Acta Biotheor (2016) 64:197–217
DOI 10.1007/s10441-016-9280-5

REGULAR ARTICLE

Generalizing Contextual Analysis

Pierrick Bourrat[1]

Received: 1 October 2015 / Accepted: 23 May 2016 / Published online: 26 May 2016

Abstract Okasha, in *Evolution and the Levels of Selection*, convincingly argues that two rival statistical decompositions of covariance, namely contextual analysis and the neighbour approach, are better causal decompositions than the hierarchical Price approach. However, he claims that this result cannot be generalized in the special case of soft selection and argues that the Price approach represents in this case a better option. He provides several arguments to substantiate this claim. In this paper, I demonstrate that these arguments are flawed and argue that neither the Price equation nor the contextual and neighbour partitionings sensu Okasha are adequate causal decompositions in cases of soft selection. The Price partitioning is generally unable to detect cross-level by-products and this naturally also applies to soft selection. Both contextual and neighbour partitionings violate the fundamental principle of determinism that the same cause always produces the same effect. I argue that a fourth partitioning widely used in the contemporary social sciences, under the generic term of 'hierarchical linear model' and related to contextual analysis understood broadly, addresses the shortcomings of the three other partitionings and thus represents a better causal decomposition. I then defend this model against the argument that because it predicts that there is some organismal selection in some specific cases of segregation distortion then it should be rejected. I show that cases of segregation distortion that intuitively seem to contradict the conclusion drawn from the hierarchical linear model are in fact cases of multilevel selection 2 while the assessment of the different partitionings are restricted to multilevel selection 1.

Keywords Soft selection · Price equation · Multilevel analysis · MLS1 · MLS2

✉ Pierrick Bourrat
p.bourrat@gmail.com

[1] Department of Philosophy, Unit for the History and Philosophy of Science, Charles Perkins Centre, The University of Sydney, Sydney, NSW 2006, Australia

1 Introduction

Multilevel selection is the view that selection can act simultaneously at different levels of organisation (Bourrat 2015a, b; Okasha 2006). The simplest case of multilevel selection is a case of a population with two levels of selection: one I refer to, following Okasha (2006), as the *particle* level and the other I refer to as the *collective* level, with collectives being constituted of particles.[1] Under a classical view on the process of natural selection (e.g., Lewontin 1970), to occur at one given level of organisation, selection requires that there are differences in fitness between the entities forming a population at that level. Following this reasoning, if variance in fitness at the collective level and variance in fitness at the particle level have different values, the idea that different levels of selection are responsible for these difference is a plausible one.

So far so good, but the history of the concept of multilevel selection is contentious and the notion of group selection, a multilevel context in which the particles are biological individuals and the collectives are groups, has been highly debated over the last 50 years (for an overview of the debate see Okasha 2006). One reason of this debate, maybe the most important one, is that in a multilevel setting, determining the extent to which each level causes a trait to spread in a population can be very tricky to establish. Many disagreements remain as to which criteria should be used to do so. One of the clearest, but also earliest discussion explaining why causality at different levels of organisation matters so much in multilevel settings can be found in Williams (1966, 16–17). Williams famously contrasted a 'herd of fleet deer' with a 'fleet herd of deer' to illustrate that some collective traits (in his example the fleetness of a herd of deer) leading prima facie to differences in particle fitness, can actually be reduced to differences in particle character (the fleetness of each deer composing the herd) leading to differences in particle fitness. The collective character and collective fitness, in this case, result merely from the summation of particle character and fitness composing the collective. Thus, according to Williams, the fleetness of the collective should not be seen as a 'group adaptation' but rather as a 'fortuitous group benefit'. It is now accepted by many that this argument is a correct one (see for instance Okasha 2006; Sober and Wilson 2011). In those cases selection only occurs at the particle level for differences in collective fitness are merely a by-product of differences in particles fitness. Okasha (2006, 5) calls these by-products 'cross-level by-products'. I will follow the same terminology. Although this argument is sound for many, it is far from obvious what the best way to causally decompose the effects of particle character and those of collective character on particle fitness is. It should be noted that the question of whether a causal decomposition between the influences of different levels of selection on particle fitness is at all possible is an important one. This question however will not be discussed here. Causal decompositions have been de facto assumed by many authors and I will follow suit.

Another reason why the concept of multilevel selection has been contentious is that different authors have generally meant two distinct things when employing the

[1] The two levels scenario will be the one I will use throughout the paper.

term ‘multilevel selection’ or related ones. The conflation of these two concepts has been a source of confusion in the debate over multilevel selection. These two notions have first been clearly distinguished by Damuth and Heisler (1988) although similar distinctions can be found in earlier discussions (see Damuth and Heisler 1988, 410; Okasha 2006, 56). Under what Damuth and Heisler call the multilevel selection 1 (MLS1) framework, the focal level from which we assess the selection process(es) is the particle level. Both particle and collective fitnesses have the same metric. Usually, but not necessarily, particle fitness will be measured as the number of offspring particle produced after some time and the collective fitness as the number of offspring *particles* produced after some time. Under what Damuth and Heisler call the multilevel selection 2 (MLS2) framework, both the particle and collective levels are the focal levels from which we assess the selection process(es). The particle fitness and collective fitness have, in this case, different metrics. Usually, but not necessarily, collective fitness will be measured in terms of offspring *collectives* produced after some time while particle fitness will be measured in terms of offspring particles produced after some time. Gardner (2015a) proposes that the fitness of a collective should be defined in terms of neither number of daughter particles nor number of daughter collectives, but in terms “of its expected long-term genetic contribution to future generations” (2015a, 310), and that this resolves the MLS1 versus MLS2 dichotomy. This is a view I regard as correct for some cases, but we will see in the last section that the distinction between MLS1 and MLS2 can involve other features than number of particles or collectives produced.

Finally, another reason why the concept of multilevel selection is still muddled is that some authors understand ‘multilevel selection’ as ‘differences in fitness between particles *within a collective* and difference in fitness of collectives *between collectives*’ (e.g., Heisler and Damuth 1987). Others such as Nunney (1985) understand it as ‘differences in fitness between particles and between collectives when *the whole population is taken into account both at the particle and collective level’*.

In his book *Evolution and the Levels of Selection*, Okasha (2006) provides a thorough analysis in formal terms of the notion of multilevel selection that attempts to address each one of these problems. Okasha’s analysis, following a tradition that started with Price and his now famous equation (Price 1970, 1972), relies on statistical techniques that partition total evolutionary change into components, each of which aims at representing selection at one level. Yet, total evolutionary change can be partitioned in many different ways and people disagree as to which partitioning is the correct one, that is, the one that represents the real causal structure of evolutionary change. Okasha proposes to compare the advantages and disadvantages of the Price equation with two other statistical rivals used in the literature, based on the multivariate regression approach initially proposed by Lande and Arnold (1983), namely contextual analysis (Damuth and Heisler 1988; Goodnight et al. 1992; Heisler and Damuth 1987) and what he labels the ‘neighbour approach’ (Nunney 1985; Okasha 2006, 192–202). Although Okasha prefers contextual analysis and the neighbour approaches over the Price equation, he claims that the formers lead to an ‘intuitively wrong’ answer in cases of soft selection since they detect selection at the collective level when intuitively there is none (Okasha

2006, 95). The Price equation, on the other hand, provides according to him the 'intuitively correct' answer (Okasha 2006, 96). This leads him to conclude that none of the three techniques represents the absolute best causal decomposition for all cases of multilevel selection.

