attention and Associative Learning From Brain to Behaviour*, eds C. J. Mitchel and M. Le Pelley (Oxford: Oxford University Press), 159–186.
- Honey, R. C., and Hall, G. (1989). The acquired equivalence and distinctiveness of cues. *J. Exp. Psychol. Anim. Behav. Process.* 15, 338–346. doi: 10.1037/0097-7403.15.4.338
- Honey, R. C., and Ward-Robinson, J. (2001). Transfer between contextual conditional discriminations: an examination of how stimulus conjunctions are represented. *J. Exp. Psychol. Anim. Behav. Process.* 27, 196–205. doi: 10.1037/0097-7403.27.3.196
- Iordanova, M. D., Killcross, A. S., and Honey, R. C. (2007). Role of the medial prefrontal cortex in acquired distinctiveness and equivalence of cues. *Behav. Neurosci.* 121, 1431–1436. doi:10.1037/0735-7044.121.6.1431
- Mackintosh, N. J. (1974). *The Psychology of Animal Learning*. London: Academic Press.
- Owen, A. M., Roberts, A. C., Polkey, C. E., Sahakian, B. J., and Robbins, T. W. (1991). Extra-dimensional versus intra-dimensional set shifting performance following frontal lobe excisions, temporal lobe excisions or amygdalo-hippocampectomy in man. *Neuropsychologia* 29, 993–1003. doi: 10.1016/0028-3932(91)90063-E
- Pearce, J. M. (2002). Evaluation and development of a connectionist theory of configural learning. *Anim. Learn. Behav.* 30, 73–95. doi: 10.3758/BF03192911
- Ringo, J. L. (1988). Seemingly discrepant data from hippocampectomized macaques are reconciled by detectability analysis. *Behav. Neurosci.* 102, 173–177. doi: 10.1037/0735-7044.102.1.173
- Rescorla, R. A. (1976). Stimulus generalization: some predictions from a model of Pavlovian conditioning. *J. Exp. Psychol. Anim. Behav. Process.* 2, 88–96. doi: 10.1037/0097-7403.2.1.88
- Ward-Robinson, J., and Hall, G. (1999). The role of mediated conditioning in acquired equivalence. *Q. J. Exp. Psychol.* 52, 335–350. doi: 10.1080/027249999393031
- Ward-Robinson, J., and Honey, R. C. (2000). A novel contextual dimension for use with an operant chamber: from simple to hierarchical forms of learning. *J. Exp. Psychol. Anim. Behav. Process.* 26, 358–363. doi: 10.1037/0097-7403.26.3.358

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: 07 May 2013; paper pending published: 12 June 2013; accepted: 20 September 2013; published online: 11 October 2013.

Citation: Robinson J and Owens E (2013) Diminished acquired equivalence yet good discrimination performance in older participants. *Front. Psychol.* 4:726. doi: 10.3389/fpsyg.2013.00726

This article was submitted to *Personality Science and Individual Differences*, a section of the journal *Frontiers in Psychology*.

Copyright © 2013 Robinson and Owens. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Individual differences in chemotherapy-induced anticipatory nausea

Marcial Rodríguez *

Laboratory of Comparative Psychology, Department of Experimental Psychology, Faculty of Education and Humanities, University of Granada, Ceuta, Spain

Edited by:

Rachel M. Msetfi, *University of Limerick, Ireland*

Reviewed by:

José C. Perales, *Universidad de Granada, Spain*

Ursula Stockhorst, *University of Osnabrueck, Germany*

***Correspondence:**

Marcial Rodríguez, *Laboratory of Comparative Psychology, Department of Experimental Psychology, Faculty of Education and Humanities, University of Granada, C/El Greco, nº 10., 51002 Ceuta, Spain*
e-mail: marcial@ugr.es

Anticipatory Nausea (AN) is a severe side effect of chemotherapy that can lead cancer patients to discontinue their treatment. This kind of nausea is usually elicited by the re-exposure of the patients to the clinical context they need to attend to be treated. There has been considerable agreement that AN represents a paradigmatic example of Pavlovian conditioning, and within this framework, several behavioral interventions have been proposed in order to prevent this phenomenon. However, some studies have questioned the validity of the Pavlovian approach, suggesting that CS-US associations are neither necessary nor sufficient for AN to occur. The data and the alternative theories behind such criticisms are discussed. Additionally, it is suggested that animal models of AN could be enriched by taking into account rats' individual differences.

Keywords: chemotherapy, nausea, classical conditioning, differences, rat

INTRODUCTION

Chemotherapy treatment leads to a wide range of harmful collateral effects which include hair loss, diarrhoea, fatigue, loss of appetite, sexual dysfunction, and cognitive deficits (e.g., Kayl and Meyers, 2006). But in addition to these distressing side effects (perhaps to be expected given that, in essence, this treatment works through poisoning), the major unpleasant symptom that patients have to cope with while undergoing chemotherapy treatment is nausea (e.g., Haiderali et al., 2011). When severe, this consequence of chemotherapy dramatically reduces the patient's quality of life, and may even lead to discontinuation of the treatment (Roscoe et al., 2011). Adequate management of nausea for these patients has not been completely achieved through pharmacological interventions (Hsu, 2010), so that behavioral and cognitive therapies are being increasingly recommended (Schiff and Ben-Arye, 2011).

In an ordinary chemotherapy schedule (e.g., Jacobsen et al., 1993) high and low doses of cytotoxic drugs (such as cisplatin, carboplatin, cyclophosphamide) are administered in cycles spaced for a period of weeks. During each of those cycles, patients may need to attend the hospital for up to six consecutive weeks to receive an infusion on each visit. The visit to the hospital for the administration of the infusion can last for hours, and the first signs of intoxication (nausea, vomiting, sweating, changes in heart rate etc.) can be experienced when patients are still in the hospital room. Later, when they return home, sporadic nausea episodes can appear during the next 24 h, and also during a following period of ~5 days. These two phases are usually referred to as acute and delayed nausea, respectively (e.g., Haiderali et al., 2011). If patients undergoing chemotherapy repeatedly experience episodes of nausea, then this can lead to a further problem known as anticipatory nausea (AN).

PAVLOVIAN CONDITIONING OF NAUSEA IN CANCER PATIENTS

In a given moment during the course of the treatment, cancer patients can experience nausea and/or vomiting before the start of a new infusion. Originally considered as a kind of neurosis, AN was finally identified in 1980 by Nesse et al. (1980) as a case of Pavlovian conditioning, an interpretation that prevails to the present day, albeit not exclusively. First, its etiology is taken to be psychological because this kind of nausea is not directly related to the infusion of the cytotoxic drugs; and secondly, it tends to occur when patients expect it on the basis of some specific environmental cues or thoughts. According to the classical conditioning model, the chemotherapy schedule can be conceptualized as a set of learning trials. Thus, in a particular context, the administration of the cytotoxic drugs would act as an unconditioned stimulus (US) with nauseating effects (the unconditioned response, UR). By virtue of association with the contextual stimuli present during the infusion sessions (conditioned stimulus, CS), these effects are subsequently elicited as a conditioned response (CR). The similarity between the UR and the CR; the fact that AN is more easy to observe as the chemotherapy treatment progresses (i.e., as the number of conditioning trials increases); that the stimuli acting as the CS are usually those related to the hospital setting (either directly perceived or imagined); and that AN persists during the follow-up visits to the hospital once the chemotherapy was completed, are four characteristics that clearly fit this interpretation (e.g., Tomoyasu et al., 1996).

The adequacy of a Pavlovian theoretical framework to account for AN has been widely accepted, and has had not only important implications for cancer patients but also for learning theorists. Patients have obtained two main benefits from the conditioning approach to AN. First, the simple fact of knowing the

reasons why they react as they do has been supportive and a source of relief for them (Nesse et al., 1980). And second, two well-established learning phenomena that reduce the efficacy of CS-US pairings in producing an association—latent inhibition and overshadowing—have been offered (Stockhorst et al., 1998; Klosterhalfen et al., 2005) as possible behavioral interventions that could help to prevent AN¹. Furthermore, this line of investigation has also been very fruitful for learning psychologists. Attempts to reproduce with laboratory animals the conditions under which AN develops in humans has helped to provide a useful paradigm for studying the laws of contextual aversion learning (see Symonds and Hall, 2012, for a review on this topic).

However, it is also necessary to recognize that there are two major problems that the Pavlovian framework needs to address. First, AN affects approximately only one in four patients (Roscoe et al., 2011), which means that factors other than CS-US contingencies may be affecting the development of AN, or in other words, that the predictive capacity of the Pavlovian model to identify those patients who are at risk of suffering from AN needs to be improved (Watson et al., 1998). Second, and more intriguingly, it has been asserted (Aapro et al., 2005) that nausea can be anticipated in patients without their having the previous experience that classical conditioning involves. In the following section we will first consider some data that do not fit with the conditioning model and the alternative theories that could account for them.

ARE ALTERNATIVE EXPLANATIONS FOR AN NECESSARY?

A good example of why some authors have questioned the validity of the associative theory as an account of AN was provided by Tyc et al. (1997). These authors observed that of the 45 children (59%) who developed AN in their sample, 11 (25%) had not previously suffered from post-treatment nausea (see also Matteson et al., 2002). This fact does not fit with learning rules in that the supposed CR could not be elicited if the subject has not previously experienced the UR. It might be possible to argue that experiencing the UR is not necessary for conditioning, i.e., that the simple fact of pairing the CS and the US could be enough for the formation of an association. However, such an argument could be considered as implausible and for the purposes of practical intervention in the clinic it seems reasonable to consider other possible explanations.

¹Latent inhibition is a very robust effect that has been observed under a great variety of preparations—presentations of a stimulus followed by no consequences will subsequently retard the acquisition of its association with any given US. Klosterhalfen et al. (2005) showed for example that AN induced by a rotation chair was lower in subjects preexposed to that apparatus. Thus, within the context of chemotherapy treatment, these authors suggested that preexposures to the clinical cues could be useful to prevent chemotherapy-induced AN. Overshadowing is also a well-established phenomenon in associative learning—a salient cue is presented in combination with a target CS, with the result that the capacity of that CS to predict the US is reduced. Stockhorst et al. (1998) gave a group of cancer patients different novel combinations of flavors during their first two infusions. At their third clinical visit none of the 8 experimental subjects showed AN, but it was found in two patients of the control group that had drunk water instead of the novel tastes. It is supposed that the readiness of the flavors to become associated with the gastric discomfort overshadowed the conditioning of the clinical cues.

Tyc et al. (1997) proposed two possible accounts for their data. Firstly, nausea could have been directly elicited by an acute attack of anxiety, a phenomenon known as “psychogenic” or “nervous” nausea (Yugin, 1989). Secondly, it is known that a person can get sick through observational learning, i.e., by viewing other people vomiting. Given that subjects in the sample by Tyc et al., shared the chemotherapy room, this possibility seems more than plausible (see also Cohen et al., 1986). Finally, a third possibility, which has been increasingly analyzed during the last decade, is that the expectancies of the patients play an important role in the development of AN. Response expectancy theory supposes that patients might anticipate nauseating symptoms on the basis of their previous thoughts or beliefs and that this can be a direct cause of the occurrence of these symptoms (e.g., Montgomery et al., 1998; Sohl et al., 2009). In this case, AN would be governed by the same general mechanisms that operate in producing the placebo effect, and could be observed without the necessary mediation of a previous Pavlovian association (see Stewart-Williams and Podd, 2004, for a discussion about the relationship between the placebo effect and Pavlovian conditioning).

It is suggested then, that patients who expect to experience nausea, and even those who are uncertain about it, are more likely to develop AN than patients who clearly do not expect to get sick during the course of chemotherapy (Hickok et al., 2001). Given that the incidence of cancer among the population has increased over recent decades, patients may have acquired some knowledge from the media (through films, news, or documentaries), as well as from their friends and relatives, about the collateral side effects induced by the chemotherapy. The information provided by such unofficial sources, as well as that provided by medical staff, could influence the patient's expectations, thus producing a kind of “nocebo” effect (Colloca and Miller, 2011). Unfortunately, the few experimental attempts that have sought to confirm that the cancer patient's expectancies are a causal factor for nausea have reported inconsistent results. In one study, Shelke et al. (2008) showed that patients in an experimental group who trusted more than controls in the power of a new antiemetic medication, showed almost as much nauseating symptoms as the control. (Shelke et al., acknowledged, however, that their intervention may not have been enough to counteract the patient's previous expectancies, a possibility supported by the fact that the response expectancies assessed before the start of the experimental manipulation correlated with both the frequency and the severity of the post-treatment nausea). In contrast, Roscoe et al. (2010) did succeed in reducing the attacks of nausea by emphasizing the benefits of an acupuncture technique in a previously identified “high-expectancy” group. But, unexpectedly, this manipulation also resulted in a significant augmentation of the occurrence of nausea in a group of patients who initially had low expectancies about it.

IDENTIFYING AN RISK FACTORS BEYOND CS-US CONTINGENCIES

We should now turn to the apparent inability of the Pavlovian model to explain why some patients are at much higher risk of suffering AN than others. It has already been noted that,

according to the contingency rules that govern the formation of associations, one might expect that all patients undergoing equivalent emetogenic chemotherapy schedules—i.e., a similar number of infusions with analogous cytotoxic drug doses—would suffer from similar AN symptoms. However, the validity of the contingency principle as the only factor that could account for AN is questioned by certain empirical findings. Firstly, it has been noted that the problem can occur under distinct chemotherapy regimes that, by employing the advised cytotoxic drugs, differ in their emetogenic capacity (van Komen and Redd, 1985; Andrykowski et al., 1988). In addition, in some cases no differences in the number of infusions, or in the severity of postchemotherapy nausea, have been found between those subjects who develop AN and those who do not (Fredrikson et al., 1993; Tyc et al., 1997). Finally, Andrykowski et al. (1988) noted that the consistency of AN was lower than would be expected on the basis of the Pavlovian model: in their study just 40% of patients who initially developed AN showed this response during the next 15 infusion sessions. Considering all of these facts, it seems necessary to accept that some variables other than those traditionally considered by learning theorists must be modulating the conditioning of nausea in cancer patients. Regression analyses have identified several factors, some of which can be classified as environmental or external, and others that refer to internal differences.

EXTERNAL VARIABLES

It seems reasonable to assume that many environmental features might affect the capacity of the patients to cope with nausea. Certainly, the challenge that cancer patients must meet is severe and can push them almost to their limits. Under such extreme circumstances, it might be the case that some details that might otherwise be regarded as irrelevant could become much more significant in terms of managing the unwanted side effects. For example, several studies have pointed out that the family characteristics of the patients may help them to deal with the collateral emetic effects of chemotherapy. Patients who have, for instance, non-conflicting and balanced families that allow them to speak openly about their suffering are less likely to experience AN (see, e.g., Youngmee and Morrow, 2007). In addition, Cohen et al. (1986) found that the characteristics of the treatment center strongly predicted the presence of AN, and suggested that some fine points pertaining to the chemotherapy room, such as being within sight of basins or the absence of entertainment or comfortable chairs, could facilitate the manifestation of emetic anticipatory symptoms. Further support for this suggestion comes from recent studies (e.g., McCarthy et al., 2012) claiming that cozy waiting and treatment rooms are necessary to reduce both anticipatory anxiety and pain in cancer patients.

Another hospital-related difference affecting patients is the antiemetic protocol dispensed by nurses and doctors. Clearly this practice is likely to be important as it constitutes the first pharmacological line of defense against the emetic syndrome induced by the chemotherapy. It is, however, far from uniform and the unification of intervention protocols still remains an unreach goal (e.g., Schwartzberg, 2011). Furthermore, it has been asserted that antiemetic treatments are often incorrectly

employed by doctors and nurses (e.g., Burmeister et al., 2012; Fernández-Ortega et al., 2012), perhaps because medical staff fail to appreciate fully the severity of the symptoms (Foubert and Vaessen, 2005; Majem et al., 2011). Given the close relationship between post-treatment nausea and AN, and that reports on this topic often recruit their sample from different hospitals, the adequacy of the antiemetic intervention can be an important factor in generating differences in AN. In this regard it should be noted that such variations could be argued by associative theorists to explain why similar chemotherapy regimes do not always produce equivalent anticipatory symptoms. An adequate use of this prophylactic medication could avoid, at least in some degree, the emetic capacity of chemotherapy and hence would reduce the possibilities of nausea conditioning. Thus, in a study including patients from different hospitals, a similar number of infusions of equivalent cytotoxic drugs should only be taken as a comparable contingency program if there are evidences that the antiemetic protocol is also equivalent.

INTERNAL VARIABLES

Personality variables are known to affect conditioning in humans but, in spite of this, they are not usually theoretically integrated into traditional associative learning models. It seems reasonable, particularly from a practical point of view, to know if cancer patients who develop AN do share some characteristics. Thus, in addition to hospital and family differences, regression models have isolated some other variables that provide a profile of those patients who are at greatest risk of suffering from AN. These studies indicate that variables such as being less than 50 years old, having susceptibility to motion sickness or nausea during pregnancy, and being under a state of anxiety, hostility or depression, can account for part of the variability of the occurrence of AN (e.g., Roscoe et al., 2011). This risk profile can be extended by taking into account some personality traits associated with AN. For instance, Challis and Stam (1992) found that AN correlated with higher scores in scales measuring suggestibility, and van Komen and Redd (1985) have reported similar correlations with traits such as future despair, social alienation, inhibited personality style, and anxiety. Additionally, Hursti et al. (1992) asked relapse-free cancer patients to complete several personality scales and to report how they experienced nausea when attending chemotherapy. Their results showed that neuroticism and inhibiting style were two dimensions that correlated with AN. Interestingly, this same sample of subjects was further explored by Fredrikson et al. (1993) in an attempt to determine if AN patients were more susceptible to conditioning. Subjects in this sample were classified as AN or Non-AN, and treated as independent groups. Their capacity to associate visual figures with a mild electric shock was assessed using a heart-rate measure. They found that patients who developed AN were more easily conditioned than those in the Non-AN group. Taken together, these latter two studies suggest a relationship between personality traits (a high score in neuroticism or introversion) and the aversive conditioning of both nausea and fear. This conclusion, however, needs to be treated with caution. First, the groups were compared as if their distribution was randomized when it was not, and secondly, the possibility exists that in a retrospective study of this

sort, the effects were generated by the chemotherapy treatment itself.

THE USE OF INDIVIDUAL DIFFERENCES IN ANIMAL MODELS OF AN

The difficulties of carrying out an experimental assessment of the role of personality traits in the development of AN might be resolved by using an animal model. Individual differences are not only to be found in our species—several studies have demonstrated consistent individual differences in rats and, moreover, that some of these can be used to predict some Pavlovian related phenomena (e.g., Robinson and Flagel, 2009). Of course there are features of human personality that cannot be modeled in animals as they are exclusively revealed as verbal thoughts. However, other animal behaviors parallel reasonably well some individual characteristics that, as it has just been mentioned, are present in those cancer patients that develop AN.

On the other hand, contextual aversions modeling AN can easily be reproduced in the rat by simply pairing a novel environment with the effects of an emetic drug. Rodríguez et al. (2000) showed that exposures to a novel place following an injection of lithium chloride (a fast-acting emetic drug) produce a learned aversion that can be assessed by simply measuring the ingestion of a novel flavor offered in that place—after a few of those trials a novel palatable solution is consumed unwillingly. The aversive properties acquired by the context after such training have been evaluated through other more accurate aversive measures such as the taste reactivity test (e.g., Limebeer et al., 2006), supporting the validity of the model. In the next sections the possibilities of using some animal differences to improve the modeling of AN will be discussed.

ANIMAL DIFFERENCES IN ANXIETY

Traits labeled anxious-neurotic in humans can be assessed in rats by using tests such as the elevated maze or defensive burying behavior. In particular, a reluctance to enter open arms, or a failure to cover dangerous or disgusting objects, can be considered as a sign that a rat is anxious (Ho et al., 2002). A possible way of testing the influence of personality traits on Pavlovian conditioning of nausea, therefore, could be to assess if anxious rats show better acquisition of a context-illness association.

To our knowledge this specific investigation has not yet been carried out, but there are some studies that support its viability. Borta et al. (2006), for example, observed that rats with a low tendency to enter open arms, i.e., those supposed to be more anxious, seemed to learn more readily an association between a tone and a shock. In another study, Walker et al. (2008) found that results obtained in the test of defensive burying behavior predicted stronger aversive learning in which a new cage (a contextual CS) was paired with attacks by a male (bites as the US) living in that context: 44% of the variability in locomotor activity in that cage during the test (carried out when the resident male was not present) was explained by the previous defensive burying related behaviors that the rats showed in response to prods that had been placed in their home cages.

The role of anxiety differences in the success of associative interventions intended to prevent AN could also be analyzed

using animal models. Latent inhibition and overshadowing (see Note 1), which have already been shown to reduce AN in rats (see Symonds and Hall, 2012), demand attentional processes (e.g., Granger et al., 2012) that could be affected by a state of acute anxiety (Braunstein-Bercovitz et al., 2002). If anxiety disrupts the capacity to select those more reliable environmental stimuli, the retardation in aversive context conditioning derived from these interventions could be in question for these more anxious subjects. This can be analyzed by testing if, in a preexposed or overshadowed context, anxious rats consume less of a novel taste than normal rats.

OTHER ANIMAL DIFFERENCES

Regressive methodology can also readily be used to assess, in a highly controlled way, the relationship between experimentally induced AN and other individual characteristics, even though, unlike anxiety, these are not facilitators of aversive conditioning in non-human animals. For example, the application of unpredictable chronic mild stress (UCMS) in rats to model human depression has no effect on contextual aversion but, interestingly, it does impair place preference. In particular, UCMS appears to produce anhedonia, i.e., a reduction in the capacity of the subjects to appreciate the appetitive effects of the rewarding drugs. (see Willner and Mitchell, 2002 for a review of these models.)

Anhedonia is a core symptom of human depression that can be identified in rats by simply registering its ingestion of sucrose (e.g., Strekalova and Steinbusch, 2010; but see Matthews et al., 1995). Thus, a possible investigation to test if depression is related to AN could be to analyze if subjects displaying a lower preference for sucrose are later more likely to develop AN. Similarly, the connection between hostility and AN could be studied in animals by using some rat strains that differ in their latency to attack an opponent male—the shorter the latency of attack the higher the supposed level of aggression (e.g., de Boer et al., 2003). Considering these examples, it seems promising to use this methodological strategy to complete a better model of AN, by seeing whether any other differential characteristic present in the AN population can also be validated in rodents.

CONCLUSION

It is widely acknowledged that animal models can be a useful first step in testing the validity of any novel antiemetic intervention. And it is now clear that developing animal models of AN that take into account individual differences will bestow certain advantages. First, it will clearly be more efficient to test the reliability of any given prophylactic intervention in animals known to have a particular sensitivity to the conditioning of nausea. And success in identifying the personality variables that make some people particularly vulnerable to AN would then allow us to focus our antiemetic efforts on such people, providing them with more accurate treatment that is appropriate for their specific profile. For example, relaxation techniques may be necessary in those patients with high scores in anxiety before the application of associative interventions in order to guarantee its maximum efficiency. The costs involved in pharmaceutical and psychosocial interventions are substantial, providing a further reason for

concentrating our efforts on these more vulnerable patients and giving them more adequate therapy. Finally, such studies could pay theoretical dividends by confirming the relevance of certain traits or response tendencies as predictors of the development of AN.

REFERENCES

- Aapro, M. S., Molassiotis, A., and Olver, I. (2005). Anticipatory nausea and vomiting. *Support. Care Cancer* 13, 117–121. doi: 10.1007/s00520-004-0745-8
- Andrykowski, M. A., Jacobsen, P. B., Marks, E., Gorfinkle, K., Hakes, T. B., Kaufman, R. J., et al. (1988). Prevalence, predictors, and course of anticipatory nausea in women receiving adjuvant chemotherapy for breast cancer. *Cancer* 62, 2607–2613.
- Borta, A., Wöhr, M., and Schwarting, R. K. W. (2006). Rat ultrasonic vocalization in aversively motivated situations and the role of individual differences in anxiety-related behavior. *Behav. Brain Res.* 166, 271–280. doi: 10.1016/j.bbr.2005.08.009
- Braunstein-Bercovitz, H., Rammsayer, T., Gibbons, H., and Lubow, R. E. (2002). Latent inhibition deficits in high-schizotypal normals: symptom-specific or anxiety-related? *Schizophr. Res.* 53, 109–121.
- Burmeister, H., Aebi, S., Studer, C., Fey, M. F., and Gautschi, O. (2012). Adherence to ESMO clinical recommendations for prophylaxis of chemotherapy-induced nausea and vomiting. *Support. Care Cancer* 20, 141–147. doi: 10.1007/s00520-010-1079-3
- Challis, G. B., and Stam, H. J. (1992). A longitudinal study of the development of anticipatory nausea and vomiting in cancer chemotherapy patients: the role of absorption and autonomic perception. *Health Psychol.* 11, 181–189. doi: 10.1037/0278-6133.11.3.181
- Cohen, R. E., Blanchard, E. B., Ruckdeschel, J. C., and Smolen, R. C. (1986). Prevalence and correlates of posttreatment and anticipatory nausea and vomiting in cancer chemotherapy. *J. Psychosom. Res.* 30, 643–654. doi: 10.1016/0022-3999(86)90098-X
- Colloca, L., and Miller, F. G. (2011). The nocebo effect and its relevance for clinical practice. *Psychosom. Med.* 73, 598–603. doi: 10.1097/PSY.0b013e3182294a50
- de Boer, S. F., van der Vegt, B. J., and Koolhaas, J. M. (2003). Individual variation in aggression of feral rodents strains: a standard for the genetics of aggression and violence? *Behav. Genet.* 33, 485–501.
- Fernández-Ortega, P., Caloto, M. T., Chirveches, E., Marquilles, R., San Francisco, J., Quesada, E., et al. (2012). Chemotherapy-induced nausea and vomiting in clinical practice: impact on patients' quality of life. *Support. Care Cancer* 20, 3141–3148. doi: 10.1007/s00520-012-1448-1
- Foubert, J., and Vaessen, G. (2005). Nausea: the neglected symptom? *Eur. J. Oncol. Nurs.* 9, 21–32. doi: 10.1016/j.ejon.2004.03.006
- Fredrikson, M., Hursti, T., Salmi, P., Börjeson, S., Fürst, C.-J., Peterson, C., et al. (1993). Conditioned nausea after cancer chemotherapy and autonomic nervous system conditionability. *Scand. J. Psychol.* 34, 318–327. doi: 10.1111/j.1467-9450.1993.tb01128.x
- Granger, K. T., Prados, J., and Young, A. M. J. (2012). Disruption of overshadowing and latent inhibition in high schizotypy individuals. *Behav. Brain Res.* 233, 201–208. doi: 10.1016/j.bbr.2012.05.003
- Haiderali, A., Menditto, L., Good, M., Teitelbaum, A., and Wegner, J. (2011). Impact on daily functioning and indirect/direct costs associated with chemotherapy-induced nausea and vomiting (CINV) in a US population. *Support. Care Cancer* 19, 843–851. doi: 10.1007/s00520-010-0915-9
- Hickok, J. T., Roscoe, J. A., and Morrow, G. R. (2001). The role of patients' expectations in the development of anticipatory nausea related to chemotherapy for cancer. *J. Pain Symptom Manage.* 22, 843–850. doi: 10.1016/S0885-3924(01)00317-7
- Ho, Y.-J., Eichendorff, J., and Schwarting, R. K. W. (2002). Individual response profiles of male wistar rats in animal models for anxiety and depression. *Behav. Brain Res.* 136, 1–12. doi: 10.1016/S0166-4328(02)00089-X
- Hursti, T., Fredrikson, M., Börjeson, S., Fürst, C. J., Peterson, C., and Steineck, G. (1992). Association between personality characteristics and the prevalence and extinction of conditioned nausea after chemotherapy. *J. Psychosoc. Oncol.* 10, 59–77. doi: 10.1300/J077v10n02_04
- Hsu, E. S. (2010). A review of granisetron, 5-Hydroxytryptamine (3) receptor antagonists, and other antiemetics. *Am. J. Ther.* 17, 476–486. doi: 10.1097/MJT.0b013e3181ea7821
- Jacobsen, P. B., Bovbjerg, D. H., and Redd, W. H. (1993). Anticipatory anxiety in women receiving chemotherapy for breast cancer. *Health Psychol.* 12, 469–475. doi: 10.1037/0278-6133.12.6.469
- Kayl, A. E., and Meyers, C. A. (2006). Side-effects of chemotherapy and quality of life in ovarian and breast cancer patients. *Curr. Opin. Obstet. Gynecol.* 18, 24–24. doi: 10.1097/01.gco.00000192996.20040.24
- Klosterhalfen, S., Kellermann, S., Stockhorst, U., Wolf, J., Kirschbaum, C., Hall, G., et al. (2005). Latent inhibition of rotation chair-induced nausea in healthy male and female volunteers. *Psychosom. Med.* 67, 335–340. doi: 10.1097/01.psy.0000156930.00201.e0
- Limebeer, C. L., Hall, G., and Parker, L. (2006). Exposure to a lithium-paired context elicits gaping in rats: a model of anticipatory nausea. *Physiol. Behav.* 88, 398–403. doi: 10.1016/j.physbeh.2006.04.014
- McCarthy, M., Glick, R., Green, J., Plummer, K., Peters, K., Johnsey, L., et al. (2012). Comfort first: an evaluation of a procedural pain management programme for children with cancer. *Psychooncology* 22, 775–782. doi: 10.1002/pon.3061
- Majem, M., Moreno, M. E., Calvo, N., Feliu, A., Pérez, J., Mangués, A. M., et al. (2011). Perception of health care providers versus patient reported incidence of chemotherapy-induced nausea and vomiting after the addition of NK-1 receptor antagonists. *Support. Care Cancer* 19, 1983–1990. doi: 10.1007/s00520-010-1042-3
- Matteson, S., Roscoe, J., Hickok, J., and Morrow, G. R. (2002). The role of behavioral conditioning in the development of nausea. *Am. J. Obstet. Gynecol.* 186, S239–S243.
- Matthews, K., Forbes, N., and Reid, I. (1995). Sucrose consumption as an hedonic measure following chronic unpredictable mild stress. *Physiol. Behav.* 57, 241–248. doi: 10.1016/0031-9384(94)00286-E
- Montgomery, G. H., Tomoyasu, N., Bovbjerg, D. H., Andrykowski, M. A., Currie, V. E., Jacobsen, P. B., et al. (1998). Patient's pretreatment expectations of chemotherapy-related nausea are an independent predictor of anticipatory nausea. *Ann. Behav. Med.* 20, 104–109. doi: 10.1007/BF02884456
- Nesse, R. N., Carli, T., Curtis, G. C., and Kleinman, P. D. (1980). Pretreatment nausea in cancer chemotherapy: a conditioned response? *Psychosom. Med.* 42, 33–36.
- Robinson, T. E., and Flagel, S. B. (2009). Dissociating the predictive and incentive motivational properties of reward-related cues through the study of individual differences. *Biol. Psychiatry* 65, 869–873. doi: 10.1016/j.biopsych.2008.09.006
- Rodríguez, M., López, M., Symonds, M., and Hall, G. (2000). Lithium-induced context aversion in rats as a model of anticipatory nausea in humans. *Physiol. Behav.* 71, 571–579. doi: 10.1016/S0031-9384(00)00376-0
- Roscoe, J. A., Morrow, G. R., Aapro, M. S., Molassiotis, A., and Olver, I. (2011). Anticipatory nausea and vomiting. *Support. Care Cancer* 19, 1533–1538. doi: 10.1007/s00520-010-0980-0
- Roscoe, J. A., O'Neill, M., Jean-Pierre, P., Heckler, C. E., Kaptchuk, T. J., Bushnow, P., et al. (2010). An exploratory study on the effects of an expectancy manipulation on chemotherapy-related nausea. *J. Pain Symptom Manage.* 40, 379–390.
- Schiff, E., and Ben-Arye, E. (2011). Complementary therapies for side effects of chemotherapy and radiotherapy in the upper gastrointestinal system. *Eur. J. Integr. Med.* 3, 11–16. doi: 10.1016/j.eujim.2011.02.006
- Schwartzberg, L. S. (2011). Incorporating chemotherapy-induced nausea and vomiting guidelines into clinical practice. *Clin. Adv. Hematol. Oncol.* 9, 11(Suppl. 27), 9–10.
- Shelke, A. R., Roscoe, J. A., Morrow, G. R., Colman, L. K., Banerjee, T. K., and Kirshner, J. J. (2008). Effect of a nausea expectancy manipulation on chemotherapy-induced nausea: a University of Rochester

- cancer center community clinical oncology program study. *J. Pain Symptom Manage.* 35, 381–387. doi: 10.1016/j.jpainsymman.2007.05.008
- Sohl, S. J., Schnur, J. B., and Montgomery, G. H. (2009). A meta-analysis of the relationship between response expectancies and cancer treatment-related side effects. *J. Pain Symptom Manage.* 38, 775–784. doi: 10.1016/j.jpainsymman.2009.01.008
- Stewart-Williams, S., and Podd, J. (2004). The placebo effect: dissolving the expectancy versus conditioning debate. *Psychol. Bull.* 130, 324–340. doi: 10.1037/0033-2909.130.2.324
- Stockhorst, U., Wiener, A., Klosterhalfen, S., Klosterhalfen, W., Aul, C., and Steingruber, H.-J. (1998). Effects of overshadowing on conditioning nausea in cancer patients: an experimental study. *Physiol. Behav.* 64, 743–753. doi: 10.1016/S0031-9384(98)00135-8
- Strekalova, T., and Steinbusch, W. M. (2010). Measuring behavior in mice with chronic stress depression paradigm. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 34, 348–361.
- Symonds, M., and Hall, G. (2012). “Avoidance, anxiety, and aversion in the clinical setting: the role of classical conditioning,” in *Clinical Applications of Learning Theory*, eds M. Haselgrove and L. Hogarth (New York; Hove: Psychology Press), 27–47.
- Tomoyasu, N., Bovbjerg, D. H., and Jacobsen, P. B. (1996). Conditioned reactions to cancer chemotherapy: percent reinforcement predicts anticipatory nausea. *Physiol. Behav.* 59, 273–276. doi: 10.1016/0031-9384(95)02072-1
- Tyc, V. L., Mulhern, R. K., Barclay, D. R., Smith, B. F., and Bieberich, A. A. (1997). Variables associated with anticipatory nausea and vomiting in pediatric cancer patients receiving ondansetron antiemetic therapy. *J. Pediatr. Psychol.* 22, 45–58. doi: 10.1093/jpepsy/22.1.45
- van Komen, R. W., and Redd, W. H. (1985). Personality factors associated with anticipatory nausea/vomiting in patients receiving cancer chemotherapy. *Health Psychol.* 4, 189–202. doi: 10.1037/0278-6133.4.3.189
- Walker, F. R., Hinwood, M., Masters, L., Deilenberg, R. A., and Day, T. A. (2008). Individual differences predict susceptibility to conditioned fear arising from psychosocial trauma. *J. Psychiatr. Res.* 42, 371–383. doi: 10.1016/j.jpsychires.2007.01.007
- Watson, M., Meyer, L., Thomson, A., and Osofsky, S. (1998). Psychological factors predicting nausea and vomiting in breast cancer patients on chemotherapy. *Eur. J. Cancer* 34, 831–837. doi: 10.1016/S0959-8049(97)10146-0
- Willner, P., and Mitchell, P. J. (2002). The validity of animal models of predisposition to depression. *Behav. Pharmacol.* 13, 169–188. doi: 10.1097/00008877-200205000-00001
- Youngmee, K., and Morrow, G. R. (2007). The effects of family support, anxiety, and post-treatment nausea on the development of anticipatory nausea: a latent growth model. *J. Pain Symptom Manage.* 34, 265–276.
- Yugin, L. (1989). Treating patients with nervous vomiting in the dental office by point-stimulating therapy. *Spec. Care Dentist.* 9, 27–28. doi: 10.1111/j.1754-4505.1989.tb01017.x

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: 19 December 2012; accepted: 17 July 2013; published online: 09 August 2013.

