



# Identifying Causes of Social Evolution: Contextual Analysis, the Price Approach, and Multilevel Selection

Christoph Thies\* and Richard A. Watson

School of Electronics and Computer Science, Faculty of Engineering and Physical Sciences, University of Southampton, Southampton, United Kingdom

## OPEN ACCESS

### Edited by:

Peter Nonacs,  
University of California, Los Angeles,  
United States

### Reviewed by:

Samir Okasha,  
University of Bristol, United Kingdom  
Michael Taborsky,  
University of Bern, Switzerland

### \*Correspondence:

Christoph Thies  
c.thies@soton.ac.uk

### Specialty section:

This article was submitted to  
Social Evolution,  
a section of the journal  
Frontiers in Ecology and Evolution

**Received:** 21 September 2021

**Accepted:** 06 December 2021

**Published:** 24 December 2021

### Citation:

Thies C and Watson RA (2021)  
Identifying Causes of Social Evolution:  
Contextual Analysis, the Price  
Approach, and Multilevel Selection.  
*Front. Ecol. Evol.* 9:780508.  
doi: 10.3389/fevo.2021.780508

Kin selection theory and multilevel selection theory are distinct approaches to explaining the evolution of social traits. The latter claims that it is useful to regard selection as a process that can occur on multiple levels of organisation such as the level of individuals and the level of groups. This is reflected in a decomposition of fitness into an individual component and a group component. This multilevel view is central to understanding and characterising evolutionary transitions in individuality, e.g., from unicellular life to multicellular organisms, but currently suffers from the lack of a consistent, quantifiable measure. Specifically, the two major statistical tools to determine the coefficients of such a decomposition, the multilevel Price equation and contextual analysis, are inconsistent and may disagree on whether group selection is present. Here we show that the reason for the discrepancies is that underlying the multilevel Price equation and contextual analysis are two non-equivalent causal models for the generation of individual fitness effects (thus leaving different “remainders” explained by group effects). While the multilevel Price equation assumes that the individual effect of a trait determines an individual’s relative success within a group, contextual analysis posits that the individual effect is context-independent. Since these different assumptions reflect claims about the causal structure of the system, the correct approach cannot be determined on general theoretical or statistical grounds but must be identified by experimental intervention. We outline interventions that reveal the underlying causal structure and thus facilitate choosing the appropriate approach. We note that kin selection theory with its focus on the individual is immune to such inconsistency because it does not address causal structure with respect to levels of organisation. In contrast, our analysis of the two approaches to measuring group selection demonstrates that multilevel selection theory adds meaningful (falsifiable) causal structure to explain the sources of individual fitness and thereby constitutes a proper refinement of kin selection theory. Taking such refined causal structure into account seems indispensable for studying evolutionary transitions in individuality because these transitions are characterised by changes in the selection pressures that act on the respective levels.

**Keywords:** social evolution, multilevel selection, causal model, Price equation, kin selection

## 1. INTRODUCTION

When individual traits have effects on other individuals, individual fitness depends not only on self but also on the social environment, i.e., interaction partners. Kin selection theory (KS) deals with this problem by regarding the social environment as an external factor that, together with direct fitness effects of a trait, determines evolutionary dynamics with respect to selection. By assuming a certain correlation between trait value of an individual and average trait value of its social environment, e.g., through relatedness, Hamilton's rule can be formulated and answers the question of whether a trait with direct and indirect effects increases or decreases in frequency given the organisation of the population, i.e., the parameter of relatedness  $r$  (Frank, 1997). In short, KS acknowledges indirect effects (for which it was developed) but focuses on how relatedness affects individual fitness and is indifferent to the levels on which selection acts.

Multilevel selection theory (MLS) differs from this picture in that it posits the social environment as a unit, e.g., the group, that can be subject to selection acting at a level above that of individuals (Wilson, 1975; Wade, 1976, 1978; Uyenoyama and Feldman, 1980; Wilson and Sober, 1989). The theory thus promotes the concept of a group from a mere collection of individuals targeted by similar selection pressures to a unit that has a causal role in the selection process. More precisely, MLS theory understands a group as a unit whose interaction with the selective environment—through properties of the group as a whole—causally affects the fitness of its individual subunits (Wade and Kalisz, 1990). This means that individual fitness is a composite quantity determined by two factors: the individual effect of the trait and an effect on the group that an individual is a part of, and via this group effect, on the individual itself. The MLS view is not in opposition with KS but merely highlights that selection at the group level may be part of a causal mechanism resulting in individual fitness differences and must be taken into account if we want to understand the source of individual fitness differences (Dugatkin and Reeve, 1994; Sober and Wilson, 1994). Put differently, while KS is content with determining inclusive fitness at the individual level, MLS claims that individual traits can have effects that are best understood as group effects. Note that KS and MLS make the same predictions about trait-frequency dynamics on the individual level because selection at higher levels entails selection at lower levels and KS interprets all selection in individual terms. The explanatory goals of KS itself (Okasha, 2015; Marshall, 2016) derive largely from the goal of establishing inclusive fitness as a quantity maximised by evolution (Hamilton, 1964). Here, we refer to KS as a model that is free of assumptions regarding the level of selection in the sense that KS subsumes all selection at the individual level while MLS deviates from this model by assigning selection to several levels. While MLS aims to analyse the proximate causal structure of selection at multiple levels of organisation, KS establishes the direction of trait-frequency change based on individual fitness consequences of the trait and the relatedness structure of the population.

The distinction between individual effects and group effects of individual traits presents MLS with a problem not encountered

by KS: how can the presence of a group effect be detected empirically/statistically and how can the strength of the group effect be quantified in comparison to the individual effect of the trait. After all, the claim that group effects determine individual fitness can only be of use if such effects can be detected empirically. To give an example, Eldakar et al. (2010) claim that the fitness of male water striders *Aquarius remigis* organised into patches (also referred to as social environments or groups) depends on two components that are both affected by an aggressiveness trait individually expressed by the males. The individual component is given by the positive effect of aggressiveness on fitness mediated by mating success which is higher for more aggressive males that secure more mating opportunities than less aggressive males (Watters and Sih, 2005). The group component of individual fitness, on the other hand, arises from a different causal pathway and represents a negative effect of aggressiveness on fitness. Since the harassment experienced by females on a patch reflects the cumulative male aggression level on that patch and females tend to avoid harassment by escaping their current patch, the trait has a negative effect on patch productivity by decreasing the number of females on the patch and therefore the reproductive resources of all males on that patch. If such a decomposition into causes of individual fitness is to be useful, this decomposition must be empirically accessible in the sense that fitness is quantitatively given as a function of an individual component and a group component. This is possible only with a valid method of measuring the decomposition in empirical data.

Two methods for carrying out a quantitative decomposition of individual fitness into an individual component and a group component have received particular attention in the literature (Heisler and Damuth, 1987; Goodnight et al., 1992; Frank, 1998; Okasha, 2006; Sober, 2011; McLoone, 2015): the multilevel Price equation and contextual analysis which, following Okasha, we refer to as the “Price approach” and the “contextual approach,” respectively. However, the partitions of individual fitness given by the two methods are different in general. In particular, there are cases in which the multilevel Price equation claims the absence of group effects while contextual analysis claims their presence and vice versa.

The inconsistency between the two approaches is problematic because proponents of MLS argue that the distinction between individual effects and group effects is not just a statistical exercise but reflects a separation of causal pathways in the biological system under study as described above. While one causal pathway emanating from the individual trait is proposed to affect only individual aspects of fitness (the fitnesses of the bearer and its interaction partners), a different pathway is claimed to relate the trait with properties of the group as a whole and hence with a group component of individual fitness. Since the desired decomposition must reflect the underlying biological reality, two methods of decomposition that yield different answers cannot both be correct (Sober, 2011). Previous attempts at resolving these discrepancies have been inconclusive, leaving theorists and empiricists applying multilevel selection theory in the unfortunate situation that, even among proponents of multilevel selection theory, there is no unanimously agreed upon

method for measuring the strength of group selection in the simplest additive cases (Eldakar et al., 2010; Clarke, 2016). Given that multilevel selection is sometimes viewed as contentious in traditional evolutionary theory, and even the proponents of MLS have been unable to agree on a measure (and thus even unable to agree whether a group effect is present or not in a particular case), this may suggest that MLS is not well-understood and should be abandoned in favour of kin selection theory.

The aim of this paper is to show that the essential difference between the Price approach and contextual analysis lies in the causal structure each method posits as underlying the observed measurements of individual fitness. Briefly, while contextual analysis assumes that the individual component is determined by direct fitness effects of the trait only, the Price approach sees the individual component as a result of within-group competition and duly assumes it to be affected by the trait values of group mates. Put differently, contextual analysis assumes that the individual effect of the trait is absolute in the sense that it is independent of the social environment. The Price approach, on the other hand, assumes that the trait affects the competitiveness of its bearer so that its fitness effect is relative in the sense that it depends on the social environment. This difference leads to different remainders to be explained by group effects and thereby to different measurements of the strength of group selection. Recognising that the difference between the two approaches arises from a difference in the underlying model of reality enables us to see how to determine which of the two approaches is correct in a given case, i.e., the one whose underlying model reflects the causal structure of the system that is being studied. In particular, the applicability of the two approaches depends on the biological scenario at hand and cannot be made on theoretical grounds.

Our analysis demonstrates how the application of MLS to a biological scenario requires and formalises an understanding of the system that is not implied by KS. More precisely, MLS introduces a layer in the causal structure that cannot be deduced from the reduced theory. For evolutionary transitions in individuality, understanding this additional causal structure is crucial because how it changes over evolutionary time reflects the transition from independent individuals to integrated groups as selection shifts from the individual to the group level (Godfrey-Smith, 2009; Clarke, 2016).