Importantly, his analysis is only made in the context of MLS1 for contextual analysis and the neighbour approach have both been designed to deal only with MLS1. In this paper, I re-evaluate the claim that contextual and neighbour partitionings provide the wrong answer in MSL1 cases of soft selection while the Price equation provides the right one. My conclusions are different from those of Okasha. I argue that each of these three statistical partitionings provide the wrong causal decomposition (in as much as the Price equation can be seen as a causal decomposition, which is contentious for some) for cases of soft selection and show that a fourth one, also based on a linear regression model and stemming from the up-to-date multilevel analysis literature in the social sciences (e.g., Goldstein 2011; Hox 2010; Snijders and Bosker 1999), is absolutely better than the three others.

2 The Hierarchical Form of the Price Equation

Historically, the first statistical partitioning of total evolutionary change for multilevel settings has been introduced by Price (1972) as a variant of his now famous equation. For this reason, it is naturally the starting point of most theoretically oriented discussions on multilevel selection and the first one introduced by Okasha (2006). The non-hierarchical form of the Price equation (Price 1970; see also Robertson 1966) provides a way to decompose total evolutionary change on a continuous character in a population of entities reproducing in discrete generations into two components as follows[2]:

$$\bar{w}\Delta\bar{z} = Cov(w_i, z_i) + E(w_i\Delta z_i) \tag{1}$$

where $\bar{w}$ is the average fitness of the parent population; $\Delta\bar{z}$ is the change in average character from one generation to another; w_i is the absolute fitness of the ith entity; z_i is the character value of the ith entity; Δz_i is the difference between the character value of the ith entity and the average for its offspring.[3]

The first term on the RHS, $Cov(w_i, z_i)$, is classically interpreted in the literature as the 'change due to selection' while the second term on the RHS, $E(w_i\Delta z_i)$, is classically interpreted as the 'transmission bias' (see Okasha 2006, Chapter 1 for more details on the non-hierarchical form of the Price equation). If the transmission bias is nil, Eq. (1) can be simplified as follow:

$$\bar{w}\Delta\bar{z} = Cov(w_i, z_i) \tag{2}$$

[2] For simplicity discrete generations are assumed.

[3] For an example of how Eq. (1) can be derived see Okasha (2006, 22). An alternative interpretation of this equation is to suppose that the focus of attention is the action of selection rather than total evolutionary change.

If we suppose now that the entities are nested in collectives (which I assume for simplicity have all the same size), after some rearrangements and definitions of new terms referring to the collective level, Eq. (1) can be rewritten as follows:

$$\bar{w}\Delta\bar{z} = Cov(W_k, Z_k) + E[Cov_k(w_{jk}, z_{jk})] \tag{3}$$

where w_{jk} is the absolute fitness of the jth particle in the kth collective; z_{jk} is the character value of the jth particle in the kth collective; W_k is the fitness of the kth collective and defined as the average fitness w of its particles and Z_k is the character value of the collective which is defined as the average character value z of its particles. Equation (3) is the Price equation in its hierarchical form.[4]

The first term on the RHS of Eq. (3), $Cov(W_j, Z_j)$, is the covariance between collective fitness W and collective character Z. It is classically interpreted as the selection between collectives. The second term of the RHS, $E[Cov_k(w_{jk}, z_{jk})]$, is the average of covariances between particle fitness w and particle character z within collectives. It is classically interpreted as selection between particles within collectives.

Although, Eq. (3) gives some traction to the concept of multilevel selection, one problem with it, extensively detailed in Okasha (2006, Chapter 3), is that $Cov(W_j, Z_j)$ does not allow discriminating causally the part of the covariance between collective fitness and collective character that results from differences in fitness at the particle level from the part that is due to differences in fitness at the collective level. In other words, the hierarchical form of the Price equation does not permit to separate causally the effects of cross-level by-products from those of direct selection at the collective level as it leaves causation implicit. This means, taking Williams' example presented in the Introduction, that using Eq. (3) in causal terms, one would not be able to conclude whether the herd character 'fleetness' causally influences the fitness of deer or not.

Before going further one remarks is in order. Some consider that the project of comparing different partitionings to the Price equation is misplaced as the Price equation can be seen, following the words of Frank (2012, 1014), as "an abstract placeholder" and does not pretend in and of itself to provide a causal decomposition of evolutionary change. Although, this criticism is well taken, the Price equation has de facto be used to make causal claims about selection occurring at different levels of organisation and it is important to show why this might be problematic [see also the recent debate on this topic between Goodnight (2015) and Gardner (2015b)].

3 Contextual Analysis

In order to address the shortcoming of Eq. (3), Heisler and Damuth (1987) propose, using a variant of the linear regression model put forward by Lande and Arnold (1983) to study selection in cases of correlated characters, that the fitness of a particle in a collective can be causally influenced by two factors, namely its own

[4] For a full derivation of the hierarchical form of the Price equation from the non-hierarchical form see for example Price (1972), Frank (1998) or Wade (1985).

character and a "contextual character" that will be the same for all the particles of a collective and thus by transitivity represents the indirect effect of the collective character on individual fitness. Damuth and Heisler borrow this technique from the social sciences which is known under the name of 'contextual analysis' (Boyd and Iversen 1979). In formal terms the contextual analysis regression model can be written as follows:

$$w_{jk} = \alpha + \beta_1 z_{jk} + \beta_2 Z_k + e_{jk}$$

where α is the intercept; β_1 is the partial regression coefficient of particle fitness on particle character, controlling for collective character[5] and thus measures the direct effect of particle character on particle fitness, controlling for collective character; β_2 is the partial regression coefficient of particle fitness on collective character, controlling for particle character and thus measures the indirect effect of collective character on particle fitness, controlling for particle character and e_{jk} is the residual whose variance is to be minimized.