Citation: Rodríguez M (2013) Individual differences in chemotherapy-induced anticipatory nausea. *Front. Psychol.* 4:502. doi: 10.3389/fpsyg.2013.00502
This article was submitted to *Frontiers in Personality Science and Individual Differences*, a specialty of *Frontiers in Psychology*.

Copyright © 2013 Rodríguez. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Pathological gamblers are more vulnerable to the illusion of control in a standard associative learning task

Cristina Orgaz¹, Ana Estévez² and Helena Matute^{2*}

¹ Department of Psychology, Universidad Nacional de Educación a Distancia, Madrid, Spain

² Department of Psychology, Universidad de Deusto, Bilbao, Spain

Edited by:

Rachel M. Msetfi, University of Limerick, Ireland

Reviewed by:

Tom Beckers, KU Leuven, Belgium
Mark Haselgrove, The University of Nottingham, UK
Irina Baetu, University of Adelaide, Australia

*Correspondence:

Helena Matute, Departamento de Fundamentos y Métodos de la, Psicología, Universidad de Deusto, Apartado 1, 48080 Bilbao, Spain
e-mail: matute@deusto.es

An illusion of control is said to occur when a person believes that he or she controls an outcome that is uncontrollable. Pathological gambling has often been related to an illusion of control, but the assessment of the illusion has generally used introspective methods in domain-specific (i.e., gambling) situations. The illusion of control of pathological gamblers, however, could be a more general problem, affecting other aspects of their daily life. Thus, we tested them using a standard associative learning task which is known to produce illusions of control in most people under certain conditions. The results showed that the illusion was significantly stronger in pathological gamblers than in a control undiagnosed sample. This suggests (1) that the experimental tasks used in basic associative learning research could be used to detect illusions of control in gamblers in a more indirect way, as compared to introspective and domain-specific questionnaires; and (2), that in addition to gambling-specific problems, pathological gamblers may have a higher-than-normal illusion of control in their daily life.

Keywords: gambling, illusion of control, associative learning, contingency learning, contingency judgments, causal learning

The perception of control over important events in our lives has been studied from many different perspectives in psychology. It allows us to predict the consequences of our actions and the actions of others, which adaptively can imply the difference between surviving and perishing. Sometimes, however, perceived control is not real. People often fail to distinguish those events that are controllable from those that are not, which gives rise to the illusion of control (Langer, 1975). The illusion of control can be defined as the tendency to believe that our behavior is the cause of the occurrence of desired events that occur independently of our own actions (Alloy and Abramson, 1979; Taylor and Brown, 1988; Matute, 1996).

The illusion of control is a universal phenomenon which has been observed to occur in most people and under many different conditions. Many laboratory experiments have shown that college students develop the illusion that they are controlling uncontrollable lights or tones or lottery tickets (e.g., Langer, 1975; Alloy and Abramson, 1979; Wasserman et al., 1983; Matute, 1996; Aeschleman et al., 2003; Msetfi et al., 2005). Illusions of control have also been reported in students trying to cure fictitious patients in a medical decision task (Blanco et al., 2011), or in Internet users who are trying to obtain points in an otherwise uncontrollable computer game (Matute et al., 2007). The illusion of control is also well-known among athletes and sports players, who often feel that a given ritual or lucky charm is necessary for success (Bleak and Frederick, 1998), or even in sport spectators, who tend to feel that supporting (or not) their favorite team through their TV at home contributes to the happy (or disastrous) score of the team (Pronin et al., 2006). Trading and consumer behavior have also been shown to be vulnerable to the illusion of control (Fenton-O'Creevy et al., 2003; Kramer and

Block, 2011), as have companies and organizations themselves (Durand, 2003).

Finding out which conditions modulate the development and maintenance of the illusion of control is therefore important, given that it affects almost anyone and almost any decision or aspect in our daily life. Thus, at the same time that there is an extensive scientific literature which has highlighted the universality of this bias, there is also an important research agenda which explores the degree to which the illusion of control is sensitive to individual differences among humans.

The study of individual differences in the illusion of control has been concerned with gender (with women generally showing stronger illusions of control than men; see Alloy and Abramson, 1979; Wong, 1982; Vyse, 1997; Wolfradt, 1997; Dag, 1999), superstitious attitudes (Rudski, 2004), psychopathology (e.g., Wolfradt, 1997; Dag, 1999); cooperative behavior (Morris et al., 1998; Goldberg et al., 2005), or even sports (Laurendeau, 2006). It is also possible to come across studies about the illusion of control in psychological disorders such as depression (with depressed people generally being less vulnerable to the illusion of control; see Alloy and Abramson, 1979; Vázquez, 1987; Blanco et al., 2009, 2012), obsessive-compulsive disorder (Reuven-Magril and Reuven, 2008) and physical health (Harris and Middleton, 1994).

Pathological gambling is one of several psychological disorders with which the illusion of control has been most strongly associated (Ladouceur et al., 1984; Wolfgang et al., 1984; Coventry and Norman, 1998; Källmén et al., 2008; Lingyuan and Austin, 2008). It is a disorder of impulse control in which cognitive distortions are assumed to play an important role (Myrseth et al., 2010). According to some researchers (Sharpe, 2008; Lund, 2011) people

with pathological gambling disorder bet because they hold wrong or irrational beliefs about the game and their ability to influence its outcome. Pathological gambling is closely related to the perception of the player that, to some extent, he or she can control the outcome of his or her bets (Goodie, 2005).

According to some, however, the lack of valid measures has impeded the systematic investigation of cognitive biases in gamblers (MacKillop et al., 2006). Data suggestive of an illusion of control in gamblers have often been obtained through talk-aloud methods and self-reporting measures and almost always in domain-specific (i.e., gambling) conditions (e.g., Dickerson, 1993; Strickland et al., 2006). As is already well-known in the literature, this type of data collection can be subject to a series of social desirability biases, avoidance of cognitive dissonance, or even investigator biases, particularly when the questions are related to the variable under study (i.e., in this case, gambling). Recent reviews have shown that the contribution of cognitive distortions to the development and maintenance of pathological gambling behavior is still in need of further scrutiny (Fortune and Goodie, 2011). Furthermore, some researchers from the clinical domain have argued, against the view of many others (e.g., Coventry and Norman, 1998; Källmén et al., 2008; Hudgens-Haney et al., 2013), that the illusion of control has only a limited influence in the maintenance of gambling behavior (Labrador et al., 2002; Mañoso et al., 2004). A better understanding of gamblers' subjective judgments of control seems therefore a necessary step in clarifying the etiology and maintenance of pathological gambling behavior (Matheson et al., 2009).

A question of particular interest is whether pathological gamblers actually suffer from a general distortion in their perception of control or is, by contrast, a domain-specific problem what they suffer. If it were a generalized distortion, then they should show a stronger than normal illusion of control in tasks and activities which are unrelated to gambling. Therefore, it seems important to rely on a more indirect methodology which is unrelated to gambling and which can collect indicators of the illusion of control that are not mediated by introspection. For all these reasons, we propose that the study of pathological gamblers should benefit from using the same assessment techniques that are typically used in the study of contingency judgments and illusions of control in general associative learning theory and research. Of particular interest, from our point of view, is that this methodology will allow us to test, not whether gamblers develop illusions of control during gambling, but, most importantly, whether they tend to overestimate cause-effect relationships in other areas of their life as well.

CONTINGENCY LEARNING AND THE ILLUSION OF CONTROL

The clinical and social psychology approach to the illusion of control has typically explained this illusion as a means to protect self-esteem (e.g., Taylor and Brown, 1988; Alloy and Clements, 1992). However, these illusions have also been reported in many cases in which participants are not personally involved and their self-esteem is not at risk, as when participants ask somebody else to roll a dice for them (e.g., Wohl and Enzle, 2009), when participants are just spectators in a sports competition and believe they influence their team's results (e.g., Pronin et al., 2006), or

when participants develop the illusion by just observing or being told that someone took a (fake) medicine and reported feeling better (Matute et al., 2011). Associative learning researchers have explained the illusion of control as a special case of the illusion of causality, a cognitive bias that takes place in most people when associating causes and effects in null contingency situations (Matute et al., 2011). In this framework, being personally involved or trying to protect self-esteem is not critical, as the illusion is thought to be the output of the way our cognitive system interacts with the world and extracts contingency and causal information from it (e.g., Matute, 1996; Msetfi et al., 2005, 2007; Allan et al., 2008; Matute et al., 2011).

In order to infer that a causal relationship exists, the potential cause (the participants' action, in the case of the illusion of control) and the outcome should be contingent to each other. A commonly used index of contingency is the Δp index (Jenkins and Ward, 1965; Allan and Jenkins, 1983). It is calculated as the probability of the outcome occurring when the potential cause (i.e., the response, in the case of illusion of control) has been presented $P(O|C)$, minus the probability of the outcome occurring when the cause is absent, $P(O|\neg C)$. That is, $\Delta p = P(O|C) - P(O|\neg C)$.

A zero contingency relationship between our behavior and an outcome would be that in which the outcome occurs with the same probability regardless of whether we perform the response. Thus, a value of Δp of 0 means that our behavior does not cause the outcome. An illusion of control is said to occur in a zero contingency situation whenever people report a subjective judgment of contingency significantly higher than 0. This is a very common illusion of causality which has been shown in many different experiments in the associative learning literature (e.g., Alloy and Abramson, 1979; Wasserman et al., 1983; Matute, 1996; Allan et al., 2005; Msetfi et al., 2005, 2007; Matute et al., 2007; Hannah and Beneteau, 2009; Blanco et al., 2012). According to associative theories, this illusion is a consequence of the associative learning mechanism constantly trying to associate causes and effects. It sometimes overestimates the relationship between potential causes and effects, particularly under certain conditions.

One of the variables that has been most clearly established to affect the development of the illusion of causality is the probability of the outcome, for instance, the probability with which spontaneous remissions of pain occur (e.g., Alloy and Abramson, 1979; Allan and Jenkins, 1983; Matute, 1995; Wasserman et al., 1996; Buehner et al., 2003; Allan et al., 2005, 2008; Msetfi et al., 2005, 2007; Musca et al., 2010). Another variable that is known to affect this illusion is the probability of responding (or, more generally, the probability with which the potential cause occurs; e.g., Allan and Jenkins, 1983; Matute, 1996; Wasserman et al., 1996; Perales et al., 2005; Hannah and Beneteau, 2009; Matute et al., 2011; Vadillo et al., 2011). The higher these two probabilities, the higher the probability that coincidences will occur between the potential cause and the outcome, and thus, the higher the probability than an illusion of control will develop (see Blanco et al., 2011, 2013; Hannah and Beneteau, 2009).

Therefore, we used a standard task that measures perceived contingency with respect to an actual null contingency in a fictitious medical scenario. The outcome was programmed to occur

at high rate (i.e., high frequency of spontaneous recovery in fictitious patients), so that control participants would develop the illusion, particularly if they responded frequently. This procedure should be low on biases inherent to introspective and domain-specific measures, but should nevertheless induce an illusion of control in most participants. If pathological gamblers suffer from a stronger-than-normal distortion in their general perception of contingency, their bias should manifest in this standard medical judgments task as compared to the control group. If this were the case, this would mean that the gamblers' misperception of control is not restricted to their gambling activities, but could possibly be generalizable to other aspects of their daily life.

METHODS

PARTICIPANTS AND APPARATUS

One hundred anonymous participants took part in this experiment. The gambler group was recruited through FEJAR (Spanish Federation of Rehabilitated Gamblers). It consisted of 49 participants (42 men and 7 women, mean age = 40.4, $SD = 11.31$) who had been diagnosed of pathological gambling using the South Oak Gambling Screen Questionnaire (i.e., SOGS, see Leiseur and Blume, 1987, Spanish adaptation by Echeburua et al., 1994). They were currently in rehabilitation stage. Their voluntary and anonymous participation was requested through FEJAR. The experiment was available during 6 months at our online laboratory, <http://www.labpsico.deusto.es>, so that participants in both groups could access the experiment at their convenience.

The control group consisted of 51 anonymous Internet users (27 men and 24 women, mean age = 37.04, $SD = 10.54$) who happened to visit our online laboratory (because they were visiting a web site or social network that linked our laboratory or because they were searching the Internet for concepts related to information published in our laboratory, or because of other reasons) during the time the experiment was available, and voluntarily decided to participate. To increase participation and following ethical standards for human research over the Internet (Frankel and Siang, 1999), we never ask participants in our online laboratory to provide additional personal or demographic data, nor do we use cookies or software to obtain information without their consent.

Internet experiments could be in principle suspect to providing noisy data, but they have been shown to yield results that are as reliable as those observed in the laboratory if certain cautionary measures are taken (e.g., Kraut et al., 2004; Germine et al., 2012; Ryan et al., 2013). Most importantly for our present purposes, illusion of control effects have already been replicated both in the laboratory and through the Internet using associative learning procedures similar to the one we are using here (e.g., Matute et al., 2007; Blanco et al., 2013).

PROCEDURE AND DESIGN

Participants performed a task known as the "Contingency Judgments Task," which, under different variations and versions, is frequently used in the study of associative learning (e.g., Allan et al., 2005; Msetfi et al., 2005; Blanco et al., 2011). In our procedure, participants were asked to imagine being a medical doctor who was using an experimental medicine, Batatrim, which might

cure painful crises produced by a fictitious disease called Lindsay Syndrome. They were also told that the effectiveness of Batatrim had not been proven yet and that this medicine produced some secondary effects, so that they needed to use it with caution (this instruction was given so that participants would not administer Batatrim at every opportunity to their fictitious patients). Participants were exposed to the records of 100 fictitious patients suffering from Lindsay's crises, one patient per trial. In each trial, the screen was divided in three horizontal panels. In the upper panel, participants were informed that that patient was suffering a crisis. In the second panel, participants could choose between giving or not giving Batatrim to this particular patient. Responses to this question were given by clicking on one of two bottoms, "Yes" or "No." The lower panel of each trial was presented immediately after participants entered their response. It showed whether the fictitious patient overcame the crisis. It also showed a "click to continue" bottom that participants could click at their pace in order to continue to the next trial.

After the 100 training trials, participants were asked to rate the efficacy of Batatrim in healing the crises. For this purpose the following question was presented in the middle of the screen: to what extent do you believe that Batatrim has been effective in healing the crises of the patients you have seen? This test question was answered in a scale ranging from 0 (labeled "Definitely not") to 100 (labeled "Definitely").

The outcome (healings) occurred with a probability of 0.80, but following a pseudorandom order which was independent of the participants' behavior. As mentioned in the Introduction, the reason we are using a high probability of the outcome is because this has been shown to favor the development of the illusion in most people in previous reports (Alloy and Abramson, 1979; Allan and Jenkins, 1983; Matute, 1995; Hannah and Beneteau, 2009). Thus, even though it occurred very frequently, the outcome was absolutely independent of the participants' behavior, which means that any subjective estimation of control that is significantly greater than 0 can be considered an illusion of control. Most importantly, the critical question of this experiment is whether the gambler group will show a stronger illusion than the control group under this high-outcome procedure.

RESULTS

As could be expected from previous reports on the illusion of control using a high probability of the outcome, both groups of participants overestimated the contingency between their behavior and the outcome. Student's *t*-tests confirmed that in both groups the judgments of contingency were significantly higher than 0, $t_{(48)} = 24.42$, $p < 0.01$ for the gamblers group, and $t_{(50)} = 13.85$, $p < 0.01$ for the control group. The critical result in this experiment, however, is the stronger illusion of control that was observed in the gamblers group ($M = 70.61$, $SE = 2.89$) as compared to the control group ($M = 57.20$, $SE = 4.12$). A *t*-test revealed that this difference was statistically significant, $t_{(98)} = 2.643$, $p = 0.010$. These data indicate that pathological gamblers perceived a stronger illusory relationship between their behavior and the desired outcome in a medical diagnostic task commonly used to assess associative learning and contingency judgments in laboratory settings.

In addition, and in line with previous reports (Matute, 1996; Blanco et al., 2009, 2011; Hannah and Beneteau, 2009), the results of this experiment showed a significant correlation between the probability with which participants administered the medicine to their fictitious patients and their judgment of control, Pearson's $r = 0.418$, $p < 0.01$. That is, the higher the probability of responding, the higher the illusion of control. Interestingly, however, there were no significant differences in the probability with which the gamblers group ($M = 63.27$, $SE = 0.054$) and the control group ($M = 59.69$, $SE = 0.048$) administered the medicine to their patients, $t_{(98)} = 0.492$, $p > 0.05$. Thus, as expected, participants' judgments of control were highly correlated with their probability of responding, so that they developed stronger illusions as responding increased. However, the higher illusion observed in the gamblers group was not due to stronger responding in this group. Thus, a genuine difference in the way they process causal information seems to be responsible for the stronger illusion shown by this group.

Before we finish this section some comment is in order in relation to the possible influence of demographic variables such as age and gender on the observed results. First, no significant differences were observed between the two groups with respect to age, $t_{(98)} = 1.373$, $p > 0.05$, thus, the observed differences cannot be attributed to this variable. However, and despite the experiment being available online during 6 months, the gamblers group was composed mainly of men. Thus, the effect of gender cannot be properly analyzed. Nevertheless, the results of the present experiment are exactly opposite to what should be expected if the effect of group and gender had been confounded. Previous research had shown that women are more vulnerable to the illusion of control than men (Alloy and Abramson, 1979; Wong, 1982; Vyse, 1997; Wolfradt, 1997; Dag, 1999). Thus, if anything, a group composed mainly of men should have shown a weaker, rather than a stronger illusion.

DISCUSSION

The present results show that in a standard medical judgmental task in which the outcome occurs at a high rate and most people develop an illusion of control, pathological gamblers show an illusion that is even stronger than that of control participants. That is, in this experiment, the actual causal relationship between the participants' administering a medicine to the fictitious patients and the healing of the patients was non-existent, but even so, gambler participants perceived it as highly contingent (compared to a control group without diagnosed pathology who also developed the illusion but less intensely). The use of this associative learning task, which also reflects a feasible situation in daily life (i.e., using medication to reduce pain or illness), suggests that the illusion of control may be a generalized problem in gamblers' daily life and is, therefore, not restricted to their gambling behavior. This has implications for our understanding of the way pathological gamblers process causal information outside of the gambling domain, and may provide hints for a more general assessment and treatment of their problem.

As previously mentioned, the illusion of control is explained from most associative learning theories as a misperception of

contingencies that takes place in most people under certain conditions. There are some differences between different theories, and this misperception could occur through several different mechanisms. For instance, it could be due to people giving more weight to cases that confirm that their behavior is followed by the desired outcome and less weight to other information such as, for example, those cases in which the result occurs when they do not act (e.g., cases in which the health crises are also overcome even when the patient is not given the medicine). Several associative theories have contemplated a weighted Δp rule, in which people would weight differently the different types of information that can be encountered, with maximal weight given to those cases in which both the potential cause and the outcome are present, and minimum weight to cases in which neither one is present (e.g., Wasserman et al., 1996). A related but slightly different approach has been taken by associative theories that emphasize the differential perception of contextual information or, in other words, the way participants perceive what happens during the time in which no cues or outcomes are being presented and therefore they are just exposed to the experimental context (Msetfi et al., 2005, 2007). There are also theories that propose that the locus of the distortion does not reside at the perception (or encoding) stage, but at the subsequent judgmental stage (e.g., Allan et al., 2008). Yet, other theories have emphasized the role of the probability of responding (or, more generally, of the potential cause), so that, for instance, participants have been shown to expose themselves to more (adventitious) cause-effect coincidences when they respond frequently to obtain the outcome (i.e., assuming the outcome also occurs with high frequency, which is usually the case in situations in which illusions occur; see e.g., Matute, 1996; Hannah and Beneteau, 2009; Matute et al., 2011; Blanco et al., 2012, 2013). In these cases the number of accidental coincidences increases as the probability of the cause increases (e.g., as participants respond more), thereby the illusory perception of causality becomes stronger as well. The locus of the illusion of control here is therefore behavioral: the more participants respond, the greater their illusion.

This latter view is the one we have favored in many previous reports, and the basic finding that the probability of responding influences the illusion has been replicated in many experiments (e.g., Matute, 1996; Blanco et al., 2009, 2011, 2012; Hannah and Beneteau, 2009). The general effect of the probability of responding has also been replicated in the present experiment, in which the outcome was frequent and the results showed that the higher the probability of responding, the higher the illusion of control. Importantly, however, the probability-of-response effect was clearly not responsible for the stronger illusion of control developed by the gamblers group, as differences in response probability were not observed between the two groups. Thus, even though the present experiment was not designed to discriminate among the different theories of the illusion of control, it seems clear that the locus of the stronger illusion observed in the gamblers group in this medical task resides, not at the behavioral level, but at the perceptual or the judgmental stages.

The present results also suggest that the development of programs and strategies to help people be more accurate in their general detection of contingencies could be a good complement

to clinical therapies designed to eliminate gambling behavior. Cognitive interventions for pathological gambling usually focus on cognitive behavioral therapy (Gooding and Tarrier, 2009) and on identifying and restructuring cognitive distortions (Ledgerwood and Petry, 2005; Fortune and Goodie, 2011). Making use of strategies developed under the general associative learning framework to reduce the likelihood of overestimation of contingencies could probably be a helpful addition. Indeed, proper training in recognizing the actual relationships between actions and outcomes in different non-gambling situations could possibly help patients learn to detect the lack of control in situations where there is no contingency between the events, at least in non-gambling conditions (see e.g., Wasserman et al., 1983; Matute, 1996; Msetfi et al., 2005, 2007; Hannah and Beneteau, 2009; Matute et al., 2011; Blanco et al., 2012).

As mentioned in the Introduction, many experiments have shown that a high outcome probability favors the development of the illusion of control. This outcome density effect was documented in the 1980's and 1990's (Alloy and Abramson, 1979; Allan and Jenkins, 1983; Matute, 1995) and is still a topic of high relevance in the experimental study of contingency learning (Buehner et al., 2003; Allan et al., 2008, 2005). For this reason, we used a high-outcome schedule. Control participants should develop the illusion and in this way a potentially stronger illusion could be observed in the gamblers group when confronting this standardized procedure. Thus, it is important to note that the present research does not speak to the issue of how the illusion of control operates during the low outcome conditions which are common during gambling. As many authors have already stated, other factors are critical in explaining the origin and maintenance of gambling behavior, such as, for instance, variable schedules of reinforcement (Ferster and Skinner, 1957) and the fact that gamblers often mention that their gambling behavior was reinforced during the early trials (Molde et al., 2009). Our research is silent with respect to those factors and to gambling behavior itself. What it shows is that gamblers are more vulnerable to the illusion of control than control participants in other areas of their life.

Our findings raise several questions about the role the illusion of control plays in pathological gambling (Myrseth et al., 2010). It is possible that people who are more vulnerable to the illusion of control have a greater risk of falling into gambling behavior, though it might also be that it is gambling behavior what increases vulnerability to the illusion of control. In this respect, the experimental assessment of the illusion of control that we propose can provide a richer and more complete assessment in different situations and could serve therefore as predictor or detector of the appearance of the pathology. Longitudinal studies could therefore be of use in future research to test this view.

On the limitation side of our experiment, the fact that the gamblers group was composed mostly of men could be problematic. Despite our leaving the experiment online for 6 months we were unable to obtain more female gamblers to participate in the study. This might reflect, on the one hand, their lower proportion in the general population (Desai et al., 2005; Blanco et al., 2006), and on the other one, their greater reluctance to publicly acknowledge and discuss their condition, perhaps because gambling has been

traditionally regarded as a male activity (Potenza et al., 2001). Thus, given the asymmetrical distribution of participants we were unable to analyze the effect of gender on the observed results. Nevertheless, previous research had shown that men tend to show weaker illusions of control than women (Alloy and Abramson, 1979; Wong, 1982; Vyse, 1997; Wolfradt, 1997; Dag, 1999), which suggests that, if anything, our gamblers group should have shown a weaker, rather than a stronger, illusion than the control group, had the results been confounded by gender. In any case, it will be necessary to achieve better control of this variable in future research.

Another potential problem is that we did not ask our participants to provide any demographic or personal information in addition to their age and gender. We always do it this way in order to increase participation and to comply with ethical standards on anonymity and privacy in our online experiments. However, because in the current experiment group assignment was not random, it might have occurred that the two groups differed by chance in some particular variable, such as, for instance, number of years of formal education, and this might have affected the development of the illusion of control. We believe this is unlikely, and reviews on related effects, such as superstitious beliefs, have concluded that there are no consistent results on the effects of variables such as years of education, or even general intelligence, on the development of these types of biased thinking (Wiseman and Watt, 2006). Nevertheless, it would also be desirable to obtain information on a larger number of demographic variables in future experiments.

To sum up, we believe that the use of the medical associative learning task may be an appropriate way to measure the illusion of control, not only in the general population but also in people with pathology such as gambling, which seems to be especially vulnerable to this type of illusion. One advantage of the associative learning task that we used is that it can easily detect the illusion of control in more general conditions and life areas, not necessarily related to pathology. Moreover, this procedure provides a less clinical perspective and is more focused on general associative-learning skills, so that the biases it detects are, at least in principle, domain-independent and common to most people, though some people are more vulnerable than others. This should allow researchers and therapists to use the large amount of already published evidence on contingency learning to test new and innovative strategies to reduce these biases in pathological gamblers (and other) populations under clinical treatment.

ACKNOWLEDGMENTS

Support for this research was provided by Grant 2011-26965 from Dirección General de Investigación of the Spanish Government, and Grant IT363-10 from Departamento de Educación, Universidades e Investigación of the Basque Government. We would like to thank the Spanish Federation of Rehabilitated Gamblers (FEJAR) for their valuable support on this research. Correspondence concerning this article should be addressed to Cristina Orgaz, Departamento de Psicología Básica I, UNED, Juan del Rosal 10, Madrid (scorgaz@psi.uned.es) or to Helena Matute, Departamento de Psicología, Universidad de Deusto, Apartado 1, 48080 Bilbao, Spain (matute@deusto.es).