This paper is organised as follows. First, we show that both contextual analysis and the Price approach can be interpreted in terms of causal graphs that describe how each of the approaches models the dependence of individual and group component of fitness on individual and group trait. We then compare the two approaches using these causal graphs. This allows us to illustrate very clearly why the two approaches give different answers with respect to the strength of group selection. In addition to identifying the source of the discrepancy, our analysis identifies an experimental intervention that reveals which, if any, of the two approaches is correct and shows that neither is always correct. Indeed, the correct approach depends on causal mechanisms in the biological system that cannot be determined based on the distribution of individual fitness over individual trait and group trait without experimental intervention.

## 2. MODEL

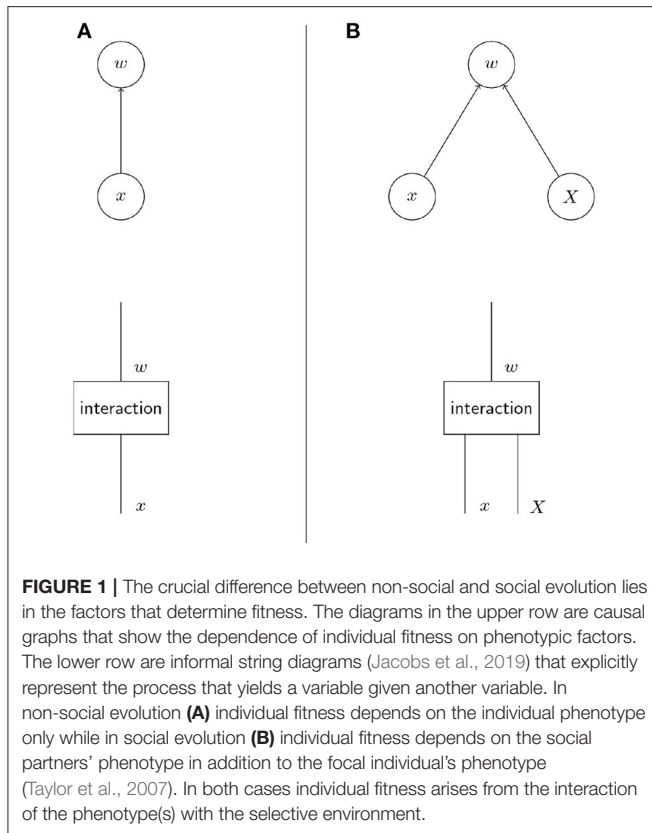
### 2.1. Fitness, Selection, and the Price Equation

The evolutionary model in which we frame our arguments is as simple as possible whilst being able to support the features we set out to discuss. Individuals are defined by their allele at a biallelic locus, with the two alleles representing the presence and absence of a trait, which also defines their phenotype (denoted by  $x$ ) and replicate asexually without mutation. A population of individuals is partitioned into non-overlapping groups of equal size such that an individual interacts equally with all members of its group (the assumptions on group size and disjointness are made for convenience only). The absolute fitness of an individual determines per capita growth rate and is a function of its own trait as well as the group trait, but not a function of other properties of the population (absence of, e.g., global frequency dependence; moreover we do not assume a mechanism of global population regulation). The group phenotype is defined as the average phenotype of the individuals within the group. The following considerations concern on bout of selection so that individuals reproduce within their given social environment. We therefore take the groups for granted and do not consider the mechanism of their formation. Taking a causalist stance, we assume that the fitness function is deterministic rather than a statistical abstraction from data (Otsuka, 2016) and stable in its functional form (i.e., the selective environment that determines fitness in interaction with the phenotype is not changing, Wade and Kalisz, 1990). Moreover, we assume that the fitness function is additive such that

$$w = c_1x + c_2X, \quad (1)$$

where  $w$  denotes the fitness of an individual with phenotype  $x$  and group phenotype  $X$ , and  $c_1, c_2 \in \mathbb{R}$  denote the coefficients of the functional representation of fitness in this simple hypothetical example (fitness and individual phenotype are centered at the population mean). This notation corresponds to the method of direct fitness or neighbour-modulated fitness in KS (Taylor et al., 2007).  $c_1x$  represents the direct fitness effect of the trait on its bearer,  $c_2X$  the indirect fitness effect on trait bearers' interaction partners. However, in contrast to approaches using inclusive fitness, we emphasise a causal viewpoint in this paper by regarding equations such as Equation (1) as structural equations that mirror causal assumptions about the system rather than as regression equations. The assumption of additivity in Equation (1) is a gross simplification that has been criticised because it ignores synergistic effects and therefore only applies to rare cases (Allen et al., 2013; Van Cleve and Akçay, 2014). The aim of this paper, however, is not biological generality but to demonstrate causal distinctions made by contextual analysis and the Price approach. The additive fitness function (1) suffices to show that the two approaches to multilevel selection discussed in this paper yield causally non-equivalent structural equations in this case.

The upper row of diagrams in **Figure 1** shows the crucial difference between non-social and social evolution in terms



of causal graphs. While individual fitness  $w$  depends only on the individual phenotype  $x$  in non-social evolution, social evolution introduces a dependency on social partners' phenotype  $X$  (Wolf et al., 1999). In Pearl's causal modelling framework, causal graphs as in the upper row of **Figure 1** represent the graphical counterpart of structural equations (Pearl, 2012). These structural equations model the causal process that yields the output variables, here  $w$ , given the input variables, here  $x$  and  $X$ . In contrast to multiple regression these structural equations reflect assumptions about the causal structure of the system. To make the causal process explicit we make informal use of string diagrams (Jacobs et al., 2019) to represent the causal graphs (i.e., Bayesian networks) in the lower row of diagrams in **Figure 1**. These diagrams are read from bottom to top and depict the variables as strings and the structural equations that transform the variables (i.e., the modelled processes) as boxes. The process of interaction of the phenotype with the selective environment is represented as box *interaction* that determines individual fitness based on individual and group phenotype. The diagrams in **Figure 1** are graphical and non-parametric versions of the structural Equation (1). Both the diagrams and the structural equation describe general features of the causal process proposed to determine the fitness of individuals with phenotype  $x$  within a group with phenotype  $X$ .

The process of selection in a population is given by the change in trait frequency according to the Price equation without

mutation

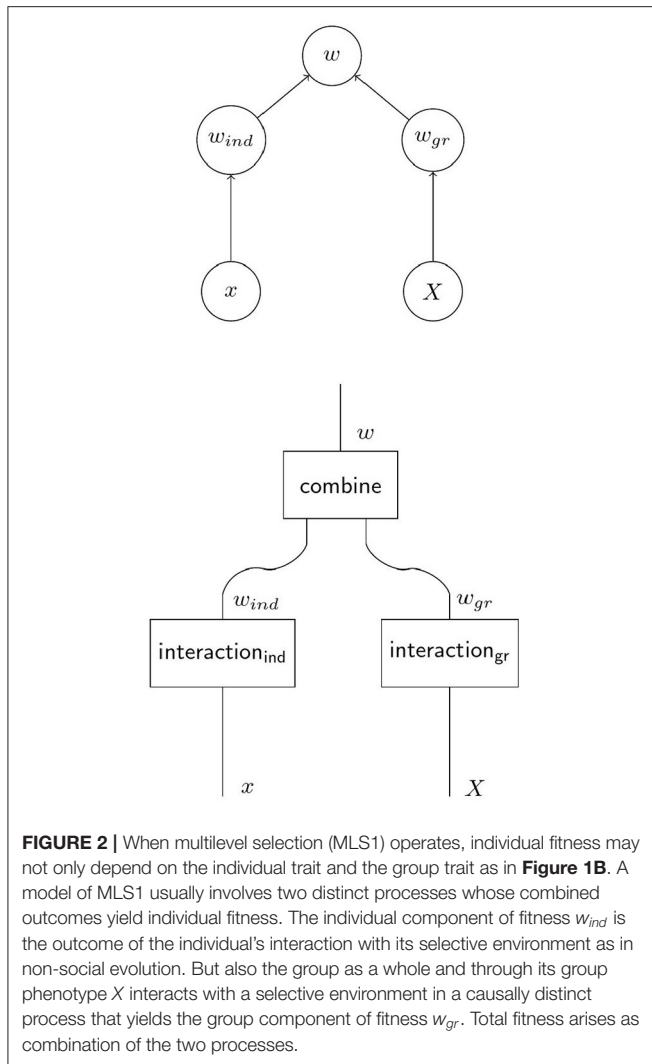
$$\bar{w}\Delta\bar{x} = \text{Cov}(w, x). \quad (2)$$

Note that we do not assume that groups themselves replicate or can be assigned group fitness over and above the fitness of the individuals that constitute a group. Our model is therefore of MLS1 type in the sense of Heisler and Damuth (1987), i.e., the focus of the analysis is on individuals, group trait and group fitness are averages of the corresponding quantities of the individuals within the group.

