Biol Theory (2015) 10:311–321
DOI 10.1007/s13752-015-0210-6

ORGINAL ARTICLE

Distinguishing Natural Selection from Other Evolutionary Processes in the Evolution of Altruism

Pierrick Bourrat[1]

Received: 8 October 2013 / Accepted: 28 April 2015 / Published online: 4 June 2015

Abstract Altruism is one of the most studied topics in theoretical evolutionary biology. The debate surrounding the evolution of altruism has generally focused on the conditions under which altruism can evolve and whether it is better explained by kin selection or multilevel selection. This debate has occupied the forefront of the stage and left behind a number of equally important questions. One of them, which is the subject of this article, is whether the word "selection" in "kin selection" and "multilevel selection" necessarily refers to "evolution by natural selection." I show, using a simple individual-centered model, that once clear conditions for natural selection and altruism are specified, one can distinguish two kinds of evolution of altruism, only one of which corresponds to the evolution of altruism by natural selection, the other resulting from other evolutionary processes.

Keywords Altruism · Evolution · Hamilton's rule · Kin selection · Multilevel selection · Natural selection

Introduction

Altruism, which can be defined as "a behavior costly to the actor and beneficial to the recipient" (West et al. 2007, p. 416), is one of the most studied topics in evolutionary theory. Yet it is still a concept that is hard to pin down and the subject of heated debates. In fact, altruism looks like a puzzle from an evolutionary perspective (Sterelny and Griffiths 1999, p. 153). "Standard" Darwinian reasoning tells us that only beneficial traits can evolve by natural selection. However, although being altruistic is costly, altruism is found everywhere around us, and some of the most successful lineages on Earth display altruistic behaviors (e.g., humans, bees, ants).

Several approaches have been proposed to solve this puzzle and delimit the conditions under which altruism can evolve. Although the different approaches generally agree on these conditions, there is still a lot of discord over which one is best suited to study altruism. Hamilton (1963, 1964a, b) has been the first to propose a clear solution to the problem of altruism with the notion of inclusive fitness, which has led to the "kin selection" approach to altruism with many followers (e.g., Grafen 1984; Taylor and Frank 1996; Rousset 2004; West et al. 2007; Bourke 2011). Under this framework, although altruistic individuals pay a cost, the behavior benefits preferentially their kin who have a higher probability to have the same genes for altruism than other individuals. As a result, if the benefit is superior to the cost and the probability to interact with individuals with the same genes high enough, altruism can evolve. Another solution has been put forward by David Wilson (e.g., Wilson 1980; Sober and Wilson 1998; Wilson and Wilson 2007) with the notion of trait-group or more generally multilevel selection. Under this latter approach, in populations structured in groups, although selfish individuals beat altruistic ones within groups, altruistic groups beat selfish ones (Wilson and Wilson 2007, p. 345), and thus altruism can evolve in the general population. Many consider the two approaches to be formally equivalent (Okasha 2006; West et al. 2007; Wilson and Wilson 2007) but for group selectionists the multilevel approach represents the best causal structure of the phenomenon, while the kin selectionists disagree that the

✉ Pierrick Bourrat
p.bourrat@gmail.com

[1] Department of Philosophy, University of Sydney, Sydney, NSW, Australia

notion of trait-group substantially enlightens any aspect that would not be captured by the notion of kin selection (West et al. 2007, 2008). This has led some to advocate for pluralism on this question (Dugatkin and Reeve 1994; Sterelny 1996; Kerr and Godfrey-Smith 2002).[1] More recently, Fletcher and Doebeli (2009) have proposed what they consider to be a simpler and more general alternative individual-level explanation of the evolution of altruism, in which positive assortment of phenotypes rather than genotypes, as is supposed in kin selection theory, is sufficient for altruism to evolve.

Although reaching a consensus over whether one approach (if any) is the best to understand the evolution of altruism is an important and useful project, this question has occupied the forefront of the stage and left behind a number of equally important questions. One of them, which is the subject of this paper, is whether the word "selection" in "kin selection" and "multilevel selection" necessarily refers to "evolution by natural selection."