REFERENCES

- Aeschleman, S., Rosen, C., and Williams, M. (2003). The effect of non-contingent negative and positive reinforcement operations on the acquisition of superstitious behaviors. *Behav. Processes* 61, 37–45. doi: 10.1016/S0376-6357(02)00158-4
- Allan, L., Hannah, S., Crump, M., and Siegel, S. (2008). The psychophysics of contingency assessment. *J. Exp. Psychol. Gen.* 137, 226–243. doi: 10.1037/0096-3445.137.2.226
- Allan, L., and Jenkins, H. M. (1983). The effect of representations of binary variables on judgment of influence. *Learn. Motiv.* 14, 381–405. doi: 10.1016/0023-9690(83)90024-3
- Allan, L., Siegel, S., and Tangen, J. (2005). A signal detection analysis of contingency data. *Learn. Behav.* 33, 250–263. doi: 10.3758/BF03196067
- Alloy, L. B., and Abramson, L. Y. (1979). Judgments of contingency in depressed and nondepressed students: sadder but wiser? *J. Exp. Psychol. Gen.* 108, 441–485. doi: 10.1037/0096-3445.108.4.441
- Alloy, L. B., and Clements, C. M. (1992). Illusion of control: vulnerability to negative affect and depressive symptoms after laboratory and natural stressors. *J. Abnorm. Psychol.* 101, 234–245. doi: 10.1037/0021-843X.101.2.234
- Bleak, J. L., and Frederick, C. M. (1998). Superstitious behavior in sport: levels of effectiveness and determinants of use in three collegiate sports. *J. Sport Behav.* 21, 1–15. doi: 10.1177/0956797610372631
- Blanco, C., Hasin, D. S., Petry, N., Stinson, F. S., and Grant, B. F. (2006). Sex differences in subclinical and DSM-IV pathological gambling: results from the national epidemiologic survey on alcohol and related conditions. *Psychol. Med.* 36, 943–953. doi: 10.1017/S0033291706007410
- Blanco, F., Matute, H., and Vadillo, M. A. (2009). Depressive realism: wiser or quieter? *Psychol. Rec.* 59, 551–562.
- Blanco, F., Matute, H., and Vadillo, M. A. (2011). Making the uncontrollable seem controllable: the role of action in the illusion of control. *Q. J. Exp. Psychol.* 64, 1290–1304. doi: 10.1080/17470218.2011.552727
- Blanco, F., Matute, H., and Vadillo, M. A. (2012). Mediating role of activity in the depressive realism effect. *PLoS ONE* 7:e46203. doi: 10.1371/journal.pone.0046203
- Blanco, F., Matute, H., and Vadillo, M. A. (2013). Interactive effects of the probability of the cue and the probability of the outcome on the overestimation of null contingency. *Learn. Behav.* doi: 10.3758/s13420-013-0108-8. [Epub ahead of print].
- Buehner, M., Cheng, P., and Clifford, D. (2003). From covariation to causation: a test of the assumption of causal power. *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 1119–1140. doi: 10.1037/0278-7393.29.6.1119
- Coventry, K., and Norman, A. (1998). Arousal, erroneous verbalizations and the illusion of control during a computer-generated gambling task. *Br. J. Psychol.* 89, 629–645. doi: 10.1111/j.2044-8295.1998.tb02707.x
- Dag, I. (1999). The relationships among paranormal beliefs, locus of control and psychopathology in a Turkish college sample. *Pers. Individ. Dif.* 26, 723–737. doi: 10.1016/S0191-8869(98)00184-6
- Desai, R. A., Maciejewski, P. K., Pantalon, M. V., and Potenza, M. N. (2005). Gender differences in adolescent gambling. *Ann. Clin. Psychiatry* 17, 249–258. doi: 10.1080/10401230500295636
- Dickerson, M. (1993). Internal and external determinants of persistent gambling: problems in generalizing from one form of gambling to another. *J. Gambl. Stud.* 9, 225–245. doi: 10.1007/BF01015920
- Durand, R. (2003). Predicting a firm's forecasting ability: the roles of organizational illusion of control and organizational attention. *Strateg. Manage. J.* 24, 821–838. doi: 10.1002/smj.339
- Echeburua, E., Báez, C., Fernández-Montalvo, J., and Páez, D. (1994). Cuestionario de Juego Patológico de South Oaks (SOGS): validación española [South Oaks Gambling Screen: Spanish validation]. *Análisis y Modificación de Conducta* 20, 769–791. doi: 10.1037/t03938-000
- Fenton-O'Creevy, M., Nicholson, N., Soane, E., and Willman, P. (2003). Trading on illusions: unrealistic perceptions of control and trading performance. *J. Occup. Organ. Psychol.* 76, 53–68. doi: 10.1348/096317903321208880
- Ferster, C. B., and Skinner, B. F. (1957). *Schedules of Reinforcement*. New York, NY: Appleton-Century-Crofts. doi: 10.1037/10627-000
- Fortune, E., and Goodie, A. S. (2011). Cognitive distortions as a component and treatment focus of pathological gambling: a review. *Psychol. Addict. Behav.* 26, 298–310. doi: 10.1037/a0026422
- Frankel, M. S., and Siang, S. (1999). "Ethical and legal aspects of human subjects research on the Internet," in *Report of a Workshop Convened by the American Association for the Advancement of Science, Program on Scientific Freedom, Responsibility, and Law* (Washington, DC). Retrieved December 26, 2012. Available online at: <http://www.aaas.org/spp/sfrl/projects/intres/rep.ort.pdf>
- Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., and Wilmer, J. B. (2012). Is the web as good as the lab? Comparable performance from web and lab in cognitive/perceptual experiments. *Psychon. Bull. Rev.* 19, 847–857. doi: 10.3758/s13423-012-0296-9
- Goldberg, J., Markóczy, L., and Zahn, G. (2005). Symmetry and the illusion of control as bases for cooperative behavior. *Ration. Soc.* 17, 243–270. doi: 10.1177/1043463105055431
- Goodie, A. (2005). The role of perceived control and overconfidence in pathological gambling. *J. Gambl. Stud.* 21, 481–502. doi: 10.1007/s10899-005-5559-1
- Gooding, P., and Tarrier, N. (2009). A systematic review and meta-analysis of cognitive-behavioral interventions to reduce problem gambling: hedging our bets? *Behav. Res. Ther.* 47, 592–607. doi: 10.1016/j.brat.2009.04.002
- Hannah, S. D., and Beneteau, J. L. (2009). Just tell me what to do: bringing back experimenter control in active contingency tasks with the command-performance procedure and finding cue density effects along the way. *Can. J. Exp. Psychol.* 63, 59–73. doi: 10.1037/a0013403
- Harris, P., and Middleton, W. (1994). The illusion of control and optimism about health: on being less at risk but no more in control than others. *Br. J. Soc. Psychol.* 33, 369–386. doi: 10.1111/j.2044-8309.1994.tb01035.x
- Hudgens-Haney, M. E., Hamm, J. P., Goodie, A. S., Krusemark, E. A., McDowell, J. E., and Clementz, B. A. (2013). Neural correlates of the impact of control on decision making in pathological gambling. *Biol. Psychol.* 92, 365–372. doi: 10.1016/j.biopsycho.2012.11.015
- Jenkins, H. M., and Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychol. Monogr.* 79, 1–17. doi: 10.1037/h0093874
- Källmén, H., Andersson, P., and Andren, A. (2008). Are irrational beliefs and depressive mood more common among problem gamblers than non-gamblers? A survey study of Swedish problem gamblers and controls. *J. Gambl. Stud.* 24, 441–450. doi: 10.1007/s10899-008-9101-0
- Kramer, T., and Block, L. (2011). Nonconscious effects of peculiar beliefs on consumer psychology and choice. *J. Consum. Psychol.* 21, 101–111. doi: 10.1016/j.jcps.2010.09.009
- Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., and Couper, M. (2004). Psychological research online: report of board of scientific affairs' advisory group on the conduct of research on the Internet. *Am. Psychol.* 59, 105–117. doi: 10.1037/0003-066X.59.2.105
- Labrador, F. J., Fernández-Alba, A., and y Mañoso, V. (2002). Relación entre la reducción de las distorsiones cognitivas referidas al azar y la consecución de éxito terapéutico en jugadores patológicos. *Psicothema* 14, 551–557. doi: 10.5218/prts.2013.0003
- Ladouceur, R., Mayrand, M., Dussault, R., Letarte, A., and Tremblay, J. (1984). Illusion of control: effects of participation and involvement. *J. Psychol.* 117, 47–52. doi: 10.1080/00223980.1984.9923657
- Langer, E. J. (1975). The illusion of control. *J. Pers. Soc. Psychol.* 32, 311–328. doi: 10.1037/0022-3514.32.2.311
- Laurendeau, J. (2006). "He didn't go in doing a skydive": sustaining the illusion of control in an edge-work activity. *Sociol. Perspect.* 49, 583–605. doi: 10.1080/0267257X.2011.627366
- Ledgerwood, D. M., and Petry, N. M. (2005). Current trends and future directions in the study of psychosocial treatments for pathological gambling. *Curr. Dir. Psychol. Sci.* 14, 89–94. doi: 10.1007/s11920-010-0141-7
- Leiseur, H. R., and Blume, S. B. (1987). The South Oaks Gambling Screen (SOGS): a new instrument for the identification of pathological gamblers. *Am. J. Psychiatry* 144, 1184–1188. doi: 10.1192/apt.116.450
- Lingyan, W., and Austin, J. (2008). Investigating illusion of control in experienced and non-experienced gamblers: replication and extension. *Anal. Gambl. Behav.* 2, 12–24.
- Lund, I. (2011). Irrational beliefs revisited: exploring the role of gambling preferences in the development

- of misconceptions in gamblers. *Addict. Res. Theory* 19, 40–46. doi: 10.3109/16066359.2010.493979
- MacKillop, J., Anderson, E. J., Castelda, B. A., Mattson, R. E., and Donovan, P. J. (2006). Convergent validity of measures of cognitive distortions, impulsivity, and time perspective with pathological gambling. *Psychol. Addict. Behav.* 20, 75–79. doi: 10.1037/0893-164X.20.1.75
- Mañoso, V., Labrador, F., and Fernández-Alba, A. (2004). Tipo de distorsiones cognitivas durante el juego en jugadores patológicos y no jugadores. *Psicothema* 16, 576–581.
- Matheson, K., Wohl, M. A., and Anisman, H. (2009). The interplay of appraisals, specific coping styles, and depressive symptoms among young male and female gamblers. *Soc. Psychol.* 40, 212–221. doi: 10.1027/1864-9335.40.4.212
- Matute, H. (1995). Human reactions to uncontrollable outcomes: further evidence for superstitions rather than helplessness. *Q. J. Exp. Psychol.* 48B, 142–157. doi: 10.1177/0146167202281009
- Matute, H. (1996). Illusion of control: detecting response-outcome independence in analytic but not in naturalistic conditions. *Psychol. Sci.* 7, 325–331. doi: 10.1111/j.1467-9280.1996.tb00376
- Matute, H., Vadillo, M. A., Vegas, S., and Blanco, F. (2007). Illusion of control in Internet users and college students. *CyberPsychol. Behav.* 10, 176–181. doi: 10.1089/cpb.2006.9971
- Matute, H., Yarritu, I., and Vadillo, M. A. (2011). Illusions of causality at the heart of pseudoscience. *Br. J. Psychol.* 102, 392–405. doi: 10.1348/000712610X532210
- Molde, H., Pallesen, S., Bartone, P., Hystad, S., and Johnsen, B. (2009). Prevalence and correlates of gambling among 16 to 19-year-old adolescents in Norway. *Scand. J. Psychol.* 50, 55–64. doi: 10.1111/j.1467-9450.2008.00667.x
- Morris, M. W., Sim, D., and Giroto, V. (1998). Distinguishing sources of cooperation in the one-round prisoner's dilemma: evidence for cooperative decisions based on the illusion of control. *J. Exp. Soc. Psychol.* 34, 494–512. doi: 10.1006/jesp.1998.1361
- Msetfi, R. M., Murphy, R. A., and Simpson, J. (2007). Depressive realism and the effect of intertrial interval on judgements of zero, positive, and negative contingencies. *Q. J. Exp. Psychol.* 60, 461–481. doi: 10.1080/17470210601002595
- Msetfi, R. M., Murphy, R. A., Simpson, J., and Kornbrot, D. E. (2005). Depressive realism and outcome density bias in contingency judgments: the effect of the context and intertrial interval. *J. Exp. Psychol. Gen.* 134, 10–22. doi: 10.1037/0096-3445.134.1.10
- Musca, S. C., Vadillo, M. A., Blanco, F., and Matute, H. (2010). The role of cue information in the outcome-density effect: evidence from neural network simulations and a causal learning experiment. *Connect. Sci.* 22, 177–192. doi: 10.3758/s13420-011-0015-9
- Myrseth, H., Brunborg, G., and Eidem, M. (2010). Differences in cognitive distortions between pathological and non-pathological gamblers with preferences for chance or skill games. *J. Gambl. Stud.* 26, 561–569. doi: 10.1007/s10899-010-9180-6
- Perales, J. C., Catena, A., Shanks, D. R., and González, J. A. (2005). Dissociation between judgments and outcome-expectancy measures in covariation learning: a signal detection theory approach. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 1105–1120. doi: 10.1037/0278-7393.31.5.1105
- Potenza, M. N., Steinberg, M. A., McLaughlin, S. D., Wu, R., Rounsaville, B. J., and O'Malley, S. S. (2001). Gender-related differences in the characteristics of problem gamblers using a gambling help-line. *Am. J. Psychiatry* 158, 1500–1505. doi: 10.1007/s10899-006-9013-9
- Pronin, E., Wegner, D. M., McCarthy, K., and Rodriguez, S. (2006). Everyday magical powers: the role of apparent mental causation in the overestimation of personal influence. *J. Pers. Soc. Psychol.* 91, 218–231. doi: 10.1037/0022-3514.91.2.218
- Reuven-Magril, O. D., and Reuven, N. (2008). Illusion of control and behavioral control attempts in obsessive-compulsive disorder. *J. Abnorm. Psychol.* 117, 334–341. doi: 10.1037/0021-843X.117.2.334
- Rudski, J. (2004). The illusion of control, superstitious belief, and optimism. *Curr. Psychol.* 22, 306–315. doi: 10.1007/s12144-004-1036-8
- Ryan, R. S., Wilde, M., and Crist, S. (2013). Compared to a small, supervised lab experiment, a large, unsupervised web-based experiment on a previously unknown effect has benefits that outweigh its potential costs. *Comput. Hum. Behav.* 29, 1295–1301. doi: 10.1016/j.chb.2013.01.024
- Sharpe, L. (2008). “Understanding pathological gambling: distinct pathways or individual formulations?” in *Psychology of Gambling*, ed M. J. Esposito (New York, NY: Nova Science Publishers Inc.), 147–162. doi: 10.1080/00223980903356081
- Strickland, C.-J. R., Taylor, A., Hendon, K. J., Provost, S., and Bizo, L. A. (2006). Erroneous beliefs among frequent fruit-machine gamblers. *Gambl. Res.* 18, 42–54.
- Taylor, S. E., and Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychol. Bull.* 103, 192–210. doi: 10.1037/0033-2909.103.2.193
- Vadillo, M. A., Musca, S. C., Blanco, F., and Matute, H. (2011). Contrasting cue-density effects in causal and prediction judgments. *Psychon. Bull. Rev.* 18, 110–115. doi: 10.3758/s13423-010-0032-2
- Vázquez, C. (1987). Judgment of contingency: cognitive biases in depressed and non-depressed subjects. *J. Pers. Soc. Psychol.* 52, 419–431. doi: 10.1037/0022-3514.52.2.419
- Vyse, S. A. (1997). *Believing in Magic: The Psychology of Superstition*. New York, NY: Oxford University Press.
- Wasserman, E. A., Chatlosh, D. L., and Neunaber, D. J. (1983). Perception of causal relations in humans: factors affecting judgments of response-outcome contingencies under free-operant procedures. *Learn. Motiv.* 14, 406–432. doi: 10.1016/0023-9690(83)90025-5
- Wasserman, E. A., Kao, S.-F., Van Hamme, L. J., Katagari, M., and Young, M. E. (1996). “Causation and association,” in *The Psychology of Learning and Motivation*, Vol. 34. *Causal learning*, eds D. R. Shanks, K. J. Holyoak, and D. L. Medin (San Diego, CA: Academic Press), 207–264.
- Wiseman, R., and Watt, C. (2006). Belief in psychic ability and the misattribution hypothesis: a qualitative review. *Br. J. Psychol.* 97, 323–338. doi: 10.1348/000712605X72523
- Wohl, M. J. A., and Enzle, M. E. (2009). Illusion of control by proxy: placing one's fate in the hands of another. *Br. J. Soc. Psychol.* 48, 183–200. doi: 10.1348/014466607X258696
- Wolfgang, A. K., Zenker, S. I., and Viscusi, T. (1984). Control motivation and the illusion of control in betting on dice. *J. Psychol.* 116, 67–72. doi: 10.1080/00223980.1984.9923618
- Wolfradt, U. (1997). Dissociative experiences, trait anxiety and paranormal beliefs. *Pers. Individ. Dif.* 23, 15–19. doi: 10.1016/S0191-8869(97)00043-3
- Wong, P. (1982). Sex differences in performance attribution and contingency judgement. *Sex Roles* 8, 381–388. doi: 10.1007/BF00287277

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 December 2012; accepted: 13 May 2013; published online: 17 June 2013.

Citation: Orgaz C, Estévez A and Matute H (2013) Pathological gamblers are more vulnerable to the illusion of control in a standard associative learning task. *Front. Psychol.* 4:306. doi: 10.3389/fpsyg.2013.00306

This article was submitted to *Frontiers in Personality Science and Individual Differences*, a specialty of *Frontiers in Psychology*.

Copyright © 2013 Orgaz, Estévez and Matute. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.



Do personality traits predict individual differences in excitatory and inhibitory learning?

Zhimin He, Helen J. Cassaday*, Charlotte Bonardi and Peter A. Bibby

School of Psychology, University of Nottingham, Nottingham, UK

Edited by:

Robin A. Murphy, University of Oxford, UK

Reviewed by:

Alexander Weiss, The University of Edinburgh, UK

Louis Matzel, Rutgers University, USA

*Correspondence:

Helen J. Cassaday, School of Psychology, University of Nottingham, University Park, Nottingham NG7 2RD, UK.

e-mail: helen.cassaday@nottingham.ac.uk

Conditioned inhibition (CI) is demonstrated in classical conditioning when a stimulus is used to signal the omission of an otherwise expected outcome. This basic learning ability is involved in a wide range of normal behavior – and thus its disruption could produce a correspondingly wide range of behavioral deficits. The present study employed a computer-based task to measure conditioned excitation and inhibition in the same discrimination procedure. CI by summation test was clearly demonstrated. Additionally summary measures of excitatory and inhibitory learning (difference scores) were calculated in order to explore how performance related to individual differences in a large sample of normal participants ($n = 176$ following exclusion of those not meeting the basic learning criterion). The individual difference measures selected derive from two biologically based personality theories, Gray's (1982) reinforcement sensitivity theory and Eysenck and Eysenck (1991) psychoticism, extraversion, and neuroticism theory. Following the behavioral tasks, participants completed the behavioral inhibition system/behavioral activation system (BIS/BAS) scales and the Eysenck personality questionnaire revised short scale (EPQ-RS). Analyses of the relationship between scores on each of the scales and summary measures of excitatory and inhibitory learning suggested that those with higher BAS (specifically the drive sub-scale) and higher EPQ-RS neuroticism showed reduced levels of excitatory conditioning. Inhibitory conditioning was similarly attenuated in those with higher EPQ-RS neuroticism, as well as in those with higher BIS scores. Thus the findings are consistent with higher levels of neuroticism being accompanied by generally impaired associative learning, both inhibitory and excitatory. There was also evidence for some dissociation in the effects of behavioral activation and behavioral inhibition on excitatory and inhibitory learning respectively.

Keywords: conditioned inhibition, behavioral activation, behavioral inhibition, neuroticism

INTRODUCTION

Conditioned inhibition (CI) is an associative learning phenomenon in which a stimulus (known as a conditioned inhibitor) is used to signal the omission of an otherwise expected outcome. For example, if a conditioned stimulus (CS) A signals a reinforcing unconditioned stimulus (US), and then after a number of training trials A is presented with another CS B, but now the expected US does not follow, participants learn that B indicates no US; in other words B is a conditioned inhibitor (Pavlov, 1927). Associative learning is a ubiquitous process of evolutionary advantage. It is not only fundamental, being found in all vertebrates, but has been argued to underlie many more sophisticated cognitive processes in both animals and humans. CI is therefore likely to be involved in a broad range of normal behavior – and thus its disruption could produce a wide range of behavioral deficits.

Lack of inhibitory control has been argued to lie at the heart of impulsivity (Buss and Plomin, 1975), which is a core feature of a number of psychological conditions, such as schizophrenia, and personality disorders (PDs), especially within forensic populations (Hare et al., 1991; Munro et al., 2007). Highly impulsive individuals

have difficulty withholding responding, as demonstrated by poor performance in laboratory-based behavioral tasks such as Go/No-Go (Visser et al., 1996; Logan et al., 1997; Enticott et al., 2006). However, these established tasks measure participants' ability to inhibit pre-potent motor responses, and are generally thought to involve the inhibition of stimulus-response associations. In contrast, relatively little research has explored the inhibition of stimulus-stimulus (CS-US) associations (formally CI) in populations likely to differ in impulsivity. To our knowledge, the only exception is evidence from our own work – we have reported individual variation in CI in relation to medication (Kantini et al., 2011a,b), level of dangerousness and severity of PDs (He et al., 2011), as well as in relation to symptom profile in schizophrenia (He et al., 2012).

However, such clinical samples are difficult to recruit in large numbers, and it is especially hard to isolate larger samples "uncontaminated" by confounded conditions – such as participants with Tourette syndrome in the absence of ADHD (Kantini et al., 2011a) or vice versa (Kantini et al., 2011b; see also He et al., 2011, 2012). Thus an alternative approach would be to examine the relationship between CI learning and individual differences in personality traits

in the general population (Migo et al., 2006). This previous study used the behavioral inhibition system/behavioral activation system (BIS/BAS) scale (Gray, 1981; Carver and White, 1994), as well as a measure of schizotypy, and CI was measured using an earlier task variant without full behavioral controls (as here). Probably the most widely used model of normal personality is the “Big Five” (Costa and McCrae, 1992) which includes extraversion and neuroticism, but not psychoticism which we wished to examine given our findings in clinical groups (He et al., 2011, 2012). The present study set out to examine CI in a large sample of normal participants using questionnaires designed to tap personality traits relating to comparative analyses of brain function, specifically in terms of differences in conditionability. Accordingly, participants were administered the Eysenck personality questionnaire revised short scale (EPQ-RS; Eysenck et al., 1985), as well as the BIS/BAS (Gray, 1981; Carver and White, 1994).

Eysenck's personality scales initially captured impulsivity in relation to extraversion and, in the revised version of the theory, as a core feature of its psychoticism dimension (Eysenck and Eysenck, 1991). Building on Eysenck's theory, the BIS/BAS scales were devised as orthogonal measures of anxiety and impulsivity respectively (Gray, 1981; Carver and White, 1994; Pickering and Gray, 1999). More specifically, Gray (1970, 1972, 1982, 1990) argued that the BAS measures activity in a system sensitive to signals of reward, which may, in predisposed individuals, elicit impulsive or antisocial tendencies. Consistent with this analysis, impulsivity has been related to enhanced learning about signals for reward (Avila et al., 2008), and neuroimaging evidence suggests that BAS activation is associated with the processing of positive stimuli in reward-related areas (albeit with some inconsistencies which may relate to the relative salience of the images in use for different individuals; Beaver et al., 2006; Avila et al., 2008). In contrast, the BIS relates to activity in a system responding to signals for non-reward, punishment, and novelty, producing inhibition of movement toward goals and other symptoms of anxiety. According to Gray's theory, BIS and BAS activity are independent, and dissociations in the relationship between anxiety and impulsivity and (for example) the processing of threat-relevant stimuli have in fact been demonstrated (Putman et al., 2004). Moreover, in anxiety disorders, aspects of impulsivity are negatively related to behavioral inhibition (Pierò, 2010; Snorrason et al., 2011); as would be expected, impulsivity has been suggested to result from deficient behavioral inhibition (Fowles, 1987). Thus there are both theoretical and empirical grounds to suggest that anxiety and impulsivity are inversely related.

Later refinement of the original behavioral inhibition theory (Gray and McNaughton, 2000) resulted in the introduction of sub-scales to the BIS (Carver and White, 1994), to capture the distinction between fear and anxiety (with BIS-anxiety and BIS-FFFS sub-scales; Gray and McNaughton, 2000; Smillie et al., 2006). Confirmatory factor analysis supports this revision to the theory and shows how the new model (with BIS-anxiety and BIS-FFFS sub-scales) relates to Eysenck's theory; for example, neuroticism relates to BIS-anxiety as well as the BIS-FFFS sub-scale, whereas psychoticism relates to BIS-anxiety and BAS (Heym et al., 2008).

Thus, although they do not measure it directly, impulsivity is nonetheless captured by these general theories of personality. The broader predispositions measured by the EPQ-RS and the BIS/BAS also relate to disorder, in that EPQ-RS neuroticism and BIS scores specifically measure susceptibility to anxiety-related conditions (Eysenck, 1957, 1967; Eysenck and Eysenck, 1976a,b). More generally, disinhibition as a mechanism for impulsivity could potentially apply to a variety of behavioral disorders to which anxiety is less central, including antisocial behavior, and psychopathy (He et al., 2011). Although psychopathy is a clinical condition rather than a personality trait, it is nonetheless related to the personality trait of psychoticism (Eysenck and Eysenck, 1976b). In relation to underlying neuropsychological substrates, both have been argued to result from dysfunction in the BIS (Gray, 1972, 1982).

This relationship has been further specified in terms of the BIS-FFFS, which mediates avoidance or escape in response to fear (Gray and McNaughton, 2000; Smillie et al., 2006). Low and high BIS-FFFS activity have been suggested to characterize primary and secondary psychopathy respectively, while secondary psychopaths are said also to be characterized by high BAS activity (Corr, 2010). Relatedly, statistical analyses of scores from a normal population have recently confirmed that high psychoticism scores are associated with reduced fear and anxiety (also characteristic of primary psychopathy) and increased impulsivity (more characteristic of secondary psychopathy), and this psychoticism-impulsivity link is stronger in individuals with elevated BIS-FFFS scores (Heym and Lawrence, 2010). In the present study, the use of EPQ-RS enabled us to test whether psychoticism is negatively related to CI learning, as might be expected based on the fact that, using the same task variant, CI was found to be abolished in offenders with PDs (He et al., 2011).

Further predictions follow from Eysenck's (1957, 1967) theory: for example, it suggests that the tendency for introverts to condition more readily than extraverts should be exacerbated by high neuroticism. This theory has been modified to take the nature of the US into account (Gray, 1970, 1972). For positive stimuli (as used in the present study), Eysenck's theory predicts that conditioning will be better in those with higher levels of introversion, whereas Gray's (1970) theory predicts that conditioning will be better in those with higher levels of extraversion. These predictions have been tested many times, but not in relation to CI.

In a previous study using a different inhibitory learning procedure, participants with higher BAS scores (specifically reward responsiveness, but not the other sub-scales) unexpectedly showed more rather than less CI (Migo et al., 2006). From a theoretical perspective, this is surprising in that higher BAS activity is predicted to increase conditioning to reward-related stimuli, and higher BIS activity conditioning to signals of non-reward (Corr et al., 1995; Pickering, 1997) – such as the absence of the expected rewarding outcome learned about in the CI task. Therefore we would predict that CI should have increased with BIS scores in this task – yet no such relationship was found (Migo et al., 2006). The present study used a larger sample to further explore the direction of the relationship between CI and those aspects of impulsivity measured by the BAS scales, and to reevaluate the prediction that increased BIS scores should be associated with higher levels of CI.

MATERIALS AND METHODS

DESIGN

The overall design of the experiment was identical to that used in previous studies (He et al., 2011, 2012), and employed Lego blocks as neutral CSs and positive and neutral International Affective Picture System (IAPS) pictures as reinforcement and non-reinforcement respectively. There were three stages: (1) pre-test; (2) training with elemental and compound stimuli; and (3) the test stage (Table 2). In the pre-test stage, participants were required to rate the stimuli and stimulus compounds to be used in the training and test stages, to establish whether differences in responding to the stimuli at test could be due to biases present before the start of training.

In the elemental training stage two CSs, A and C, were paired with reinforcement (A+ and C+ trials), while a further two, U and V, were paired with non-reinforcement. This training provided a measure of participants' simple associative learning. It also established A and C as excitatory CSs signaling a positive outcome, which facilitated the subsequent establishment and detection of CI. An *a priori* exclusion criterion was applied based on elemental training performance: participants who failed to learn the simple discrimination between C+ and V− trials [i.e., rating scores (C−V) = <0¹] were excluded from all subsequent analyses (with the exception of the correlational analyses performed to examine the relationships between the level of excitatory or of inhibitory learning and the age of the participants).

During the compound training stage, the AZ compound signaled reinforcement (AZ+), whereas AP signaled non-reinforcement (AP−). As A had been paired with reinforcement in the previous stage, presenting AP allowed P to signal the absence of the reinforcement otherwise indicated by A, and was thus expected to establish P as a conditioned inhibitor. Two additional stimulus compounds, CY and BX, were reinforced and non-reinforced respectively.

Although successful discrimination between AZ and AP would be consistent with the proposal that P was a conditioned inhibitor, it is not sufficient. For example, participants might respond more to AZ simply because Z was reinforced on every trial. In order to establish unequivocally that P was a conditioned inhibitor we conducted a summation test – more specifically, we examined whether P would suppress responding to a different excitatory stimulus more than would a suitable control stimulus (cf. Rescorla, 1969). The continued excitatory training with C on CY+ trials (C had also been reinforced in the previous stage) means it provided an excitatory test stimulus against which the inhibitory effects of P could be evaluated. The BX− trials were designed to establish X as a control stimulus which was presented the same number of times as P, and in a similar manner (in compound with another stimulus, and paired with non-reinforcement). However, the stimulus with which X was presented was novel so that X, unlike P, did not signal the absence of reinforcement during this training stage. Therefore X should not have acquired any inhibitory properties.

The test stage, like the pre-test, compared ratings of the stimuli and stimulus compounds that had signaled reinforcement (A, C, AZ, CY) and non-reinforcement (AP, BX), and also the test compounds (CP, CX). The critical comparison was between the test compounds CP and CX. Stimulus C was excitatory, and was predicted to elicit high ratings indicating expectation of reinforcement. If P was a conditioned inhibitor it should reduce this high rating to C, whereas the critical comparison stimulus, X, should not. CI would therefore be evident as lower ratings to CP than to CX. The identities of the stimuli used as P and X were counterbalanced across the participants, as were those of A and B (and C and V, see above).

PARTICIPANTS

A total of 194 healthy participants took part in the computer-based learning task, all of whom completed the EPQ-RS and BIS/BAS questionnaires. The participants were recruited from the University of Nottingham (UK campus) and the local community. The participants included 98 males and 96 females, and the mean age of participants was 24.85, range 18–56. Eighteen out of 194 participants failed the excitatory associative learning task during the elemental training stage [i.e., rating scores (C−V) = <0 – see below], which was used as an exclusion criterion. The study was approved by the University of Nottingham, School of Psychology Ethics Committee. Participants received an inconvenience allowance of £3 cash to cover their travel expenses.

STIMULI

Lego block pictures ($n = 9$) were used as the CSs (Figure 1). The USs were selected by a pilot study from the IAPS (Lang et al., 2005). The IAPS provides a set of images, standardized on the basis of participants' ratings, on the dimensions of valence and arousal from 1 to 9, 1 representing a low rating on each dimension and 9 a high rating (i.e., 1 as low pleasure, low arousal). The USs in the present study included 10 positive pictures and 10 neutral pictures, excluding erotic pictures (see Table 1 for mean valence and arousal ratings of the images in use). Conditioning was measured using a rating scale: participants were asked to guess or predict what kind of picture would follow presentation of the Lego blocks using a rating scale from 1 (neutral) to 9 (positive), with the rating 5 to reflect uncertainty as to what kind of image was expected to follow.

QUESTIONNAIRES

The following were administered to the participants after the CI learning computer task.

Eysenck personality questionnaire revised short scale

The EPQ-RS is a 48 item yes/no questionnaire, suitable for the age range 16–70 years (Eysenck et al., 1985). It is used to assess dimensions of personality in relation to four factors: extraversion (E), psychoticism (P), neuroticism (N), and the response distortion (Lie) scale. There are 12 items for each factor.

Behavioral inhibition system/behavioral activation system scale

This consists of a list of 20 items for which participants use a four-point response scale to express whether the statement is true or

¹ Only C and V were used for this purpose as the identities of Lego blocks serving as C and V were fully counterbalanced, whereas those of A and U were not.

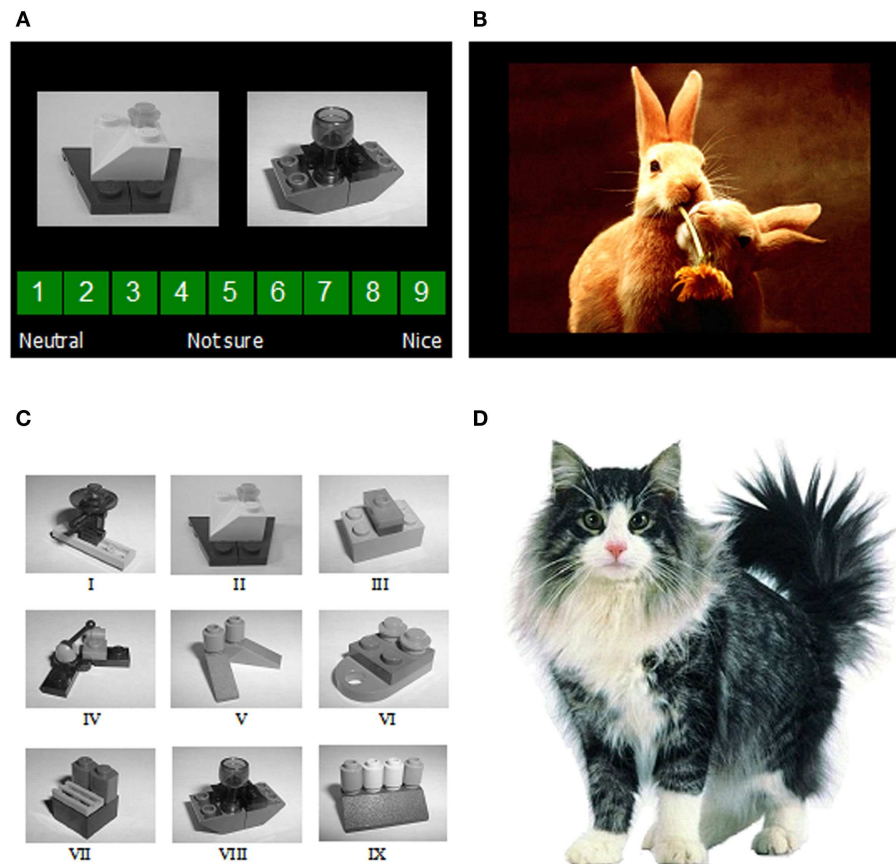


FIGURE 1 | (A) Two examples of the image presentations used as conditioned stimuli, shown together with the rating scale used to guess or predict what valence of unconditioned stimulus (a positive or neutral IAPS image) Mogwai would bring; (B) an example of one of

the image presentations used as the unconditioned stimuli; (C) the nine images of Lego blocks used as conditioned stimuli; (D) Mogwai the cat as presented prior to the unconditioned stimuli in the training stages.

Table 1 | The valence and arousal ratings of the IAPS images used.

Images	Mean valence (SD; range)	Mean arousal (SD; range)
10 Neutral	4.94 (0.08; 4.86–5.08)	2.79 (0.54; 1.72–3.46)
10 Positive	7.80 (0.27; 7.49–8.28)	4.93 (1.07; 3.08–6.73)

false for them (Carver and White, 1994). The questionnaire divides in five sub-scales: BIS-anxiety, BIS-FFFS, BAS-drive, BAS-fun seeking, and BAS-reward responsiveness.

PROCEDURE

This was the same as that used previously (He et al., 2011, 2012) with some minor variations (reported in full below). Participants were invited to take part in a research study on learning using a computer-based task. Before the task, each participant had to read the information sheet and sign a consent form. The task instructions were that a cat “Mogwai” would bring participants either a positive picture or a neutral, boring picture, depending on what kind of Lego blocks she found in her basket (Figure 1). Participants were asked to guess or predict what kind of picture would follow

presentation of the Lego blocks using the rating scale described above. Reminder instructions were presented on-screen at each stage of the procedure.

Before the start of the pre-test phase, participants were shown some example CSs and USs and further explanation was given as necessary. The samples of CS and US images were individually color printed on a 4.5 cm × 6 cm card and these pictures were representative of, but not subsequently used as, stimuli during the experiment. Participants were told that the whole computer-based experimental session would last about 20 min and comprise three stages. At the same time, they were shown an example of CS presentations with the rating scale, and were told that during the experiment they would need to click the corresponding number to guess or predict the valence of the US (a positive or a neutral picture) according to the different Lego blocks that had been presented. Participants were encouraged to ask questions at this stage. The three stages of the computer-based experimental session then followed.