The starting point for the analysis of selection in a population in terms of MLS is the observation that an aspect of selection acts on groups as a whole. This means that individual selection is in part determined by the group trait  $X$  because selection favours groups with high (or low) group trait. In particular, this aspect of selection is the same for all members of a group and is captured by the process by which some groups contribute more offspring to the next generation than others due to differential proliferation and extinction (Uyenoyama and Feldman, 1980; Wade and Goodnight, 1998). Note that it makes no difference to the change in trait frequency whether (an aspect of) selection acts on the group as a whole or on all group members individually but in the same way. However, the aim of MLS is not only the prediction of outcomes but also the attainment of a causal understanding of the selection process (Sober and Wilson, 1994). This understanding comprises selective processes at the individual and the group level: individual fitness not only *depends* on a group trait in addition to the individual trait but fitness also *arises* as consequence of a process that affects the group as a whole in addition to a process that affects each individual specifically. Acknowledging the latter of these dual viewpoints is characteristic of multilevel selection theory as opposed to kin selection theory. The neighbour-modulated approach to kin selection, for instance, formalises the fitness effects of the social environment as factor that alters individual fitness but doesn't view the group as interacting with its own selective environment (Taylor et al., 2007).

Explanations for the evolution of cooperative traits, i.e., individual traits that are costly for their bearers in comparison with non-bearers, often rely on the interplay between two processes of this kind. In microcolonies of the bacterium *Pseudomonas aeruginosa*, for instance, the production of siderophores puts individuals at a fitness disadvantage because the process of producing the metabolite binds resources that could otherwise be used for reproduction (Weigert and Kümmerli, 2017). However, the secreted siderophores are shared within the microcolony and thus increase colony fitness due to their iron-scavenging function. Total fitness of individuals results as the combination of the two processes acting on the individual directly and on its group.

In **Figure 1**, the causal graph for social evolution in **Figure 2** is refined to a causal graph and a corresponding string diagram that explicitly represent the distinct processes of interaction with the selective environments on the individual level and the group level. The box "combine" in the lower diagram in **Figure 2** corresponds to a function that combines the outcomes



of the two processes into total individual fitness. For a complete specification of the model in terms of structural equations this function must be specified in addition to structural equations for the processes on the individual and the group levels. As we will argue in the next section 2.2, the causal models corresponding to contextual analysis and the Price approach combine their respective structural equations for the individual and group level interaction with the selective environment additively, so that

$$w = w_{ind} + w_{gr}. \quad (3)$$

The purpose of the decomposition in Equation (3) is to explicitly and formally acknowledge the basic tenet of MLS that fitness (here at the individual level) is determined not only by how the individual interacts with its selective environment but also by how the individual's group interacts with the selective environment on a level above that of the individual. We introduce  $w_{gr}$  to formally capture fitness effects that result from the interaction of the group as a whole with the selective environment. The quantities  $w_{ind}$  and  $w_{gr}$  are proxies for

the effects of causal processes, the former for processes that affect individual fitness specifically for each individual, the latter for processes that affect the group. The decomposition in Equation (3) is additive because of the simple fitness function chosen in Equation (1). While an MLS analysis always rests on a decomposition of fitness into contributions from various levels, this decomposition is, generally, not additive. The difference between the two approaches with respect to the arena of individual selection, which is the global population for contextual analysis and the local group for the Price approach, holds more generally, regardless of whether individual and group selection are combined additively.

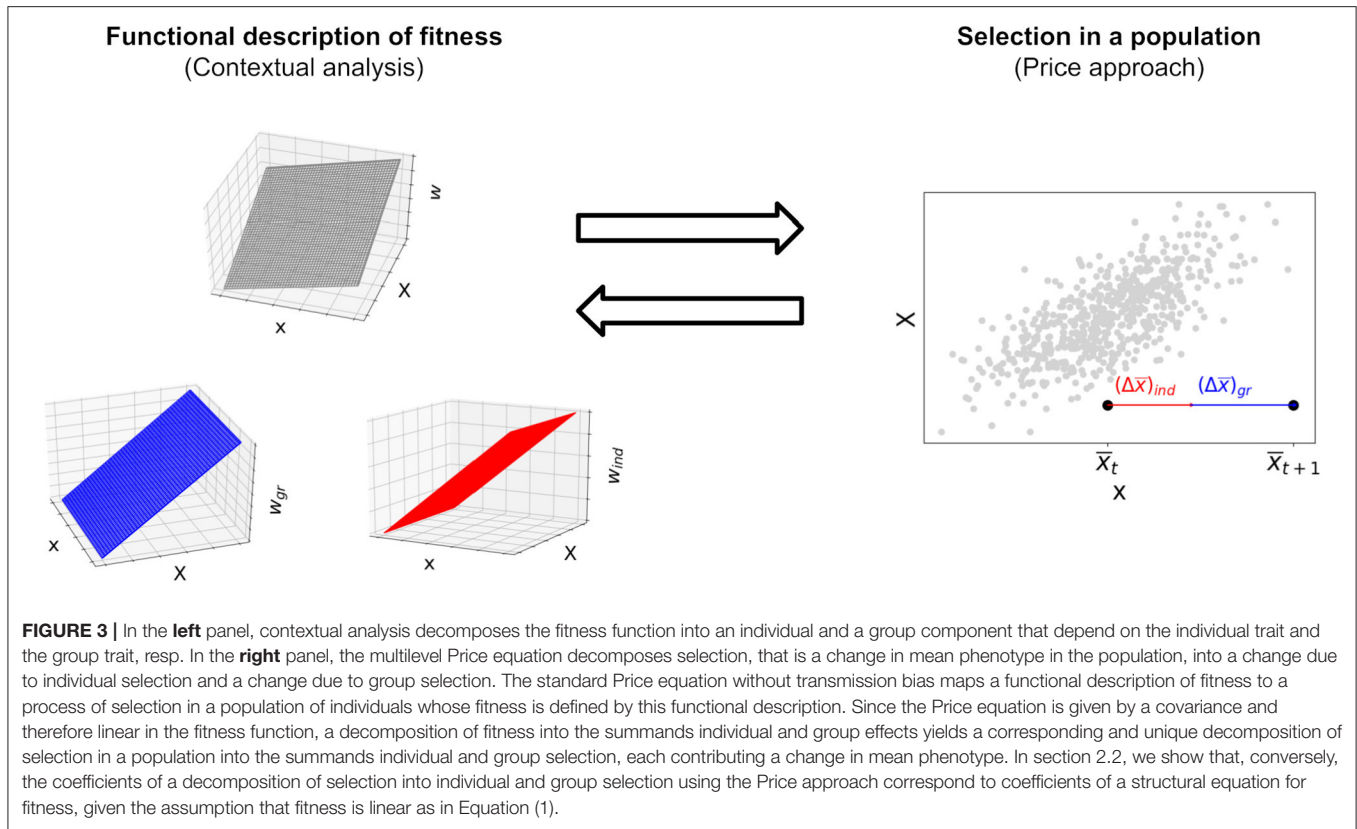
Since the Price equation is linear in the fitness argument, the decomposition expressed in Equation (3) corresponds to a decomposition of the strength of selection itself

$$\bar{w}\Delta\bar{x} = \bar{w}(\Delta\bar{x})_{ind} + \bar{w}(\Delta\bar{x})_{gr} = \text{Cov}(w_{ind}, x) + \text{Cov}(w_{gr}, x)$$

In order to make quantitative statements about the strengths of group selection vs. individual selection, an MLS analysis must determine the components in this decomposition. However, while individual trait and fitness as well as aggregates thereof can be measured directly, individual effect and group effect, or their covariance with the individual trait, are generally not amenable to direct measurement. The multilevel Price equation and contextual analysis are two methods of obtaining  $w_{ind}$  and  $w_{gr}$  by statistical means given individual traits and fitnesses (Okasha, 2006).

## 2.2. Contextual Analysis and the Price Approach

Equation (1) partitions individual fitness into the effects of the individual trait and the group trait. It describes how two phenotypic traits combine to yield another trait of the individual, namely absolute fitness. Contextual analysis (Heisler and Damuth, 1987; Okasha, 2006; note that here and the following we refer by "contextual analysis" to the standard structural equation with the untransformed variables  $x$  and  $X$ , as is customary in discussions on the issues reported here) takes effects of the group trait in Equation (1) as indicating group selection. Strictly speaking,  $c_2 \neq 0$  in Equation (1) implies the potential of the trait to undergo group selection conditional on the existence of group-trait variation between groups (Wolf et al., 1999, see McLoone, 2015 for a discussion of this difference). We regard group effects on fitness as more fundamental than a concept of group selection itself as the former do not depend on properties of a population but reflect causal processes that increase or decrease reproductive success of an individual situated in a group context vis-à-vis a specific selection regime that in turn determines individual fitness. Group effects can lead to group selection if, in a specific population, they generate fitness differences between individuals. This requires  $\text{Var}(X) \neq 0$ , for if  $\text{Var}(X) = 0$  all individuals have the same group trait and are therefore subject to the same group effects. Given a population of individuals and a fitness function that yields individual fitness as superposition of fitness effects of the variables that causally determine fitness, the Price equation yields



the effect of selection on this population, that is the change in mean phenotype over one generation (**Figure 3**).