To answer this question, let us first remark that one of the first lessons taught in population genetics is that evolution does not necessarily imply evolution by natural selection. Natural selection is only one possible "force" that might drive evolutionary change (Sober 1984). Although every protagonist in the debate on the best approach to the evolution of altruism claims to defend a view on how *natural selection* can operate to favor altruism over selfishness in some particular settings, a striking problem with this claim is that the term *natural selection* is usually not defined precisely and/or contrasted with other evolutionary processes that could lead to the same outcomes. Using an approach similar to Fletcher and Doebeli's (Fletcher and Doebeli 2009), I will show that positive assortment can occur systematically between altruistic individuals and lead altruism to evolve for reasons that challenge a theoretically worked out concept of natural selection (which I will detail below) and should thus be considered as resulting from other evolutionary processes.

To do so, the paper is divided in three sections. In the first section, I start by delimiting consistent concepts of altruism and natural selection. In the second section, I expose the problem of altruism starting from a very simple individual-centered model inspired from Sober and Wilson (1998). In a population in which altruistic and selfish individuals interact perfectly randomly altruism cannot evolve by natural selection. From there, using a method similar to that developed by Fletcher and Doebeli (2009), I show that for altruism to evolve by natural selection, positive assortment between altruistic individuals is a necessary condition, and I relate this to Hamilton's rule (Hamilton 1963, 1964a, b). In the third section, I compare this result to the definition of natural selection given in the first section and show, using a simple setup, that positive assortment between altruistic individuals can broadly have two different causes. One of the two causes has to do with being altruistic, while the other is contingent on this fact. I argue, following the condition for natural selection provided in the first section, that only when positive assortment is tied to the fact of being an altruist, the resulting evolution is evolution by natural selection. When positive assortment is not tied to the fact of being an altruist, the claim that the evolution results from evolution by natural selection is unwarranted because the difference in reproductive output between the two types results from environmental contingencies.

Semantic Issues on the Evolution of Altruism

Before considering under which conditions altruism can evolve by natural selection, one needs a clear understanding of the concepts of "altruism" and "natural selection." Both the notion of altruism and of natural selection I propose below would be objected to by a hostile reader with different interpretations of these terms. However, my claim is not that altruism and natural selection *should* necessarily be understood as I conceive of them below. Rather, I aim to draw the consequences for the notion of evolution of altruism by natural selection *if* one understands natural selection and altruism as such. I believe the notions I propose below to be consistent with much of evolutionary theory and logically coherent, which renders them legitimate.

What is Altruism?

One reason why altruism is still highly debated in evolutionary theory is that it is often used inconsistently across disciplines (see Kerr et al. 2004 for the different notions of altruism used in evolutionary theory). Second there are no standardized conventions over the terms of fitness, cost, and benefit. West et al. (2007) attempt to integrate the literature on altruism around a few conventions. They propose a simple and general definition of altruism as "[A] behaviour which is costly to the actor and beneficial to the recipient; … cost and benefit are defined on the basis of the lifetime direct fitness consequences of a behaviour" (2007, p. 416).

This is the definition I will use throughout the article after having made a few important points about it. First, this definition calls for more definitions (see Table 1 for the different definitions and conditions for the concepts used throughout the paper). Direct fitness is defined by West et al. (2007, p. 416) as "the component of fitness gained

[1] For a recent discussion on the relation between multilevel selection and kin selection see Okasha (2015).

Table 1 Key definitions and necessary conditions of the terms used throughout the article

Concept	Definition or necessary conditions
Altruism	"[A] behaviour which is costly to the actor and beneficial to the recipient; … cost and benefit are defined on the basis of the lifetime direct fitness consequences of a behaviour" West et al. (2007), p. 416
Direct fitness	"[T]he component of fitness gained through the impact of an individual's behaviour on the production of offspring" West et al. (2007), p. 416
Indirect fitness	"[T]he component of fitness gained from aiding the reproduction of related individuals" West et al. (2007), p. 416
Relatedness	"[A] measure of genetic similarity" West et al. (2007), p. 416
Inclusive fitness	"[T]he sum of direct and indirect fitness" West et al. (2007), p. 416
Neighbor-modulated fitness	"[T]otal personal fitness, including the effects of one's own behaviour and the behaviours of social partners" West et al. (2007), p. 416
Natural selection	Natural selection results from differences in intrinsic-invariable properties between the individuals of a population that lead to differences in reproductive output
Other evolutionary processes	Other evolutionary processes result from differences in extrinsic and intrinsic-variable properties between the individuals of a population that lead to differences in reproductive output (e.g., mutation, correlated response to selection, drift)
Intrinsic property	Property that does not depend on the existence and arrangement of other objects (Godfrey-Smith 2009)
Intrinsic-invariable property	Property that does not depend on the existence and arrangement of other objects and that does not change over a certain range of environmental conditions at a certain grain of description
Intrinsic-variable property	Property that does not depend on the structure and arrangement of other objects and that can change over a certain range of environmental conditions at a certain grain of description
Extrinsic property	Property that depends on the existence and arrangement of other objects (Godfrey-Smith 2009)
Fitness	Expected number of individuals of one's type produced at the next generation that an individual is causally (either directly or indirectly) responsible for