Pre-test stage

During the first (pre-test) stage of the experiment, participants were told they must guess what kind of picture the cat might

bring based on the Lego blocks presented, although the instructions specified that no pictures would follow. A Lego block CS was presented with the rating scale, until the participants clicked on a number button to guess the US valence; this triggered the next CS presentation, which followed immediately. In this and all subsequent stages of the experiment CS presentations were counterbalanced for right/left position on the screen across participants, and the various trial types were presented in a semi-random sequence (i.e., constrained only by the total number of trials of a particular type scheduled in each stage). In this stage there was a total of 16 presentations, two of each stimulus or stimulus combination presented (these being A, C, AZ, AP, BX, CY, CP, and CX; see **Table 2**).

Training stages

On completion of the pre-test, the conditioning trials commenced and US presentations were introduced. The instructions were as before, but with the exception that participants were advised that following their guess they would be shown the picture that the cat had brought. The first training stage used the CS elements, and comprised six training blocks, each with two of each of the four kinds of trial (A+, U–, V–, and C+). As in the pre-test, the Lego block was presented until the participant clicked a number button to predict the valence of the US to follow, at which point a US, randomly selected from the pool of positive or neutral USs as appropriate, was shown on the screen for 1 s. This was followed by a 1 s gap, during which a picture of the cat Mogwai (around 6 cm × 6 cm) was presented in the middle of the screen on a white background. This sequence of events comprised a trial. The second, compound training stage followed directly after this training with the CS elements, and comprised four kinds of trial (AZ+, AP–, BX–, and CY+). There was a total of eight excitatory trials of each type in this stage; the number of inhibitory trials depended on the task variant (see below). The different trial types were analyzed in four equivalent blocks of trials.

Test stage

The test stage was exactly the same as the pre-test stage, except that there were four rather than two presentations of each of the critical test compounds CP and CX. As in the earlier stages of the experiment, there were on-screen reminders of the task instructions. Throughout the experiment, whenever participants asked questions or made comments they were asked to try to focus on the task and to try to remember or guess which outcome (positive or neutral picture) was predicted by the Lego blocks.

PROCEDURAL VARIANTS

There were three variants on the experimental procedure used to test CI in the present study. In the first ($n = 43$) the pictures of the CSs were colored and the number of presentations of the non-reinforced compounds was eight (rather than 12 as shown in **Table 2**). The second refinement was identical to the first ($n = 19$), except that the colored CS images were changed to black and white pictures. The final variant ($n = 132$) differed only in that the number of non-reinforced compound presentations was increased from 8 to 12 (as in **Table 2**). This final version was that used in our previously published reports (He et al., 2011, 2012). These three procedural variants did not result in equivalent levels of CI, the third being the most effective. However, variation in the level of CI does not preclude investigation of its relationship to individual differences variables and – as would be expected – CI was clearly demonstrated over the sample as a whole.

ANALYSIS

The dependent variable was the mean rating given for each particular trial type, which was assessed in each training block of each stage. Statistical analyses of overall learning were by analysis of variance (ANOVA), with discrimination (e.g., A+ vs. U– and C+ vs. V–), reinforcement (reinforced or not), and trial block as within-subjects factors. Additionally, a summary measure of excitatory learning was provided by the difference in mean ratings on C

Table 2 | The design of the experiment used in the third variant of the task.

Pre-test		Elemental training		Compound training		Test	
CSs	No. of trials	CSs	No. of trials ¹	CSs ± outcome	No. of trials	CSs	No. of trials
PHASE							
A	2	A+	12	AZ+	8	A	2
C	2	U–	12	AP–	12	C	2
AZ	2	V–	12	BX–	12	AZ	2
AP	2	C+	12	CY+	8	AP	2
BX	2					BX	2
CY	2					CY	2
CP	2					CP	4
CX	2					CX	4

In the pre-test all participants gave baseline ratings of the various stimuli. Letters denote the nine CSs (pictures of Lego blocks) which were counterbalanced (see text). “+” Denotes reinforcement (a positive IAPS picture) and “–” non-reinforcement (a neutral IAPS picture). ¹Sixty two participants were tested with 8 rather than 12 elemental training trials. Compound training established P as a signal for the absence of reinforcement, rendering it inhibitory. In addition CY was reinforced, and BX non-reinforced. Thus C served as an excitatory cue against which the effect of the inhibitory P could be examined, while X served as a control for P. At test CP and CX were presented: to the extent that P was inhibitory, it would successfully counteract the tendency of C to predict reinforcement, relative to X.

and V trials during the initial training stage, i.e., C–V. As C was the excitatory stimulus, the greater the C–V score, the higher the level of excitatory learning. A summary measure of CI was provided by the difference between the mean ratings on CX and CP trials given during the test stage, i.e., CX–CP. P was the putative inhibitor, and thus supposed to suppress evaluation of C more than X; thus the higher the CX–CP score, the greater the inhibitory learning. Significant two-way interactions were explored with simple main effects analysis. Comparison of the summary learning scores in males vs. females was by *t*-test.

Correlational analyses were used to compare overall learning and questionnaire scores for EPQ and BIS/BAS sub-scales. Bonferroni adjustments can be employed to reduce the possibility of Type I errors when examining multiple correlation coefficients (Larzelere and Mulaik, 1977; Holm, 1979; Rice, 1989). However, particularly for statistically small effects, the likelihood of Type II error is increased (Perneger, 1998; Jennions and Möller, 2003; Nakagawa, 2004). Thus, unless otherwise stated, the correlations reported in this paper are corrected using Benjamini and Hochberg's (1995) procedure, rather than Bonferroni which has less statistical power (so the uncorrected *p* values are reported in Table 3).

RESULTS

CONDITIONED INHIBITION CONFIRMED BY SUMMATION TEST

Pre-test stage

There was little difference on the rating scores of the stimuli prior to conditioning (all being around five). Importantly, there was no significant difference in responding to the two critical test compounds (CP vs. CX), $F < 1$.

Pre-training stage and training stage

During the pre-training stage, the ratings of A and C steadily increased, while those to the U and V stimuli fell gradually, suggesting that the participants learned both discriminations in this phase (see Figure 2). This impression was supported by statistical analysis. ANOVA with discrimination (A/U vs. C/V), reinforcement and pre-training block (1–6) as factors revealed a significant three-way interaction, $F(5, 875) = 2.70$, $p = 0.02$, $\eta_p^2 = 0.015$. The main effects of block and reinforcement were significant, $F(5, 875) = 4.80$, $p < 0.001$, $\eta_p^2 = 0.027$, and $F(5, 175) = 465.68$, $p < 0.001$, $\eta_p^2 = 0.727$, respectively. Moreover, these two factors interacted significantly, $F(5, 875) = 119.07$, $p < 0.001$, $\eta_p^2 = 0.405$. The effect of discrimination was not significant, $F < 1$, nor the interaction between block and discrimination, $F(5, 875) = 1.77$, $p = 0.12$, $\eta_p^2 = 0.01$. The interaction between discrimination and reinforcement was not significant, $F(1, 175) = 1.57$, $p = 0.211$, $\eta_p^2 = 0.009$.

To explore the three-way interaction further ANOVAs were performed separately on the two discriminations. These revealed a significant interaction between reinforcement and discrimination for both the A/U and C/V discriminations, $F(5, 875) = 355.05$, $p < 0.001$, $\eta_p^2 = 0.239$, and $F(5, 875) = 83.51$, $p < 0.001$, $\eta_p^2 = 0.323$, respectively. Simple main effects analysis revealed that the effect of reinforcement was highly significant on all training blocks in both discriminations, smallest $F(1, 175) = 12.36$, $p = 0.001$, $\eta_p^2 = 0.066$, for block 1 of the C/V discrimination.

Table 3 | Correlations between the EPQ-RS and BIS/BAS variables, the demographic variables, and the measures of excitatory and inhibitory associative learning.

	C-V	CX-CP	Age	Sex	L	P	E	N	FFFS	BIS	BAS-D	BAS-FS	BAS-RR
C-V		-0.01	0.22**	-0.10	0.03	0.07	0.07	-0.17*	0.01	-0.07	-0.21**	-0.11	-0.08
CX-CP	-0.14, 0.12		0.11	-0.18*	0.17	0.00	-0.03	-0.19*	-0.17*	-0.19*	0.08	0.03	-0.13
Age	0.08, 0.35	-0.02, 0.24		-0.19*	0.13	-0.03	-0.16	-0.36**	-0.10	-0.23**	-0.22**	-0.29**	-0.37**
Sex	-0.23, 0.03	-0.32, -0.04	-0.32, -0.04		0.18*	-0.16	0.19*	0.28**	0.32**	0.41**	0.05	0.00	0.10
L	-0.10, 0.16	-0.01, 0.24	-0.01, 0.26	0.04, 0.31		-0.08	0.04	-0.22**	-0.14	-0.11	-0.17	-0.16	-0.18*
P	-0.07, 0.19	-0.13, 0.13	-0.16, 0.10	-0.28, -0.03	-0.20, 0.05		0.04	-0.02	-0.12	-0.22**	0.08	0.27**	0.09
E	-0.07, 0.20	-0.16, 0.10	-0.28, -0.02	0.06, 0.32	-0.09, 0.16	-0.08, 0.16		-0.31**	-0.22**	-0.07	0.30**	0.43**	0.22**
N	-0.31, -0.03	-0.32, -0.05	-0.45, -0.19	0.14, 0.40	-0.34, -0.08	-0.14, 0.10	-0.43, -0.18		0.47**	0.54**	-0.04	-0.04	0.13
FFFS	-0.13, 0.14	-0.31, -0.03	-0.23, 0.04	0.19, 0.44	-0.26, -0.01	-0.24, 0.00	-0.35, -0.09	0.36, 0.57		0.53**	-0.18*	-0.12	0.08
BIS	-0.20, 0.06	-0.32, -0.05	-0.36, -0.09	0.29, 0.52	-0.23, 0.01	-0.34, -0.09	-0.19, 0.06	0.44, 0.63	0.42, 0.62		0.00	0.00	0.24**
BAS-D	-0.35, -0.07	-0.05, 0.20	-0.35, -0.07	-0.07, 0.18	-0.29, -0.04	-0.04, 0.20	0.16, 0.42	-0.16, 0.09	-0.31, -0.04	-0.14, 0.13		0.51**	0.53**
BAS-FS	-0.24, 0.02	-0.10, 0.16	-0.42, -0.15	-0.12, 0.13	-0.28, -0.04	0.14, 0.39	0.30, 0.53	-0.16, 0.08	-0.24, 0.01	-0.09, 0.16	0.40, 0.61		0.49**
BAS-RR	-0.21, 0.05	-0.25, 0.00	-0.49, -0.23	-0.03, 0.22	-0.30, -0.04	-0.04, 0.21	0.09, 0.32	0.00, 0.25	-0.05, 0.21	0.10, 0.37	0.42, 0.63	0.37, 0.59	

N = 176. * $p < 0.05$, ** $p < 0.01$ (uncorrected). Above the diagonal the correlation coefficient, *r*. Below the diagonal, the lower and upper limits of the 95% confidence interval of the correlation coefficient. C-V, excitatory associative learning; CX-CP, inhibitory associative learning; L, Lie (response distortion); P, psychoticism; E, extraversion; N, neuroticism; FFFS, BIS fight, flight, freeze sub-scale; BIS, behavioral inhibition system anxiety sub-scale; BAS-D, BAS-drive sub-scale; BAS-FS, BAS-fun seeking; BAS-RR, BAS-reward responsiveness.

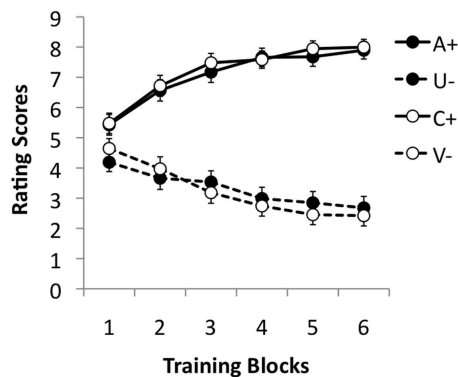


FIGURE 2 | Mean rating scores for A+, U-, V-, and C+ during the six training blocks of the pre-training stage. A rating of 9 reflects expectation of a positive image, 1 of a neutral image, and 5 uncertainty; 95% confidence intervals are presented.

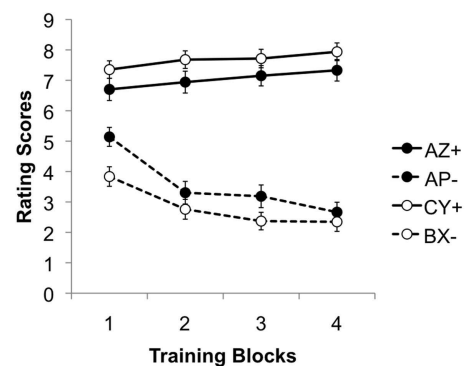


FIGURE 3 | Mean rating scores for AZ+, AP-, BX-, and CY+ during the four blocks of the training stage. A rating of 9 reflects expectation of a positive image, 1 of a neutral image, and 5 uncertainty; 95% confidence intervals are presented.

The main effect of block was also significant for both reinforced and non-reinforced trials in both discriminations, smallest $F(5, 875) = 16.07$, $p < 0.001$, $\eta_p^2 = 0.084$, for U trials.

During the training stage, the ratings of AZ and CY steadily increased, while those of AP and BX fell gradually (see **Figure 3**), again suggesting that both discriminations were learned successfully. This impression was again confirmed by statistical analysis. An ANOVA with discrimination (AZ/AP vs. CY/BX), reinforcement and training block (1–4) as factors, revealed a significant three-way interaction, $F(3, 525) = 74.54$, $p < 0.001$, $\eta_p^2 = 0.299$. The main effects of block and reinforcement were significant, $F(3, 525) = 29.80$, $p < 0.001$, $\eta_p^2 = 0.146$, and $F(1, 175) = 45.58$, $p < 0.001$, $\eta_p^2 = 0.214$, respectively. Moreover, these two factors interacted significantly, $F(3, 525) = 3.15$, $p = 0.025$, $\eta_p^2 = 0.018$. The effect of discrimination was not significant, $F < 1$, but the interactions between discrimination and both block and reinforcement were significant, $F(3, 525) = 3.53$, $p = 0.015$, $\eta_p^2 = 0.02$, and $F(1, 175) = 480.34$, $p < 0.001$, $\eta_p^2 = 0.733$ respectively.

Further ANOVAs were conducted to explore the three-way interaction further. These confirmed a significant interaction between block and reinforcement for both discriminations, smallest $F(3, 525) = 33.95$, $p < 0.001$, $\eta_p^2 = 0.162$, for the CY/BX discrimination. Simple main effects analysis revealed that the effect of reinforcement was significant for both discriminations on every block, smallest $F(1, 175) = 39.57$, $p < 0.001$, $\eta_p^2 = 0.184$, for the first block of the AZ/AP discrimination. In addition the effect of blocks was significant for both reinforced and non-reinforced trials in both discriminations, smallest $F(3, 525) = 3.96$, $p = 0.008$, $\eta_p^2 = 0.022$ for AP trials.

Test stage

Figure 4 shows the rating scores during the test stage. Here the critical comparison was between ratings of CP and CX during the pre-test and the test stages. It can be seen from **Figure 4** that the rating of CP was noticeably lower than CX during the test. This difference was confirmed by statistical analysis: an ANOVA with stage (pre-test and test), and stimulus (CP vs. CX) as

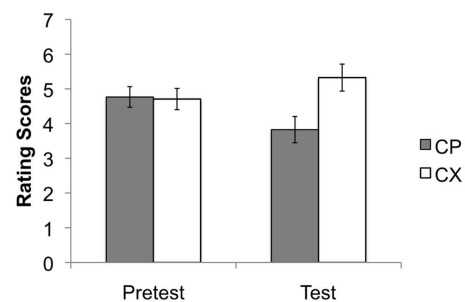


FIGURE 4 | Mean rating scores for the key comparison stimulus compounds CP and CX during the pre-test and test stages. A rating of 9 reflects expectation of a positive image, 1 of a neutral image, and 5 uncertainty; 95% confidence intervals are presented. The stimulus compounds elicited similar ratings prior to conditioning, but the test ratings confirmed the presence of conditioned inhibition, evident as lower ratings to CP than to CX.

factors revealed no effect of stage, $F < 1$, but a significant effect of stimulus, $F(1, 175) = 22.95$, $p < 0.001$, $\eta_p^2 = 0.116$. There was also a significant interaction between these two factors, $F(1, 175) = 22.65$, $p < 0.001$, $\eta_p^2 = 0.115$. Simple main effects confirmed that participants gave significantly lower rating scores to CP than to CX during the test stage, $F(1, 175) = 49.79$, $p < 0.001$, $\eta_p^2 = 0.183$ but not at the pre-test stage, $F < 1$. The results confirm the overall conclusion that P had become a conditioned inhibitor.

DEMOGRAPHIC CHARACTERISTICS AND LEARNING DIFFERENCES

In general, males performed better than females, as reflected in the summary measures of both excitatory, $t(192) = 2.08$, $p = 0.04$, and inhibitory learning, $t(174) = 2.44$, $p = 0.02$. There was also a significant correlation between the age of the participants and the summary measure of excitatory learning (C–V), $r(194) = 0.18$, $p = 0.01$. However, there was no correlation between age and the summary measure of inhibitory learning, $r(174) = 0.11$, $p = 0.14$.

THE RELATIONSHIP BETWEEN EXCITATORY AND INHIBITORY LEARNING

The correlation between the rating scores for (C–V) and (CX–CP) was examined directly. The results showed that there was no significant correlation between the two ratings, $r(194) = 0.12$, $p = 0.09$, suggesting that – despite their inevitable interdependence – individual differences in inhibitory learning are not entirely dependent on differences in excitatory learning.

INDIVIDUAL DIFFERENCES IN EXCITATORY LEARNING

Eysenck personality questionnaire revised short scale

There was a significant negative correlation between the EPQ-RS neuroticism scores and the summary measure of excitatory learning (C–V), $r = -0.17$, $p = 0.021$ (see **Table 3**). However, the correlations between excitatory learning and psychoticism and extraversion were not significant.

Behavioral inhibition system/behavioral activation system scale

There was a significant negative correlation between the BAS-drive scores and the summary measure of excitatory learning (C–V), $r = -0.21$, $p = 0.004$. However, there were no further significant correlations between the other sub-scales of the BIS/BAS and excitatory learning (C–V, see **Table 3**).

INDIVIDUAL DIFFERENCES IN INHIBITORY LEARNING

Eysenck personality questionnaire revised short scale

There was a significant negative correlation between the EPQ-RS neuroticism scores and the summary measure of inhibitory learning (CX–CP), $r = -0.19$, $p = 0.013$. However, there were no significant correlations between the other sub-scales of the EPQ-RS and CX–CP (see **Table 3**).

Behavioral inhibition system/behavioral activation system scale

There were significant negative correlations between the BIS-anxiety scores ($r = -0.19$, $p = 0.013$) and BIS-FFFS ($r = -0.17$, $p = 0.021$) scores and the summary measure of inhibitory learning (CX–CP). However, there were no significant correlations for the BAS sub-scales and CX–CP (see **Table 3**).

DEMOGRAPHIC AND INDIVIDUAL DIFFERENCES VARIABLES JOINT EFFECTS ON EXCITATORY AND INHIBITORY LEARNING

To take into account the observation that both age and sex are related to the individual difference variables as well as the learning measures two multiple linear regressions were conducted using the inhibitory and excitatory learning measures as the criterion variables. The predictor variables were the demographic variables and the individual difference variables associated with the EPQ-RS and BIS/BAS measures.

Taken together the multiple-R for the measure of excitatory learning was 0.37 ($R^2 = 0.13$) which was significant ($p = 0.007$). However, only BAS-drive had a statistically significant unique relationship with the excitatory learning measure ($\beta = -0.24$, $r_p^2 = 0.04$, $p = 0.01$), accounting for less than one third of the variability that the overall equation accounts for. The reason for neuroticism not showing a unique relationship is likely to be because of its relatively high correlations with both BIS-revised and FFFS as well as age and sex of the participants (see **Table 3**).

For the measure of inhibitory learning the multiple-R was 0.31 ($R^2 = 0.10$). This was not statistically significant ($p = 0.07$). Similarly, none of the demographic, EPQ-RS or BIS/BAS variables was individually statistically significant. This suggests that while the zero order correlations demonstrate relationships between some of the demographic and individual difference variables and the inhibitory learning measures the covariance of subsets of the predictor variables is sufficiently high to be partialled out as part of the linear regression procedure, leading to an underestimation of the relationship between individual predictor variables and the criterion variable.

DISCUSSION

As might be expected, using an established procedure (He et al., 2011, 2012) CI was robustly demonstrated in this large sample of participants in a summation test. What the present study adds to this prior work is clarification of how individual variations in inhibitory and excitatory learning relate to established individual difference measures. Specifically we examined participants' neuroticism, extraversion, and psychoticism, as well as behavioral inhibition and behavioral activation, as proposed by the personality theories of Eysenck (1957, 1967, 1981), Eysenck et al. (1985), Gray (1972, 1982), and Gray and McNaughton (2000). These biologically based personality theories should most closely relate to associative learning theories derived from the study of animal behavior.

We found that those with higher EPQ-RS neuroticism showed reduced levels of both excitatory and inhibitory conditioning (as reflected in the C–V and CX–CP scores respectively). Reduced excitatory learning was also found in those with higher BAS-drive, but here there was a dissociation, in that inhibitory learning was not affected by this measure but was instead negatively related to both BIS-FFFS and BIS-anxiety.

Thus, as might be expected given the dependence between excitatory and inhibitory learning, both were attenuated in those with higher neuroticism. Similarly, as might be expected given the relationship between neuroticism and BIS, inhibitory learning was also related to the BIS scores. The correlations found here between the EPQ-RS and the BIS/BAS sub-scales largely replicate those earlier reported (**Table 3**; Heym et al., 2008). Thus the findings are consistent with higher levels of neuroticism being accompanied by generally impaired associative learning. There was also evidence for some dissociation in the effects of behavioral activation and behavioral inhibition on excitatory and inhibitory learning respectively.

However, contrary to what might seem to follow from the original version of Gray's (1972, 1982) theory, we found that higher scores on the BIS scale were correlated with *impaired* rather than facilitated inhibitory learning. Clinical observations are consistent with elevated behavioral inhibition in anxiety disorders (Barlow, 2000), and according to Gray (1972, 1982) the BIS is activated by signals of punishment, signals of non-reward, and innate fear stimuli. It should be noted that Gray's behavioral inhibition theory is not a theory of Pavlovian CI as such. However, there is overlap in the sense that signals of non-reward should excite the BIS (whereas signals of non-punishment excite the behavioral activation system and result in an emotional state more akin to relief). Since the present task was appetitively motivated (using positive

IAPS images), the conditioned inhibitor is equivalent to a signal of non-reward and would be expected to engage the BIS.

Thus in a general sense, the present results suggest that habitual overactivity in the BIS in those high in the related temperamental trait can impair its normal function. According to the revised version of the theory (Gray and McNaughton, 2000; Corr, 2010) BIS-anxiety mediates the detection and resolution of goal conflict (for example between approach and avoidance, by way of “risk assessment” behaviors) rather than reactions to conditioned aversive stimuli, which are mediated by the BIS-FFFS. Signals of non-reward are secondarily aversive, but are a less likely trigger for the BIS-FFFS than are signals of punishment, and are more likely to engage the BIS-anxiety system. In any event, in the present study both BIS-FFFS and BIS-anxiety were negatively related to inhibitory learning, so the general conclusion still stands: temperamentally high levels of BIS activation were associated with impaired rather than enhanced BIS functioning.

Another surprising finding was the lack of any correlation between measures of excitatory or inhibitory learning and extraversion, which is inconsistent with Eysenck’s (1957, 1967) theory of how differences in conditionability give rise to differences in personality. There are grounds to suppose that conditioning differences will also depend on the nature of the US for positive stimuli (as used in the present study), but this should just affect the direction of difference, with higher rather than lower conditioning predicted in extraverts (Gray, 1970, 1972).

The results of the present study are likely to be robust in that the sample size was relatively large. However, to draw stronger conclusions ideally the experiment should be replicated using a different task variant, to exclude the possibility that there could be some artifact in consequence of the use of a single procedure. In particular, the inhibitory learning procedure used in the present study uses positive IAPS images as the US. The negative images are both more salient and would be predicted to show a different pattern of interrelationships with BIS/BAS scores.

Finally, males generally performed better than females, as reflected in their higher overall scores for both excitatory and inhibitory learning. This sex difference is consistent with the finding that both excitatory and inhibitory learning are reduced in those with higher neuroticism scores – as it is very well-established that females show higher levels of neuroticism (Jorm, 1987; Francis, 1993; Lynn and Martin, 1997), as well as higher levels of BIS-anxiety (Gray, 1971). Both of these sex differences were confirmed in the correlational analyses reported in **Table 3** (the correlations go in the predicted direction in that females are coded higher than males in the data file). Thus the females tested in the present sample were more neurotic and showed higher behavioural inhibition than did the males.

There was also a significant correlation between age and associative learning, in that older participants showed relatively better excitatory learning, although inhibitory conditioning did not vary with age (also it should be noted that this was a relatively young sample – in the range 18–56 years).

COMPARISON WITH EARLIER STUDIES

The overall pattern of results is consistent with a role for impulsivity, as measured by BAS-drive, in excitatory but not inhibitory

learning, and for behavioral inhibition in inhibitory but not excitatory learning. A number of previous studies have demonstrated apparently opponent effects using measures of impulsivity and behavioral inhibition, e.g., using the Go/No-Go task and the Stop Signal task (Visser et al., 1996; Logan et al., 1997; Enticott et al., 2006). However, to date there has been little systematic examination of the relationship between impulsivity and associative learning. The present results are consistent with the possibility that impaired associative learning processes could be responsible for aspects of impulsive behavior and disorders (He et al., 2011, 2012).

However, contrary to our predictions, the present study did not find any correlation between impulsivity (as measured by the BAS) and inhibitory learning performance, although inhibitory learning was related to BIS scores. This contrasts with our previous findings using a different task variant (Migo et al., 2006), where we found a negative correlation between inhibitory learning and BAS-reward responsiveness, but none with behavioral inhibition as measured by BIS scores. There are several possible explanations of these discrepancies. First, the sample was much smaller in the earlier study (Migo et al., 2006, which used 60 participants), thus there was less statistical power. Moreover, not only are the correlations between paper-and-pencil questionnaire measures and behavioral measures of impulsivity relatively low (Paulsen and Johnson, 1980; Milich and Kramer, 1984; Helmers et al., 1995; Claes et al., 2006), but it has also been argued that the low arousal conditions typical of laboratory testing underestimate impulsivity (Helmers et al., 1997). There were also procedural differences: in the earlier variant, stimuli were presented serially and included distractors, to reduce the potential role of external inhibition as an alternative explanation of disrupted responding when the inhibitory stimulus was introduced (Migo et al., 2006). By contrast, the present design controlled for external inhibition explicitly with the non-reinforced control stimulus, X.

SMALL EFFECT SIZES FOR PERSONALITY

Although statistically some associations were demonstrated, the effect sizes were relatively small. Yet the experimental design used in the present study has been used to demonstrate CI deficits in disordered groups with much smaller sample sizes. Specifically CI was clearly impaired in a sample of 24 non-psychotic offenders with PDs (He et al., 2011). We also found CI to be significantly reduced in a sample of 25 community-based schizophrenic participants, although with a different profile to that seen in offenders in that excitatory learning was also reduced (He et al., 2012). The study of offenders included dimensional scores from the International PD Examination (Loranger et al., 1994) and the Psychopathy Check List-Revised (Hare, 1991). There was no significant correlation between any of the available measures of personality or behavioral traits and the summary measures of excitatory and inhibitory learning. However, some of the effect sizes for these non-significant correlations were moderate and – despite the relatively modest sample sizes – clear group differences in relation to dangerousness and severity were demonstrated (He et al., 2011). In the study of CI in relation to schizophrenia, individual differences in symptomatology were captured by the Positive and Negative Syndrome Scale (PANSS; Kay et al., 1987). We found a significant correlation between the negative symptoms sub-scales

of this measure and the summary measure of inhibitory learning, and also a marginally significant correlation with the excitatory learning score. In both cases the effect size was medium-large – this despite the fact that PANSS scores were not available for all participants (He et al., 2012).

IMPLICATIONS FOR DISORDER

The results of the present study can be related to earlier studies of anxiety-related disorders. For example, the significant negative correlation between excitatory learning performance and EPQ-RS neuroticism suggests that individuals who are prone to suffer strong, changeable mood, and to overreact in emotional situations, show poorer excitatory learning ability. People who score higher on neuroticism have been argued to be more likely to experience anxiety (Eysenck, 1957, 1967), particularly if their extraversion scores are also low (Gray, 1970, 1972). In this sense, the results of the current study are consistent with the impaired associative learning processes seen in anxiety and depressive disorders (Fowles, 1980, 1993; Gray, 1985; Davey, 1992; Grillon, 2002). The present study extends the demonstration of impaired associative learning processes to inhibitory conditioning, which was also reduced in those with higher EPQ-RS neuroticism and higher BIS scores. Thus, the results point to (susceptibility to) anxiety as a predictor of impaired CI.

To date, we have been unable to recruit participants with clinical levels of anxiety disorder in sufficient numbers. However, the apparent relationship to anxiety demonstrated in the present study of normal participants is consistent with our finding of reduced inhibitory and excitatory learning in participants with schizophrenia (He et al., 2012). Patients with schizophrenia have been found to have relatively high BIS scores. Moreover, this questionnaire study showed that higher BIS sensitivity correlated with duration of illness (Scholten et al., 2006). However, we have no basis to comment on anxiety levels in the group of offenders we studied using this same task (He et al., 2011), and in the present study there was no relationship between inhibitory learning scores and psychotism (which has been argued to predict psychopathic tendencies, Eysenck and Eysenck, 1976b; Eysenck, 1992).

THE RELATIONSHIP BETWEEN EXCITATORY AND INHIBITORY LEARNING

Inhibitory and excitatory learning are inevitably inter-dependent, since a conditioned inhibitor signals the absence of an outcome predicted by an excitatory stimulus. Thus excitatory learning must first be established before inhibitory learning is introduced. Indeed in the present study, in total 18 participants were excluded from the CI test because they did not meet the excitatory learning criterion. Given this background, some commonalities in the individual differences profile predicting better excitatory and those predicting better inhibitory learning is to be expected.

However, animal studies nonetheless suggest that inhibitory and excitatory learning are dissociable (Rescorla, 1969; Daw et al., 2002), and that positive and negative prediction error are coded oppositely at the neuronal level (Tobler et al., 2003). Thus distinct neural substrates could underlie the variation in excitatory and inhibitory learning accompanying differences in neuroticism and behavioral inhibition in the present study (see also He et al.,

2011, 2012). Moreover, the overall correlation between excitatory and inhibitory learning scores was not significant in the present study, suggesting that – despite their inevitable dependence on earlier excitatory conditioning – individual differences in inhibitory learning are not entirely dependent on those seen in excitatory learning.

IMPLICATIONS FOR GENERAL THEORIES OF ASSOCIATIVE LEARNING

Variations in excitatory and inhibitory learning could in principle be used to account for differences between people, but the available learning theories are monolithic. In other words, theories of associative learning are not yet sufficiently articulate to accommodate the effects of individual differences in information processing, in turn based in individual differences in nervous system function. The results reported in the present study underscore the importance of this kind of theoretical development, but the work needed is more complex than modeling a group difference in terms of an existing theory. Temperamental traits are measured as scores on continuous variables and the full complexity of an individual's personality can only be captured as a profile of scores on a variety of measures, some of which are orthogonal, some of which are inter-dependent. Thus, for example, neuroticism and extraversion were originally conceived as orthogonal factors (Eysenck, 1957, 1967; 1981; Eysenck et al., 1985; Eysenck and Eysenck, 1991), as were behavioral inhibition and activation (Gray, 1972, 1982). However, since the latter reflect a rotation of Eysenck's personality dimensions, neuroticism is correlated with behavioral inhibition and extraversion is correlated with behavioral activation (Gray, 1972, 1982). Similarly, as might be expected given that they are derived from a single scale, BIS-anxiety and BIS-FFFS are inter-dependent (Heym et al., 2008). Thus the formal inclusion of individual differences into contemporary theories of associative learning will require the introduction of multi-factorial moderating variables, to specify their effects on learning rate parameters such as the CS and US factors which influence associability.

Historically the aim has been to establish general laws of learning. The observed dissociation in the effects of behavioral activation and behavioral inhibition on excitatory vs. inhibitory learning could in principle be incorporated into learning theories which make formal predictions about inhibitory as well as excitatory learning (e.g., Rescorla and Wagner, 1972). This would not affect the generality of the theories and could improve their predictive power. However, the formal inclusion of reinforcement sensitivity theory (Gray, 1972, 1982; Gray and McNaughton, 2000) would suggest the need for different variants of the models to be applied to learning situations which use appetitive vs. aversive USs. Moreover, any such learning models would need to be weighted to take effect size into account, and effect sizes of the magnitude reported here could be too small to warrant what might be viewed as unnecessary complication. Ultimately, dynamic interactionist models would be necessary to capture the three-way interaction between personality, conditionability, and environmental context (Ferguson et al., 2012; Haslam et al., 2012).