The Price approach to multilevel selection (Price, 1972; Okasha, 2006; Gardner, 2015) rests on the partition of selection itself given by the multilevel expansion of the Price Equation (2)

$$\bar{w}\Delta\bar{x} = \text{Cov}(W, X) + E[\text{Cov}_{\text{wg}}(w, x)] \quad (4)$$

and posits that a population is undergoing group selection if the first term in Equation (4) is non-zero. In light of our remarks regarding group effects and group selection above, the Price approach and contextual analysis therefore decompose different quantities and are not directly comparable. However, this difference is superficial as the partition of fitness effects given by contextual analysis corresponds to a partition of selection and the partition of selection given by the multilevel Price equation corresponds to a partition of fitness effects. Contextual analysis, i.e., the functional representation of fitness in Equation (1), determines selection according to Equation (2) for a population that is partitioned into groups: given a population of individuals  $i \in 1, \dots, n$  with fitnesses

$$w_i = c_1 x_i + c_2 X_i,$$

where  $X_i$  is the trait of the group the  $i^{\text{th}}$  individual is part of, the change in mean trait value in the population follows from

Equation (2) as (Okasha, 2004),

$$\bar{w}\Delta\bar{x} = c_1 \text{Var}(x) + c_2 \text{Var}(X). \quad (5)$$

Thus the decomposition of fitness into individual and group effects given by contextual analysis corresponds to a decomposition of selection whose components, according to contextual analysis, represent the component of individual selection  $c_1 \text{Var}(x)$  and the component of group selection  $c_2 \text{Var}(X)$ . Conversely, the components of individual selection and group selection according to the Price approach for a population with non-vanishing variance within and between groups correspond to a decomposition of individual fitness into a component of individual effects and group effects. To see how, note that with  $w = c_1 x + c_2 X$  (Equation 1),

$$\text{Cov}(W, X) = (c_1 + c_2) \text{Var}(X) \quad (6)$$

(Okasha, 2006; p. 89). Using Equations (5) and (6), the decomposition according to Equation (4) is

$$\begin{aligned} \bar{w}\Delta\bar{x} &= \text{Cov}(W, X) \\ &+ E[\text{Cov}_{\text{wg}}(w, x)] = (c_1 + c_2) \text{Var}(X) + E[\text{Cov}_{\text{wg}}(w, x)] \\ &= c_1 \text{Var}(x) + c_2 \text{Var}(X) \end{aligned}$$

and therefore

$$E[\text{Cov}_{\text{wg}}(w, x)] = c_1 (\text{Var}(x) - \text{Var}(X)).$$

Hence the decomposition of fitness

$$w = c'_1(x - X) + c'_2X \quad (7)$$

corresponds to the decomposition of selection

$$\begin{aligned} \text{Cov}(w, x) &= \text{Cov}(c'_1(x - X) + c'_2X, x) = c'_1(\text{Var}(x) - \text{Var}(X)) \\ &\quad + c'_2\text{Var}(X), \end{aligned}$$

where  $c'_1 = c_1$  and  $c'_2 = c_1 + c_2$ .

Through this correspondence both contextual analysis and the Price approach yield decompositions of fitness effects as well as of selection (see **Table 1**). Note that the possibility of conducting contextual analysis with respect to the variables  $x - X$  and  $X$  rather than  $x$  and  $X$ —the former choice of variables being equivalent to the Price approach, the latter to contextual analysis—is discussed in Heisler and Damuth (1987) along with examples of circumstances under which this might be causally adequate.

It should be noted that while the interpretation of contextual analysis as structural equation based on causal assumptions is natural, the Price equation and its multilevel expansion are usually viewed as mere statistical identities. However, a causal interpretation of the coefficients of the multilevel Price equation as indicators of selection on the individual and group level as in (Okasha, 2006) requires the assumption of a causal process that gives rise to the measured coefficients. As demonstrated in this section above for the additive fitness function given by Equation (1) the structural equation for a process corresponding to the multilevel Price equation is given by Equation (7).

### 2.3. Causal Intuitions Underlying an MLS Analysis

A core idea of social evolution is that an individual trait of social organisms has fitness effects not only on its bearer but also on the social environment of the bearer. Common to paradigmatic examples of group selection is an individual trait with effects that change individual fitness homogeneously across the group such that these effects are best viewed as group effects (Sober, 1980). For the water striders described in Eldakar et al. (2010) the exodus of females from patches with high levels of aggressiveness is a group effect of the trait “aggressiveness” in males. This group effect is negative because group productivity is assumed to decrease with the number of females on a patch as females provide reproductive resources. On the other hand, aggressiveness has a positive individual effect because aggressive males secure more mating opportunities. Whether contextual analysis or the Price approach is appropriate depends on details of this latter mechanism. If, for instance, less aggressive males generally have lower reproduction rates due to female behaviour and independent of other males on the local patch the contextual analysis model may be more appropriate. If, on the other hand, male reproduction is subject to competition within patches where, for example, successful reproduction depends on the ability of males to guard their mates the model suggested by the Price approach may correspond more closely to the actual process. In either case, the causal interpretation of the trait refers

to proximate fitness effects of the trait and involves the individual as well as the group it is in but not other groups or the population as a whole. Therefore the causal interpretation takes place on the fitness side rather than on the selection side in **Figure 3**.

Since we assume that fitness is an effect of the individual/group trait an individual exhibits, we can read the equations in the left column of **Table 1** as structural equations that determine fitness given the traits. By the assumption on the additivity of interactions these equations are linear. The interpretation of structural equations is aided by the use of causal graphs, more precisely, directed acyclic graphs with causal rather than correlational interpretation (Pearl, 2009). **Figure 4** shows the causal graphs corresponding to the structural equations in **Table 1**. Since the components  $w_{\text{ind}}$  and  $w_{\text{gr}}$  reflect quantities that refer to processes occurring in the biological system studied, the causal graphs constitute models of the underlying reality. For example, the group effect of the aggressiveness trait in water striders is given by the propensity of females to remain on the focal patch and this propensity is a function of mean male aggressiveness in the patch (this function is linear by assumption), i.e., the group trait  $X$ . The non-equivalence of the causal graphs (**Figures 4A,B**) reflects a difference in how the individual/group components of individual fitness depend on individual/group trait. It should be noted that the factors  $x$  and  $X$  are not strictly independent as suggested by omitted arrows between  $x$  and  $X$  in **Figure 4**. Since the group phenotype is generated collectively by all individuals within a group,  $x$  does affect  $X$ . The arrows are omitted in **Figure 4** because our arguments focus on that part of the causal structure that determines fitness. Details of how the interaction of individual phenotypes gives rise to the group phenotype are not relevant for the present discussion.

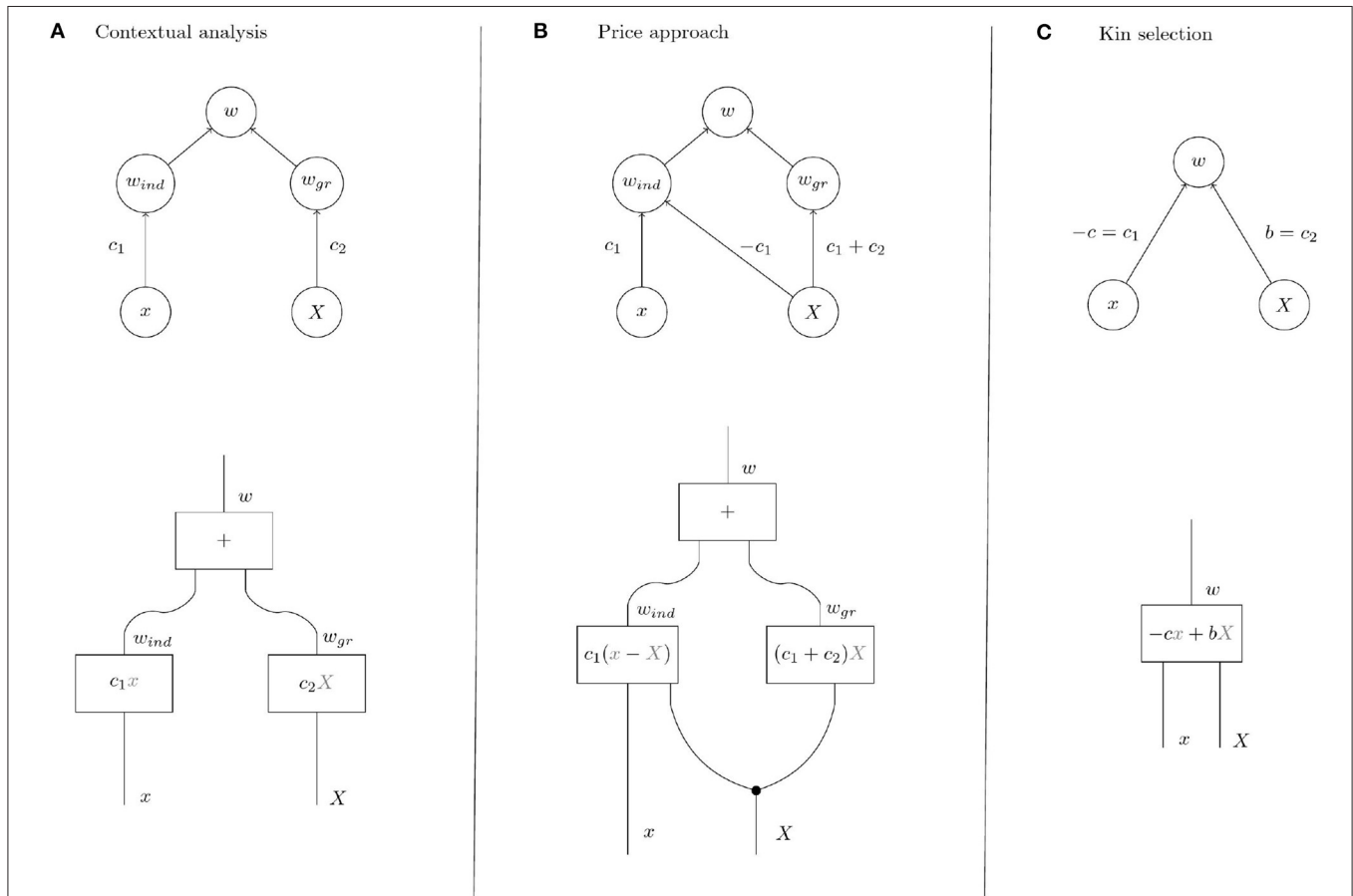
The model of fitness underlying contextual analysis (**Figure 4A**) is based on the assumption that the individual component and the group component of fitness are determined only by the individual trait and the group trait, respectively. This means that fitness differences within groups, i.e., differences in the individual component, are due to the individual trait and independent of the group trait. In that sense contextual analysis assumes the individual effects of the trait to be absolute, i.e., independent of group context. In contrast, the Price approach assumes that the group trait also affects the individual component of fitness in a specific way (see the path coefficients in **Figure 4**). This effect of the group trait on the individual component is equivalent to the assumption that fitness differences within groups are due to competition between group members in which the individual trait determines competitiveness of an individual. Indeed, the functional representation of fitness according to the Price approach from **Table 1**

$$w = c'_1(x - X) + c'_2X$$

shows that the individual component sums to zero over each group and that individuals with higher-than-average trait have a positive individual component (negative if  $c'_1 < 0$ ).