If we now substitute w by this decomposed form in Eq. (2), this leads to:

$$\begin{aligned}\bar{w}\Delta\bar{z} &= Cov(\alpha + \beta_1 z_{jk} + \beta_2 Z_k, z_{jk} + e_{jk}) \\ &= Cov(\alpha, z_{jk}) + \beta_1 Cov(z_{jk}, z_{jk}) + \beta_2 Cov(Z_k, z_{jk}) + Cov(e_{jk}, z_{jk})\end{aligned}$$

This equation simplifies for $Cov(\alpha, z_{jk})$ is by definition nil because α is a constant, $Cov(z_{jk}, z_{jk})$ is by definition equal to $Var(z_{jk})$, $Cov(Z_k, z_{jk})$ is equal to the variance $Var(Z_k)$ and by virtue of what the least-squares regression analysis is doing $Cov(e_{jk}, z_{jk})$ is nil. We thus get:

$$\bar{w}\Delta\bar{z} = \beta_1 Var(z_{jk}) + \beta_2 Var(Z_k) \qquad (4)$$

In Eq. (4), the first term of the RHS, $\beta_1 Var(z_{jk})$, can be interpreted as the direct selection on particles. The second term of the RHS, $\beta_2 Var(Z_k)$ can be interpreted as the cross-level by-product that results from direct selection on the collectives.

The contextual approach is an improvement over the Price approach, when interpreted causally, for it allows distinguishing direct selection at the collective level from the by-product of selection at the particle level. Yet, it is not fully satisfactory. This is because from the point of view of a particle, the effect of the collective character on the particle fitness also includes the effects of this particle on its own fitness by contributing to the collective character. This means that contextual analysis does not completely eliminate the cross-level by-product problem. To fully eliminate it, the decomposition in the regression model for particle fitness should not be between particle character and collective character but between particle character (including the effects[6] of particle character on collective

[5] It is in fact the particle's contextual character, but since a perfect mapping between the contextual and collective character exists, for simplicity, I will use "collective character" in the reminder of the paper in places where it should be "contextual character".

[6] Note that the term 'effect' is understood here in a metaphysical sense, not a causal one. It thus includes supervenience relations.

character) and collective character minus the effect of the particle on collective character.[7]

4 The Neighbour Approach

To address the foregoing issue Okasha (2004, 2005, 2006), borrowing a reasoning from Nunney (1985) proposes the following alternative regression model:

$$w_{jk} = \alpha + \beta_3 z_{jk} + \beta_4 X_{jk} + e_{jk}$$

where X_k is what Okasha calls the neighbourhood character[8] of the jth particle in the kth collective. The neighbourhood character measures the collective character minus the effect of the particle on collective character. The term β_3 is the partial regression of fitness on particle character, controlling for neighbourhood character, while β_4 is the partial regression of fitness on neighbourhood character, controlling for particle character. I provide in the Appendix definitions of β_3 and β_4 in terms of β_1 and β_2 to make clear the straightforward links between the notion of collective and that of neighbourhood. It also demonstrates that most claims made about contextual analysis, can straightforwardly be applied to the neighbour partitioning.

As was done in the previous section with the contextual regression model, we can now replace w by this alternative model in Eq. (2). Because the covariance of the residual with individual trait value is nil, this leads to:

$$\bar{w}\Delta\bar{z} = \beta_3 Var(z_{jk}) + \beta_4 Cov(z_{jk}, X_{jk}) \quad (5)$$

In Eq. (5), the first term of the RHS, $\beta_3 Var(z_{jk})$ can be interpreted as the direct effect of the particle character on its own fitness. The second term of the RHS, $\beta_4 Cov(z_{jk}, X_{jk})$ can be interpreted as the cross-level by-product of selection on the neighbourhood character on particle fitness.

5 Rethinking the Advantages and Disadvantages of Each Approach

Because both the contextual and neighbour partitionings provide (different) solutions to the problems of cross-level by-products, Okasha is inclined to claim that they represent superior causal decompositions, that is, carve selection processes at their joints more effectively.[9] In fact both contextual analysis and the neighbour partitioning do not detect a collective (or neighbourhood) level component of selection, when the fitness of the particle depends solely on its character, which seems to be the right causal decomposition. The Price equation does not allow us

[7] Note that Okasha (2006, 201) points out that in some case of emergent collective character it might be worth including these effects in the collective level character. For the purpose of this paper I will not consider those cases.

[8] See Appendix for a formal definition of neighbourhood character.

[9] Each of these two approaches has different advantages and disadvantages, but they are unimportant for the main purpose of this paper.

making this discrimination. But Okasha also argues (using several examples) that in cases of 'soft-selection' (more on what exactly soft selection is in a moment) the Price equation provides the correct answer while both the contextual and neighbour partitionings provide very counter-intuitive answers. Thus, Okasha claims, the Price equation should be favoured in such cases.

In this section, I demonstrate *pace* Okasha and in spite of the fact the Price equation leads to the intuitively right answer, this is for the wrong reasons. The Price equation does not represent a good causal decomposition in general and this also applies to cases of soft selection. I then present Okasha's assessment of the adequacy of the contextual partitioning[10] for cases of soft selection. I concur with him that it does not represent a good causal decomposition for those cases. Yet, my arguments are different from his. I argue that although its makes precise causal predictions, these predictions are incompatible with the fundamental principle of determinism that the same cause always produces the same effect. Since no indeterminacy is supposed in those models, this simply means that the two models are inadequate causal partitions for cases of soft selection.

5.1 A Generic Case of Soft Selection and the Price Partitioning

Soft selection refers to any case in which all the collectives have the same fitness (due for example to some resource constraints) in spite of variation in collective character (see Fig. 1 for a theoretical example). In such a scenario, the fitness of a particle depends both on its own character and its relative ranking within its collective. This means that the collective influences particle fitness as noted by Okasha (2006, 95), and before him Goodnight et al. (1992) and Heisler and Damuth (1987).

Concretely, in Eq. (3) (Price partitioning), a case of soft selection implies that the first term of the RHS is nil, $Cov(W_k, Z_k) = 0$. Thus Eq. (3) can be rewritten as:

$$\bar{w}\Delta\bar{z} = E[Cov_k(w_{jk}, z_{jk})] \tag{3'}$$

which means that the Price equation partitioning predicts that the total evolutionary change depends solely on the covariances of particle fitnesses and particle character. This leads Okasha to claim that in a case of soft selection, the Price equation leads to the intuitively correct prediction that there is no collective level selection. But this interpretation, I claim, is not warranted. In fact, if we give a causal interpretation of the Price partitioning in a case of soft selection, the fact that the evolutionary change depends solely on covariance between particle fitness and particle character within collectives so that $Cov(W_k, Z_k) = 0$ does not necessarily imply that no selection between collectives would be detected had cross-level by-products been eliminated.

To see why, let us go back to one remark I made in Sect. 2. I noted that one problem with interpreting causally the hierarchical form of the Price equation is that it does not allow discriminating selection at the collective level from a by-product of

[10] *Mutatis mutandis*, the same can be argued for neighbour partitioning, for there is a straightforward relation between contextual and neighbour partitionings (see Appendix).