ACKNOWLEDGMENTS

Zhimin He was supported by a University of Nottingham School of Psychology Studentship.

REFERENCES

- Avila, C., Parcet, M. A., and Barros-Loscertales, A. (2008). A cognitive neuroscience approach to individual differences in sensitivity to reward. *Neurotox. Res.* 14, 191–203.
- Barlow, D. H. (2000). Unraveling the mysteries of anxiety and its disorders from the perspective of emotion theory. *Am. Psychol.* 55, 1247–1263.
- Beaver, J. D., Lawrence, A. D., Van Ditzhuijzen, J., Davis, M. H., Woods, A., and Calder, A. J. (2006). Individual differences in reward drive predict neural responses to images of food. *J. Neurosci.* 26, 5160–5166.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 57, 289–300.
- Buss, A. H., and Plomin, R. (1975). *A Temperament Theory of Personality Development*. New York: Wiley-Interscience.
- Carver, C. S., and White, T. L. (1994). Behavioural inhibition, behavioural activation, and affective responses to impending reward and punishment: the BIS/BAS scales. *J. Pers. Soc. Psychol.* 67, 319–333.
- Claes, L., Nederkoorn, C., Vandereycken, W., Guerrieri, R., and Vertommen, H. (2006). Impulsiveness and lack of inhibitory control in eating disorders. *Eat. Behav.* 7, 196–203.
- Corr, P. J. (2010). The psychoticism-psychopathy continuum: a neuropsychological model of core deficits. *Pers. Individ. Dif.* 48, 695–703.
- Corr, P. J., Pickering, A. D., and Gray, J. A. (1995). Personality and reinforcement in associative and instrumental learning. *Pers. Individ. Dif.* 19, 47–71.
- Costa, P. T., and McCrae, R. R. (1992). Four ways five factors are basic. *Pers. Individ. Dif.* 13, 653–665.
- Davey, G. C. L. (1992). Classical conditioning and the acquisition of human fears and phobias: a review and synthesis of the literature. *Adv. Behav. Res. Ther.* 14, 29–66.
- Daw, N. D., Kakade, S., and Dayan, P. (2002). Opponent interactions between serotonin and dopamine. *Neural Netw.* 15, 603–616.
- Enticott, P. G., Ogloff, J. R. P., and Bradshaw, J. L. (2006). Associations between laboratory measures of executive inhibitory control and self-reported impulsivity. *Pers. Individ. Dif.* 41, 285–294.
- Eysenck, H. J. (1957). *Dynamics of Anxiety and Hysteria*. London: Routledge and Kegan Paul.
- Eysenck, H. J. (1967). *The Biological Basis of Personality*. Springfield, IL: Charles C. Thomas.
- Eysenck, H. J. (1981). *A Model for Personality*. Berlin: Springer.
- Eysenck, H. J. (1992). The definition and measurement of psychoticism. *Pers. Individ. Dif.* 13, 757–785.
- Eysenck, H. J., and Eysenck, S. B. G. (1976a). *Manual of the Eysenck Personality Questionnaire*. London: Hodder & Stoughton.
- Eysenck, H. J., and Eysenck, S. B. G. (1976b). *Psychoticism as a Dimension of Personality*. London: Hodder & Stoughton.
- Eysenck, H. J., and Eysenck, S. B. G. (1991). *Manual of the Eysenck Personality Scales*. London: Hodder & Stoughton.
- Eysenck, S. B. G., Eysenck, H. J., and Barrett, P. (1985). A revised version of the psychoticism scale. *Pers. Individ. Dif.* 6, 21–29.
- Ferguson, E., Ward, J. W., Skatova, A., Cassaday, H. J., Bibby, P. A., and Lawrence, C. (2012). Health specific traits beyond the Five Factor Model, cognitive processes and trait expression: replies to Watson (2012), Matthews (2012) and Haslam, Jetten, Reynolds and Reicher (2012). *Health Psychol. Rev.* doi:10.1080/17437199.2012.701061
- Fowles, D. C. (1980). The three arousal model: implications of Gray's two-factor learning theory for heart rate, electrodermal activity, and psychopathy. *Psychophysiology* 17, 87–104.
- Fowles, D. C. (1987). Application of a behavioural theory of motivation to the concepts of anxiety and impulsivity. *J. Res. Pers.* 21, 417–435.
- Fowles, D. C. (1993). "Biological variables in psychopathology: a psychobiological perspective," in *Comprehensive Handbook of Psychopathology*, 2nd Edn, eds P. B. Sutker and H. E. Adams (New York: Plenum), 57–82.
- Francis, L. J. (1993). The dual nature of the Eysenckian neuroticism scales: a question of sex differences. *Pers. Individ. Dif.* 15, 43–49.
- Gray, J. A. (1970). The psychophysiological basis of introversion-extraversion. *Behav. Res. Ther.* 8, 249–266.
- Gray, J. A. (1971). Sex differences in emotional behaviour in mammals including man: endocrine bases. *Acta Psychol. (Amst.)* 35, 29–46.
- Gray, J. A. (1972). "The psychophysiological basis of introversion-extraversion: a modification of Eysenck's theory," in *The Biological Bases of Individual Behaviour*, eds V. D. Nebylitsyn and J. A. Gray (San Diego, CA: Academic Press), 182–205.
- Gray, J. A. (1981). "A critique of Eysenck's theory of personality," in *A Model for Personality*, ed. H. J. Eysenck (Berlin: Springer-Verlag), 246–276.
- Gray, J. A. (1982). *The Neuropsychology of Anxiety: An Enquiry into the Functions of the Septo-Hippocampal System*. New York: Oxford University Press.
- Gray, J. A. (1985). "Issues in the neuropsychology of anxiety," in *Anxiety and the Anxiety Disorders*, eds A. H. Tuma and J. D. Maser (Hillsdale, NJ: Erlbaum), 5–25.
- Gray, J. A. (1990). Brain systems that mediate both emotion and cognition. *Cogn. Emot.* 4, 269–288.
- Gray, J. A., and McNaughton, N. (2000). *The Neuropsychology of Anxiety: An Enquiry into the Functions of the Septo-Hippocampal System*, 2nd Edn. Oxford: Oxford University Press.
- Grillon, C. (2002). Associative learning deficits increase symptoms of anxiety in humans. *Biol. Psychiatry* 51, 851–858.
- Hare, R. D. (1991). *Manual for the Hare Psychopathy Checklist-Revised*. Toronto, ON: Multi-Health Systems.
- Hare, R. D., Hart, S. D., and Harpur, T. J. (1991). Psychopathy and the proposed DSM-IV criteria for antisocial personality disorder. *J. Abnorm. Psychol.* 100, 391–398.
- Haslam, S. A., Jetten, J., Reynolds, K. J., and Reicher, S. D. (2012). The dangers of isolating the individual: the need for a dynamic and socially structured model of personality – commentary on Ferguson (2011). *Health Psychol. Rev.* doi:10.1080/17437199.2011.606097
- He, Z., Cassaday, H. J., Howard, R. C., Khalifa, N., and Bonardi, C. M. (2011). Impaired Pavlovian conditioned inhibition in offenders with personality disorders. *Q. J. Exp. Psychol.* 64, 2334–2351.
- He, Z., Cassaday, H. J., Park, S. B. G., and Bonardi, C. M. (2012). When to hold that thought: an experimental study showing reduced inhibition of pre-trained associations in schizophrenia. *PLoS ONE* 7:e42175. doi:10.1371/journal.pone.0042175
- Helmers, K. F., Young, S. N., and Pihl, R. O. (1995). Assessment of measures of impulsivity in healthy male volunteers. *Pers. Individ. Dif.* 19, 927–935.
- Helmers, K. F., Young, S. N., and Pihl, R. O. (1997). Extraversion and behavioral impulsivity. *Pers. Individ. Dif.* 23, 441–452.
- Heym, N., Ferguson, E., and Lawrence, C. (2008). An evaluation of the relationship between Gray's revised RST and Eysenck's PEN: distinguishing BIS and FFFS in Carver and White's BIS/BAS scales. *Pers. Individ. Dif.* 45, 709–715.
- Heym, N., and Lawrence, C. (2010). The role of Gray's revised RST in the P-psychopathy continuum: the relationships of Psychoticism with a lack of fear and anxiety, and increased impulsivity (2010). *Pers. Individ. Dif.* 49, 874–879.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* 6, 65–70.
- Jennions, M. D., and Möller, A. P. (2003). A survey of the statistical power of research in behavioural ecology and animal behaviour. *Behav. Ecol.* 14, 438–445.
- Jorm, A. F. (1987). Sex differences in neuroticism: a quantitative synthesis of published research. *Aust. N. Z. J. Psychiatry* 21, 501–506.
- Kantini, E., Cassaday, H. J., Hollis, C., and Jackson, G. M. (2011a). The normal inhibition of associations is impaired by clonidine in Tourette Syndrome. *J. Can. Acad. Child Adolesc. Psychiatry* 20, 96–106.
- Kantini, E., Cassaday, H. J., Batty, M. J., Hollis, C., and Jackson, G. M. (2011b). Associative learning in ADHD: improved expression under methylphenidate. *Open J. Psychiatry* 1, 20–29.
- Kay, S. R., Fiszbein, A., and Opler, L. (1987). The Positive and Negative Syndrome Scale for schizophrenia. *Schizophr. Bull.* 13, 261–276.
- Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (2005). *International Affective Picture System (IAPS): Instruction Manual and Affective Ratings*. Technical Report A-6. Gainesville, FL: University of Florida.
- Larzelere, R. E., and Mulaik, S. A. (1977). Single-sample tests for many correlations. *Psychol. Bull.* 84, 557–569.
- Logan, G. D., Schachar, R. J., and Tan-nock, R. (1997). Impulsivity and inhibitory control. *Psychol. Sci.* 8, 60–64.
- Loranger, A. W., Sartorius, N., Andreoli, A., Berger, P., Buchheim, P., Channabasavanna, S. M., et al. (1994). The international personality disorder examination, IPDE. The WHO/ADAMHA International Pilot Study of personality disorders. *Arch. Gen. Psychiatry* 51, 215–224.
- Lynn, R., and Martin, T. (1997). Gender differences in extraversion, neuroticism, and psychoticism in

- 37 nations. *J. Soc. Psychol.* 137, 369–373.
- Migo, E. M., Corbett, K., Graham, J., Smith, S., Tate, S., Moran, P. M., et al. (2006). A novel test of conditioned inhibition correlates with personality measures of schizotypy and reward sensitivity. *Behav. Brain Res.* 168, 299–306.
- Milich, R., and Kramer, J. (1984). “Reflections on impulsivity: an empirical investigation of impulsivity as a construct,” in *Advances in Learning and Behavioral Disabilities*, Vol. 3, ed. K. Gadow (Greenwich, CT: JAI Press), 57–93.
- Munro, G., Dywan, J., Harris, G., McKee, S., Unsal, A., and Segalowitz, S. (2007). Response inhibition in psychopathy: the frontal N2 and P3. *Neurosci. Lett.* 418, 149–153.
- Nakagawa, S. (2004). A farewell to Bonferroni: the problems of low statistical power and publication bias. *Behav. Ecol.* 15, 1044–1045.
- Paulsen, K., and Johnson, M. (1980). Impulsivity: a multidimensional concept with developmental aspects. *J. Abnorm. Child Psychol.* 8, 269–277.
- Pavlov, I. P. (1927). *Conditioned Reflexes*. London: Oxford University Press.
- Perneger, T. V. (1998). What’s wrong with Bonferroni adjustments. *Br. Med. J.* 316, 1236–1238.
- Pickering, A. D. (1997). The conceptual nervous system and personality: from Pavlov to neural networks. *Eur. Psychol.* 2, 139–163.
- Pickering, A. D., and Gray, J. A. (1999). “The neuroscience of personality,” in *Handbook of Personality: Theory and Research*, eds L. A. Pervin and O. P. John (New York: Guilford Press), 277–299.
- Pierò, A. (2010). Personality correlates of impulsivity in subjects with generalized anxiety disorders. *Compr. Psychiatry* 51, 538–545.
- Putman, P., Hermans, E., and van Honk, J. (2004). Emotional Stroop performance for masked angry faces: it’s BAS, not BIS. *Emotion* 4, 305–311.
- Rescorla, R. A. (1969). Pavlovian conditioned inhibition. *Psychol. Bull.* 72, 77–94.
- Rescorla, R. A., and Wagner, A. R. (1972). “A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and non-reinforcement,” in *Classical Conditioning II*, eds A. H. Black, and W. F. Prokasy (New York: Appleton-Century-Crofts), 64–99.
- Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution* 43, 223–225.
- Scholten, M. R. M., van Honk, J., Aleman, A., and Kahn, R. S. (2006). Behavioural inhibition system (BIS), behavioural activation system (BAS) and schizophrenia: relationship with psychopathology and physiology. *J. Psychiatr. Res.* 40, 638–645.
- Smillie, L. D., Pickering, A. D., and Jackson, C. J. (2006). The new reinforcement sensitivity theory: implications for personality measurement. *Pers. Soc. Psychol. Rev.* 10, 320–335.
- Snorrason, Í., Smári, J., and Ólafsson, R. P. (2011). The interaction effect of impulsivity and responsibility in relation to obsessive – compulsive symptoms. *Cognit. Ther. Res.* 35, 79–86.
- Tobler, P. N., Dickinson, A., and Schultz, W. (2003). Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. *J. Neurosci.* 23, 10402–10410.
- Visser, M., Das-Smaal, E., and Kwakman, H. (1996). Impulsivity and negative priming: evidence for diminished cognitive inhibition in impulsive children. *Br. J. Psychol.* 87, 131–140.

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: 07 January 2013; paper pending published: 14 February 2013; accepted: 14 April 2013; published online: 08 May 2013.

Citation: He Z, Cassaday HJ, Bonardi C and Bibby PA (2013) Do personality traits predict individual differences in excitatory and inhibitory learning? *Front. Psychol.* 4:245. doi: 10.3389/fpsyg.2013.00245

This article was submitted to *Frontiers in Personality Science and Individual Differences*, a specialty of *Frontiers in Psychology*.

Copyright © 2013 He, Cassaday, Bonardi and Bibby. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.



Cocaine dependent individuals and gamblers present different associative learning anomalies in feedback-driven decision making: a behavioral and ERP study

Ana Torres, Andrés Catena, Antonio Cándido*, Antonio Maldonado, Alberto Megías and José C. Perales

Learning, Emotion and Decision Research Group, Mind, Brain and Behavior Research Center/Centro de Investigación Mente, Cerebro y Comportamiento, University of Granada, Granada, Spain

Edited by:

Rachel M. Msetfi, University of Limerick, Ireland

Reviewed by:

Caroline N. Wade, Oxford University, UK

Henry W. Chase, University of Pittsburgh, USA

*Correspondence:

Antonio Cándido, Departamento de Psicología Experimental, Universidad de Granada, Campus de Cartuja s/n, 18071 Granada, Spain.
e-mail: acandido@ugr.es

Several recent studies have demonstrated that addicts behave less flexibly than healthy controls in the probabilistic reversal learning task (PRLT), in which participants must gradually learn to choose between a probably rewarded option and an improbably rewarded one, on the basis of corrective feedback, and in which preferences must adjust to abrupt reward contingency changes (reversals). In the present study, pathological gamblers (PG) and cocaine dependent individuals (CDI) showed different learning curves in the PRLT. PG also showed a reduced electroencephalographic response to feedback (Feedback-Related Negativity, FRN) when compared to controls. CDI's FRN was not significantly different either from PG or from healthy controls. Additionally, according to Standardized Low-Resolution Electromagnetic Tomography analysis, cortical activity in regions of interest (previously selected by virtue of their involvement in FRN generation in controls) strongly differed between CDI and PG. However, the nature of such anomalies varied within-groups across individuals. Cocaine use severity had a strong deleterious impact on the learning asymptote, whereas gambling intensity significantly increased reversal cost. These two effects have remained confounded in most previous studies, which can be hiding important associative learning differences between different populations of addicts.

Keywords: addiction, cocaine, gambling, reversal learning, feedback-related negativity, decision-making

INTRODUCTION

Response-outcome association learning tasks have been widely used to explore the cognitive and biological underpinnings of neuropsychiatric disorders (e.g., Everitt et al., 2001; Clark et al., 2004; Redish et al., 2007). The *probabilistic reversal learning task* (PRLT; Swainson et al., 2000) is a dynamic decision-making test (Hastie and Dawes, 2009) in which participants must learn to choose between two response options, one frequently rewarded (and infrequently punished), and the other infrequently rewarded (and frequently punished). Payoffs are administered in the form of real or play money, or virtual points. Once preferences are stable, reward/punishment contingencies reverse, in such a way that the advantageous option becomes disadvantageous, and *vice versa*, and learners must retune their preferences in accordance with the new contingencies.

Addicted individuals (and patients from other psychopathological and neurological populations) have been observed to display abnormal performance patterns in the reversal learning task. Some types of patients are slower than normal to readjust their preferences after a reversal. This *increased reversal cost* has been interpreted as a sign of goal-disengaged, habit-driven, error-insensitive, or perseverative behavior (Clarke et al., 2005, 2008). In other cases, pre-reversal learning asymptote has been observed to be abnormally low (e.g., Fernández-Serrano et al., 2012), or abnormally high (e.g., Verdejo-García et al., 2010). Although in many studies pre-asymptotic and asymptotic effects have not been dissociated

(see Tsuchida et al., 2010; Torres et al., submitted; for similar arguments), there is broad consensus that the sort of dynamic decision-making processes involved in reversal learning tasks is crucial to understand the neuropsychology of addictive disorders (Ersche et al., 2008; Camchong et al., 2011; Izquierdo and Jentsch, 2012; Leeman and Potenza, 2012; Lucantonio et al., 2012). As also shown in this work, reversal learning tasks tackle on the type of balanced feedback sensitivity and learning flexibility that are needed for adaptive decision making in real life.

In spite of that, abnormal PRLT performance patterns do not seem to fully generalize across addictions. Separate sources of evidence seem to show that heavy gambling is preferentially linked to the increase of reversal cost (de Ruiter et al., 2009), whereas cumulative toxicity of cocaine generates more unspecific performance deviations and, particularly, less accurate decision making once asymptotic learning has been reached, prior to contingency reversal (accompanied by working memory and planning dysfunction; Fernández-Serrano et al., 2012). In a recent review, Leeman and Potenza (2012) have integrated these independent pieces of evidence, and have concluded that increased reversal costs are more frequent and robust in pathological gamblers (PG) than in drug-dependent individuals (see also Ersche et al., 2008).

The present study focuses on the coincidences and divergences between gambling and cocaine addiction, with regard to the anomalies they generate in reversal learning performance. A sample of cocaine dependent individuals (CDI) was compared against one

of pathological gamblers (PG), and a group of matched healthy controls (HC) in a reversal learning task, at the behavioral and the electroencephalographic levels. To our knowledge, only two studies have directly compared cocaine users against PG in a battery of personality and neuropsychological tests (Albein-Urios et al., 2012a; Torres et al., 2013). However, no studies have directly compared matched groups of patients with the two disorders, between them and against a group of HC, in the PRLT.

There are several reasons to jointly study PG and CDI samples, but also to draw conclusions with some caution. First, parallelisms between these two addictions have been known for a long time. Some studies have found behavioral similarities, high comorbidity rates, and a partially common neurobiological and genetic etiology (see Hall et al., 2000; Potenza, 2008). For example, prospective family studies have observed that the percentage of future gamblers among children of gamblers doubles the population baseline (8 versus 4%). And, in parallel, children of gamblers tend to show a preference for stimulant drugs, so that the proportion of future cocaine users among children of gamblers doubles the population baseline (10 versus 5%; Jacobs et al., 1989). Complementarily, in a sample of 298 treatment-seeking cocaine abusers, Steinberg et al. (1992) found a prevalence of pathological gambling approximately 10 times larger than the rate of gamblers found in community samples.

As noted by Albein-Urios et al. (2012a), “the two disorders have also notable similarities in terms of subjective effects, reinforcing schedules, and temporal patterns of consumption [. . .]. In these respects, cocaine addiction is arguably more similar to pathological gambling than other forms of drug dependence.” Moreover, direct experimental evidence shows that a game of chance can serve as an alternative reinforcer to smoking cocaine (Vosburg et al., 2010).

Second, these similarities seem to indicate that a comparison between PG and CDI could be helpful to disentangle vulnerability and toxicity effects of cocaine use in group comparison studies. This argument is based on the assumption that the neurobehavioral anomalies observed in PG samples are equivalent to those found in CDI samples *minus* the neurotoxic effects of cocaine. Still, this rationale is problematic, as far as it assumes that gambling does not have a cumulative impact on brain function (an assumption that goes against current evidence; see Robinson and Berridge, 2003; van Holst et al., 2010). Mere between-groups comparisons do not strictly allow such a type of conclusions.

And third, although only prospective and longitudinal studies can strictly discriminate vulnerability from cocaine/gambling exposure factors, studies comparing samples of addicts against HC can be informative if they meet some criteria. On the one hand, although complete matching between samples is virtually unattainable, it is important to select samples carefully. They must be completely separated in terms of key addictive behaviors (gamblers do not use cocaine, cocaine users do not gamble, and controls neither gamble nor use cocaine), and matched in terms sociodemographic variables, intellectual functioning, and absence of any other psychiatric disorders. And, on the other hand, chronic exposure to cocaine/gambling must be estimated on an individual basis. In this type of studies, the degree of exposure can be measured only retrospectively, but there exist interview-based methods to approximate it. These

methods allow for the estimation of exposure-dependent effects on neurobehavioral anomalies (Verdejo-García et al., 2005). Estimation of exposure-dependent effects can help us to identify *acquired* individual differences caused by the progressive course of the addictive processes, that is, by toxicity, neuroadaptation, or sensitization.

In this work, we also recorded feedback-evoked electroencephalographic activity during reversal learning. The analysis of this activity is valuable in several senses. Evoked-related potentials (ERP) are sometimes more sensitive to between-condition differences than behavioral measures (see, for example, Karayanidis et al., 2000; Hajcak et al., 2005b). Accordingly, convergent psychophysiological and behavioral evidence is more conclusive than behavioral results alone, especially when behavioral effects are subtle. Furthermore, in the present case, there are also evidence-driven hypotheses about the potential biological substrate of reversal learning anomalies, and the candidate ERP components that best reveal such anomalies. Our interest in the feedback-related negativity (FRN) and its potential relation with reversal costs, is grounded on previous experimental evidence (Chase et al., 2011; Bismark et al., 2012; Hampshire et al., 2012). Finally, our attempts to identify the most likely anatomical origins of addiction-related FRN anomalies can be useful to link such anomalies to the malfunctioning of specific circuits in the brain (Schoenbaum et al., 2006).

In summary, in the present work we analyze in detail some dynamic features of reversal learning performance in PG and CDI, matched in potentially confounding factors, and compared against non-addicts. Our main aim is threefold: (1) To check for the existence of anomalies in reversal learning in both types of addicts. On this regard, we expect reversal cost to be more evident in PG than in cocaine users. (2) To explore the roles of gambling and cocaine exposure on specific components of reversal learning (specifically, reversal costs and asymptotic learning levels). As measures of chronic and acute exposure, *severity* (the estimation of the lifetime total amount gambled, or the total quantity of cocaine consumed) and *intensity* (mean amount of drug consumed/money gambled per month) scores will be obtained for all participants in the clinical groups. On the basis of the abovementioned evidence, we expect learning anomalies in the CDI group to be attributable to cocaine dosage exposure (and thus to correlate with cocaine use severity). Whether or not reversal cost depends on gambling intensity or severity remains an open question. And (3), to analyze the electroencephalographical response to feedback in the three groups. Although we can foretell the presence of FRN anomalies in the clinical groups (and associated abnormal brain activations), whether or not such anomalies differ across the clinical groups also remains to be tested.

MATERIALS AND METHODS

PARTICIPANTS AND PROCEDURE

Cocaine dependent individuals ($n = 20$) were recruited from the *Proyecto Hombre* rehabilitation centers in Granada and Málaga (Spain) between January 2011 and December 2012. PG ($n = 21$) were recruited from AGRAJER (Granadian Association of Gamblers in Rehabilitation, Granada, Spain) between October 2010 and December 2012. Controls ($n = 23$) were recruited by incidental

sampling, in such a way that their sociodemographic characteristics were not far from the clinical groups.

The inclusion criteria were (i) meeting DSM-IV criteria for cocaine dependence (CDI group) or pathological gambling (PG group) – as assessed by the Structured Clinical Interview for DSM-IV Disorders – Clinician Version (SCID; First et al., 1997); (ii) having a minimum abstinence interval of 15 days for all substances of abuse except nicotine, as determined by weekly urine toxicological tests (CDI) or cross validated therapist- and self-reports (PG). Exclusion criteria were: (i) the presence of any other Axis I or Axis II comorbid disorders with the exception of nicotine dependence; (ii) the presence of history of head injury or any diseases affecting the central nervous system. The study counted with explicit permission from the University of Granada's Ethics Committee. Prior to psychological and neuropsychological assessment, all participants were informed about the objectives and characteristics of the study, and signed an informed consent form. All of them were compensated with 36€ for their participation, independently of performance.

In order to assess the degree of matching between-groups, participants were also assessed using the Kaufman Brief Intelligence Test (K-BIT), and were questioned about their age, and number of education years. **Table 1** displays main descriptive data for all the relevant variables in the three groups. The three groups were matched on sociodemographic variables, but not on usage of other drugs. As shown in the table, the group differences in alcohol and cannabis use were globally significant, with CDI being the group with larger alcohol and cannabis consumption.

The procedure went as follows: upon consent, participants were instructed about the general procedure, and then questioned about the abovementioned sociodemographic variables. The K-BIT, Interview for Research on Addictive Behavior (IRAB), and the IRAB-equivalent gambling-related questions were administered together, in a random order. A fourth psychometric instrument,

the UPPS-P questionnaire on impulsive behavior, was administered intertwined with these, also in a random position. The PRLT and a second neuropsychological task (the Go/No-go motor inhibition task) were administered together. Half the participants performed the neuropsychological tasks first, in a random order, followed by the psychometric instruments. The other half were assessed with the psychometric tools first, and then performed the two neuropsychological tasks.

The UPPS-P and Go/No-go tests were included in this procedure as part of a different study, carried out with the same participants, on the role of impulsivity in addiction and motor inhibition (Go/No-go). UPPS-P scores did not exert any effect on reversal learning performance, either by itself or in combination with group (minimum $p = 0.22$). In addition, UPPS-P scores are confounded with addictive behaviors (addictive behaviors are by definition impulsive) so, they were not taken into account for the present study. Still, between-group differences in impulsivity, and the impact of impulsivity on other decision-making tasks have been reported in Torres et al. (2013).

INSTRUMENTS

Interview for research on addictive behaviors (Spanish version; Verdejo-García et al., 2005)

As noted in the introduction, a key factor in the present study is the degree of dosage-like exposure to cocaine and gambling activities (in the CDI and PG groups, respectively). Psychometric tools developed for clinical purposes do not measure exposure in an isolated manner (disregarding craving intensity, perception of lack of control over the addictive behavior, social and family problems, financial problems, and other symptoms and consequences of addiction).

All of those side factors are irrelevant to the current study. Actually, they would blur drug/gambling exposure effects. Hence, information about lifetime amount and duration of use of the

Table 1 | Sociodemographic, psychometric, and drug use differences between healthy controls, HC; pathological gamblers, PG; and cocaine dependent individuals, CDI.

	Group							
	HC		PG		CDI			
SOCIODEMOGRAPHIC VARIABLES								
<i>n</i>	23		21		20			
Proportion of females	0.09		0.10		0.00			
	Mean	SD	Mean	SD	Mean	SD	<i>F</i>	<i>p</i>
Age	30.13	8.63	31.43	5.92	34.75	6.51	2.31	0.11
Education years	14.55	3.16	13.90	4.66	15.05	4.21	0.42	0.66
INTELLECTUAL PERFORMANCE								
IQ (K-BIT)	106.25	10.22	101.10	9.07	105.35	9.39	1.77	0.18
DRUG USE PATTERNS								
Alcohol monthly use (ethanol units/month)	44.02	42.81	99.14	73.87	158.20	118.88	10.18	<0.01*
Cannabis monthly use (joints/month)	13.00	26.61	8.39	25.96	67.00	63.71	12.54	<0.01*
Addiction course duration (years)			7.79	5.51	10.48	5.04	2.65	0.11
Abstinence duration (months)			5.64	3.51	6.41	4.32	0.39	0.53

Addiction and abstinence durations refer to the clinically significant addictive behavior (gambling for PG, and cocaine use for CDI). *p*-values in bold are statistically significant.

different drugs was collected using the IRAB (Verdejo-García et al., 2005). The IRAB is inspired by applied and experimental behavior analysis, and was not developed to estimate the clinical significance of addiction, but to quantify the most important parameters of drug use behaviors (frequency, duration, amount), independently of the clinical status of the participant and the accompanying symptomatology. All the participants in the three groups went through the full IRAB interview. Here, we will consider the answers to three questions included in the interview: the average frequency of use (times/month), the average amount consumed per episode (in grams or units), and the total duration of the usage period (in months). In accordance with standardized instructions, these parameters were used to compute two composite measures: (1) average monthly amount of each drug consumed (amount \times frequency), in grams/month, and (2) severity, or estimated lifetime amount of drug consumed (amount \times frequency \times duration), in grams or units.

In order to avoid extremely skewed distributions, monthly amount and severity were translated into within-design rank scores for all analyses. A more detailed display of (non-transformed) IRAB results for HC, PG, and CDI can be found in **Table A1** in the Appendix.

Average monthly use is customarily interpreted as an estimate of the *intensity* of addiction during its course (acute exposure). Severity, on the other hand, is attributed the cumulative effect of addiction (chronic exposure). In the case of drugs of abuse, severity is customarily assumed to correlate with the long-term neurotoxic or neuroadaptive effects of that drug (Albein-Urios et al., 2012a).

The IRAB has not been yet developed for gambling activities. Thus, in order to have equivalent measures for gambling and cocaine use, gamblers were asked the same above-mentioned IRAB questions (amount, frequency, lifetime duration of usage), but referred to gambling activities. That is, the same questions used in the IRAB for registering drug use, were adapted to estimate the two key gambling parameters (intensity and severity), and then translated into within-design rank scores. In this case, as no toxic substance is involved, severity would correlate with cumulative neuroadaptive or practice-dependent effects of gambling activities (Robinson and Berridge, 2003).

The Kaufman brief intelligence test (Kaufman and Kaufman, 1990)

The K-BIT has been standardized and utilized widely, both in clinical and research settings, to assess cognitive abilities. It comprises measures of verbal and non-verbal intelligence and takes 10–30 min to administer. For our purposes, we will use only the compound IQ total score.

Probabilistic reversal learning task (Verdejo-García et al., 2010)

The reversal learning task used here is based on the PROB task described in Swainson et al. (2000). A graphical description of the task can be found in Verdejo-García et al. (2010). In each trial of the task, there was a simultaneous presentation of two squares, drawn in different colored lines. The task consisted of four phases in total. In each phase, one stimulus is considered the “correct” one, as choosing it (i.e., mouse-clicking on it) provides reward in most

cases, and the other is the “wrong” one, as choosing it was penalized most of the times. This means that, on some trials, the computer provided false feedback, i.e., selecting the correct stimulus was followed by false negative feedback (NF) and selecting the incorrect one was followed by false positive feedback. Positions of stimuli were randomly shifted to avoid motor perseveration. Both negative and positive feedbacks were presented visually, and involved winning or losing five points. The total amount of points accrued was continuously viewed just below the center of the screen. Crucially, the color corresponding to the correct choice and the one corresponding to the wrong choice shifted after every phase (40 trials), that is, the stimulus that was previously correct became incorrect, and vice versa.

For half the participants in each group, the percentage of rewarded clicks on the good option was 75% in Phases 1 and 2, and 87.5% in Phases 3 and 4 (the task became slightly easier in its second half). The other way round, for the other half of participants, the percentage of rewarded clicks on the good option was 87.5% in Phases 1 and 2, and 75% in Phases 3 and 4 (the task became slightly more difficult in its second half). In other words, the order of contingencies was a balanced factor.

STATISTICAL ANALYSIS OF PRLT PERFORMANCE

The main dependent variable for global PRLT performance analysis was the number of correct choices (clicks on the highly rewarded option) per each 10-trial block within each 40-trial phase. In a first, full-task analysis, correct choices per phase and block were submitted to a mixed three-factor ANOVA, with phase (1–4) and block (1–4) as within-group factors, and group (HC, PG, CDI) as between-group factor.

Secondly, theory-driven analyses will focus on the number of correct choices in the first block of each phase, and the number of correct choices in the last two blocks of each phase (collapsed). It is important to note that (1) only the number of correct choices in the first block of phases 2–4 can be interpreted as an index of reversal cost. However, block 1 from the first phase will be also included in analyses for design completeness reasons (the inclusion or exclusion of that block does not significantly influence the results of those analyses, nor the main conclusions drawn from them). And (2) the number of correct choices in the two last blocks of each phase can be interpreted as an estimate of asymptotic learning level.¹

This second series of analysis will be restricted to the impact of chronic and acute exposure to gambling/cocaine in the clinical groups. Four ANCOVAs (with intensity and severity as covariates) were separately carried out for the PG and the CDI groups, with the phase-wise first and last (two) blocks' correct choices as separate dependent measures.