**TABLE 1** | The decompositions of contextual analysis and the Price approach as individual and group fitness effects  $w = w_{ind} + w_{gr}$  and as components of selection. The parameters of contextual analysis and the Price approach are linked by the equations  $c'_1 = c_1$  and  $c'_2 = c_1 + c_2$ .

	Fitness Individual component + group component	Selection Individual selection + group selection
Contextual analysis	$w = w_{ind} + w_{gr}$ $c_1x + c_2X$	$\Delta\bar{x} = (\Delta\bar{x})_{ind} + (\Delta\bar{x})_{gr}$ $c_1\text{Var}(x) + c_2\text{Var}(X)$
Price approach	$c'_1(x - X) + c'_2X$	$c'_1(\text{Var}(x) - \text{Var}(X)) + c'_2\text{Var}(X)$



**FIGURE 4** | Upper row: Causal graphs showing the interdependence of the variables  $x$  (individual trait),  $X$  (group trait, i.e., group mean of individual trait),  $w$  (individual fitness),  $w_{ind}$  (individual component of individual fitness), and  $w_{gr}$  (group component of individual fitness). Group fitness  $W$  is given as average over the individual fitnesses in a population. **(A)** Contextual analysis assumes an absolute individual effect of the trait. **(B)** In the Price approach, the trait is assumed to have a relative effect in the sense that the trait affects fitness depending on the trait expression of other members of the group. **(C)** In contrast, kin selection theory acknowledges the possibility of indirect effects in addition to direct effects but makes no further assumptions on the causal structure. In KS, it is customary to denote the direct effect of the trait on its bearer by  $-c$  and the indirect effect by  $b$ . The parameter of relatedness  $r$  represents the correlation between  $x$  and  $X$  and is not pictured in the graph because we focus on selection rather than on properties of group composition. Also the effect of individual phenotype on group phenotype has been omitted, see text. Lower row: String diagrams making the processes that yield the output variables given the input variables explicit. Inside the boxes are the structural equations that mirror the respective processes numerically. In diagrams **(A,B)**, the fitness effect of the group process  $int_{gr}$  is proportional to the group phenotype. In diagram **(B)** for the Price approach, the group phenotype  $X$  is “copied” at the black dot because it is involved in both processes. In  $int_{ind}$  instantiated with the structural equation  $c_1(x - X)$ , the group phenotype renders the effect of the individual phenotype relative to the group so that individuals, via their individual phenotype, compete within groups.

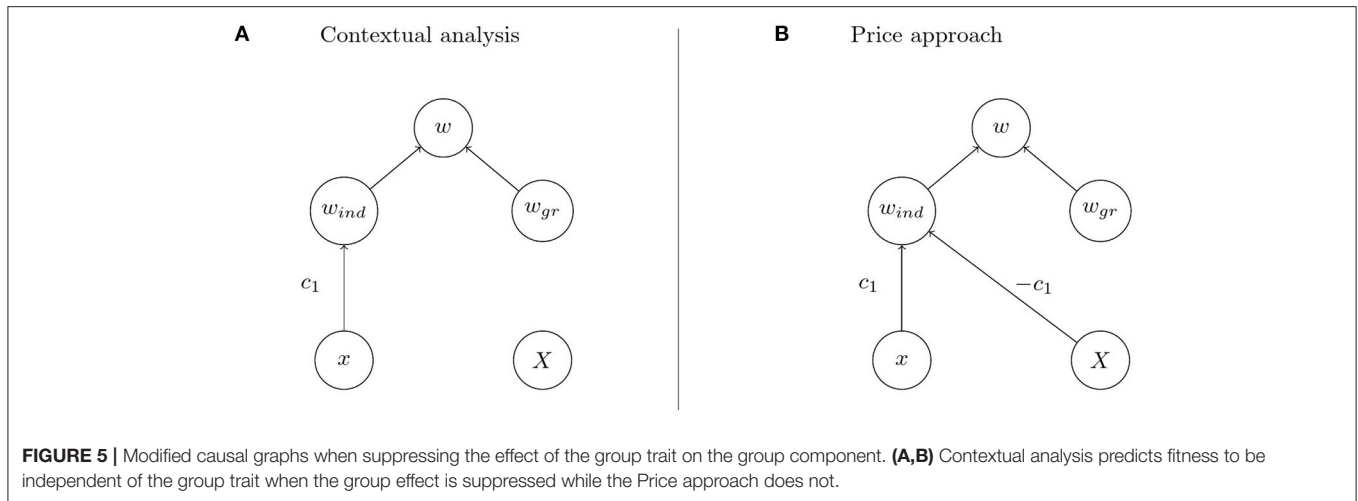
In other words, the trait affects individual fitness not by generally increasing or decreasing its bearer’s fitness but by increasing or decreasing its bearer’s competitive ability within the group. We discuss examples of these differences in the next section.

### 3. RESULTS

#### 3.1. Cases of Disagreement

When comparing the Price approach and contextual analysis it should be kept in mind that both aim to quantify group





selection and therefore start with the intuitive identification of an effect of the trait on a group-level property that affects fitness of all individuals within a group homogeneously. In the water strider example, an MLS analysis is based on the assertion, intuitively acquired by inspection of the empirical system, that the group mean of the trait “aggressiveness” affects the number of females in a group and therefore the productivity of the group as a whole. This assertion is independent of the subsequent choice of statistical approach to quantifying the strength of group selection. Contextual analysis and the Price approach therefore agree on the nature of the group effect on fitness  $w_{gr}$  and on the mechanism bringing forth this effect, though not on its magnitude. The difference between the two approaches lies in the question of which factors affect the individual component of fitness, i.e., which factors are responsible for within-group differences in fitness.

The problem cases for contextual analysis and the Price approach discussed by Okasha (2006) and others (Heisler and Damuth, 1987; Sober, 2011; Goodnight, 2015) reveal issues with the two approaches because the intuition about the level on which fitness effects occur is inconsistent with the verdict of one of the approaches with respect to the strengths of individual and group selection. This intuition is best understood in terms of fitness effects and not in terms of selection because it is based on a mechanism that mediates an effect of the trait on the group component of absolute fitness and is therefore independent of composition and organisation of the population as a whole. Changing a patch of water striders to exhibit a lower level of the group trait “aggressiveness” increases group fitness because less females will flee the patch. This causal explanation of the biological scenario is the core of an MLS analysis and it is independent of other patches and selection dynamics in the overall population. We conclude that the intuition with regard to the levels on which selection acts is about the mechanisms and not about frequency changes in the population. Accordingly, the following discussion is couched in terms of the left-hand side of **Figure 3**.

In the following examples, we determine the coefficients  $c_1, c_2$  of the kin selection model (**Figure 4C**) and discuss their interpretation in terms of the refined models provided by contextual analysis and the Price approach (models in **Figures 4A,B**).

### 3.1.1. Non-social Trait

A trait is non-social if the fitness of an individual does not depend on the trait values of its interaction partners (group mates) (Okasha, 2006) so that  $c_2 = 0$  and  $w = c_1x$  ( $c_1 \neq 0$  unless the trait is altogether neutral) in **Figure 4**. Intuitively, a trait of this type cannot be subject to group selection, because it has no fitness effects on its bearer’s interaction partners and therefore cannot affect the group component of fitness. However, the causal graph that represents the assumptions of the Price approach (**Figure 4B**) shows an effect of group trait on group component of fitness with weight  $c_1 + c_2 = c_1$  and therefore detects group selection where intuitively there is none. Group effects of this type have been called cross-level by-products (Okasha, 2006) and will be discussed in a later section. Note that the causal graph underlying contextual analysis correctly shows the absence of group effects.

### 3.1.2. Soft Selection

The tension between the Price approach and contextual analysis is reversed in the case of soft selection (Wade, 1985; Goodnight et al., 1992; Débarre and Gandon, 2011). Briefly, soft selection occurs in a group-structured population if mean individual fitness is homogeneous across groups, i.e., if all groups have the same reproductive output. Soft selection models situations in which individuals of each group share a fixed resource and the trait under soft selection determines how an individual fares in the within-group competition for this resource. The group trait determines competitiveness of the group, i.e., mean competitiveness of its members, in the sense that an individual has lower fitness in a competitive group than in a group with low group trait. Soft selection is intuitively considered to be free of group selection (Wade, 1985; Okasha, 2006; Sober, 2011).