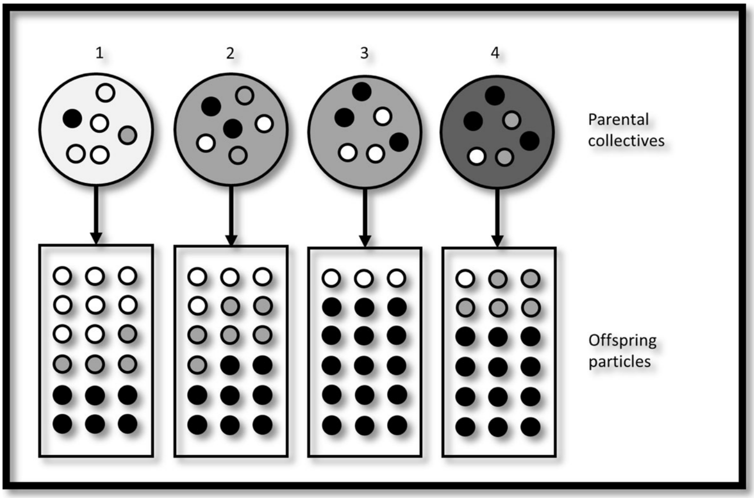

Fig. 1 A case of soft selection: each collective produces the same number of particles but the *black particles* are fitter than the *grey particles* within each collective, and the *greys particles* are fitter than the *white particles* within each collective. The colour of the parental collective reflects collective character, that is, the average of its particle characters (inspired from Okasha 2006, 95)

selection at the particle level. What I did not mention but which is discussed by Okasha (2006, 84–94) is that cross-level by-products can also run in the direction collective → particle, that is, when differences in fitness at the particle level result from direct selection at the collective level. In those cases, some or all the selection will be attributed to the particle level while the collective level character certainly plays an indirect causal role in determining particle fitness due to direct selection on collectives. Thus, although some covariance between particle character and particle fitness within collectives are observed in cases of soft selection, it could be in principle the case that they result, at least partly, from the effects of collective character on particle fitness within each collective.

It is also important to note the Price equation does not allow discriminating the *absence* of selection at the collective level from the combination of a cross-level by-product at the collective level of selection at the particle level going in one direction and direct selection at the collective level going in the other direction with the same magnitude and resulting in $Cov(W_k, Z_k) = 0$. In this case, in the absence of the indirect effect of particle character on collective fitness, a non-nil covariance would be observed between collective character and collective fitness which would be evidence that there is collective level selection in spite of $Cov(W_k, Z_k)$ being nil.

Thus because of the possibility of cross-level by product running in the two directions (that could explain $E[Cov_k(w_{jk}, z_{jk})] \neq 0$) and the possible cancellation of selection going in the opposite direction at each level (that could explain $Cov(W_k, Z_k) = 0$), using the Price approach should not be regarded as a correct causal decomposition on the basis that it gives an intuitively correct answer in cases of selection and more generally.

5.2 Contextual Analysis and Soft Selection

Let us turn now to the predictions made by contextual analysis in cases of soft selection. Those are quite different from that of the Price equation. But before going further, note that although my arguments will only be provided using the contextual partitioning, *mutatis mutandis* the same arguments can be made using the neighbour approach. This is because of the straightforward relation between β_2 and β_4 (see Appendix). Starting from a close version of Eq. (4), Goodnight et al. (1992, 752–753) have shown that for any case of soft selection, because we have: $Cov(W_k, Z_k) = 0$, it follows that $\beta_2 = -\beta_1$.

This result implies two things for cases of soft selection. First, if there is direct particle selection, that is, if β_1 is non-zero, then it immediately follows, under this interpretation, that there is direct collective selection since β_2 will non-zero. Second, if there is direct collective selection, then it immediately follows that there is direct particle selection. There is thus interdependence between particle and collective level selection. Note that interdependence *does not* imply that there is a relation of causality between the two levels. The interdependence is mathematical only.

With the interdependence between particle and collective selection, the contextual regression model presented earlier in Sect. 3 becomes thus:

$$w_{jk} = \alpha + \beta_1 z_{jk} - \beta_1 Z_k + e_{jk}$$

If we now replace w by this decomposed form in Eq. (2), because the covariance of the residual with individual trait value is nil, this leads to:

$$\bar{w}\Delta\bar{z} = \beta_1 Var(z_{jk}) + \beta_2 Var(Z_k). \tag{4'}$$

Okasha argues that the non-nullity of β_2 in Eq. (4)—or β_1 in Eq. (4′)—for cases of soft selection is a counterintuitive result, for many theorists have argued that there can only be collective-level selection if there are differences in fitness between collectives. This point, Okasha stresses, applies indistinguishably to MLS1 and MLS2 cases (Okasha 2006, 96–97).

Although it is correct that most theorists would argue that there can only be selection if there are differences in fitness between collectives, I do not believe we can draw the conclusion from contextual analysis that the fitness of all the collectives is the same *because* there is no selection at the collective level. This is because, as I demonstrate below, in cases of soft selection in which there is collective character variation, contextual analysis violates the fundamental principle of determinism that the same cause always produce the same effect in the following way: the effects of the independent variables in the linear regression do not remain invariant for different tokens of the same type of cause even in cases where Laplacian determinism is supposed.[11] I consider that invariance of effects for the same type of cause to be a fundamental criterion to determine the causal adequacy of equation in a deterministic setting.

[11] Note that Laplacian determinism can be supposed in all the equations of this paper and it is also what has been supposed by Okasha throughout his book. I will follow suit.

To see how contextual analysis fails in this criterion, let us suppose two particles j and l that belong respectively to the collective k and m of the same population. Following contextual analysis, we can write the linear model predicting their fitness in a case of soft selection as follows:

$$w_{jk} = \alpha + \beta_1 z_{jk} - \beta_1 Z_k + e_{jk}$$
$$w_{lm} = \alpha + \beta_1 z_{lm} - \beta_1 Z_m + e_{lm}$$

If $\beta_1 z_{jk}$ represents the additive effect in the model of the jth particle character of the kth collective on its own fitness, then following the fundamental principle of determinism, the additive effect of the particle character of the lth particle of the mth collective on its own fitness, $\beta_1 z_{lm}$, should be the same if the two particles have the same particle character, that is, if $z_{jk} = z_{lm}$. This should remain true irrespective of the values of collective character Z of the kth collective and the mth collective.

Similarly, if $-\beta_1 Z_k$ represents the additive effect in the model of kth collective character on the fitness of the jth particle of the kth collective, then applying the fundamental principle of determinism as defined above, the additive effect of the collective character of the mth collective on the fitness of the lth particle of the mth collective, $-\beta_1 Z_m$, should be the same if the two collectives have the same collective character, that is, if $Z_k = Z_m$.[12] This should remain true irrespective of the values of particle character z of the jth particle of the kth collective and the lth particle of the mth collective.