¹The use of the averaged two last blocks as an index of asymptotic learning rests on the assumption that no further learning occurs in any of the three groups after block 3 (so learning can be considered maximal in blocks 3 and 4). Actually, taking into account those two blocks only, there was neither effect of block [$F(1, 61) = 2.49$, $MSE = 1.73$, $p = 0.12$] nor block \times group interaction [$F(1, 61) = 1.65$, $MSE = 1.73$, $p = 0.20$]. Mean (SE) number of correct choices were 8.02 (0.26), 7.80 (0.26), 7.28 (0.32), 7.91 (0.27), 7.92 (0.26), 8.03 (0.23), 7.72 (0.23), and 7.89 (0.25) for phase 1 block 3, phase 1 block 4, phase 2 block 3, phase 2 block 4, phase 3 block 3, phase 3 block 4, phase 4 block 3, and phase 4 block 4, respectively.

Similarly to correct choices, decision latencies (measured as reaction times from presentation of the two choice options to the decision made by the participant, averaged for each block) were submitted, firstly, to a block \times phase \times group global analyses. Subsequently, separate ANCOVAs for PG and CDI groups, with monthly use and severity as joint continuous predictors, were also carried out. Although decision latencies have not been customarily taken into account in reversal learning tasks, we will include them here as complementary evidence.

Given that groups differ in alcohol and cannabis use, prior to all analysis involving the group factor (HC, PG, CDI) we carried out an ANCOVA disregarding the factor group, but including alcohol and cannabis monthly use (translated into rank scores) as continuous covariate predictors, and the same dependent measure used in the corresponding between-group analysis. These pre-analyses were thus carried out for the phase- and block-wise number of correct choices, decision latencies, and FRN magnitudes. As shown in ANCOVAs for the potential effects of cannabis and alcohol use on relevant dependent measures in Appendix, none of the potential confounders (alcohol use, cannabis use, and their interaction) had a significant impact on the abovementioned measures.

For all tests, the significance level was set at 0.05, after Greenhouse–Geisser correction of degrees of freedom where it was necessary.

EEG RECORDING

EEGs were recorded from 62 scalp locations using tin electrodes arranged according to the extended 10–20 system mounted on an elastic cap (Brain Products, Inc), and referenced online to FCz. Vertical and horizontal eye activity were recorded from one monopolar electrode placed below the left eye, and one monopolar electrode located in a straight line at the outer canthi of the right eye. Two scalp electrodes were attached to mastoids. All electrode impedances during recording were below 5 k Ω . EEG and EOG were sampled at 1000 Hz and amplified using a 0.016–1000 Hz band-pass filter. Subsequently, all EEG recordings were downsampled to 250 Hz, band-pass filtered using a 0.1–25 Hz 12 dB/octave, re-referenced offline to average activity of the mastoids electrodes, and FCz activity was recovered. Offline signal preprocessing was done using EEGLAB software (Delorme and Makeig, 2004) freely available at <http://scn.ucsd.edu/eeGLab>.

ERP EXTRACTION AND ANALYSIS

EEG recordings were segmented from -200 to $+350$ ms, time-locked to the feedback onset. Epochs were corrected for ocular artifacts by computing the SOBI ICA decomposition (Belouchrani et al., 1993, 1997; Cardoso and Souloumiac, 1996, see also Tang et al., 2004), as identified by the ADJUST algorithm (Mognon et al., 2011). Other artifacts were subsequently removed using an automatic rejection procedure: segments were excluded for the remaining analyses when amplitudes were outside the ± 100 μ V range. Afterward, segments were categorized as belonging to positive- or negative-feedback trials (PF, NF). After the artifact correction procedure, a minimum of 27 trials for the NF and 51 for PF segments were retained for further processing.

Next, the FRN was computed for each participant and feedback condition, as the difference between the average amplitude in

the 220–350 ms post-feedback interval, and the preceding positive peak in the 150–220 ms interval. The magnitude of that difference is normally larger for negative than for positive feedback (Hajcak et al., 2005a,b), so a differential FRN score (henceforth, simply FRN score) was computed as the difference between the FRN for PF and the FRN for NF.

Statistical analyses were carried out on FRN scores for Fz and FCz electrodes. The Pz channel was also included to test whether observed effects could be attributed to P3 (as it has been observed that P3 amplitude can affect FRN, and that it is affected by contingency changes; Barcelo et al., 2006). P3 was thus extracted from Fz, FCz, and Pz. However, the time window in which P3 is normally observed includes, in our task, activity evoked by the following trial in the sequence. In order to avoid signal contamination, we carried out P3 analyses on a score computed as the average amplitude for the last 50 ms of each segment referred to the average amplitude during the immediately preceding 100 ms time window (see Chase et al., 2011, for a similar procedure). As we did with the FRN, a differential P3 score (henceforth simply P3 score) was computed as the difference between the P3 scores for NF and PF.

Feedback-related negativity scores were submitted to a 3 (group: CDI, PG, and HC) \times 2 (channel: Fz, FCz) repeated-measures analysis of variance. P3 scores were submitted to a 3 (group) \times 3 (channel: Fz, FCz, Pz) repeated-measures ANOVA. The Bonferroni procedure was used to correct for multiple comparisons. A 0.05 p -level was used for all the statistical decisions. Two participants from the PG group and two from the HC group were excluded from the analysis due to equipment malfunctioning.

BRAIN LOCALIZATION

Standardized low-resolution electromagnetic tomography (sLORETA) was used for estimating the 3D cortical distribution of current density underlying scalp activity. sLORETA, computations were done using the MNI152 template, with the 3D space solution restricted to cortical gray matter, according to the probabilistic Talairach atlas. The cortical gray matter is partitioned in 6239 voxels at 5 mm spatial resolution. Brodmann anatomical labels are reported using MNI space. Standardized sLORETA current source densities with no regularization method were obtained from 61 channels (after recovering FCz) for each participant in each condition, and for each time point in each feedback condition. A discussion on the technical details of sLORETA and, specifically, on the necessary restrictions for a viable solution to the inverse problem can be found in Pascual-Marqui (2002).

The identification of the sources with a differential involvement in the generation of FRN across groups followed the rationale recently described by Catena et al. (2012 see also Silton et al., 2009; Torres et al., 2013). A significant correlation across participants between current source density (i.e., estimated activation) at a certain voxel and the magnitude of FRN observed at FCz can be interpreted as indicative of the involvement of such a voxel in the generation of FRN. In other words, the correlations between voxelwise current densities and FRN magnitudes can be used to identify the brain areas involved in the generation of FRN.

Under such an assumption, brain localization analysis was carried out according to the following steps: first, a representative measure of the activation of each voxel for the FRN interval was computed, by averaging voxel activations across the 220–330 post-feedback time window. Second, we computed the correlation (across participants) between that averaged current density and the magnitude of the FRN effect, for each voxel. And third, those areas in which at least 10 voxels were found to significantly correlate with the FRN score were singled out as candidate areas with a functional role in its generation.

RESULTS

BEHAVIORAL RESULTS

PRLT: Decision making

The main dependent variable in the global analysis was the number of correct choices (clicks on the highly rewarded option) per each 10-trial block and each 40-trial phase. **Figure 1** shows the mean number of correct choices per phase, block, and group. The mixed ANOVA, with phase (1–4) and block (1–4) as within-group factors, and group (HC, PG, CDI) as the between-group factor, yielded a significant block \times phase \times group interaction, $F(18, 549) = 1.86$, $MSE = 2.347$, $p = 0.03$, $\eta^2 = 0.06$. As expected, there were also significant effects of phase, $F(3, 183) = 3.06$, $MSE = 8.88$, $p = 0.04$, $\eta^2 = 0.05$, block, $F(3, 183) = 85.62$, $MSE = 3.68$, $p < 0.01$, $\eta^2 = 0.58$, and phase \times block, $F(9, 549) = 5.416$, $MSE = 2.35$, $p < 0.03$, $\eta^2 = 0.08$, showing within-phase learning effects, and between-phases reversal costs in all groups.

Despite the significant three-way interaction, differences across groups were not significant for any phase and block of the task according to Bonferroni *post hoc* tests. Applying a non-corrected *post hoc* LSD approach, differences between HC and CDI were observed on the second block of the first phase, $t(41) = 2.40$, $p = 0.02$, and between PG and CDI on blocks 2, $t(39) = 2.07$, $p = 0.04$, and 3, $t(39) = 2.12$, $p = 0.04$, of Phase 3.

Results were clearer after taking monthly use and severity scores into account. Given that monthly use and severity refer to different

addictive behaviors for PG and CDI, and that HC participants have neither monthly use nor severity scores, we carried out separate repeated-measures ANCOVAs for PG and CDI groups, using severity and monthly amount as covariates, and the number of correct choices in the first block of each phase, and the number of correct choices in the last two blocks of each phase as dependent measures (see Statistical Analysis and footnote 1).

In search of reversal cost effects, we carried out separate ANCOVAs for the two clinical groups, with the number of correct choices in the first block of each phase as dependent measure. In the PG group, the analysis yielded a main effect of monthly amount gambled, $F(1, 18) = 4.42$, $MSE = 5.66$, $p = 0.05$, $\eta^2 = 0.19$. No other marginal or interactive effects involving monthly amount gambled or gambling severity were close to significance (minimum $p = 0.16$). In the CDI group however, an identical analysis carried out with cocaine monthly use and cocaine use severity as covariates, did not yield any main or interactive significant effect (minimum $p = 0.44$).

Similarly, two ANCOVAs were carried out with asymptotic learning scores as the dependent measure. In the PG group, the analysis did not yield any marginal or interactive significant effect (minimum $p = 0.18$). In the CDI group, on the contrary, the analysis yielded now a significant main effect of cocaine severity, $F(1, 17) = 4.71$, $MSE = 4.71$, $p = 0.04$, $\eta^2 = 0.22$.

Table 2 shows where the effects yielded by these ANCOVAs originate. The table displays partial correlations – in the PG group – between monthly amount gambled and the number of correct choices in block 1 (phases 1–4), with gambling severity as variable of control; and – in the CDI group – between cocaine use severity and the asymptotic learning measure, with cocaine monthly use as control variable. The effect of monthly amount gambled on first block correct choices was actually restricted to phases 2 and 4, namely, to the first and the third reversals of the task. Cocaine use severity, in turn, exerted its effect on phases 3 and 4.

DECISION LATENCIES

Finally, we analyzed the effects of group, monthly use, and severity on decision latencies. The main dependent measure was the mean decision latency per phase and block. The group (HC, PG, CDI) \times phase (1–4) \times block (1–4) ANOVA did not show any significant marginal or interactive effect of group (minimum $p = 0.328$).

The ANCOVA for the PG group, with block and phase as within-group variables, and monthly amount gambled and severity scores as continuous predictors, showed significant main effects of the monthly amount gambled, $F(1, 18) = 5.66$, $p = 0.03$, $\eta^2 = 0.24$ and gambling severity, $F(1, 18) = 4.81$, $p = 0.04$, $\eta^2 = 0.21$. No other marginal or interactive effect was close to significance (minimum $p = 0.24$). An analogous ANCOVA on CDI decision latencies, and cocaine monthly use and cocaine severity as covariates, did not show any significant effect of monthly use, or severity (all $p > 0.10$).

Figure 2 shows a graphical depiction of the monthly amount effect observed in the PG group (coefficients represent partial correlations between monthly amount gambled and decision latency for each phase and block, computed while controlling for

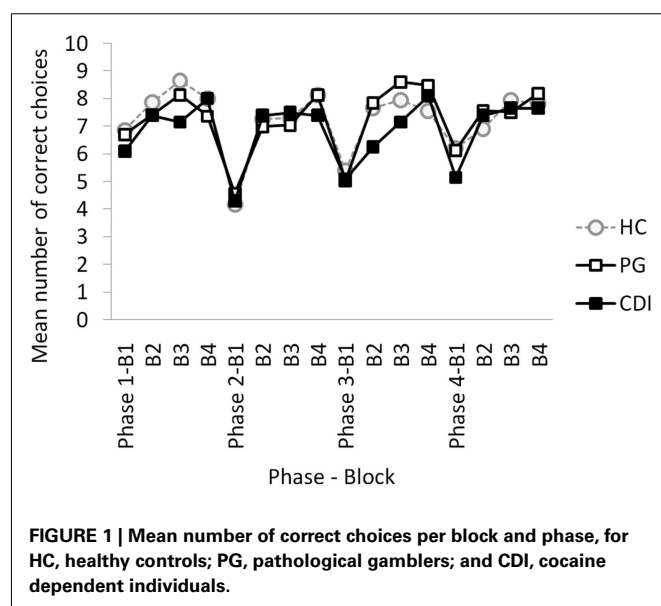
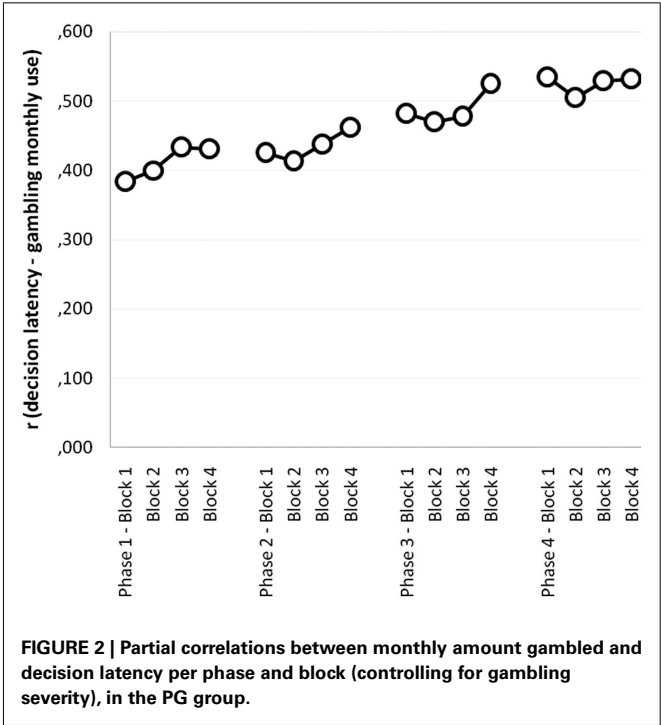


Table 2 | Partial correlations between monthly amount gambled and number of correct choices in block 1 (phases 1–4), and between cocaine use severity and number of correct choices in blocks 3/4 (phases 1–4).

	Phase 1		Phase 2		Phase 3		Phase 4	
	partial <i>r</i>	α	partial <i>r</i>	α	partial <i>r</i>	α	partial <i>r</i>	α
Gambling MU – correct choices (block 1)	–0.12	0.63	–0.50	0.03*	–0.17	0.49	–0.46	0.05*
Cocaine severity – correct choices (blocks 3/4)	–0.31	0.21	–0.24	0.33	–0.69	0.01*	–0.51	0.03*

Correlations larger than 0.46 (in absolute terms) are bilaterally significant. Values in bold stand for statistically significant correlations and their corresponding *p*-values.



severity). Consistently across the task, the monthly amount gambled positively covaried with decision latency, which means that the intensity of addiction slowed decisions down (all correlations above 0.445 are bilaterally significant).

In summary, the clinical groups seem to show different learning dynamics in the reversal learning task when compared to matched controls. However, such differences cannot be fully characterized if addictions are not considered from an idiosyncratic point of view (i.e., taking chronic and acute exposure into account).

Gambling intensity, measured as the monthly amount gambled, emerges as a powerful mediator of learning-driven decision-making: heavier gamblers tend to show signs of enhanced reversal cost, and, additionally, tend to make significantly slower predictions. Increased latency in decision-making tasks is customarily interpreted as a sign of decisional difficulty (Spinoza-Varas and Watson, 1994), although this measure has been paid no attention at all in reversal learning studies.

The severity of gambling did not exert any significant effect on reversal cost, which implies that the effect of gambling on that

particular aspect of reversal learning is not cumulative. On the other hand, cocaine use severity, but not intensity, interfered with asymptotic-level decision making. In this case, the potential effects on decision latencies were negligible.

EEG RESULTS
Feedback-related negativity

Figure 3 displays ERP waveforms for each group in each feedback condition. The 3 (group: HC, PG, CDI) × 2 (channel: Fz, FCz) mixed ANOVA on the FRN score yielded significant main effects of group, $F(2, 57) = 4.04$, $MSE = 1.39$, $p < 0.03$, $\eta^2 = 0.12$, and Channel, $F(1, 57) = 20.19$, $MSE = 0.45$, $p < 0.01$, $\eta^2 = 0.26$, being the largest FRN score observed at FCz. There was no interaction between the two factors, $F(2, 57) = 0.51$. Bonferroni-corrected *post hoc* comparisons showed that the FRN score was larger for HC than for PG ($p = 0.02$). No other effects were significant. With regard to P3, there was a theoretically irrelevant effect of channel, $F(2, 116) = 3.19$, $MSE = 1.17$, $p < 0.05$, $\eta^2 = 0.05$, but both the group effect, $F(2, 58) = 0.92$, and the group × channel interaction, $F(4, 58) = 0.97$, were very far from the significance level.

Source location

Using the bootstrapping approach (included in the sLoreta package) we observed several right hemisphere clusters of voxels that significantly correlated with FRN scores in the control group (Table 3): the inferior (BA46) and middle (BA9 and BA10) frontal gyri, the insula (BA13), and the posterior cingulate gyrus (BA23). As noted in the Section “Materials and Methods,” we take this as evidence of the involvement of these areas in the generation of FRN in normal conditions (please note that negative correlations imply that the larger the activation in these areas, the larger – in absolute values – the FRN score). These areas were established as regions of interest to detect differences between the clinical groups.

Feedback-related negativity-current density correlations in those same areas for the two clinical groups are reported in Table 4. Not surprisingly, those correlations differed from the ones in the control group ($p = 0.14, 0.06, 0.19, 0.06, 0.04$ for the HC versus CDI contrasts; and $p < 0.01$ for all HC versus PG contrasts across areas). The deviation was thus larger for PG than for CDI. Density-FRN correlations in the CDI group, although lower, were in the same direction than the ones observed in the HC group. Correlations in the PG group were mostly in the opposite direction, and (according to the Bonferroni correction) significantly differed from CDI’s in BA9, BA10, BA13, and BA23 (Table 4, rightmost column).

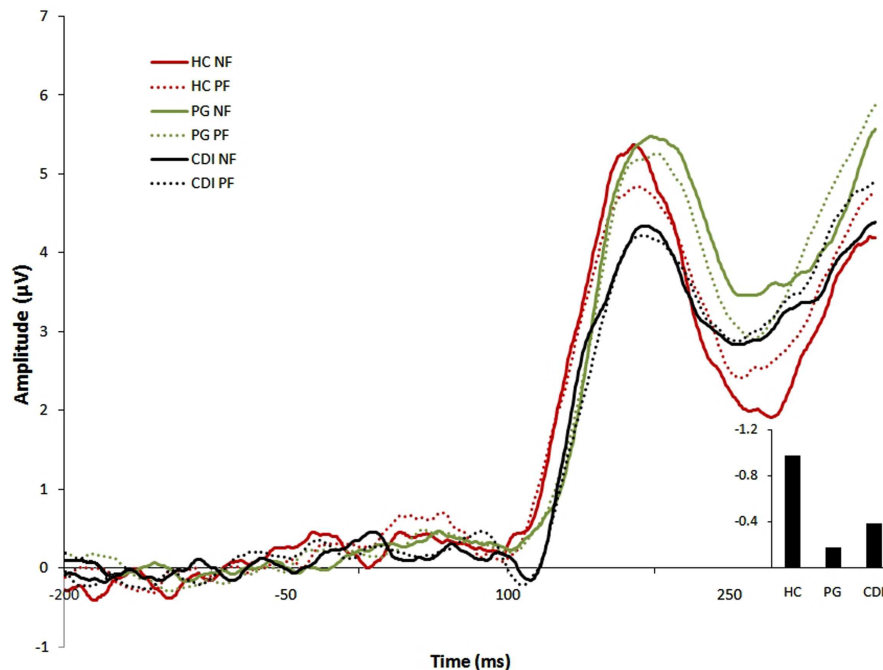


FIGURE 3 | ERP waveforms for each group in each feedback condition. Right bottom panel: differential FRN effect for the three groups. All amplitudes are in μV . HC, Healthy controls; PG, Pathological gamblers; CDI, Cocaine dependent individuals; NF, Negative feedback; PF, Positive feedback.

Table 3 | Brain areas significantly correlated to the FRN score in the control group.

Lobe	Structure	BA	k	X	Y	Z	CSD-FRN Correlation
Frontal	Middle frontal	9	10	20	35	20	-0.70
Frontal	Middle frontal	10	16	35	40	15	-0.75
Sub-lobar	Insula	13	9	40	15	15	-0.71
Limbic	Posterior cingulate	23	4	5	-40	25	-0.75
Frontal	Inferior frontal gyrus	46	16	35	35	15	-0.74

BA, Brodmann area; k, cluster size in voxels; X, Y, and Z are in MNI space.

DISCUSSION

Our first research aim was to check for the existence of anomalies in reversal learning in two groups of CDI and PG, when compared against HC. Such anomalies have been only partially corroborated. The group \times block \times phase interaction effect on correct decisions indicates that learning progressed differently in the three groups. Such a difference is, however, subtle. In specific points of the task, individuals in the CDI group performed worse than controls (phase 1, block 2), or than PG participants (phase 3, blocks 2 and 3). The observation that cocaine addicts are globally (although only slightly) more hampered than other groups in PRLT performance is fully coincident with the results reported by Fernández-Serrano et al. (2012). Additionally, the significance of such a difference is strengthened by the existence of differences at the electroencephalographic level, as discussed later.

Table 4 | FRN-current density correlation coefficients for the key areas involved in FRN generation (as detected in controls), and significance of Bonferroni-corrected contrasts between correlation coefficients across groups (PG, Pathological gamblers; CDI, Cocaine dependent individuals).

Structure	BA	CDI	PG	p (PG versus CDI)
Middle frontal	9	-0.47	0.35	<0.01*
Middle frontal	10	-0.43	0.34	0.01*
Insula	13	-0.53	0.72	<0.01*
Posterior cingulate	23	-0.42	0.78	<0.01*
Inferior frontal gyrus	46	-0.36	0.07	0.10°

°Non-significant; * $p < 0.05$. p values in bold correspond to significant differences between PGs and CDIs.

In relation to our second research aim, group analyses demonstrate that the difficulty to interpret between-group PRLT performance differences can be due – at least in part – to differences within the clinical groups. On the one hand, asymptotic learning, as measured by the averaged number of correct choices in the two last blocks of each phase, was significantly affected by cocaine severity, that is, by the estimated cumulative exposure to cocaine during the course of the addictive process.

On the other hand, reversal costs (as observed in phases 2 and 4; see Table 2) were specifically associated to gambling intensity, namely, to the averaged amount gambled per unit of time. Those gamblers who spend more money in gambling activities also tend to show larger reversal costs. This is compatible with

Leeman and Potenza's (2012) proposal that there is a privileged link between gambling and learning inflexibility². Additionally, we provide evidence that gambling, but not cocaine use, slows decisions down. Increased latency in decision-making tasks is customarily interpreted as a sign of decision difficulty (Spinoza-Varas and Watson, 1994). So, this finding supports the idea that gambling is specifically linked to the decisional aspects of reversal learning. This association between gambling intensity, reversal cost, and increased decision difficulty probably deserves further research.

The fact that the monthly amount gambled (i.e., gambling intensity), but not gambling severity, exerts a significant impact on phase-by-phase first block correct choices implies that the gambling effect on such measure is not cumulative, that is, not due to practice with gambling scenarios, or chronic gambling-induced neuroadaptation. In other words, it is unlikely that increased reversal costs are attributable to practice or sensitization. Conversely, the evidence that cocaine use severity, but not monthly use (i.e., intensity), exerts an impact on asymptotic reversal learning seems to prove that the cumulative effect of cocaine exposure (neurotoxicity) is exerted on a different component of reversal learning performance, not necessarily involving learning inflexibility. Relatedly, Albein-Urios et al. (2012a) and Torres et al. (2013) have recently demonstrated that some other well-known neuropsychological anomalies observed in CDI (e.g., working memory and motor inhibition deficits) are also attributable to cocaine neurotoxic effects.

Still, the interpretation of our PRLT behavioral results requires some further considerations. Firstly, recent evidence (van Holst et al., 2010; Shaffer and Martin, 2011) shows that there exist non-trivial psychological differences underlying differential preferences for low-rate high-stakes gambling modalities (casino games, sport bets), versus high-rate low-stakes ones (Video-lottery terminals, slot machines). Our sample mostly consisted of male slot machine gamblers, and was not large enough to segregate these two categories. At this moment, the role of gambling preferences in PRLT performance remains open. And secondly, in most implementations of the PRLT reversals do not occur at fixed times, but when the participant have reached a pre-established learning criterion (for example, five-correct choices in a row; Franken et al., 2008). Performance is then assessed as the total number of reversals during the task, the total number of incorrect choices, the mean number of trials-to-criterion, or the mean number of incorrect choices in a row after a reversal (perseverative series; Franken et al., 2008; de Ruiter et al., 2009; Camchong et al., 2011; Lucantonio et al., 2012). These measures are customarily interpreted as measures of reversal cost or reversal learning (in)flexibility.

In our version of the task, phase length was fixed (40 trials), to ensure comparability of learning curves across groups in global analyses (and consequently to make the distinction between reversal cost and asymptotic learning possible). In addition, our main flexibility-related dependent measure was the number of correct

choices in phase-by-phase first blocks. The reason underlying the use of such measure (instead of the more common perseverative series mean length), is strictly statistical: given that the PRLT provides probabilistic false feedback (punishment for a correct choice) the length of error series tends to be very variable within each participant, depending on the particular ordering of trials in the series. Most PRLT implementations do not warrant asymptotic learning, but allow for a high number of reversals, so variability can be reduced by means of averaging. In our case, the task ensures asymptotic learning in each phase (see footnote 1), but contains only three contingency reversal points, and thus a more stable measure is required. This particularity, however, does not compromise the interpretation of the measure in terms of reversal cost/learning inflexibility (at least for blocks 2–4).

Our third and last research aim was to analyze electroencephalographic differences between-groups (and, particularly, the differences between the two clinical groups) with regard to their response to feedback during reversal learning. We have observed abnormal feedback-evoked cortical activity in the PG group. If we take the magnitude and sign of the differential FRN score in the control group as a reference of normality, the deviation from that reference was maximal for gamblers (the FRN was visually smaller for CDI than for HC, but CDI did not statistically differ either from HC or from the PG).

According to Hajcak et al. (2005a,b), the FRN is mainly elicited by unexpected negative outcomes (see also Holroyd et al., 2004), and reflects the binary evaluation of good versus bad outcomes. If that interpretation is taken as correct, it implies that gamblers are particularly hampered to adequately ponder the impact of NF. Consequently, we can assume that they are also hampered to learn to make decisions on the individual history of losses. This is fully coincident with the finding that gambling slows decisions down, and also with our separate finding that recreational gamblers, at non-pathological levels, are less sensitive to losses than non-gamblers (Torres et al., submitted).

Results regarding source location point in the same direction. In accordance with the results we have obtained with HC, Hampshire et al. (2012) found several areas to be particularly active when reversal events were compared against other switch events (i.e., changes in the set of stimuli). These areas included the most posterior extent of the right inferior frontal gyrus, extending into the anterior insula, and the frontopolar portion of the middle frontal gyrus (see also Mitchell et al., 2008), and were more active when NF led to a change in the response pattern. In that study, the dorso-lateral prefrontal cortex was also found to be involved in reversal events, whereas other studies have attributed to it more general higher-order executive functions involving attention (Remijne et al., 2005), and coordination of search behavior (Hampshire and Owen, 2006). In any case, this set of anatomical areas is almost fully coincident with the ones found to be involved in the generation of the FRN in HC in the present study.

The only discordance between the present and previous results seems to be the involvement of the posterior portion of the cingulate gyrus in the generation of FRN. D'Cruz et al. (2011), and Robinson et al. (2010) found the activation of posterior cingulate cortex after positive feedback in the reversal learning task

²On the contrary, Ersche et al. (2008) made a very careful evaluation of a number of performance indices and found true reversal cost in cocaine users (but not in amphetamine users). Importantly, that effect was found in *current* cocaine users, but not in abstinent ex-addicts. At difference with gambling, the existence of reversal costs associated to chronic effects of cocaine addiction remains undemonstrated.

to depend on whether it was expected or not. In a work by Gläscher et al. (2009), activation in the same area was associated to the experienced value of the chosen option. Relatedly, Nashiro et al. (2012) found it to be more active when feedback was emotional than when it was neutral. And finally, a study by Albein-Urios et al. (2012b), found it to be involved in regulation of negative emotions. So, it is plausible that variability in the magnitude of FRN is associated to emotional aspects of feedback valuation.

Most importantly when these areas were taken as regions of interest for the clinical groups, PG strikingly differed from CDI. Although, as noted above, the magnitude of FRN did not differ between the clinical groups, sLORETA analyses unveiled differences in the involvement of these areas in FRN generation. There is ample evidence that addictive processes are associated to abnormal response to feedback and abnormal activation of prefrontal and orbitofrontal areas (see Schoenbaum et al., 2006, for a review). However, our results provide the first direct neuroanatomical evidence in favor of Leeman and Potenza's (2012) proposal that reversal learning deficits are particularly severe in gamblers (when compared against other populations of addicted individuals).

Despite its several specific strengths (the careful selection of participants, the close matching between-groups in intellectual functioning and sociodemographic variables and the absence of comorbidities in them), this study has also some limitations that are worth mentioning. One of them is undoubtedly the absence of enough trials in the reversal learning task to track changes in the FRN across the task, and, more specifically, to clearly separate between reversal errors (those occurring in the first trials after reversal points) and errors spontaneously occurring during other parts of the task. We have shown that learning dynamics are behaviorally relevant; further research is needed to describe in a similarly detailed way the evolving changes in cortical activity occurring in parallel with such learning dynamics. A second limitation is the impossibility to separate gamblers with preferences for different games of chance. Third, despite the careful selection of participants, it is virtually impossible to match groups in every potentially relevant factor. Specifically, potentially addictive behaviors tend to show complex correlation patterns. In our case, cocaine users were also more likely to use alcohol and cannabis than gamblers and controls. Although cannabis and alcohol use did not exert any

direct effect on PRLT performance or cortical activity, the possibility exists that these drugs modulated the chronic effects of gambling/cocaine. This limitation is common to virtually all studies in which the group comparison methodology is used. And finally, despite the recording of drug/gambling exposure measures, group comparison studies are less conclusive than prospective studies, with regard to the possibility to establish directional causal links between neuropsychological abnormalities and addictive behaviors. These four potential weaknesses warrant further research.

FINAL REMARKS

To date, results regarding reversal learning deficits in addicts have been elusive. This work confirms previous proposals that feedback-based instrumental learning is more inflexible in pathological gambling than in other forms of addiction, such as cocaine dependence. At the same time, however, it raises important questions about the causes of such inflexibility and the role within-group variability in clinical samples. At the behavioral level, the main findings of this research point out that gambling severity slows decisions and increases reversal cost, whereas cocaine addiction affects asymptotic scores in reversal learning tasks. More importantly, psychophysiological and neuroanatomical data provided the first direct evidence that reversal learning deficits in gamblers differ from drug addicts' and are related to abnormal activity in specific prefrontal and orbitofrontal areas.

ACKNOWLEDGMENTS

The research presented here has been funded by grants from the Ministerio de Ciencia e Innovación, MICINN (Spain), for Ana Torres and José C. Perales (ref. # PSI2009-13133), and Andrés Catena and Antonio Maldonado (ref. # PSI2009-12217); by a Junta de Andalucía (Spain) grant (ref. # PB09-SEJ4752) for Antonio Cándido; and by a RETICS (Redes Temáticas de Investigación Cooperativa en Salud) subprogramme grant (ref. # RD12/0028/0017) from the Ministerio de Sanidad, Políticas Sociales e Igualdad for José C. Perales. We would like to thank *Proyecto Hombre's* Málaga and Granada centers, and AGRAJER (Asociación Granadina de Jugadores de Azar en Rehabilitación) for their invaluable and disinterested collaboration.