through the impact of an individual's behaviour on the production of offspring" and is opposed to indirect fitness which is defined as "the component of fitness gained from aiding the reproduction of related individuals." They define relatedness as "a measure of genetic similarity" (2007, p. 416). The sum of direct and indirect fitness represents the inclusive fitness (2007, p. 416). The inclusive fitness approach represents one method of fitness accounting in social evolution. It has been very popular among behavioral ecologists, mainly because it is easily associable to an agential view of evolution such as the one found in Dawkins (1976). Yet, in recent years another approach to study social evolution has gained the favor of theorists, namely the neighbor-modulated fitness or personal fitness approach (Taylor and Frank 1996; West et al. 2007; Gardner and Foster 2008). Under this approach, instead of calculating the inclusive fitness of the individual, its "total personal fitness" is calculated as the sum of its direct fitness and the effects of the behaviors of social partners (West et al. 2007, p. 416). Although the two approaches are widely believed to be formally equivalent, that is, merely different methods of accounting (Rousset 2004; Gardner and Foster 2008; Wenseleers et al. 2010), the neighbor-modulated approach makes causation easier to track down. Because in a population the indirect fitness resulting from an individual helping its neighbors is on average the same as the one obtained by this individual via its neighbors, instead of partitioning the total effect of an actor's behavior into its direct and indirect fitness as is done with the inclusive fitness approach, the partitioning here is made as the direct fitness and the part of the personal fitness due to the help of other individuals (for more details on the two approaches see West et al. 2007; Rosas 2010; Wenseleers et al. 2010). For the purpose of this article I will use a form of the neighbor-modulated fitness accounting.

Note that if we are to speak of altruistic traits (or any other social trait), the actor and recipient need to be two individuals. Thus, throughout the article, when I refer to an actor and a recipient, I will consistently refer to two different individuals. Before going further, it should also be noted that among the different notions of altruism used in the literature (see Kerr et al. 2004), two particular notions have been debated, namely *weak* altruism and *strong* altruism (Wilson 1980; Okasha 2006). A weakly altruistic behavior is a behavior that benefits everyone interacting with the actor and the recipient equally. Because the actor pays a cost of producing this behavior, the cost-benefits balance is lower than for its interactors. Yet, because weakly altruistic individuals gain more than the average individual in the population, they have been considered by some as selfish. A strongly altruistic behavior is a behavior that is strictly costly to the actor, i.e., the direct cost

incurred is higher than the direct benefit obtained from this cost. The evolution of strong altruism is the most challenging one to explain and represents the worst-case scenario. I will thus not consider weak altruism in the remainder of paper and focus only on strong altruism.

Finally, because the notion of fitness can generally be understood either as direct fitness or as total fitness (inclusive or neighbor-modulated fitness), one should note that the cost and benefit terms are relative to the direct fitness of the actor and recipient. If one uses the notion of total fitness instead of direct fitness when referring to cost and benefit, then altruism cannot, by definition, be selectively advantageous. This is because if a total fitness cost is paid by altruistic individuals, the net relative production of altruistic offspring due to the altruistic behavior (total fitness) will always be lower than the relative production of selfish offspring, leaving no room for altruism to ever evolve by natural selection. However when the cost and benefit are defined on the basis of the actor's lifetime direct fitness consequences of the behavior, the indirect consequences of the behavior on the lifetime direct fitness consequence of the recipient(s) are not taken into account, leaving some room for the evolution of altruism by natural selection.

What is Natural Selection?