Finally, if e_{jk} represents the deviation from linearity, so that it can represent the joint non-additive effects of the jth particle character and kth collective character on the fitness of the jth particle of the kth collective, then applying the fundamental principle of determinism as defined above, the non-linear effects of the lth particle character of the mth collective and the non-linear effects of the mth collective, on the fitness of the lth particle of the mth collective, e_{lm}, should be the same if the two collectives have the same collective character ($Z_k = Z_m$) and the same particle character ($z_{jk} = z_{lm}$).

From these three pieces of reasoning, we can make two predictions. First, two particles of the same type in two different collectives with the same collective character value should *always* have the same fitness irrespective of the collective they are found in. This is because, in virtue of the fundamental principle of determinism the particle additive effect and non-additive effect on their own fitness should be same. Second, two collectives with the same collective character, should *always* have the same effects on the particle fitness of the same type. This is because, in virtue of the fundamental principle of determinism the collective additive effect and non-additive effect on particle fitness should be same.

One way to 'test' wether these two predictions are verified with contextual analysis in cases of soft selection is to compare the fitness of two particles of the same type that belong to two collectives with the same collective character. A situation of this sort is illustrated in Fig. 1, which represents a pure case of soft selection, with the two collectives of the centre (collectives 2 and 3) having the

[12] Note that I am talking here about the same *type* of collective character which might be realized by different tokens as is the case in Fig. 1 with the collective characters of collectives 2 and 3 which are the same but realized in two different ways.

same collective character (same shade of grey) but a very different composition of particles although both the black type and the white type are present in both collectives. If the fundamental principle of determinism is respected in this case, the two particles should have the same fitness, because in both cases the particle character and the collective character (plus their interaction) should lead to the same effects on particle fitness. If this is verified, we can safely conclude that contextual partitioning is an accurate causal partition of multilevel selection in cases of soft selection. If not, then, following a causal interpretation of contextual analysis, the difference should be attributed to a different additive effect in each collective coming from the same particle character, a different additive effect coming from the same collective character or again to a different non-additive effect in each collective coming from the same interaction between particle-character and collective-character.[13] If such was the case, this *reductio ad absurdum* would demonstrate that contextual analysis does not represent an accurate causal decomposition of multilevel selection in cases of soft selection.

On Fig. 1, we can see that a black particle has on average 4 offspring particles in collective 2 and 5 in collective 3. Similarly a white particles has on average 2 particles in collective 2 and 1 in collective 3. Since in both situations the two particles compared have the same particle and collective characters, following our reasoning, they should have the same fitness in our deterministic setup in which particle and collective characters (plus their interaction) are the only difference makers for particle fitness. Yet, in both cases, the two particles have different fitnesses. Using contextual analysis, we must thus attribute this difference either to a difference in the effect of particle character on particle fitness, in the effect of collective character on particle or again in their interaction which is absurd because we are, by stipulation in a deterministic setting, and doing so violates precisely this assumption. For that reason, contextual analysis should be rejected as an adequate causal model for cases of soft selection in which there is variation in collective character.[14]

6 An Alternative and Superior Causal Model: The Intercepts-as-outcomes and Slopes-as-outcomes Hierarchical Linear Models

Multilevel analysis in its modern form has been designed in the social sciences to address several drawbacks of contextual analysis as understood by Boyd and Iversen (1979) and thus Goodnight et al. (1992) and Okasha (2006). I will refer from now on to contextual analysis as 'classical contextual analysis'.[15] One of the problems of classical contextual analysis is that does not consider that particle and collective

[13] Note that these do not represent mutually exclusive situations: the overall difference could be attributed a combination of these three causes.

[14] Note that other complex scenario of multilevel selection, some involving soft selection while others not, are expected to violate the fundamental principle of determinism in deterministic settings when modelised by contextual analysis. Therefore my demonstration is not intended to apply solely to soft selection cases, but these are the ones I take issue with in this paper.

[15] As previously, everything said about contextual analysis can be transposed to neighbour partitioning (see Appendix). To classical contextual analysis, corresponds a classical form of neighbour partitioning.

characters are distinct sources of variability (Snijders and Bosker 1999, 2). An instance of this problem arose when educational researchers wanted to understand if there was a dependence between performance of students and the school they belong too (De Leeuw and Meijer 2008; Goldstein 2011). Classical contextual analysis treats the performance of a school (collective) solely as an aggregation of the performance of its students (particle). Yet this assumption is quite problematic because in this case there are causes of the dependent variable (performance of the student) other than the individual-level variable and their school-level aggregates. Furthermore these additional causes are correlated with the school-level aggregates. Taking into these correlated school-level variables is one of the main theoretical motivation of modern multilevel analysis and why it has now supplanted classical contextual analysis, as it is stressed in all recent textbooks on multilevel modeling (e.g., De Leeuw and Meijer 2008; Goldstein 2011; Hox 2010; Snijders and Bosker 1999). It should be noted that the exact same problem is present in biological populations as exemplified by Gardner (2015a).

Going back to the cases of soft selection, since classical contextual analysis fails in the fundamental principle of determinism it might be worth asking why this is the case. One possible answer is that cases of soft selection are in some ways inherently similar to the ones of the schools. It is perfectly plausible that particle fitness is causally explained by particle character, collective character and a third (or more) factor(s) affecting all the members of a particular collective in the same way (or at least in more similar way within collectives than between collectives) and that although is not *explicitly* stated in the description of deterministic soft selection cases is inherent to all of them.

But how should this factor be integrated in our linear model? The hierarchical linear model, which is the main tool used in modern multilevel analysis, comes in two flavours and assumes in simple cases that the intercept for each collective is a variable in the same way particle character and collective character are. Snijders and Bosker (1999) call this model the 'intercepts-as-outcomes model'. Under more complex models, it is assumed that the slope of each independent variable within each collective is itself a variable. Snijders and Bosker (1999) call this model the 'slopes-as-outcomes model'. (See Fig. 2 for a comparison of these two approaches with classical contextual analysis using rudimentary causal graphs.) In this latter case that would imply, in causal terms, that the effect of particle character and/or collective character varies within each collective due to some causal factor influencing the relationship between particle and/or collective character and particle fitness (see Fig. 2c, where the third factor influences the relationship between collective character and particle fitness). In the former case this would simply mean that this third factor directly influences particle fitness without influencing the effect of particle and/or collective character (see Fig. 2b). Note the curved, dashed and double-headed arrows in Fig. 2b, c that represent a correlation between the collective character and the third factor *within each collective*. It is because of these correlations (known as intraclass correlations in the context of analyses of variance) that the hierarchical-linear model is different from classical contextual analysis which assumes independence of all observations. Thus the hierarchical linear model does not represent the simple addition of an explanatory variable in a multivariate