REFERENCES

- Albein-Urios, N., Martínez-González, J. M., Lozano, O., Clark, L., and Verdejo-García, A. (2012a). Comparison of impulsivity and working memory in cocaine addiction and pathological gambling: implications for cocaine-induced neurotoxicity. *Drug Alcohol Depend.* 126, 112.
- Albein-Urios, N., Verdejo-Román, J., Asensio, S., Soriano-Mas, C., Martínez-González, J. M., and Verdejo-García, A. (2012b). Re-appraisal of negative emotions in cocaine dependence: dysfunctional cortic limbic activation and connectivity. *Addict. Biol.* doi:10.1111/j.1369-1600.2012.00497.x
- Barcelo, F., Escera, C., Corral, M. J., and Perianez, J. A. (2006). Task switching and novelty processing activate a common neural network for cognitive control. *J. Cogn. Neurosci.* 18, 1734–1748.
- Belouchrani, A., Abed Meraim, K., Cardoso, J.-F., and Moulines, E. (1993). "Second-order blind separation of temporally correlated sources," in *Proceeding of the International Conference on Digital Signal Processing*, Cyprus, 346–351.
- Belouchrani, A., Abed Meraim, K., Cardoso, J.-F., and Moulines, E. (1997). A blind source separation technique based on second order statistics. *IEEE Trans. Signal Process.* 45, 434–444.
- Bismark, A. W., Hajcak, G., Whitworth, N. M., and Allen, J. J. B. (2012). The role of outcome expectations in the generation of the feedback-related negativity. *Psychophysiology* 50, 125–133.
- Camchong, J., MacDonald, A. W., Nelson, B., Bell, C., Mueller, B. A., Specker, S., et al. (2011). Frontal hyperconnectivity related to discounting and reversal learning in cocaine subjects. *Biol. Psychiatry* 69, 1117–1123.
- Cardoso, J. F., and Souloumiac, A. (1996). Jacobi angles for simultaneous diagonalization. *SIAM J. Mat. Anal. Appl.* 17, 161–164.
- Catena, A., Perales, J. C., Megías, A., Cándido, A., Jara, E., and Maldonado, A. (2012). The brain network of expectancy and uncertainty processing. *PLoS ONE* 7:e40252. doi:10.1371/journal.pone.0040252

- Chase, H. W., Swainson, R., Durham, L., Benham, L., and Cools, R. (2011). Feedback-related negativity codes prediction error but not behavioral adjustment during probabilistic reversal learning. *J. Cogn. Neurosci.* 23, 936–946.
- Clark, L., Cools, R., and Robbins, T. W. (2004). The neuropsychology of ventral prefrontal cortex: decision-making and reversal learning. *Brain Cogn.* 55, 41–53.
- Clarke, H. F., Robbins, T. W., and Roberts, A. C. (2008). Lesions of the medial striatum in monkeys produce perseverative impairments during reversal learning similar to those produced by lesions of the orbitofrontal cortex. *J. Neurosci.* 28, 10972–10982.
- Clarke, H. F., Walker, S. C., Crofts, H. S., Dalley, J. W., Robbins, T. W., and Roberts, A. C. (2005). Prefrontal serotonin depletion affects reversal learning but not attentional set shifting. *J. Neurosci.* 25, 532–538.
- D'Cruz, A. M., Ragozzino, M. E., Mosconi, M. W., Pavuluri, M. N., and Sweeney, J. A. (2011). Human reversal learning under conditions of certain versus uncertain outcomes. *Neuroimage* 56, 315–322.
- de Ruiter, M. B., Veltman, D. J., Goudriaan, A. E., Oosterlaan, J., Sjoerds, Z., and Van den Brink, W. (2009). Response perseveration and ventral prefrontal sensitivity to reward and punishment in male problem gamblers and smokers. *Neuropsychopharmacology* 34, 1027–1038.
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *J. Neurosci. Methods* 134, 9–21.
- Ersche, K. D., Roiser, J. P., Robbins, T. W., and Sahakian, B. J. (2008). Chronic cocaine but not chronic amphetamine use is associated with perseverative responding in humans. *Psychopharmacology (Berl.)* 197, 421–431.
- Everitt, B. J., Dickinson, A., and Robbins, T. W. (2001). The neuropsychological basis of addictive behavior. *Brain Res. Rev.* 36, 129–138.
- Fernández-Serrano, M. J., Perales, J. C., Moreno-López, L., Pérez-García, M., and Verdejo-García, A. (2012). Neuropsychological profiling of impulsivity and compulsivity in cocaine dependent individuals. *Psychopharmacology (Berl.)* 219, 673–683.
- First, M. B., Spitzer, R. L., Gibbon, M., and Williams, J. B. W. (1997). *Structured Clinical Interview for DSM-IV Axis I Disorders (SCID I)*. New York: Biometric Research Department.
- Franken, I. H. A., Van Strien, J. W., Nijs, I., and Muris, P. (2008). Impulsivity is associated with behavioral decision-making deficits. *Psychiatry Res.* 158, 155–163.
- Gläscher, J., Hampton, A. N., and O'Doherty, J. P. (2009). Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cereb. Cortex* 19, 483–495.
- Hajcak, G., Holroyd, C. B., Moser, J. S., and Simon, R. F. (2005a). Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology* 42, 161–170.
- Hajcak, G., Moser, J. S., Yeung, N., and Simons, R. F. (2005b). On the ERN and the significance of errors. *Psychophysiology* 42, 151–160.
- Hall, G. W., Carriero, N. J., Takushi, R. Y., Montoya, I. D., Preston, K. L., David, A., et al. (2000). Pathological gambling among cocaine-dependent outpatients. *Am. J. Psychiatry* 157, 1127–1133.
- Hampshire, A., and Owen, A. M. (2006). Fractionating attentional control using event-related fMRI. *Cereb. Cortex* 16, 1679–1689.
- Hampshire, A. D., Chaudhry, A. M., Owen, A. M., and Roberts, A. C. (2012). Dissociable roles for lateral orbitofrontal cortex and lateral prefrontal cortex during preference driven reversal learning. *Neuroimage* 59, 4102–4112.
- Hastie, R., and Dawes, R. M. (2009). *Rational Choice in an Uncertain World*, 2nd Edn. London: Sage Publications.
- Holroyd, C. B., Larsen, J. T., and Cole, J. D. (2004). Context dependence of the event-related potential associated with reward and punishment. *Psychophysiology* 41, 245–253.
- Izquierdo, A., and Jentsch, J. D. (2012). Reversal learning as a measure of impulsive and compulsive behavior in addictions. *Psychopharmacology (Berl.)* 219, 607–620.
- Jacobs, D. F., Marston, A., Singer, R., Widaman, K., Little, T., and Veizades, J. (1989). Children of problem gamblers. *J. Gambl. Behav.* 5, 261–268.
- Karayanidis, F., Robaey, P., Bourassa, M., de Koning, D., Geoffroy, G., and Pelletier, G. (2000). ERP differences in visual attention processing between attention-deficit hyperactivity disorder and control boys in the absence of performance differences. *Psychophysiology* 37, 319–333.
- Kaufman, A., and Kaufman, N. (1990). *Kaufman Brief Intelligence Test*. Circle Pines, MN: American Guidance Service.
- Leeman, R. F., and Potenza, M. N. (2012). Similarities and differences between pathological gambling and substance use disorders: a focus on impulsivity. *Psychopharmacology (Berl.)* 219, 469–490.
- Lucantonio, F., Stalnaker, T. A., Shaham, Y., Niv, Y., and Schoenbaum, G. (2012). The impact of orbitofrontal dysfunction on cocaine addiction. *Nat. Neurosci.* 15, 358–366.
- Mitchell, D. G. V., Richell, R. A., Pine, D. S., and Blair, R. J. R. (2008). The contribution of ventrolateral and dorsolateral prefrontal cortex to response reversal. *Behav. Brain Res.* 187, 80–87.
- Mognon, A., Bruzzone, L., Jovicich, J., and Buiatti, M. (2011). ADJUST: an automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology* 48, 229–240.
- Nashiro, K., Sakaki, M., Nga, L., and Mather, M. (2012). Differential brain activity during emotional versus nonemotional reversal learning. *J. Cogn. Neurosci.* 24, 1794–1805.
- Pascual-Marqui, R. D. (2002). Standardized low resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find. Exp. Clin. Pharmacol.* 24, 5–12.
- Potenza, M. N. (2008). The neurobiology of pathological gambling and drug addiction: an overview and new findings. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 3181–3189.
- Redish, A. D., Jensen, S., Johnson, A., and Kurth-Nelson, Z. (2007). Reconciling reinforcement learning models with behavioral extinction and renewal: implications for addiction, relapse, and problem gambling. *Psychol. Rev.* 114, 784–805.
- Reminse, P. L., Nielen, M. M. A., Uylings, H. B. M., and Veltman, D. J. (2005). Neural correlates of reversal learning task and affectively neural baseline: an event-related fMRI study. *Neuroimage* 23, 752–763.
- Robinson, O. J., Frank, M. J., Sahakian, B. J., and Cools, R. (2010). Dissociable responses to punishment in distinct striatal regions during reversal learning. *Neuroimage* 51, 1459.
- Robinson, T., and Berridge, K. (2003). Addiction. *Annu. Rev. Psychol.* 54, 25–53.
- Schoenbaum, G., Roesch, M. R., and Stalnaker, T. A. (2006). Orbitofrontal cortex, decision-making and drug addiction. *Trends Neurosci.* 29, 116–124.
- Shaffer, H. J., and Martin, R. (2011). Disordered gambling: etiology, trajectory and clinical considerations. *Annu. Rev. Clin. Psychol.* 7, 483–510.
- Silton, R. L., Heller, W., Towers, D. N., Engels, A. S., Spielberg, J. M., Edgar, J. C., et al. (2009). The time course of activity in dorsolateral prefrontal cortex and anterior cingulate cortex during top-down attentional control. *Neuroimage* 50, 1292–1302.
- Spinoza-Varas, B., and Watson, C. S. (1994). Effects of decision criterion on response latencies of binary decisions. *Percept. Psychophys.* 55, 190–203.
- Steinberg, M., Kosten, T., and Rounsaville, B. (1992). Cocaine abuse and pathological gambling. *Am. J. Addict.* 30, 929–962.
- Swainson, R., Rogers, R. D., Sahakian, B. J., Summers, B. A., Polkey, C. E., and Robbins, T. W. (2000). Probabilistic learning and reversal deficits in patients with Parkinson's disease or frontal or temporal lobe lesions: possible adverse effects of dopaminergic medication. *Neuropsychologia* 38, 596–612.
- Tang, A. C., Sutherland, M. T., and McKinney, C. J. (2004). Validation of SOBI components from high-density EEG. *Neuroimage* 25, 539–553.
- Torres, A., Catena, A., Megías, A., Maldonado, A., Cándido, A., Verdejo-García, A., et al. (2013). Emotional and non-emotional pathways to impulsive behavior and addiction. *Front. Hum. Neurosci.* 7:43. doi:10.3389/fnhum.2013.00043 (in press).
- Tsushima, A., Doll, B. B., and Fellows, L. K. (2010). Beyond reversal: a critical role for human orbitofrontal cortex in flexible learning from probabilistic feedback. *J. Neurosci.* 30, 16868–16875.
- van Holst, R. J., Van den Brink, W., Veltman, D. J., and Goudriaan, A. E. (2010). Why gamblers fail to win: a review of cognitive and neuroimaging findings in pathological gambling. *Neurosci. Biobehav. Rev.* 34, 87–107.
- Verdejo-García, A., López-Torrecillas, F., Aguilar de Arcos, F., and Pérez-García, M. (2005). Differential effects of MDMA, cocaine, and cannabis use severity on

- distinctive components of the executive functions in polysubstance users, a multiple regression analysis. *Addict. Behav.* 30, 89–101.
- Verdejo-García, A., Sánchez-Fernández, M. M., Alonso-Maroto, L. M., Fernández-Calderón, F., Perales, J. C., Lozano, O., et al. (2010). Impulsivity and executive functions in polysubstance-using rave attenders. *Psychopharmacology (Berl.)* 210, 377–392.
- Vosburg, S. K., Haney, M., Rubin, E., and Foltin, R. W. (2010). Using a novel alternative to drug choice in a human laboratory model of a cocaine binge: a game of chance. *Drug Alcohol Depend.* 110, 144–150.
- 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: 26 November 2012; accepted: 25 February 2013; published online: 18 March 2013.
- Citation: Torres A, Catena A, Cándido A, Maldonado A, Megías A and Perales JC (2013) Cocaine dependent individuals and gamblers present different associative learning anomalies in feedback-driven decision making: a behavioral and ERP study. *Front. Psychol.* 4:122. doi:10.3389/fpsyg.2013.00122
- This article was submitted to *Frontiers in Personality Science and Individual Differences*, a specialty of *Frontiers in Psychology*.
- Copyright © 2013 Torres, Catena, Cándido, Maldonado, Megías and Perales. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.

APPENDIX

The IRAB provides information on frequency of drug use, amount per drug use episode, and duration of use. The combination of frequency and amount per episode allows for an estimate of the amount used per unit of time. Duration of use is customarily expressed in months, whereas amount units vary across substances. Severity is computed as the monthly use \times duration product (note that severity scores can reach extremely high values, so correction measures are taken for analysis; e.g., translating them into ranks, computing them from standardized duration and monthly use scores).

Table A1 displays the observed IRAB results for regular use of cannabis, tobacco, alcohol, and cocaine, as well as the equivalent measures of gambling. Only those drugs of abuse used once a month by at least 15% of the individuals in any of the samples have been included. Mean consumption of all other drugs included in the IRAB (MDMA, amphetamine, methamphetamine, heroine, benzodiazepines, hallucinogens) is negligible. The number of illegal drugs used at least once in the whole lifetime, however, provides extra information about the different drug use patterns in the three groups.

ANCOVAs FOR THE POTENTIAL EFFECTS OF CANNABIS AND ALCOHOL USE ON RELEVANT DEPENDENT MEASURES

Either monthly use of cannabis and alcohol, or severity of alcohol/cannabis consumption could have been used as covariates for the analyses reported in this appendix. However, severity and monthly use scores (once translated into ranks) were virtually colinear (severity-monthly use correlation was $r=0.754$ for alcohol, and $r=0.937$ for cannabis). The joint inclusion of severity and monthly use measures in the following ANCOVAs would imply

a violation of this analysis' criteria, and would thus lead to inconsistent results.

PRLT: Decision-making

The within-subject pre-analysis ANCOVA, with block and phase as within-group factors, and alcohol and cannabis monthly use (translated into rank scores) as covariates did not yield any direct or interactive significant effects for any of the covariates. Only the block \times phase \times alcohol monthly use approached significance ($p=0.10$; all other $F < 1$).

PRLT: Decision latencies

As in the previous case, the ANCOVA pre-analysis with phase and block as within-group variables, cannabis and alcohol monthly use as continuous predictors, and phase- and block-wise decision latencies as dependent measure, did not yield any main or interactive effect of the covariates (minimum $p=0.186$).

FRN Scores

As we did with behavioral measures, we carried out a pre-analysis ANCOVA on FRN scores, with channel as the within-group factor, and alcohol and cannabis monthly use (translated into rank scores) as covariates. This analysis did not yield any significant main or interactive effects of alcohol or cannabis (all $p > 0.38$). Regarding P3 scores, only a theoretically irrelevant alcohol monthly use \times channel interaction was observed, $F(2, 112)=4.39$, $p=0.02$. Again, alcohol and cannabis use can be discarded as potential causes of group effects.

Table A1 | IRAB Results for healthy controls, pathological gamblers, and cocaine dependent individuals.

	Cannabis			Cocaine			Tobacco			Alcohol			Gambling		
	Monthly use (joints)	Duration (months)	Monthly use (grams)	Monthly use (grams)	Duration (months)	Monthly use (cigarettes)	Duration (months)	Monthly use (units)	Duration (months)	Monthly use (units)	Duration (months)	Number of illegal drugs used (lifetime)	Monthly use (euros)	Duration (months)	
HC	13.00 (26.61)	39.83 (78.14)	0.33 (1.01)	4.96 (14.28)	51.26 (73.80)	225.04 (266.50)	44.02 (42.81)	126.26 (91.52)	2.17 (1.97)	–	–	–	–	–	–
PG	8.39 (28.15)	24.06 (52.68)	0.45 (1.35)	16.72 (49.05)	112.56 (95.80)	373.06 (362.37)	99.14 (80.12)	152.94 (67.11)	1.56 (1.72)	1503.67 (693.32)	93.43 (66.18)				
CDI	67.00 (63.71)	91.95 (81.42)	18.69 (11.99)	102.90 (75.02)	156.00 (92.95)	339.40 (291.41)	158.20 (118.88)	156.95 (91.51)	4.60 (1.67)	6.45 (20.09)	13.80 (53.85)				



Individual differences in discriminatory fear learning under conditions of ambiguity: a vulnerability factor for anxiety disorders?

Inna Arnaudova¹, Angelos-Miltiadis Kryptos¹, Marieke Effting¹, Yannick Boddez², Merel Kindt¹ and Tom Beckers^{1,2*}

¹ Department of Clinical Psychology and Cognitive Science Center Amsterdam, University of Amsterdam, Amsterdam, Netherlands

² Department of Psychology, KU Leuven, Leuven, Belgium

Edited by:

Robin A. Murphy, University of Oxford, UK

Reviewed by:

Samuel P. Putnam, Bowdoin College, USA

Helena Matute, Universidad de Deusto, Spain

Nicola C. Byrom, University of Oxford, UK

*Correspondence:

Tom Beckers, Department of Psychology, KU Leuven, Tiensestraat 102, Box 3712, 3000 Leuven, Belgium.

e-mail: tom.beckers@ppw.kuleuven.be

Complex fear learning procedures might be better suited than the common differential fear-conditioning paradigm for detecting individual differences related to vulnerability for anxiety disorders. Two such procedures are the *blocking* procedure and the *protection-from-overshadowing* procedure. Their comparison allows for the examination of discriminatory fear learning under conditions of ambiguity. The present study examined the role of individual differences in such discriminatory fear learning. We hypothesized that heightened trait anxiety would be related to a deficit in discriminatory fear learning. Participants gave US-expectancy ratings as an index for the threat value of individual CSs following blocking and protection-from-overshadowing training. The difference in threat value at test between the protected-from-overshadowing conditioned stimulus (CS) and the blocked CS was negatively correlated with scores on a self-report tension-stress scale that approximates facets of generalized anxiety disorder (GAD), the Depression Anxiety Stress Scale-Stress (DASS-S), but not with other individual difference variables. In addition, a behavioral test showed that only participants scoring high on the DASS-S avoided the protected-from-overshadowing CS. This observed deficit in discriminatory fear learning for participants with high levels of tension-stress might be an underlying mechanism for fear overgeneralization in diffuse anxiety disorders such as GAD.

Keywords: individual differences, selective fear-conditioning, discriminatory fear learning, anxiety, cue competition

INTRODUCTION

According to a diathesis-stress model of anxiety disorders, only individuals with certain ingrained vulnerabilities will develop an anxiety disorder following a frightening or traumatic conditioning experience (Mineka and Zinbarg, 2006). The underlying idea of this model is that particular personality traits may predispose some individuals to enhanced fear conditionability (ease of associative fear learning; Otto et al., 2007). That is, following a real-life conditioning event, vulnerable individuals are suggested to have a maladaptive fear response, which serves as the foundation for the development of an actual anxiety disorder. Thus, an important step to truly grasping the etiology of anxiety disorders is identifying individual difference variables that influence fear conditionability in a laboratory setting (i.e., Eysenck, 1976; Zinbarg and Mohlman, 1998; Lissek et al., 2005; Mineka and Zinbarg, 2006). Despite considerable efforts to do so, research has yielded mixed empirical results (Joos et al., 2012).

Imperfections of current research methods have been pinpointed as part of the reason behind the inconclusiveness of the findings (Lissek et al., 2005). For example, one crucial aspect of conditioned fear responding that might be particularly prone to effects of individual difference variables, behavioral avoidance, has often been overlooked in research so far (Beckers et al., 2013). In addition, the commonly used *differential fear-conditioning*

paradigm has been criticized as a model for pathological fear learning (Lissek et al., 2006; Mineka and Oehlberg, 2008; Beckers et al., 2013). In this paradigm, a neutral stimulus (*conditioned stimulus*, CS+) is repeatedly paired with an aversive outcome (*unconditioned stimulus*, US; e.g., shock), resulting in a conditioned fear-like reaction to the CS. This is revealed by increased US-expectancy ratings and physiological reactivity upon presentation of the CS+. A second neutral stimulus (CS-) is never followed by the US, thus acting as a safe signal in the paradigm. A comparison of fear responding to the CS+ and the CS- allows for the assessment of *discriminatory fear learning*. Reduced discriminatory fear learning is considered maladaptive, because in such case responding is not based upon actual stimulus contingencies (Lissek et al., 2005).

This procedure essentially represents a hedonically strong situation: the CS+ clearly signals danger, while the CS- clearly signals safety (Lissek et al., 2006). Because of this threat unambiguity, responses can be expected to be relatively uniform across individuals (Lissek et al., 2006). The lack of ambiguity in this procedure obstructs the examination of interindividual variability in fear learning: mostly everyone will exhibit fear upon confrontation with the CS+ and inhibit fear upon confrontation with the CS- (Lissek et al., 2006; Beckers et al., 2013). A number of studies have actually failed to find an effect of trait anxiety (a known vulnerability factor for anxiety disorders; Spielberger and Gorsuch, 1983)

on differential fear conditioning (e.g., Joos et al., 2012; Torrents-Rodas et al., 2013; but see Baas et al., 2008; Indovina et al., 2011; Gazendam et al., 2013). When comparing clinical with non-clinical populations, reduced discriminatory fear learning has been sometimes successfully observed among participants with anxiety disorders (for a review, see Lissek et al., 2005). From these studies, however, it is not clear if discriminatory fear learning is involved in the etiology or the maintenance of the disorders, because patients are tested after they have been diagnosed with an anxiety disorder (Beckers et al., 2013).

The use of a weaker or a more ambiguous assessment situation might be better suited to study individual differences in fear conditioning, because it increases the variance of individual responses and will make the proposed maladaptive responses of vulnerable individuals more apparent (Lissek et al., 2006; Beckers et al., 2013). For example, it has been observed that relative to low-neuroticism participants, participants with high neuroticism showed increased avoidance to generalization stimuli derived from a CS+ (Lommen et al., 2010). Generalization stimuli do not have a direct link to the US; their threat value is estimated from their perceptual similarity to the CS+, which makes them essentially ambiguous. Chan and Lovibond (1996) used another ambiguous assessment method, a *conditioned inhibition paradigm* (A+ training intermixed with AB− training), and found that individuals who were high in trait anxiety and were also unaware of stimulus contingencies in the task showed an expectancy bias (increased US-expectancy) for all CSs. These results provide empirical evidence for the conceptual argument of Lissek et al. (2006) that individual differences are particularly likely to be observed in weak or ambiguous testing situations.

Following this reasoning, the optimal assessment of individual differences in discriminatory fear learning requires a comparison of an ambiguous danger and an ambiguous safe signal. This can be achieved through the use of a *selective fear-conditioning paradigm*, where multiple stimuli compete for behavioral control of the fear response, thus creating some level of ambiguity. For example, a selective conditioning procedure called *protection from overshadowing* can be regarded as the ambiguous counterpart for the learning of a danger signal (CS+) in differential fear conditioning. In protection from overshadowing, one CS (C) is presented without being followed by the US in a first *elemental conditioning* phase (C−). In a second *compound conditioning* phase, C is presented together with another CS (D) to make up a compound of two CSs (CD), which is followed by the US (CD+). Following a protection-from-overshadowing procedure (C− then CD+) in associative learning tasks, heightened responding is generally assigned to the *protected-from-overshadowing* stimulus D relative to a situation where only CD+ training is given (Vandorpe and De Houwer, 2005). The fact that C is not followed by the US in selective conditioning, when presented alone, suggests that D is probably dangerous (with a higher threat value), given that the chances of the US are clearly increased by adding D to C. However, the high threat status of D remains somewhat ambiguous and can only be inferred, because D is never observed in isolation before test.

In order to analogously create an ambiguous signal for relative safety, one CS (A) can be repeatedly followed by a US in

a first phase of conditioning (A+). In a subsequent compound conditioning phase, A can be presented together with another CS (B) to make up a compound of two CSs (AB), which is also followed by the US (AB+). Following such *blocking* procedure (A+ then AB+) in associative learning tasks, it is typically found that responding to the *blocked* CS B is reduced relative to a situation where only AB+ training is presented (Kamin, 1969; Dickinson et al., 1984). The blocking effect has been observed in a variety of learning procedures in diverse species (see Haselgrove and Evans, 2010, for an overview). Thus, in a conditioning procedure, the fact that A is followed by the US when presented alone suggests that B is probably safer (has a lower threat value) than a protected-from-overshadowing D, given that the chances or the intensity of the US following the AB compound are not increased by B. Still, the relative safety of B in comparison to D remains ambiguous and can only be inferred, given that B is never observed in isolation before test (both B and D are only ever presented in a compound that is always followed by the US; Beckers et al., 2013). Individual differences in such selective learning of relative safety might therefore be readily observed. In line with this idea, it has indeed been shown that trait anxiety is correlated with reduced blocking (thus, impaired safety learning for a blocked stimulus; Boddez et al., 2012). Therefore, a selective discrimination learning procedure, where protection-from-overshadowing and blocking training are combined, allows examining discriminatory fear learning under conditions of ambiguity and uncovering individual differences therein.

Since the early years of fear-conditioning research, most attention has been paid to the role of trait anxiety in conditionability (e.g., Spence, 1964), specifically in relation to deficient safety learning. Trait anxiety is usually assessed by means of the State and Trait Anxiety Inventory (STAI; Spielberger and Gorsuch, 1983), which has recently been questioned as a pure measure of dispositional anxiety and is now seen rather as a measure of general negative affect (Bieling et al., 1998; Grös et al., 2007; Bados et al., 2010). To address the lack of specificity of the STAI and other questionnaires, the Depression Anxiety Stress Scales (DASS; Lovibond and Lovibond, 1995) were developed. They measure three negative emotional states with good discriminative validity (Clara et al., 2001; Crawford and Henry, 2003): depression (loss of self-esteem and motivation; DASS-D), anxiety (physical arousal; DASS-A), and tension-stress (persistent tension and a low threshold for distress; DASS-S). The DASS-A has predictive validity for panic, phobia, and other anxiety disorders (Brown et al., 1997) and might be related to reactivity to threat. The DASS-S has been mainly linked to generalized anxiety disorder (GAD; Brown et al., 1997), thus possibly having a specific relationship with discriminatory fear learning [GAD patients experience chronic anxiety over a number of situations; American Psychiatric Association (APA), 2000]. DASS-S has recently been linked to worry (Szabó, 2011). Interestingly, worry has recently also emerged as a predictor for heightened conditionability (Otto et al., 2007; Gazendam and Kindt, 2012; Joos et al., 2012), making it crucial to discriminate the role of anxiety and tension-stress during fear conditioning. Other personality traits related to trait anxiety such as neuroticism and extraversion have also been implicated as potential sources for individual variability in fear learning (Eysenck, 1976) and this proposal has

received partial support from a few studies (e.g., Frederikson and Georgiades, 1992; Pineles et al., 2009).

Disentangling the web of mixed results regarding these closely related personality characteristics and their influence on discriminatory fear learning under ambiguous conditions should allow a better understanding of vulnerability factors for anxiety disorders. In the present study, participants underwent blocking and protection-from-overshadowing training (see **Table 1**) and gave trial-by-trial US-expectancy ratings as indication of the threat value of each elemental and compound CS. The difference between the US-expectancy rating for the protected-from-overshadowing CS D and the blocked CS B (D minus B) at test was used as a measure of discriminatory fear learning (analogous to the difference score between CS+ and CS– typically used as index of learning in standard differential fear-conditioning studies, e.g., Joos et al., 2012). Based on the findings of Boddez et al. (2012), we hypothesized that trait anxiety should be associated with reduced discriminatory fear learning, mainly due to insufficient safety learning of the blocked CS. Other individual difference variables that have been implicated in conditionability were assessed as well for their unique contribution to disturbed discriminatory fear learning. Further, we examined the generalization of these effects to a behavioral task and across contexts. The behavioral task, in which participants chose between chocolate bars carrying symbolic representations of the blocked CS B and the protected-from-overshadowing CS D, was used to test whether individual differences can be observed in overt behavior as well. The role of test context (same or different as training context) was explored because of the lack of empirical data on the context specificity of learning following a selective fear-conditioning paradigm; we assumed that generalization across contexts might constitute another possible source of interindividual differences.

MATERIALS AND METHODS

PARTICIPANTS

A total of 68 participants from University of Amsterdam and the surrounding areas participated for course credits or a small monetary compensation (€ 7). Fourteen participants were excluded for lack of acquisition learning¹. The remaining sample (20 males) had a mean age of 22.00 (SD = 4.48) years (see **Table 2** for further demographics). All participants gave informed consent for their participation and the experimental procedure was approved by the Faculty Ethical Committee at the University of Amsterdam.

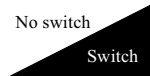
STIMULI AND MATERIALS

Images of six colored three-dimensional geometrical objects as seen from four viewing angles (computer-generated) served as CSs: a yellow stick, a blue disk, a purple cylinder, a red plane, an orange cone, and a green cube. The longest dimension (height, diameter, or internal diagonal) of all objects was 60 mm. Objects appeared on the computer screen surrounded by a white frame, measuring 106 mm × 106 mm. They were centered on the screen with either an orange or blue background, counterbalanced across participants.

Conditioned stimulus assignment was partially counterbalanced across participants. The yellow stick, blue disk, and purple cylinder were counterbalanced to serve as elemental acquisition CSs A, C, or E. During the compound conditioning phase, the

¹Excluded participants gave a positive US-expectancy rating for an elemental or compound CS never followed by the US and/or a negative US-expectancy rating for an elemental or compound CS always followed by the US on the very last trial of either elemental or compound training. These participants did not differ from the remaining sample on any of the demographic or personality variables. The conclusions of the experiment do not change when these participants are included in the analyses.

Table 1 | Conditioning contingencies.

Type of training	Elemental	Compound	Context	Test
Blocking	A+	AB+		B–, D+, F–, A+, C–, E–
Protection from overshadowing	C–	CD+		D+, B–, F–, C–, A+, E–
Control	E–	EF–		

Letters represent CSs; – represents no US was administered; + represents US was administered.

Table 2 | Mean and standard deviations (SD) for questionnaires, post-acquisition CS valence and US expectancy at CS test.

Questionnaire	STAI-S	STAI-T	DASS-D	DASS-A	DASS-S	EPQ-N	EPQ-E	IUS
Mean	32.91	36.46	4.74	3.61	8.33	8.13	14.26	61.17
SD	7.34	7.19	6.08	3.86	6.78	5.37	3.67	17.30
Post-acquisition CS valence	B	D	F	A	C	E		
Mean	0.51	–1.56	1.85	–2.34	2.20	2.71		
SD	3.22	2.77	2.72	2.93	2.54	2.19		
US expectancy at CS test	B	D	F	A	C	E		
Mean	–0.29	3.07	–3.62	4.29	–3.53	–4.44		
SD	3.76	2.57	1.74	1.65	2.66	1.27		

compound CSs were composed of the yellow stick and the red plane; the blue disk and the orange cone; the purple cylinder and the green cube (*de facto* counterbalanced to AB, CD, and EF, as a result of the counterbalancing of A, C, and E). In this phase, the two images, randomly assigned to the left or right part of the screen, appeared separated by 48 mm.

The US was an aversive 1-s 95-dB scream delivered through headphones.

ASSESSMENTS

US expectancy

Participants rated US expectancies by clicking with a mouse on a computerized 11-point Likert scale ranging from -5 (*certainly no scream*) to 5 (*certainly scream*). The validity of this measure to assess fear learning is reviewed extensively by Boddez et al. (2013).

Evaluative ratings

Valence ratings of CSs and the US were assessed on an 11-point Likert scale, with -5 indicating *very unpleasant* and 5 indicating *very pleasant*. The US was also rated on 5-category scales for intensity (*light, moderate, intense, enormous, unbearable*) and startlingness (*not, light, moderate, strong, very strong*).

Questionnaires

State and Trait Anxiety Inventory (Spielberger and Gorsuch, 1983; Dutch version by van der Ploeg, 2000) measures trait and state anxiety with 20 items each, with sum scores representing severity. The psychometric characteristics of the STAI are as follows: test-retest reliability 0.73 – 0.86 for STAI-T and 0.33 for STAI-S, internal consistency of 0.90 for STAI-T and 0.86 – 0.93 for STAI-S (Spielberger and Gorsuch, 1983) and excellent convergent validity across ethnic groups (Novy et al., 1993).

The 42-item DASS (Lovibond and Lovibond, 1995; Dutch translation by de Beurs et al., 2001) have good psychometric properties. Cronbach's alphas for internal consistency of the three subscales DASS-D, DASS-A, and DASS-S are 0.97 , 0.95 , and 0.92 , respectively (Antony et al., 1998).

Two scales of the Dutch Eysenck Personality Questionnaire (EPQ) measure neuroticism (22-item EPQ-N, Cronbach's $\alpha = 0.87$) and extraversion (19 item EPQ-E; Cronbach's $\alpha = 0.85$; Sanderman et al., 2012).

Responses to situations of ambiguity might also be influenced by dispositional intolerance of uncertainty. The 27-item Dutch version of the Intolerance of Uncertainty Scale shows good reliability with Cronbach's α of 0.88 in a student sample (IUS; Freeston et al., 1994; Dutch translation by de Bruin et al., 2006).

Forced-choice behavioral test

Participants chose among 10 chocolate bars placed randomly in an open box by the exit of the experimental room. Five of the bars had a wrapping depicting the blocked CS B, while the rest had a wrapping representing the protected-from-overshadowing CS D; thus, participants' choice reveals their preference for one or the other CS. This procedure was modeled after Blechert et al. (2007).

PROCEDURE

After signing an informed consent form, participants sat in front of a computer in a dimly lit room, where they were separated

from the experimenter by a barrier. They filled in a computerized version of STAI-T and STAI-S.