The trait has no effect on the group level because changing the trait value of an individual in a group has no homogeneous fitness effect within the group as the change has no consequences for mean group fitness but merely changes the outcome of the within-group competition. It is easy to see that a kin selection model of soft selection takes the form  $w = c_1x - c_1X$ , i.e.,  $c_2 = -c_1$ , with  $c_1 > 0$  (respectively,  $c_1 < 0$ ) if a higher trait value implies higher (respectively, lower) competitiveness. The interpretation of these parameters according to the Price approach yields that the edge from  $X$  to  $w_{gr}$  has weight  $c_1 + c_2 = 0$  in the causal graph in **Figure 4B**. The Price approach correctly detects the absence of group selection in this example. However, contextual analysis mistakes the effect of the group trait on fitness as an effect on the group component of fitness according to the causal graph in **Figure 4A**.

Though most researchers that engaged with the problem of inconsistency between contextual analysis and the Price approach seem to agree that no group selection occurs in soft selection, some have argued to the contrary. Goodnight et al. (1992) regard soft selection as an example of group selection since an individual's fitness depends on the trait of the group of which it is a member. We agree that individual fitness depends on the group trait but this effect of the group trait on fitness is an individual effect (the diagonal arrow in **Figure 4B** targets  $w_{ind}$ ) that represents within-group competition. In soft selection, there is no group effect since the trait does not influence group fitness.

### 3.1.3. Genotypic Selection With Meiotic Drive

Okasha (2004) introduces “frameshifting” as a desirable property of a general theory of multilevel selection. The theory is capable of frameshifting if it formalises features of group selection in such a way that they hold by analogy whenever the hierarchy given by the group/individual relation is instantiated at other levels of organisation. The treatment of genotypic selection with meiotic drive in MLS terms is relevant in that context because it tests the ability of MLS to frameshift to levels below the level of organisms. Following Wilson (1990) and Okasha (2004) discusses diploid population genetics as an example of multilevel selection where alleles correspond to individuals and diploid genotypes to groups. In this analogy, group effects on allelic fitness are due to genotypic fitness, i.e., organismic fitness of the organism with a specific genotype, and individual effects are due to meiotic drive that creates within-group fitness differences between alleles.

Given the intuition that individual selection as well as group selection is at work in genotypic selection with meiotic drive, the expectation with respect to a decomposition of fitness into individual and group effects is clearly that group selection is present when genotypic fitnesses differ whereas individual selection is brought about by unfair meiosis. However, it is easy to see using specific fitness functions that contextual analysis doesn't agree with intuition in this case. When only meiotic drive is acting while there is no difference in fitness between genotypes, for example, the situation is analogous to soft selection that was shown above to be captured by the Price approach rather than contextual analysis. Furthermore, Okasha (2006) gives the example of two alleles  $A$  and  $B$  such that genotypic fitnesses are given by  $w_{AA} = 4$ ,  $w_{AB} = 3$ , and  $w_{BB} = 2$

while meiotic drive causes 2 of the 3 gametes produced by an  $AB$  organism to be  $A$ . Then fitness of an  $A$  allele is 2 and that of a  $B$  allele is 1, independent of the genetic background. Thus, despite unfair meiosis and dependence of fitness on the group trait, contextual analysis concludes the absence of group selection,  $c_1 \neq 0, c_2 = 0$ . The Price approach, in contrast, reaches the correct conclusion that individual fitness is given by  $w = \underbrace{c_1(x - X)}_{\text{ind. component}} + \underbrace{c_1X}_{\text{group component}}$  and therefore that both components of selection are non-zero.

Okasha's conclusion that the “covariance approach [i.e., the Price approach] appears to frameshift down quite well, the contextual approach very badly” (Okasha, 2004, p. 498) is thus readily explained by the viewpoint developed so far: unfair meiosis corresponds to the zero-sum game of within-group competition. This is precisely the causal structure assumed by the Price approach.

## 3.2. Cross-Level By-Products

A core assumption of MLS theory is that a trait an individual expresses may affect properties of its group as a whole and therefore group fitness (i.e., mean individual fitness in a group). This effect is captured by the group component  $w_{gr}$  of individual fitness. However, group fitness, in MLS1, is the average individual fitness in a group and therefore comprises not only the group component but also the average individual component  $w_{ind}$ . This is problematic because the part of group fitness that entails selection on the group property caused by the trait is  $w_{gr}$  only. The contribution of  $w_{ind}$  to group fitness is called a cross-level by-product (Okasha, 2006) because it represents fitness of the individuals that constitute the group, i.e., the lower level, rather than fitness that is a property of the group as a whole, i.e., the higher level. Intuitively, a group with many individually fit members seems more fit than a group with few individually fit members even when the group component  $w_{gr}$  and therefore the fitness vis-à-vis group selection that is to be quantified is the same for both groups. The non-social trait case discussed above is a good example of this effect. Since there is no group property for group selection to act on in this case, group fitness comprises solely of individual fitness from the level below and therefore consists entirely of cross-level by-products.

To see how contextual analysis and the Price approach handle cross-level by-products assume that individual fitness is given by the expression  $w = c_1x + c_2X$ . The decomposition of group fitness  $W = (c_1 + c_2)X$  into a component due to group effects and a component due to individual effects depends on the causal structure and therefore differs between the two approaches. While contextual analysis partitions group fitness into individual and group component as  $W = \underbrace{c_1X}_{\text{ind. component}} + \underbrace{c_2X}_{\text{group component}}$  the decomposition according to the Price approach yields only a group component,  $W = (c_1 + c_2)X$ . For a non-social trait ( $c_1 \neq 0, c_2 = 0$ ) the Price approach mistakenly traces group fitness entirely back to a non-existing group effect, whereas contextual analysis correctly assigns group fitness to the individual effect. The fact that contextual analysis handles cross-level by-products

correctly in the non-social trait case has led Okasha to conclude that contextual analysis is “on balance preferable” (Okasha, 2006, p. 99) to the Price approach. However, it should be noted that in the soft selection case ( $c_1 = -c_2$ ) contextual analysis decomposes group fitness as

$$W = \underbrace{c_1 X}_{W_{\text{ind}}} - \underbrace{c_1 X}_{W_{\text{gr}}}$$

and hence detects cross-level by-products of magnitude  $c_1 X$  even though cross-level by-products are absent since the individual components of fitness sum to zero in each group.

In their study on multilevel selection in water striders, Eldakar et al. (2010) choose contextual analysis for quantifying group selection because contextual analysis controls for “potential cross-level byproducts” (Eldakar et al., 2010; p. 3186). However, as we have seen, contextual analysis does not correctly account for cross-level by-products automatically. Which of the two approaches is correct depends on the kind of individual selection that acts on the system, i.e., the causal structure underlying fitness. In this case, the causal graphs in **Figures 4A,B** both seem possible. Recall that aggressiveness in male water striders is hypothesized to have an effect on the individual component of fitness (aggressive males secure more mating opportunities than non-aggressive males) and on the group component of fitness (patches with higher aggression levels have fewer females). Contextual analysis assumes that the individual component is independent of the group trait: in addition to the group component shared by all males in a group each male has an individual component that is determined by its trait and independent of the group trait. Another, perhaps more plausible, assumption underlies the Price approach: the group trait determines the number of females on a patch and this reproductive resource is distributed to the males according to their competitiveness. We will discuss an experimental intervention that would reveal the correct underlying causal structure in the next section.

### 3.3. Determining the Preferable Approach

Several authors have discussed the question which of the two approaches is preferable in general (Okasha, 2006; Sober, 2011; McLoone, 2015). However, even the most extensive discussion of this question (Okasha, 2006) has been inconclusive in the sense that in light of the problematic cases discussed above neither can be endorsed unreservedly. We argued that a general preference cannot be justified as the essential difference between the two approaches lies in non-equivalent assumptions about the causal structure of the biological system which, as the problematic cases demonstrate, may be either of the two. However, our reduction of the difference between the Price approach and contextual analysis to a difference between their respective causal graphs has the benefit that experimental interventions that reveal the correct causal structure and with it the correct approach can easily be derived from the causal graphs (Pearl, 2009). Note that while we argue that the suggested interventions in principle reveal the correct structure we do not claim that such interventions are feasible for a given biological system. Moreover, while the two

approaches discussed here are the main approaches to measuring the strength of group selection, it may well be possible that neither is suitable in a given scenario. We will discuss this and other limitations of this work below.

Imagine that we have a biological system such as a population of water striders in Eldakar et al. in which intuitive inspection suggests that individual fitness depends on an individual component and a group component as in **Figures 4A,B**. Analysis reveals proposed causal pathways for individual trait and group trait to affect individual fitness via the two components. In particular, such an analysis comprises a hypothesis on the mechanism that mediates the effect of the group trait on the group component of individual fitness. For the water strider example the group trait is mean aggressiveness on a patch, the group component is proportional to the number of females on a patch, and the mechanism that mediates the effect of the former on the latter is female exodus determined by the females’ preference for low aggressiveness patches. Choosing contextual analysis or the Price approach for quantification goes hand in hand with the commitment to regard **Figures 4A,B**, respectively, as the causal structure underlying the observed phenomena. The causal structures posited by the two approaches differ in that the Price approach assumes an effect of the group trait on the individual component of fitness. This assumption is reflected in the diagonal arrow in **Figure 4B** that is missing in **Figure 4A**. The two arrows emanating from  $X$  in **Figure 4B** represent two distinct cause-effect relations between the group trait and individual fitness. But given the hypothesis on the mechanism that mediates the effect of the group trait on the group component of fitness these two distinct cause-effect relations correspond to two distinct mechanisms through which the group trait affects fitness. Consequently, it is in principle possible to separate the effects by intervening on one of the mechanisms but not the other. This intervention translates to removing the vertical arrows from  $X$  to  $w_{\text{gr}}$  in **Figures 4A,B** so that the system is described by the modified graphs in **Figure 5**. But in the system with suppressed group effects the two causal structures in **Figures 5A,B** can be distinguished on the basis of the observable quantities  $x$ ,  $X$ , and  $w$ . In particular, contextual analysis predicts individual fitness to be independent of group membership when the system is being intervened on in this way. The Price approach, however, predicts continued dependence of fitness on group trait due to within-group competition. As these predictions cannot both be true, the intervention allows the identification of one of the two approaches as being in accord with experimental observations.