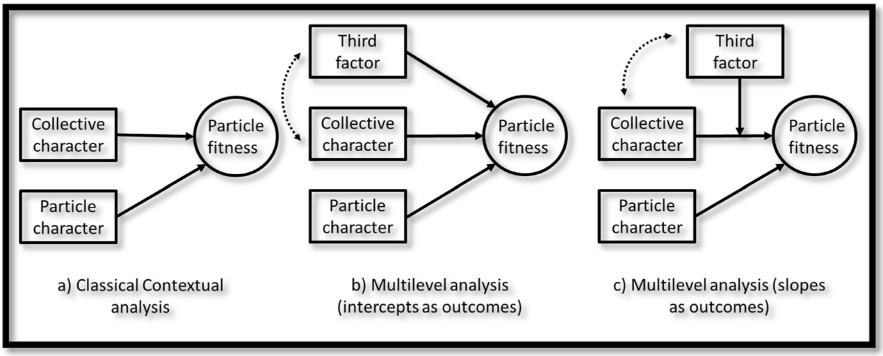

Fig. 2 Examples of simple causal graphs of the different models used in multilevel analysis. **a** Classical contextual analysis; **b** multilevel analysis with intercepts as outcomes; **c** multilevel analysis with slopes as outcomes. *Single-headed* and *straight arrows* imply causation*;* *double-headed*, *curved* and *dashed arrows* imply intraclass correlations

regression analysis as one can find in Stevens et al. (1995) for instance. It proposes a way to account for intraclass correlations.[16]

Which one of the two flavours (intercepts as outcomes or slopes as outcomes) one chooses for soft selection does not really matter for us. I will treat here the simpler case in which only the intercept within each collective can vary, that is, the intercepts-as-outcomes model (Fig. 2b), so that there is no cross-level interaction between this third factor and particle and/or collective character. Cross-level interaction occurs when the slope of an independent variable is explained by another variable (Snijders and Bosker 1999, 73–74). Note that the conclusions I draw from this simpler case (Fig. 2b) would be the same if I was considering the slope of each independent variable within each collective as an outcome (Fig. 2c) (see Snijders and Bosker 1999, chap. 5 for the details of how this can be done).

If one assumes that the intercept is a dependent variable fully explained by a third factor different from particle and collective character (residual term nil), the linear regression model for cases of soft selection (following the intercepts-as-outcomes model) can be written as follows:

$$w_{jk} = \alpha_{0k} + \beta_1 z_{jk} - \beta_1 Z_k + e_{jk}$$

where α_{0k} represents the intercept for the k th collective and is defined as:

$$\alpha_{0k} = \alpha_{00} + \alpha_{01} V_k$$

where α_{00} is the average intercept across all the collectives of the population, V_k is the independent variable explaining the intercept in the k th collective and α_{01} is the regression coefficient of the intercept of collective k on variable V in group k. For the moment, I do not provide any causal interpretation for V.

Once the substitution of α_{0k} is made in the regression model we have:

[16] Note that Gardner (2015a) develops a similar analysis in the context of the hierarchical form of the Price equation.

$$w_{jk} = \alpha_{00} + \alpha_{01} V_k + \beta_1 z_{jk} - \beta_1 Z_k + e_{jk}$$

Note that if $\alpha_{01} V_k$ is nil, then we go back to the classical form of contextual analysis.

Let us now assess whether this model is consistent with the fundamental principle of determinism using our two particles in collectives 2 and 3 represented in Fig. 1. The assessment is straightforward. We can see that although the two particles of the same type in these two collectives with the same characters have different fitnesses, we can now attribute this difference to a difference in V between the two particles' fitnesses. The intercepts-as-outcomes version of the hierarchical linear model, because it is consistent, in its generic form, with the fundamental principle of determinism in the case of soft selection is thus superior to classical contextual analysis and consequently to the neighbour approach.

If we now substitute w by this model in Eq. (2), we obtain:

$$\begin{aligned}\bar{w}\Delta\bar{z} &= Cov(\alpha_{00} + \alpha_{01} V_k + \beta_1 z_{jk} - \beta_1 Z_k + e_{jk}, z_{jk}) \\ &= Cov(\alpha_{00}, z_{jk}) + \alpha_{01} Cov(V_k, z_{jk}) + \beta_1 Cov(z_{jk}, z_{jk}) - \beta_1 Cov(Z_k, z_{jk}) \\ &\quad + Cov(e_{jk}, z_{jk})\end{aligned}$$

Since α_{00} is a constant, $Cov(\alpha_{00}, z_{jk})$ is nil and by virtue of what the least-squares regression analysis is doing $Cov(e_{jk}, z_{jk})$ is nil. This means the above equation can be simplified into:

$$\bar{w}\Delta\bar{z} = \alpha_{01} Cov(V_k, z_{jk}) + \beta_1 Var(z_{jk}) - \beta_1 Var(Z_k) \quad (6)$$

To see the difference between Eq. (4′) stemming from classical contextual analysis and Eq. (6), the term $\alpha_{01} Cov(V_k, z_{jk})$ and more particularly the factor V must be causally interpreted in Eq. (6). One possible interpretation of V is as an intrinsic factor of collectives (for instance the variance in particle character in the collective, the presence of at least one black particle in a collective, etc.). In such cases, the collective level character relevant for particle fitness is not a simple additive function of the particle character. Another possibility is that this factor does not result from the sole interaction of particles between themselves, but also includes some other parts of the collective (assuming a collective is not merely the sum of particles but also includes other material parts). This interpretation would lead us to the view that cases of soft selection are not cases of the MLS1 sort but of the MLS2 sort (see the next section for an example). If one assumes that collective fitness *is* an additive function of particle fitness, as Okasha does, this latter interpretation must be rejected.

Another interpretation compatible with MLS1 is that V is a factor extrinsic to the collective (e.g. the amount of resource available at a given time for the collective). However, in such case, selection is not the only factor determining particle reproductive output and natural selection cannot be the sole causal factor of the model. It is interesting to note that a similar point has been made by Mitchell-Olds and Shaw (1987, 1154–1155) in single level context where they argue that multilevel regression cannot rule out the existence of an unmeasured factor influencing fitness and phenotypic character that would give spurious estimates of selection bias (see also Gardner 2015a in the context of multilevel selection). The view that soft selection might be interpreted as a mixed case of selection and another factor raises the question as to whether cases of soft selection represent ‘pure’ cases of natural selection (albeit multilevel ones). It also raises the question as to whether frequency and density dependent selection can also be interpreted as cases of pure selection, for soft selection is a case of both frequency and density dependent selection (Wallace 1975). I will not pursue these questions here.

Although I have not demonstrated it here, assuming linearity and zero interaction, in non-frequency and non-density dependent cases of multilevel selection, the variable V can be assumed to be nil (or constant) without violating the fundamental principle of determinism. These cases, as argued by Okasha, are thus perfectly accountable with the classical contextual partitioning. Thus classical contextual analysis represents a special case of multilevel partitioning in which particle fitness is fully explained by particle and collective characters.