On-screen instructions informed participants that their task was to predict the occurrence of a scream based on the objects presented on the screen. The US-expectancy rating scale and the usage of the mouse were explained. The experimenter repeated the on-screen instructions and asked participants to put on the headphones.

The selective conditioning procedure consisted of three phases: an elemental and a compound training phase, followed by a test phase (Table 1). During elemental training, three individual CSs were presented four times each, with one CS always being followed by the US (4 A+, 4 C–, and 4 E–). During compound training, participants viewed four presentations of three compound CSs, with two compound CSs being followed by the US (4 AB+, 4 CD+, and 4 EF–). Thus, across phases participants received blocking (A+ then AB+), protection-from-overshadowing (C– then CD+), and filler training (E– then EF–). The filler stimuli were used in order to indicate to participants that compound stimuli can occur without the US and to discourage participants from concluding that mere compoundness predicts US occurrence. Both learning phases occurred on the same orange or blue computer background (Context A).

In the test phase, six individual CSs were presented in a fixed, counterbalanced order that included the critical CSs B and D first, followed by all other elemental CSs (either B–, D+, F–, A+, C–, E–, or D+, B–, F–, C–, A+, E–). D and A trials were reinforced at test to prevent random ratings (Lovibond, 2003). Order was partially counterbalanced across participants in order to check for the influence of the reinforced test trials on the other ratings. Test trials occurred either on the same background (Context A) or on a background different from the acquisition context (Context B). Participants were randomly assigned to the context-switch or the no-context-switch condition.

Each elemental or compound CS presentation lasted 8 s. An active US-expectancy rating scale was available at the bottom of the screen during the first 5 s. If participants failed to confirm their rating by clicking the mouse button in this time frame, the pointer position at the end of the 5-s time frame of the current trial was recorded as an indication of their response². Presentations of elemental or compound CS were randomized within the acquisition phases, with the restriction that no more than two identical trials were presented in succession. Inter-trial intervals (ITI) had an average duration of 20 s (15s, 20s, 25s). During ITIs and the last 3 s of CS presentation an inactive US-expectancy scale was present on the screen.

Following the test phase, participants took off the headphones and indicated for each elemental or compound CS presented during training whether it had been followed by the scream and the certainty in their response. After giving evaluative ratings for the CSs and the US, participants filled in the EPQ, the DASS, and the IUS. Then, participants performed the forced-choice behavioral

²Twelve percent of all trials across participants were not confirmed. The number of unconfirmed trials correlated negatively with the neuroticism scale of the EPQ, $\rho(54) = -0.29$, $p = 0.04$. Number of unconfirmed trials was not significantly related to any of the other questionnaire scores.

test. Reinforcement of D at test might have potentially affected the choices made during the following behavioral test, but this should have occurred across participants, if anything acting to reduce the influence of individual differences on behavior.

DATA ANALYSIS

As counterbalancing factors (initial background, CS assignment, and test order) had no significant effects in preliminary analyses, the data were collapsed across them. Conditioning effects during elemental and compound training phases were analyzed using a 3 (trial type: A, C, E, or AB, CD, EF) by 4 (trial number: 1–4) repeated measures analyses of variance (ANOVAs). Repeated measures ANOVA was also used to examine the ratings of the six individual CSs at test, with a Bonferroni correction for pairwise comparisons. Greenhouse-Geisser corrections were applied when the assumption of sphericity was violated. In order to test for generalization of learning across contexts, context switch was entered as a between-subject variable in the repeated measures ANOVA.

To test for individual differences in discriminatory fear learning, we calculated correlations between scores on personality measures and the D-B difference score. The normal distribution of each variable was first examined with a Kolmogorov–Smirnov test. When the data were not normally distributed, Spearman's correlations were used. Otherwise, Pearson's r is reported. Participants scoring more than two standard deviations away from the mean on a personality measure were excluded for the analyses with that particular measure ($n = 1$ for STAI-S; $n = 4$ for DASS-D; $n = 2$ for DASS-A; $n = 4$ for EPQ-E; $n = 1$ for IUS). In order to check for generalization to a behavioral task, choice data were subjected to a chi-square test to evaluate deviation from random choice.

RESULTS

VALENCE RATINGS

Mean ratings for the US were -2.80 ($SD = 1.83$) for valence, 2.76 ($SD = 0.70$) for intensity, and 2.89 ($SD = 1.04$) for startlingness, indicating that participants perceived the scream as aversive. US valence ratings were marginally correlated only with scores on STAI-T, $r(54) = 0.27$, $p = 0.047$. Post-acquisition CS valence

ratings can be seen in **Table 2**. As expected, CSs with higher threat values were given lower valence ratings compared to CSs with lower threat value.

CONDITIONING EFFECTS

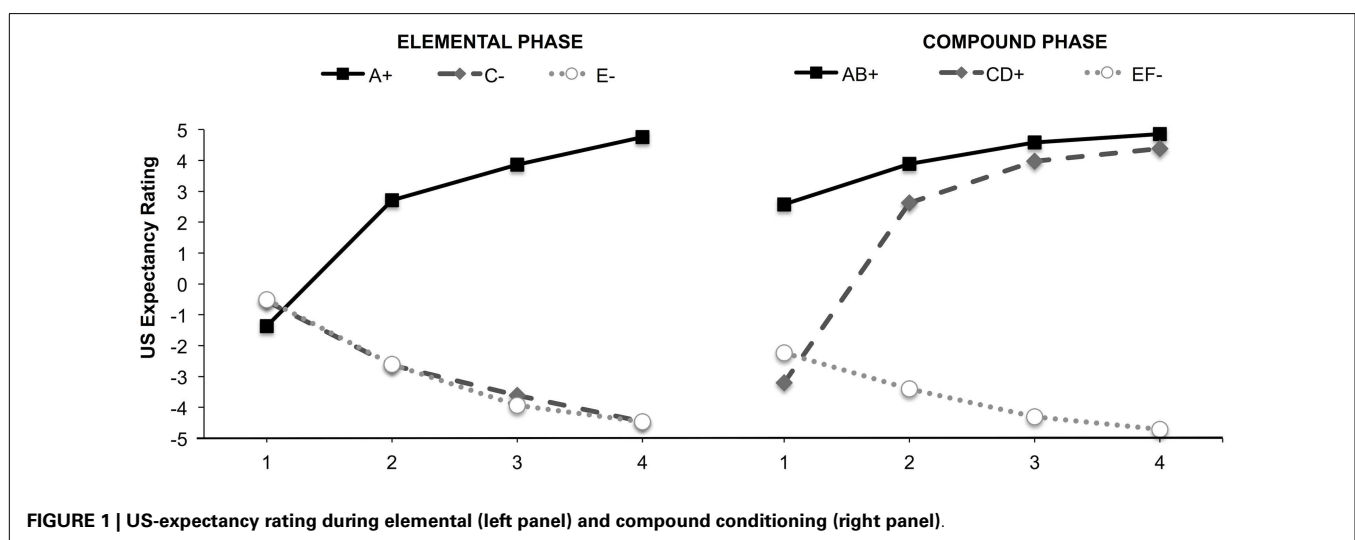
Trial-by-trial US-expectancy ratings for the CSs during both learning phases can be seen in **Figure 1**. The ANOVAs revealed significant Trial type \times Trial number interactions for both the elemental, $F(3.89, 206.36) = 133.16$, $p < 0.001$, $\eta_p^2 = 0.72$, and the compound phase, $F(4.51, 238.89) = 81.50$, $p < 0.001$, $\eta_p^2 = 0.61$. These results show that participants learned the contingencies between the specific CSs and the US across trials in both conditioning phases.

Unconditioned stimulus-expectancy ratings for the individual CSs at test can be found in **Table 2**. The six CSs elicited different ratings, $F(2.99, 158.45) = 130.45$, $p < 0.001$, $\eta_p^2 = 0.71$. All pairwise comparisons (each elemental CS with every other elemental CS) were significant ($p < 0.01$), except that US-expectancies for C were not significantly different from these for E and F ($p > 0.10$). The blocked stimulus B was rated significantly higher than the safe stimuli C, E, and F, which suggests that it remained ambiguous at test. The protected-from-overshadowing stimulus D was rated significantly lower than the dangerous stimulus A at test, which suggests it also remained somewhat ambiguous at test. However, the contrast between B and D was highly significant ($p < 0.001$). These results indicate that on average participants assigned higher threat value to the protected-from-overshadowing (relatively dangerous) CS D than the blocked (relatively safer) CS B, in line with expectations.

The main effect of CS on US-expectancy ratings was not modulated by context, $F < 1$. The test context did not affect ratings for B and D ($p = 0.83$). Our context manipulation did not affect the generalization of the assigned threat values.

INDIVIDUAL DIFFERENCES IN DISCRIMINATORY FEAR LEARNING

Contrary to our hypothesis, scores on the STAI-T did not correlate with overall discriminatory fear learning (D-B), $\rho(54) = -0.15$, $p = 0.29$. However, DASS-S scores did correlate with D-B,



$\rho(54) = -0.29, p = 0.03$, and remained significant when controlling for DASS-A scores, $\rho(49) = -0.29, p = 0.04$. This suggests that high levels of persistent tension are linked to a deficit in discriminatory fear learning under ambiguity.

Remarkably, neither STAI-T, nor DASS-S, nor any of the other scores on personality measures were correlated to the difference between the US-expectancy rating between the two elemental CSs A and C (A minus C). The results confirm that interindividual differences in discriminatory fear learning are more readily detected for the ambiguous danger and safe signals than for non-ambiguous ones.

When looking at ratings for the individual CSs, STAI-T did not correlate with any of the US-expectancy ratings at test, although a trend was observed for the filler CS E, $\rho(54) = 0.19, p = 0.07$. The DASS-A emerged as the only marginally significant predictor of ratings for the ambiguous danger CS D, $\rho(49) = -0.27, p = 0.05$. A trend was observed for a correlation between the DASS-S and both CS B, $r(54) = 0.26, p = 0.06$, and CS D, $\rho(54) = -0.26, p = 0.06$. When controlling for DASS-A, the correlation between DASS-S and B became highly significant, $r(49) = 0.45, p = 0.001$, while its correlation with D became insignificant, $\rho(49) = -0.03, p = 0.85$. When controlling for DASS-S, the correlation between DASS-A and D also became insignificant, $\rho(49) = -0.17, p = 0.22$. The correlations between DASS-S and the other cues presented at test did not reach significance (all $p > 0.10$). Further, the correlation between DASS-S and the difference score between stimulus B and F at test, which might reflect more specifically the safety value of B, did not reach significance, $r(54) = 0.159, p = 0.249$.

No significant correlations or trends emerged between other personality measures (DASS-D, EPQ, and IUS) and the threat value assigned to any of the CSs, including the two stimuli of interest: the blocked stimulus B and the protected-from-overshadowing stimulus D. This suggests that the tension-stress scale of the DASS is best suited to capture individual differences in discriminatory fear learning under conditions of ambiguity; those differences moreover appear to occur predominantly in the selective learning of safety rather than danger.

FORCED-CHOICE BEHAVIORAL TEST

Generalization of the learned threat to overt behavior was examined through the total number of participants who showed a preference toward B. Participants did not show an overall preference for B over D during the forced-choice behavioral test, $\chi^2(1) = 1.28, p = 0.26$. Since only DASS-S emerged as a predictor of the extent of discrimination learning, a median split was performed to further analyze the data. The test showed that the two groups differed in their choice behavior, $\chi^2(1) = 4.43, p = 0.04$. The high DASS-S group chose B more often than D, $\chi^2(1) = 5.26, p = 0.02$ (Figure 2), whereas the low DASS-S group was indifferent, $\chi^2(1) = 0.33, p = 0.56$. This suggests that participants with high DASS-S scores actively avoided D.

DISCUSSION

This study examined individual differences in discriminatory fear learning under conditions of ambiguity. A reduction of discriminatory fear learning between a blocked CS and a protected-from-overshadowing CS was contrary to our hypothesis

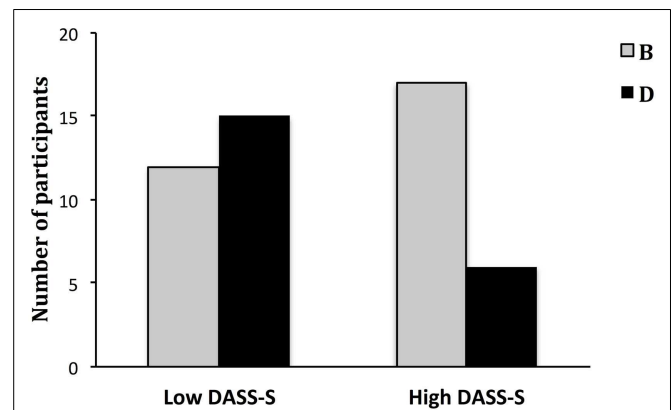


FIGURE 2 | Number of participants choosing a chocolate bar depicting either CS B or CS D in the forced-choice behavioral test according to DASS-S group.

not related to any of the trait anxiety scores (STAI-T and DASS-A), but uniquely related to higher levels of tension-stress as measured by DASS-S. This result was driven mainly by increased threat value assigned to the blocked CS B, which suggests that these participants overestimate threat for ambiguous signals with relatively low threat value (i.e., overgeneralize threat from the AB+ compound trials to B). A tendency to overgeneralize was revealed for the high tension-stress group also in their performance during a behavioral task, where the high DASS-S participants showed more behavioral avoidance to a mere depiction of the protected-from-overshadowing CS D on a food item wrapping. This suggests that these participants judge ambiguous situations with the slightest hint of threat more readily as dangerous (i.e., a better-safe-than-sorry strategy). Such overgeneralization bias has been suggested as one of the underlying mechanisms of anxiety disorders with a generalized nature (e.g., Lissek and Grillon, 2010; Lissek et al., 2010).

This bias appears also to affect avoidance behavior under circumstances where there is no source of threat (as in the behavioral task). The observed behavioral pattern of the high tension-stress individuals can be seen as a sign of threat generalization toward an innocuous stimulus (a wrapping depicting a threatening CS).

The present study did not replicate the earlier observation by Boddez et al. (2012) of a significant correlation between trait anxiety as measured by STAI-T and threat value assigned to a blocked CS. The procedural differences between the two studies might partially explain the divergence. However, the nature of the STAI-T scale should be taken into account. Recent attempts to discriminate between depression and anxiety have prompted researchers to question the ability of STAI-T to specifically capture the concept of dispositional anxiety. Its items seem to reflect depression and general negative affect, rather than anxiety itself (Bieling et al., 1998; Grös et al., 2007; Bados et al., 2010). In contrast, the anxiety and stress scales of the DASS have been shown to capture factors of anxiety that are distinct from depressive symptoms (which are captured by the depression scale), with the DASS-A indexing in particular diagnostic approximations for phobias and panic disorder and the DASS-S capturing aspects of anxious distress that relate to more free-floating anxiety disorders such as GAD

(Brown et al., 1997; Lovibond, 1998). Thus, the DASS scales offer the possibility to truly examine the divergent influence of three negative affective states upon discriminatory fear learning and to more readily draw conclusions about the link between vulnerability factors, discriminatory fear learning, and anxiety. Future research concerning individual differences in fear learning should utilize this aspect of the DASS scales to its advantage.

Only scores on the DASS-S scale were found to be linked to reduced discriminatory fear learning. One can argue that this relationship might be explained by an increased sensitivity of participants that score high on DASS-S to the aversive stimulus, but this is unlikely given the lack of correlation between US valence ratings and DASS-S scores, $r(54) = -0.05$, $p = 0.72$. Another possible interpretation of the results could be that participants with high tension-stress scores were less able to generalize from the last A+ trial in the elemental phase to the first AB+ trial in the compound phase and thus have learned more about the added stimulus B. Additional analyses, however, revealed no correlation between DASS-S scores on the one hand and expectancy ratings on the first AB+ trial, nor between DASS-S scores and generalization decrement (defined as the difference between responding on the final A+ trial and responding on the first AB+ trial), both $p > 0.7$.

Depression Anxiety Stress Scale-Stress items correspond closely to the diagnostic criteria of GAD from the DSM-IV [American Psychiatric Association (APA), 2000] and the total score on the scale has recently been empirically linked to worry behavior, a core symptom of GAD (Lovibond, 1998; Szabó, 2011). The fact that worry has been shown to be related to increased conditionability (e.g., Otto et al., 2007) combined with the present results suggest that general tension-stress might be a vulnerability factor for GAD and maybe other diffuse anxiety disorders through its effect on discriminatory fear learning under conditions of ambiguity. More research with clinical and non-clinical samples is needed to confirm this possibility. The tentative results of this study suggest that in treatment, increasing the ability of GAD patients to discriminate between safer and more dangerous signals might be worthwhile in order to decrease behavioral avoidance and to improve functioning. Indeed, therapists increasingly come to recognize that learning about safety periods is a promising route in the treatment of GAD (e.g., Woody and Rachman, 1994; Fonteyne et al., 2009).

A secondary aim of this study was to examine the context specificity of selective learning. Our results show that selective learning generalizes across contexts. However, our context manipulation might have not been salient enough, as it consisted of only a screen background switch in the absence of any explicit instructions. Other limitations of this study include the studied sample (young university students), which puts generalization to the general population under question, and the use of correlational

analyses and self-report data, which is known to be prone to demand characteristics.

Important questions remain for future research. The negative relation between selective discrimination learning and DASS-S scores might either be specific for threat-related situations (e.g., fear conditioning) or reflect a more general deficit in selective learning in people that are high in tension-stress. Future research might try to discriminate between a fear-specific versus a more general locus of the effect (e.g., by testing selective learning in neutral contingency learning tasks in relation to DASS-S scores). Also, learning theory and research suggest that several processes are involved in blocking and other forms of selective learning (Pearce and Bouton, 2001; De Houwer and Beckers, 2002; Shanks, 2010). An important challenge for future research is therefore to precisely determine the mechanisms that cause *variation* in selective (fear) learning. A deficit in selective attention (Le Pelley, 2004; Haselgrove et al., 2010) is one candidate process that could underlie the observed decrease in discrimination between protection from overshadowing and blocking in participants high in DASS-S (again, such lack of selective attention might be threat-specific or domain-general). Future research could examine this possibility by using attention measuring techniques (e.g., eye-tracking; Beesley and Le Pelley, 2011).

The present study offers empirical justification for the use of the selective fear-conditioning paradigm in the search for individual differences in discriminatory fear learning. A relationship between interindividual differences and discriminatory fear learning was observed only for ambiguous danger versus safety signals (D versus B) and not for unambiguous ones (A versus C). The present paradigm might therefore be useful for the examination of vulnerabilities to GAD. Future work should also strive toward establishing the unique contributions of anxiety, tension-stress, worry, and general negative affect to decreased discriminatory fear learning. Special attention needs to be paid to the tension-stress factor as this might predispose for the maladaptive expansion of threat toward innocuous stimuli.

ACKNOWLEDGMENTS

This work was funded by Innovation Scheme (Vidi) Grant 452-09-001 from the Netherlands Organization for Scientific Research (NWO) awarded to Tom Beckers. Merel Kindt is supported by an Innovation Scheme (Vici) Grant from NWO. Angelos-Miltiadis Kryptos is a scholar of the Alexander S. Onassis Public Benefit Foundation. Yannick Boddez is supported by KU Leuven Centre for Excellence grant PF/10/005 and Interuniversity Attraction Poles grant P7/33 of the Belgian Science Policy Office. We thank Bert Molenkamp for technical support and Pjotr van Baarle, Jeroen Butterman, Manouk Corver, Wouter Cox, and Jiri Staats for help with data collection.

REFERENCES

- American Psychiatric Association (APA). (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edn. Washington, DC: American Psychiatric Association.
- Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., and Swinson, R. P. (1998). Psychometric properties of the 42-item versions of the depression anxiety stress scales in clinical groups and a community sample. *Psychol. Assess.* 10, 176–181. doi:10.1037/1040-3590.10.2.176
- Baas, J. M. P., van Ooijen, L., Goudriaan, A., and Kenemans, J. L. (2008). Failure to condition to a cue is associated with sustained contextual fear. *Acta Psychol. (Amst.)* 127, 581–592. doi:10.1016/j.actpsy.2007.09.009
- Bados, A., Gómez-Benito, J., and Balaguer, G. (2010). The state-trait anxiety inventory, trait version: does it really measure anxiety? *J. Pers. Assess.* 92, 560–567. doi:10.1080/00223891.2010.513295
- Beckers, T., Kryptos, A.-M., Boddez, Y., Effting, M., and Kindt, M. (2013). What's wrong with fear conditioning? *Biol. Psychol.* 92, 90–96. doi:10.1016/j.biopsycho.2011.12.015

- Beesley, T., and Le Pelley, M. E. (2011). The influence of blocking on overt attention and associability in human learning. *J. Exp. Psychol. Anim. Behav. Process.* 37, 114–120. doi:10.1037/a0019526
- Bieling, P. J., Antony, M. M., and Swinson, R. P. (1998). The State-Trait Anxiety Inventory, trait version: structure and content re-examined. *Behav. Res. Ther.* 36, 777–788. doi:10.1016/S0005-7967(98)00023-0
- Blechert, J., Blechert, J., Michael, T., Vriends, N., Margraf, J., and Wilhelm, F. H. (2007). Fear conditioning in posttraumatic stress disorder: evidence for delayed extinction of autonomic, experiential and behavioural responses. *Behav. Res. Ther.* 45, 2019–2033. doi:10.1016/j.brat.2007.02.012
- Boddez, Y., Baeyens, F., Luyten, L., Vastenenwegen, D., Hermans, D., and Beckers, T. (2013). Rating data are underrated: validity of US expectancy in human fear conditioning. *J. Behav. Ther. Exp. Psychiatry* 44, 201–206. doi:10.1016/j.jbtep.2012.08.003
- Boddez, Y., Vervliet, B., Baeyens, F., Lauwers, S., Hermans, D., and Beckers, T. (2012). Expectancy bias in a selective conditioning procedure: trait anxiety increases the threat value of a blocked stimulus. *J. Behav. Ther. Exp. Psychiatry* 43, 832–837. doi:10.1016/j.jbtep.2011.11.005
- Brown, T. A., Chorpita, B. F., Korotitsch, W., and Barlow, D. H. (1997). Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. *Behav. Res. Ther.* 35, 79–89. doi:10.1016/S0005-7967(96)00068-X
- Chan, C. K. Y., and Lovibond, P. F. (1996). Expectancy bias in trait anxiety. *J. Abnorm. Psychol.* 105, 637–647. doi:10.1037/0021-843X.105.4.637
- Clara, I. P., Cox, B. J., and Enns, M. W. (2001). Confirmatory factor analysis of the Depression-Anxiety-Stress Scales in depressed and anxious patients. *J. Psychopathol. Behav. Assess.* 23, 61–67. doi:10.1023/A:1011095624717
- Crawford, J. R., and Henry, J. D. (2003). The Depression Anxiety Stress Scales (DASS): normative data and latent structure in a large non-clinical sample. *Br. J. Clin. Psychol.* 42, 111–131. doi:10.1348/014466503321903544
- de Beurs, E., Van Dyck, R., Marquenie, L. A., Lange, A., and Blonk, R. W. B. (2001). De DASS: een vragenlijst voor het meten van depressie, angst en stress. [The Dutch DASS: a questionnaire for assessment of depression, anxiety and stress]. *Gedragstherapie* 34, 35–53.
- de Bruin, G. O., Rassin, E., van der Heiden, C., and Muris, P. (2006). Psychometric properties of a Dutch version of the Intolerance of Uncertainty Scale. *Neth. J. Psychol.* 62, 87–92. doi:10.1007/BF03061055
- De Houwer, J., and Beckers, T. (2002). A review of recent developments in research and theories on human contingency learning. *Q. J. Exp. Psychol.* 55B, 289–310.
- Dickinson, A., Shanks, D., and Evenden, J. (1984). Judgment of act-outcome contingency: the role of selective attribution. *Q. J. Exp. Psychol. A* 36, 29–50.
- Eysenck, H. J. (1976). The learning theory model of neurosis – a new approach. *Behav. Res. Ther.* 14, 251–267. doi:10.1016/0005-7967(76)90001-2
- Fonteyne, R., Vervliet, B., Hermans, D., Baeyens, F., and Vansteenwegen, D. (2009). Reducing chronic anxiety by making the threatening event predictable: an experimental approach. *Behav. Res. Ther.* 47, 830–839. doi:10.1016/j.brat.2009.06.011
- Frederikson, M., and Georgiades, A. (1992). Personality dimensions and classical conditioning of autonomic nervous system reactions. *Pers. Individ. Dif.* 13, 1013–1020. doi:10.1016/0191-8869(92)90134-B
- Freeston, M. H., Rheume, J., Letarte, H., Dugas, M. J., and Ladouceur, R. (1994). Why do people worry? *Pers. Individ. Dif.* 17, 791–802. doi:10.1016/0191-8869(94)90048-5
- Gazendam, F. J., Kamphuis, J. H., and Kindt, M. (2013). Deficient safety learning characterizes high trait anxious individuals. *Biol. Psychol.* 92, 342–352. doi:10.1016/j.biopsycho.2012.11.006
- Gazendam, F. J., and Kindt, M. (2012). Worrying affects associative fear learning: a startle fear conditioning study. *PLoS ONE* 7:e34882. doi:10.1371/journal.pone.0034882
- Grös, D. F., Antony, M. M., Simms, L. J., and McCabe, R. E. (2007). Psychometric properties of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA): comparison to the State-Trait Anxiety Inventory (STAI). *Psychol. Assess.* 19, 369–381. doi:10.1037/1040-3590.19.4.369
- Haselgrove, M., Esber, G. R., Pearce, J. M., and Jones, P. M. (2010). Two kinds of attention in Pavlovian conditioning: evidence for a hybrid model of learning. *J. Exp. Psychol. Anim. Behav. Process.* 36, 456–470. doi:10.1037/a0018528
- Haselgrove, M., and Evans, L. H. (2010). Variations in selective and non-selective prediction error with the negative dimension of schizotypy. *Q. J. Exp. Psychol.* 63, 1127–1149. doi:10.1080/17470210903229979
- Indovina, I., Robbins, T. W., Núñez-Elizalde, A. O., Dunn, B. D., and Bishop, S. J. (2011). Fear-conditioning mechanisms associated with trait vulnerability to anxiety in humans. *Neuron* 69, 563–571. doi:10.1016/j.neuron.2010.12.034
- Joos, E., Vansteenwegen, D., and Hermans, D. (2012). Worry as a predictor of fear acquisition in a nonclinical sample. *Behav. Modif.* 36, 723–750. doi:10.1177/0145445512446477
- Kamin, L. J. (1969). “Predictability, surprise, attention, and conditioning,” in *Punishment and Aversive Behavior*, ed. R. M. Church (New York: Appleton-Century-Crofts), 279–296.
- Le Pelley, M. E. (2004). The role of associative history in models of associative learning: a selective review and a hybrid model. *Q. J. Exp. Psychol.* 57B, 193–243.
- Lissek, S., and Grillon, C. (2010). Overgeneralization of conditioned fear in the anxiety disorders. *J. Psychol.* 218, 146–148. doi:10.1027/0044-3409/a000022
- Lissek, S., Pine, D. S., and Grillon, C. (2006). The strong situation: a potential impediment to studying the psychobiology and pharmacology of anxiety disorders. *Biol. Psychol.* 72, 265–270. doi:10.1016/j.biopsycho.2005.11.004
- Lissek, S., Powers, A. S., McClure, E. B., Phelps, E. A., Woldehawariat, G., Grillon, C., et al. (2005). Classical fear conditioning in the anxiety disorders: a meta-analysis. *Behav. Res. Ther.* 43, 1391–1424. doi:10.1016/j.brat.2004.10.007
- Lissek, S., Rabin, S. J., Lukenbaugh, D., Geraci, M., Pine, D. S., and Grillon, C. (2010). Overgeneralization of conditioned fear as a pathogenic marker of panic disorder. *Am. J. Psychiatry* 167, 47–55. doi:10.1176/appi.ajp.2009.09030410
- Lommen, M. J. J., Engelhard, I. M., and van den Hout, M. A. (2010). Neuroticism and avoidance of ambiguous stimuli: better safe than sorry? *Pers. Individ. Dif.* 49, 1001–1006. doi:10.1016/j.paid.2010.08.012
- Lovibond, P. F. (1998). Long-term stability of depression, anxiety and stress. *J. Abnorm. Psychol.* 107, 520–526. doi:10.1037/0021-843X.107.3.520
- Lovibond, P. F. (2003). Causal beliefs and conditioned responses: retrospective reevaluation induced by experience and by instruction. *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 97–106. doi:10.1037/0278-7393.29.1.97
- Lovibond, P. F., and Lovibond, S. H. (1995). The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behav. Res. Ther.* 33, 335–343. doi:10.1016/0005-7967(94)00075-U
- Mineka, S., and Oehlberg, K. (2008). The relevance of recent developments in classical conditioning to understanding the etiology and maintenance of anxiety disorders. *Acta Psychol. (Amst.)* 127, 567–580. doi:10.1016/j.actpsy.2007.11.007
- Mineka, S., and Zinbarg, R. (2006). A contemporary learning theory perspective on the etiology of anxiety disorders: it’s not what you thought it was. *Am. Psychol.* 61, 10–26. doi:10.1037/0003-066X.61.1.10
- Novy, D. M., Nelson, D. V., Goodwin, J., and Rowzee, R. D. (1993). Psychometric comparability of the State-Trait Anxiety Inventory for different ethnic subpopulations. *Psychol. Assess.* 5, 343–349. doi:10.1037/1040-3590.5.3.343
- Otto, M. W., Leyro, T. M., Christian, K., Deveney, C. M., Reese, H., Pollack, M. H., et al. (2007). Prediction of “fear” acquisition in healthy control participants in a De Novo fear-conditioning paradigm. *Behav. Modif.* 31, 32–51. doi:10.1177/0145445506295054
- Pearce, J. M., and Bouton, M. E. (2001). Theories of associative learning in animals. *Annu. Rev. Psychol.* 52, 111–139. doi:10.1146/annurev.psych.52.1.111
- Pineles, S. L., Vogt, D. S., and Orr, S. P. (2009). Personality and fear responses during conditioning: beyond extraversion. *Pers. Individ. Dif.* 46, 48–53. doi:10.1016/j.paid.2008.09.003
- Sanderman, R., Arrindell, W. A., Ranchor, A. V., Eysenck, H. J., and Eysenck, S. B. G. (2012). *Het meten van persoonlijkheidskenmerken met de Eysenck Personality Questionnaire (EPQ), een handleiding, Tweede herziene druk.* [Manual to measuring personality with the Dutch EPQ. 2nd ed.]. Groningen: UMCG/Rijksuniversiteit Groningen.

- Shanks, D. R. (2010). Learning: from association to cognition. *Annu. Rev. Psychol.* 61, 273–301. doi:10.1146/annurev.psych.093008.100519
- Spence, K. W. (1964). Anxiety (drive) level and performance in eyelid conditioning. *Psychol. Bull.* 61, 129–139. doi:10.1037/h0042876
- Spielberger, C. D., and Gorsuch, R. L. (1983). *Manual for the State-Trait Anxiety Inventory (Form Y) "Self-Evaluation Questionnaire"*. Palo Alto, CA: Consulting Psychologists Press.
- Szabó, M. (2011). The emotional experience associated with worrying: anxiety, depression, or stress? *Anxiety Stress Coping* 24, 91–105. doi:10.1080/10615801003653430
- Torrents-Rodas, D., Fullana, M. A., Bonillo, A., Caseras, X., Andión, O., and Torrubia, R. (2013). No effect of trait anxiety on differential fear conditioning or fear generalization. *Biol. Psychol.* 92, 185–190. doi:10.1016/j.biopsycho.2012.10.006
- van der Ploeg, H. M. (2000). *Handleiding bij de Zelf-Beoordelings Vragenlijst. Een Nederlandstalige Bewerking van de Spielberger State-Trait Anxiety Inventory STAI-DY*. [Test Manual Dutch State Trait Anxiety Inventory]. Lisse: Swets & Zeitlinger.
- Vandorpe, S., and De Houwer, J. (2005). A comparison of forward blocking and reduced overshadowing in human causal learning. *Psychon. Bull. Rev.* 12, 945–949. doi:10.3758/BF03196790
- Woody, S., and Rachman, S. (1994). Generalized anxiety disorder (GAD) as an unsuccessful search for safety. *Clin. Psychol. Rev.* 14, 743–753. doi:10.1016/0272-7358(94)90040-X
- Zinbarg, R., and Mohlman, J. (1998). Individual differences in the acquisition of affectively valenced associations. *J. Pers. Soc. Psychol.* 74, 1024–1040. doi:10.1037/0022-3514.74.4.1024
- 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 January 2013; accepted: 08 May 2013; published online: 28 May 2013.
- Citation:** Arnaudova I, Kryptos A-M, Effting M, Boddez Y, Kindt M and Beckers T (2013) Individual differences in discriminatory fear learning under conditions of ambiguity: a vulnerability factor for anxiety disorders? *Front. Psychol.* 4:298. doi: 10.3389/fpsyg.2013.00298
- This article was submitted to *Frontiers in Personality Science and Individual Differences*, a specialty of *Frontiers in Psychology*.
- Copyright © 2013 Arnaudova, Kryptos, Effting, Boddez, Kindt and Beckers. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.