Corresponding to the causal models expressed in the graphs are mechanisms corresponding to each of the arrows in the model. In the water strider example, given the mechanisms corresponding to the arrows in **Figure 4**, it is now easy to see how a decision for one of the two approaches may be reached. Since the effect of group trait on fitness is mediated by female exodus, the effect can be suppressed by preventing females from leaving patches, i.e., by removing female dispersal between patches (Eldakar et al., 2009). It is crucial that this intervention leaves the diagonal arrow in **Figure 5B** intact. This is because the diagonal arrow represents a different causal pathway, namely the within-patch competition for females which is not affected by preventing

females from leaving the patch. An informed decision for contextual analysis can then be reached if fitness is independent of mean aggressiveness on a patch when female dispersal is removed, i.e., if the diagonal arrow in **Figure 5B** was not part of the underlying causal structure in the unperturbed system. The Price approach is more appropriate if fitness still depends on patch composition under this experimental condition.

Both the Price approach and contextual analysis serve the purpose to determine the quantities  $w_{ind}$  and  $w_{gr}$  in Equation (3), or equivalent quantities (see **Table 1**), from the more easily measurable variables individual trait and individual fitness. In order to achieve this, both approaches require assumptions that can be conveniently represented in terms of causal graphs as in **Figure 4**. We have shown above how, in principle, it is possible to determine which of the two approaches is more appropriate. However, we have seen that the causal structures posited are highly contrived. It seems therefore very well possible that neither of the two approaches is suitable for determining the level-specific strength of selection. This is the case when neither of the causal graphs in **Figures 4A,B** represent the causal structure underlying the biological phenomenon in question.

### 3.4. Multilevel Selection and Evolutionary Transitions in Individuality

The treatment of multilevel selection in this paper is motivated by the role multilevel selection plays in evolutionary transitions in individuality. When evolutionary units form higher level individuals, selection acts on both levels and the evolution of higher-level function is both facilitated and constrained by lower-level selection (Okasha, 2006; Godfrey-Smith, 2009; Watson and Thies, 2019). The longer-term outcome of selection depends on the interplay between the effects on the individual level and the group level, mediated by population structure. In order to understand how these factors act together, (Okasha, 2006) suggests a diachronic approach to evolutionary transitions that allows regarding the transition of a given system as entity whose parameters change over evolutionary time while it is “in progress.” From this perspective, the species of volvocine algae, for instance, that appear to be in between single-celled species and species with faithful reproduction on the group level may be seen to be similar to ancestors of obligately multicellular species, thus offering potential snapshots of the evolutionary history of evolutionary transitions in individuality (Michod, 2007). Indicators of the status of a transition as suggested in Clarke (2016) that rely on the quantification of selection pressures require causally adequate models. The refined causal viewpoint suggested in this paper allows analysis of the drivers of these changes and their dependency on the evolving system itself.

The focus of this paper is on a specific issue regarding the assignment of selection to the levels involved. The composition of groups—while central for their phenotypic properties—is taken for granted, and the models do not include a mechanism of how the groups arise. However, unlike in models of kin selection, groups are formalised as units that partake in causal processes and are assigned properties that may change over evolutionary time. In particular, processes that determine the

composition of groups may be subject to selection on multiple levels. Hamilton’s relatedness index measures how far the trait distribution within groups deviates from the distribution that would be expected in a randomly mixing population. While the description of population structure in terms of Hamilton’s index—like the description of genetic populations in terms of linkage disequilibria—is useful, it captures the notion of a population that is subdivided into groups only indirectly. The causal viewpoint described in this paper acknowledges the levels on which selection acts by promoting groups from environments of individuals to units that themselves interact with a selective environment and thus affect reproduction of their parts in a homogeneous manner. The structure of the population is then a consequence of the mechanisms with which offspring of grouped individuals in a parent generation form groups of individuals in an offspring generation. Since parameters of the mechanisms may evolve, the population structure changes over evolutionary time. Unlike Hamilton’s parameter of relatedness that is usually assumed constant for a given system, this perspective of multilevel selection allows modelling adaptations with respect to inheritance on the group level, as required for completed evolutionary transitions in individuality.

The assumption of an additive fitness function in this paper (Equation 1) is required for our argument regarding the difference in causal structure between the two approaches to multilevel selection. However, the distinction between the arenas of selection made by contextual analysis and the Price approach as well as the possibility of a group trait affecting individual selection within groups is of more general interest. Conceptualising interaction partners as part of the selective environment is central for understanding transitions in individuality as social niche construction processes in which individuals evolve to constitute and experience niches in their social environment (Powers et al., 2011; Ryan et al., 2016). During these processes, individual selection that initially acts across a population of loosely interacting individuals becomes increasingly determined by the social environment and thus internal to groups. In a similar vein, Araya-Ajoy et al. (2020) emphasise the role of feedback between selection pressures on traits that determine, for instance, phenotypic response to and impact on the social environment. Since bearers of these traits are not only subject to selection but also constitute the selective environment of their social partners the evolutionary response depends on the interplay between multiple evolving factors. Unravelling these intertwined processes theoretically and empirically requires careful analysis of the causal factors that control them as well as formal tools for their statistical analysis in biological systems. Moreover, the distinction regarding the arena of individual selection made by the two approaches to multilevel selection discussed in this paper seems crucial for understanding the interplay of ecology and evolution that occurs when populations adapt to the (social) environment they themselves create.

Another avenue of research is the role of external ecological factors in evolutionary transitions in individuality. Motivated by the question how selection on an emerging higher level may arise before it attains individual-like properties, Black et al. (2020)

discuss models of selection processes in which external factors scaffold selection on a new level. In these models, ecological factors provide the opportunity for individual cells in patchily distributed groups to determine fitness on the group level by scaffolding a pathway for the evolution of germ-like cells that reproduce the group by dispersal. The resulting division of labour between soma- and germ-like cells relies on the interplay of selection between and within groups. Including many causal factors, as in the models in Araya-Ajoy et al. (2020) and Black et al. (2020), makes it difficult to keep track of the effects they exert on each other. The causal diagrams discussed in this paper can be extended by external factors and allow the composition of “modules” of interactions between factors. We believe that a causal modelling approach with expressive diagrammatic formalism could be useful not only for the intuitive understanding but also for ensuring consistency and plausibility of these complex models.

We regard the results presented in this paper as first steps of formalising the interaction of selective forces on multiple levels in terms of causal structures. Future work combining these results with social niche construction and other mechanisms of social evolution may examine how causal structures arising in this way are themselves subject to selection and therefore to change over evolutionary time.

## 4. CONCLUSION

Group selection refines kin selection by splitting individual fitness into two components, i.e., by assuming that fitness is determined by two additional factors that are themselves determined by the variables individual trait and group trait. The causal graphs in **Figure 4** show that this means that group selection adds a layer to the causal structure of selection assumed by kin selection. This addition constitutes a proper refinement of kin selection and corresponds to avoiding averaging over the causes of individual fitness (the “averaging fallacy” described by Sober and Wilson, 1999). From this viewpoint, the tension between contextual analysis and the Price approach can be seen as an instance of the purely formal problem of connecting an additional layer of nodes to an existing graph. The connection schemes proposed by contextual analysis and the Price approach, i.e., the coefficients of the paths targeting  $w_{ind}$  and  $w_{gr}$  in **Figure 4**, are two solutions to this problem. Since omitted paths in a causal graph represent hypotheses about the absence of effects the correct approach is the approach whose hypotheses are

satisfied in the biological system at hand. Phrasing the problem in terms of causal graphs demonstrates that, even in the additive case, other refinements are in principle possible and could apply to scenarios in which the individual component is given neither by soft selection (Price approach) nor by hard selection (contextual analysis) but by intermediate selection regimes (Débarre and Gandon, 2011). Casting an MLS analysis in terms of refinements of causal graphs gives a formal argument for the non-equivalence of MLS and kin selection. We have argued that the refinement introduced by MLS is non-trivial (see difficulties with Price approach and contextual analysis) and provides a view on the system that is tailored to the levels of organisation in the system. This view is crucial when cause-effect relations that pertain to a specific level are manipulated or undergo change, such as during an evolutionary transition in individuality, and the system-level consequences of such alterations are to be predicted. Strengthening the formal core of MLS not only facilitates the application of MLS in evolutionary science but also aids in assessing benefits, limitations, and formal requirements of this approach to empirical and theoretical biological scenarios.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

CT and RW contributed to conception of the paper. CT wrote the manuscript. Both authors contributed to the article and approved the submitted version.

## FUNDING

CT received UKRI funding (award reference 1863169). This work was partially supported by Grant 62230 from the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

## ACKNOWLEDGMENTS

The authors thank Jamie Caldwell, Freddie Nash, and David Prosser for helpful discussions.