7 Segregation Distortion and Soft Selection

The result obtained by Eq. (6) suggests that if we were to keep V constant in a MLS1 setting, so that, $Cov(V_k, z_{jk}) = 0$, then we would necessarily observe a difference in fitness between collectives of different particle composition. Thus it seems that the correct conclusion to draw from the fact that $Cov(Z_k, W_k) = 0$ in the Price equation in cases of soft selection, is the result of the annulation of the sum of the effects of particle character (indirect), collective character (direct) and the third causal factor (V) (indirect) on collective fitness, not that there is no collective level selection (of the MLS1 sort). Yet, some might argue that this conclusion is at least as counterintuitive as the conclusion drawn from a causal reading of classical contextual analysis (Eq. 4), which if we recall it, is that there is collective level selection in cases of soft selection. In this section I show that the conclusion drawn from a causal reading of Eq. (6) is only counterintuitive in cases of soft selection of the MLS2 sort, a type of selection, which, I have already emphasized, is beyond the scope of the main contentions of both this article and Okasha’s version of contextual analysis.

Take the case proposed by Okasha of three genotypes AA, AB and BB that have identical fitnesses so that $W_{AA} = W_{AB} = W_{BB}$ and in which the allele A has a higher reproductive success than B when it is paired with an allele B (leading to a heterozygote genotype) rather than with a allele A (for details see Okasha 2006, 154–155).[17] The Price equation detects no selection at the genotypic level, while the intercepts-as-outcomes model does, for the same reasons given in Sect. 6, namely that the third factor V is causally responsible together with z and Z, of the nil covariance between W and Z. As in the previous case, the Price equation (Eq. $3'$) seems to provide the intuitively ‘good’ answer (but we saw in Sect. 5.1 that this intuition is misleading) while the version of hierarchical model presented in the previous section (Eq. 6) does not if one assumes that there is no extrinsic variable involved in the fitness outcome of organisms beyond particle character and

[17] See also another case of segregation distortion at the same page, which can be treated in a similar way.

collective character [that is, $\alpha_{01}Cov(V_k, z_{jk}) = 0$] and that the case is of the MLS1 sort.[18] It seems that in this case, the same fitness outcome at the multicellular level would arise even if all the organisms of the population were in the exact same environment, and thus no 'hidden variable' extrinsic to the collective or cross-level interaction could be invoked (V). There really would seem to be no organismal selection going on here. Should we thus conclude that the intercepts-as-outcomes model is faulty as Okasha did for classical contextual analysis? No so fast. Although, the segregation distortion case seems at first glance to be perfectly isomorphic to the previous case of soft selection presented in Fig. 1 (with the slight complication of sexual reproduction), a closer look reveal that it is not the case.

One reason why these two cases are not perfectly isomorphic is because 'diploid organism' is not isomorphic to 'collective' in the sense we have used the term 'collective' so far. As noted by Sarkar (1994, 1998; see also Falk and Sarkar 1992) genotypes and organisms belong to two separate biological hierarchies that we intuitively conflate, but should not. The two hierarchies are on the one hand: allele → locus → gene complex → genotype, etc., and on the other hand: molecule → organelle (including chromosome) → cell → tissue → organism → group, etc. Yet, any conclusion drawn in one hierarchy cannot be straightforwardly translated into the other. And thus from the conclusion that there is selection (assuming the analysis given earlier is correct) of *genotypes*, we cannot move to the conclusion that there would be *organismal* selection. This is because the two hierarchies are conceptually independent and a pattern of selection in one does not imply the same pattern in the other. There might be rules of equivalence between the two hierarchies that renders, in some specific conditions, the conclusion(s) drawn within one hierarchy applicable to the other(s), but without having established those rules, nothing can be said from the perspective of the hierarchical linear partitioning presented in the previous section.

Sarkar's distinction must be linked to the distinction between MLS1 and MLS2. By noticing that particles (alleles) and collectives (organisms) belong to two different hierarchies, in the case of segregation distortion, we immediately know that the fitness metric for particles and collectives are different, and that the collective character is not the average of the particle character. It seems thus that cases of segregation distortion are clearly of the MLS2 sort, not of the MLS1 one as assumed by Okasha because 'alleles' and 'organisms' belong to two different biological hierarchies. Thus it might be correct to claim, following the argument I proposed earlier, that there is genotypic selection (together with allelic selection and some other factors) in the case of segregation of distortion proposed by Okasha, when genotype are strictly compared to alleles (MLS1 and thus within the same biological hierarchy) without the contradiction that there is no organismic selection (MLS2).

I thus believe that soft selection cases of segregation distortion cannot be considered as problem cases for the hierarchical linear partitioning interpretation in a MLS1 context presented in the previous section since those cases cannot straightforwardly be considered as cases of MLS1.

[18] Effectively, as I pointed earlier, this is equivalent to a case of classical contextual analysis (Eq. 4).

8 Conclusion: Generalising Contextual Analysis?

In this paper I have shown that the claim that the Price equation represents a better causal decomposition than the contextual and neighbour partitionings is not warranted, for the Price equation will generally, and this applies also to soft selection cases, be unable to detect cross-level by-products of selection running from the particle to the collective level or from the collective to the particle level. Later on, I argued that contextual and neighbour partitionings (sensu Okasha) were also inadequate for dealing with soft selection since to be consistent they must violate the fundamental principle of determinism that the same cause always lead to the same effect. I have then proposed a fourth partitioning stemming out from the up-to-date multilevel modelling literature and more particularly the hierarchical linear model that addresses the shortcoming of the classical contextual and neighbour partitionings. I have shown how by, assuming that the intercept in the contextual model is itself a dependent variable, one can account for the evolutionary change observed in at least some cases soft selection[19] while being consistent with the fundamental principle of determinism that the same cause produces the same effect in a deterministic setting. Finally, I have shown that cases of segregation distortion cannot straightforwardly considered as MLS1 cases. Thus, soft selection cases involving segregation distortion should not be regarded as problematic cases for the version of the hierarchical-linear partitioning I proposed here.

Throughout the paper I have discussed different approaches to multilevel selection, in the particular case of soft selection. It should be stressed that fundamentally the classical contextual, neighbour and hierarchical partitionings all use the same methodology. The main difference between the three partitionings is that classical contextual analysis makes a higher number of unrealistic assumptions than the 'linear hierarchical' partitioning proposed here. Two of them are 1) considering the collective character is a mere aggregate of the particle character; and 2) considering that there is no intraclass correlation. I have shown some important limitations that come with these assumptions when one wants to interpret causally some cases of soft selection and it should be expected that similar limitations will be encountered by contextual analysis in cases different from soft selection but in which the same assumptions are violated.