The philosophical literature on natural selection is very dense, sometimes confusing (but see Pocheville 2010 for a good example of conceptual clarity) and often entangled, for good reasons, with the notion of fitness. As a starting point, one can answer the question "What is natural selection?" in the following way: natural selection is a phenomenon that can only occur when two or more types of individuals in a population found in the same selective environment have different fitnesses.[2]

This necessary condition for natural selection calls for definitions of "fitness," "selective environment," and "type." Elsewhere I provide a detailed account of fitness (Bourrat 2014, Chap. 1, 2015a, b) and show the importance of types in natural selection (Bourrat 2014, Chap. 5). In this article, I will keep things simple and assume a model of population composed of different types of individuals that reproduce asexually (with perfect transmission of their type to their offspring), synchronically, and without overlap of generations. In this model, individuals of identical types are perfectly identical and individuals of different types only vary with one property which is intrinsic-invariable (more on this notion in a moment; see Table 1) within the environment considered. These assumptions will allow me to consider that the fitness of an individual is the expected number of individuals of its type produced at the next generation that this individual is causally (either directly or indirectly) responsible for. Thus natural selection occurs when there are differences in inclusive fitness or neighbor-modulated fitness, not just direct fitness. Although the assumption that fitness is a reproductive output is very limited in scope, it will be sufficient to expose the problem of altruism in the next section. The more sophisticated notions of fitness that can be found in the literature (e.g., Mills and Beatty 1979; Bouchard and Rosenberg 2004; Bouchard 2008; Abrams 2009; Godfrey-Smith 2009) will not undermine the main points of the paper.

These different assumptions do not provide us with a notion of selective environment. Following Brandon I define the selective environment of an individual as the part of the external environment, i.e., everything that is not part of this individual or its phenotype,[3] that has *differential* consequences on the number of offspring of its type produced (directly or indirectly) by this individual when compared to an individual of a different type. Brandon (1990) as well as Nunney (1985) argue that for natural selection to occur the individuals of a population must be found in the same environment. The idea behind this requirement is quite intuitive, if two types of individuals are found on average in different contexts for reasons that have nothing to do with their biology and that, as a result, they produce on average a different number of offspring, the increase in frequency of the most successful type cannot be associated with natural selection because the causes of the increase of the type are contingent to the type.

Godfrey-Smith (2009), using the distinction between intrinsic and extrinsic property (see Table 1), makes the same point from another perspective. According to him, only differences in intrinsic properties between the individuals forming a population (e.g., their chemical composition) that lead to differences in reproductive output should be associated with the notion of natural selection. Differences in extrinsic properties (e.g., location; see Table 1) leading to differences in reproductive output cannot be associated with natural selection. This is because the differences in reproductive output they lead to are

[2] I borrow the term "selective environment" from Brandon (1990). As recognized by Brandon (2014) this concept of natural selection faces several possible objections. One of them is that any case in which the individuals of a population significantly interact with each other (altruism is only one case in which this kind of interaction occurs) will prevent these individuals from being in the same environment. I will regard this problem as non-fatal to this formulation and leave its resolution for further work.

[3] Brandon, in his original definition, refers to the environment as only external factors to the organism. For reasons that cannot be developed here, to be consistent, it should be defined in reference to a phenotype, be it expressed within or beyond the physical boundaries of the organism. See Haig (2012) for a similar notion of the environment in relation to what he calls the "strategic gene."

ultimately associated with differences that pertain to the environment of the individuals, not the individuals themselves.

I consider Brandon's and Godfrey-Smith's frameworks to be largely equivalent. That said, I will use Godfrey-Smith's framework (with some modifications) to characterize natural selection. Modifications on Godfrey-Smith's framework are necessary because although using the notions of intrinsic and extrinsic property as a way to differentiate natural selection from other selective processes seems intuitively right, it is incomplete. To see why, we can start by noting that any biological property, say for instance "height," is obviously diachronically the result of the interactions between the bearer of the property and its environment. Had a given organism been put in a different environment from birth its height might have been very different. Godfrey-Smith's distinction between intrinsic and extrinsic properties only accounts for "synchronic" dependences on reproductive output and does not explicitly account for more "diachronic" dependences on reproductive output. Yet in my view, such dependences matter a lot with respect to natural selection.

Consider for instance the following intrinsic property of an organism, "amount of fat." The amount of fat contained by each individual is generally different for each organism of a population and this might have consequences on reproductive outputs. Using Godfrey-Smith's framework, all the differences in reproductive output due to differences in amount of fat contained by organisms should be attributed to natural selection. The problem here is that there are significant cases in which the difference in reproductive output due to containing a different amount of fat should, intuitively, not be attributed to natural selection. Imagine, for instance, that two organisms have different reproductive outputs due to the fact that they contain a different amount of fat. But the difference here is the result of different life histories that cannot causally be traced back to any of their biological properties. For example, suppose that the two organisms have the same susceptibility to a disease *V*. Yet, one gets *V* due to some contingent event and has to spend more energy to eliminate it. To do so it burns a larger amount of fat than the other organism. As a result the two organisms have different amounts of fat and produce different numbers of offspring. This situation can hardly be associated with natural selection, and Godfrey-Smith's distinction is blind to this case and similar ones—I could have used a similar example with the intrinsic property "height," for instance.