## REFERENCES

- Allen, B., Nowak, M. A., and Wilson, E. O. (2013). Limitations of inclusive fitness. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20135–20139. doi: 10.1073/pnas.1317588110
- Araya-Ajoy, Y. G., Westneat, D. F., and Wright, J. (2020). Pathways to social evolution and their evolutionary feedbacks. *Evolution* 74, 1894–1907. doi: 10.1111/evo.14054
- Black, A. J., Bourrat, P., and Rainey, P. B. (2020). Ecological scaffolding and the evolution of individuality. *Nat. Ecol. Evol.* 4, 426–436. doi: 10.1038/s41559-019-1086-9
- Clarke, E. (2016). A levels-of-selection approach to evolutionary individuality. *Biol. Philos.* 31, 893–911. doi: 10.1007/s10539-016-9540-4
- Débarre, F., and Gandon, S. (2011). Evolution in heterogeneous environments: between soft and hard selection. *Am. Nat.* 177, E84–E97. doi: 10.1086/658178
- Dugatkin, L. A., and Reeve, H. K. (1994). “Behavioral ecology and levels of selection: dissolving the group selection controversy,” in *Advances in the Study of Behavior*, Vol. 23, eds P. J. B. Slater, J. S. Rosenblatt, C. T. Snowdon, and M. Milinski (San Diego, CA: Academic Press), 101–133. doi: 10.1016/S0065-3454(08)60352-6

- Eldakar, O. T., Dlugos, M. J., Pepper, J. W., and Wilson, D. S. (2009). Population structure mediates sexual conflict in water striders. *Science* 326:816. doi: 10.1126/science.1180183
- Eldakar, O. T., Wilson, D. S., Dlugos, M. J., and Pepper, J. W. (2010). The role of multilevel selection in the evolution of sexual conflict in the water strider *aquarius remigis*. *Evolution* 64, 3183–3189. doi: 10.1111/j.1558-5646.2010.01087.x
- Frank, S. A. (1997). The Price equation, Fisher's fundamental theorem, Kin selection, and causal analysis. *Evolution* 51, 1712–1729. doi: 10.1111/j.1558-5646.1997.tb05096.x
- Frank, S. A. (1998). *Foundations of Social Evolution*. Princeton, NJ: Princeton University Press. doi: 10.1515/9780691206820
- Gardner, A. (2015). The genetical theory of multilevel selection. *J. Evol. Biol.* 28, 305–319. doi: 10.1111/jeb.12566
- Godfrey-Smith, P. (2009). *Darwinian Populations and Natural Selection*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199552047.001.0001
- Goodnight, C. J. (2015). Multilevel selection theory and evidence: a critique of Gardner, 2015. *J. Evol. Biol.* 28, 1734–1746. doi: 10.1111/jeb.12685
- Goodnight, C. J., Schwartz, J. M., and Stevens, L. (1992). Contextual analysis of models of group selection, soft selection, hard selection, and the evolution of altruism. *Am. Nat.* 140, 743–761. doi: 10.1086/285438
- Hamilton, W. D. (1964). The genetical evolution of social behaviour. I. *J. Theoret. Biol.* 7, 1–16. doi: 10.1016/0022-5193(64)90038-4
- Heisler, I. L., and Damuth, J. (1987). A method for analyzing selection in hierarchically structured populations. *Am. Nat.* 130, 582–602. doi: 10.1086/284732
- Jacobs, B., Kissinger, A., and Zanasi, F. (2019). “Causal inference by string diagram surgery,” in *Lecture Notes in Computer Science*, eds M. Bojańczyk and A. Simpson (Cham: Springer International Publishing), 313–329. doi: 10.1007/978-3-030-17127-8\_18
- Marshall, J. A. R. (2016). What is inclusive fitness theory, and what is it for? *Curr. Opin. Behav. Sci.* 12, 103–108. doi: 10.1016/j.cobeha.2016.09.015
- McLoone, B. (2015). Some criticism of the contextual approach, and a few proposals. *Biol. Theory* 10, 116–124. doi: 10.1007/s13752-015-0209-z
- Michod, R. E. (2007). Evolution of individuality during the transition from unicellular to multicellular life. *Proc. Natl. Acad. Sci. U.S.A.* 104(Suppl. 1), 8613–8618. doi: 10.1073/pnas.0701489104
- Okasha, S. (2004). Multi-level selection, covariance and contextual analysis. *Brit. J. Philos. Sci.* 55, 481–504. doi: 10.1093/bjps/55.3.481
- Okasha, S. (2006). *Evolution and the Levels of Selection*. New York, NY: Clarendon Press. doi: 10.1093/acprof:oso/9780199267972.001.0001
- Okasha, S. (2015). The relation between kin and multilevel selection: an approach using causal graphs. *Brit. J. Philos. Sci.* 67, 435–470. doi: 10.1093/bjps/axu047
- Otsuka, J. (2016). Causal foundations of evolutionary genetics. *Brit. J. Philos. Sci.* 67, 247–269. doi: 10.1093/bjps/axu039
- Pearl, J. (2009). *Causality*. New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511803161
- Pearl, J. (2012). “The causal foundations of structural equation modeling,” in *Handbook of Structural Equation Modeling*, ed R. H. Hoyle (New York, NY: Guilford Press), 68–91. doi: 10.21236/ADA557445
- Powers, S. T., Penn, A. S., and Watson, R. A. (2011). The concurrent evolution of cooperation and the population structures that support it. *Evolution* 65, 1527–1543. doi: 10.1111/j.1558-5646.2011.01250.x
- Price, G. R. (1972). Extension of covariance selection mathematics. *Ann. Hum. Genet.* 35, 485–490. doi: 10.1111/j.1469-1809.1957.tb01874.x
- Ryan, P. A., Powers, S. T., and Watson, R. A. (2016). Social niche construction and evolutionary transitions in individuality. *Biol. Philos.* 31, 59–79. doi: 10.1007/s10539-015-9505-z
- Sober, E. (1980). Holism, individualism, and the units of selection. *Proc. Biennial Meet. Philos. Sci. Assoc.* 1980, 93–121. doi: 10.1086/psaprobiennmeetp.1980.2.192588
- Sober, E. (2011). Realism, conventionalism, and causal decomposition in units of selection: reflections on Samir Okasha's evolution and the levels of selection. *Philos. Phenomenol. Res.* 82, 221–231. doi: 10.1111/j.1933-1592.2010.0471.x
- Sober, E., and Wilson, D. S. (1994). A critical review of philosophical work on the units of selection problem. *Philos. Sci.* 61, 534–555. doi: 10.1086/289821
- Sober, E., and Wilson, D. S. (1999). *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.
- Taylor, P. D., Wild, G., and Gardner, A. (2007). Direct fitness or inclusive fitness: how shall we model kin selection? *J. Evol. Biol.* 20, 301–309. doi: 10.1111/j.1420-9101.2006.01196.x
- Uyenoyama, M., and Feldman, M. W. (1980). Theories of kin and group selection: a population genetics perspective. *Theoret. Popul. Biol.* 17, 380–414. doi: 10.1016/0040-5809(80)90033-7
- Van Cleve, J., and Akçay, E. (2014). Pathways to social evolution: reciprocity, relatedness, and synergy. *Evolution* 68, 2245–2258. doi: 10.1111/evo.12438
- Wade, M. J. (1976). Group selections among laboratory populations of *Tribolium*. *Proc. Natl. Acad. Sci. U.S.A.* 73, 4604–4607. doi: 10.1073/pnas.73.12.4604
- Wade, M. J. (1978). A critical review of the models of group selection. *Q. Rev. Biol.* 53, 101–114. doi: 10.1086/410450
- Wade, M. J. (1985). Soft selection, hard selection, kin selection, and group selection. *Am. Nat.* 125, 61–73. doi: 10.1086/284328
- Wade, M. J., and Goodnight, C. J. (1998). Perspective: the theories of Fisher and Wright in the context of metapopulations: when nature does many small experiments. *Evolution* 52, 1537–1553. doi: 10.1111/j.1558-5646.1998.tb02235.x
- Wade, M. J., and Kalisz, S. (1990). The causes of natural selection. *Evolution* 44:1947. doi: 10.2307/2409605
- Watson, R. A., and Thies, C. (2019). “Are developmental plasticity, Niche construction, and extended inheritance necessary for evolution by natural selection?” in *The Role of Active Phenotypes in the Minimal Criteria for Darwinian Individuality*, eds T. Uller and K. N. Laland (Cambridge, MA: MIT Press), 197–226.
- Watters, J. V., and Sih, A. (2005). The mix matters: behavioural types and group dynamics in water striders. *Behaviour* 142, 1417–1431. doi: 10.1163/156853905774539454
- Weigert, M., and Kümmerli, R. (2017). The physical boundaries of public goods cooperation between surface-attached bacterial cells. *Proc. R. Soc. B Biol. Sci.* 284:20170631. doi: 10.1098/rspb.2017.0631
- Wilson, D. S. (1975). A theory of group selection. *Proc. Natl. Acad. Sci. U.S.A.* 72, 143–146. doi: 10.1073/pnas.72.1.143
- Wilson, D. S. (1990). Weak altruism, strong group selection. *Oikos* 59:135. doi: 10.2307/3545133
- Wilson, D. S., and Sober, E. (1989). Reviving the superorganism. *J. Theoret. Biol.* 136, 337–356. doi: 10.1016/S0022-5193(89)80169-9
- Wolf, J. B., Brodie, E. D. III, and Moore, A. J. (1999). Interacting phenotypes and the evolutionary process. II. Selection resulting from social interactions. *Am. Nat.* 153, 254–266. doi: 10.1086/303168

**Conflict of Interest:** 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.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Thies and Watson. 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) and the copyright owner(s) 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.