Because contextual analysis, the neighbour approach and the hierarchical linear one all belong to the same family of models, it seems prima facie reasonable to propose a single term for all these partitionings. Some might favour the term "contextual analysis". Although from a *statistical* point of view this is perfectly justified, like Okasha (2006, 98–99) I am reluctant to do so if, given a particular multilevel setting, the goal is to find the correct *causal* story underpinning it. Once interpreted causally each type of analysis tells a structurally different causal story. This, I believe, is sufficient reason to give them a different name. But if, in a given context, everyone agrees on which statistical equation is the most appropriate, this problem becomes to a large extent a semantic one.

[19] Other cases would involve more complex models.

Acknowledgments I am thankful to Andy Gardner, Charles Goodnight, Samir Okasha and three anonymous reviewer for comments on earlier versions of the manuscript and Peter Godfrey-Smith for his advice on this topic. I am also grateful to Paul Griffiths for his support over the years. This research was supported under Australian Research Council's Discovery Projects funding scheme (Projects DP150102875).

Appendix

Okasha (2005, 718–719) provides a definition of β_4 in terms of collective characters and particle characters by demonstrating that there is the simple relation following relation between β_2 and β_4. Given that neighbourhood character is defined as $X_{jk} = \frac{nZ_k - z_{jk}}{n-1}$, we can deduct that

$$\beta_4 = \frac{n-1}{n}\beta_2$$

where n is the number of particles in a collective.

Although Okasha does not provide a demonstration of it, it is also useful to express β_3 in terms of collective and particle characters. In fact, this will allows us to highlight the difference between the direct effect of particle character on particle fitness, controlling for *collective* character and the direct effect of particle character on particle fitness, controlling for *neighbourhood* character. This also highlights the straightforward mathematical link between contextual and neighbour partitionings. This can be done as follows. Assuming e_{jk} is nil we have:

$$w_{jk} = \beta_3 z_{jk} + \beta_4 X_k = \beta_3 z_{jk} + \beta_2 \frac{nZ_k - z_{jk}}{n}$$

This expression can be rearranged as follows:

$$w_{jk} = \left(\beta_3 - \frac{\beta_2}{n}\right) z_{jk} + \beta_2 Z_k$$

Because both the contextual and Okasha's version of the neighbourhood regression models are models for particle fitness, we know that:

$$w_{jk} = \beta_1 z_{jk} + \beta_2 Z_k = \beta_3 z_{jk} + \beta_4 X_k$$

And thus it follows that:

$$\beta_1 z_{jk} + \beta_2 Z_k = \left(\beta_3 - \frac{\beta_2}{n}\right) z_{jk} + \beta_2 Z_k$$

This implies that:

$$\beta_1 = \beta_3 - \frac{\beta_2}{n}$$

and therefore that:

$$\beta_3 = \beta_1 + \frac{\beta_2}{n}$$

Recall that Okasha defines β_3 as the partial regression coefficient of fitness on particle character, controlling for neighbourhood character. We can now express it verbally in terms of particle and collective characters. Following the definitions of β_1, β_2 and n provided in the main text, β_3 is the sum of the partial regression coefficient of particle fitness on particle character, controlling for collective character and the partial regression coefficient of fitness on collective character, controlling for particle character, divided by the number of particles in the collective.

References

Bourrat P (2015a) Levels, time and fitness in evolutionary transitions in individuality. Philosophy & Theory in Biology 7. doi:10.3998/ptb.6959004.0007.001

Bourrat P (2015b) Levels of selection are artefacts of different fitness temporal measures. Ratio 28 (1):40–50. doi:10.1111/rati.12053

Boyd LH, Iversen GR (1979) Contextual analysis: concepts and statistical techniques. Wadsworth Publishing Company, Belmont

Damuth J, Heisler IL (1988) Alternative formulations of multilevel selection. Biol Philos 3(4):407–430

De Leeuw J, Meijer E (2008) Introduction to multilevel analysis. In: De Leeuw J, Meijer E (eds) Handbook of multilevel analysis. Springer, New York, pp 1–75

Falk R, Sarkar S (1992) Harmony from discord. Biol Philos 7(4):463–472

Frank SA (1998) Foundations of social evolution. Princeton University Press, Princeton

Frank SA (2012) Natural selection. IV. The Price equation. J Evol Biol 25(6):1002–1019

Gardner A (2015a) The genetical theory of multilevel selection. J Evol Biol 28(2):305–319

Gardner A (2015b) More on the genetical theory of multilevel selection. J Evol Biol 28(9):1747–1751

Goldstein H (2011) Multilevel statistical models, 4th edn. Wiley, Chichester

Goodnight CJ (2015) Multilevel selection theory and evidence: a critique of Gardner, 2015. J Evol Biol 28(9):1734–1746

Goodnight CJ, Schwartz JM, Stevens L (1992) Contextual analysis of models of group selection, soft selection, hard selection and the evolution of altruism. Am Nat 140(5):743–761

Heisler L, Damuth J (1987) A method for analyzing selection in hierarchically structured populations. Am Nat 130(4):582–602

Hox JJ (2010) Multilevel analysis: techniques and applications. Routledge, New York

Lande R, Arnold SJ (1983) The measurement of selection on correlated characters. Evolution 37(6):1210–1226

Lewontin RC (1970) The units of selection. Annu Rev Ecol Syst 1(1):1–18

Mitchell-Olds T, Shaw RG (1987) Regression analysis of natural selection: statistical inference and biological interpretation. Evolution 41(6):1149–1161

Nunney L (1985) Group selection, altruism, and structured-deme models. Am Nat 126(2):212–230

Okasha S (2004) Multilevel selection and the partitioning of covariance: a comparison of three approaches. Evolution 58(3):486–494

Okasha S (2005) Altruism, group selection and correlated interaction. Br J Philos Sci 56(4):703–725

Okasha S (2006) Evolution and the levels of selection. Oxford University Press, New York

Price GR (1970) Selection and covariance. Nature 227:520–521

Price GR (1972) Extension of covariance selection mathematics. Ann Hum Genet 35(4):485–490

Robertson A (1966) A mathematical model of the culling process in dairy cattle. Anim Prod 8(95):108

Sarkar S (1994) The selection of alleles and the additivity of variance. In: PSA: proceedings of the biennial meeting of the Philosophy of Science Association, pp 3–12

Sarkar S (1998) Genetics and reductionism. Cambridge University Press, Cambridge

Snijders TAB, Bosker RJ (1999) Multilevel analysis: an introduction to basic and advanced multilevel modeling. Sage, London

Sober E, Wilson DS (2011) Adaptation and natural selection revisited. J Evol Biol 24(2):462–468

Stevens L, Goodnight CJ, Kalisz S (1995) Multilevel selection in natural populations of impatiens capensis. Am Nat 145(4):513–526

Wade MJ (1985) Soft selection, hard selection, kin selection, and group selection. Am Nat 125(1):61–73

Wallace B (1975) Hard and soft selection revisited. Evolution 29(3):465–473

Williams GC (1966) Adaptation and natural selection: a critique of some current evolutionary thought. Princeton University Press, Princeton