If the reasoning above is correct, intrinsic properties should thus be decomposed into two subtypes that will account for diachronicity in relation to natural selection, namely, intrinsic-*invariable* properties, such as having a particular gene, and intrinsic-*variable* properties, such as having a certain amount of fat or height due to a particular life history causally independent from any intrinsic-invariable properties of the individual (see Table 1). Both intrinsic-variable and intrinsic-invariable properties should be understood as such while specifying a range of possible environmental conditions, a grain of description, and a given period of time. Specifying a range of environmental conditions over a specific period of time is crucial, since what is invariable now and here might not be at a later time or under different conditions. It is possible to imagine that a property such as height, for instance, that does not vary under a range of specific conditions would do so under other conditions if organisms were subjected to those different conditions (e.g., a different gravitational force over time).

With this distinction in mind, individual differences in intrinsic-invariable properties within an environmental background leading to some differences in reproductive output are the only ones to be attributed to natural selection. Differences in reproductive output due to differences between members of the population in extrinsic properties and intrinsic-variable properties, because they ultimately depend on extrinsic properties, within an environmental background should be attributed to evolutionary processes different from natural selection.[4] Thus, for evolution to be the result of natural selection, it should either be the *direct* result of differences in intrinsic-invariable properties that lead to differences in reproductive output, or if it results from differences in intrinsic-variable or extrinsic properties, these differences should themselves be the result of differences in intrinsic-invariable properties that lead to differences in reproductive output, that is, an *indirect* result of differences in intrinsic-invariable properties on reproductive output.

My aim in the rest of the article is to show that some classical cases of evolution of altruism, usually conceived as the result of natural selection, should not always be understood as such if one embraces the notions of altruism and natural selection provided in this section. More importantly, even if one disagrees with the necessary conditions for natural selection I propose, I will show that the evolution of altruism can be the result of two conceptually different sorts of evolutionary processes. Whether or not one wants to call them both "natural selection" is a semantic rather than a fundamental issue that I will not pursue here.

[4] I leave for further work to determine which evolutionary force(s) each kind of difference should be attributed to.

The Problem of the Evolution of Altruism and Its Solution

With these conditions in place, I can now fully specify the simple model I have initiated in the previous section (a population of types that reproduce asexually and in discrete generations), in order to spell out the problem of the evolution of altruism. The model is kept voluntarily simple to be able to grasp clearly the issues at stake. Furthermore, there is a priori no reason to believe that the same issues would not be found in more complex cases, and thus that the conclusion, once drawn with sufficient prudence, could not be exported to real cases. I remind the reader that the assumptions made so far in the model help to compute fitness in the simplest possible way, that is, the average reproductive output (direct and indirect) of an individual after one generation.

With the basic model specified, let us imagine now that our population is composed of two types: A "altruist" and S "selfish" individuals. Suppose that the population has N individuals and that p_A is the frequency of altruistic individuals in the population. Let us also suppose that, at each generation, each individual interacts randomly with only one other individual in the population and that the pairs are formed randomly and synchronically. Let us call X the baseline reproductive output for our two types (i.e., the direct number of offspring they would have with no social interaction), c the cost paid by the actor by producing an altruistic behavior, and b the benefit received by the actor from the individual it interacts with. X, c, and b are all measured in number of direct offspring (unit of direct fitness) produced. We can now calculate the expected total quantity of offspring (O) produced at each generation by both types as follows:[5]

$$O_A = X + \frac{p_A N - 1}{N - 1} b - c \tag{1}$$

$$O_S = X + \frac{p_A N}{N - 1} b \tag{2}$$

The benefit received by each type depends on the individual they are interacting with, which in turn depends only on the frequency of altruistic individuals in the population (minus the focal individual in the case of the altruistic type, hence the value "$p_A N - 1$"). If we now calculate the expected difference in offspring produced between the two types, we have:

$$O_A - O_S = \frac{p_A N - 1}{N - 1} b - c - \frac{p_A N}{N - 1} b = -\frac{1}{N - 1} b - c \tag{3}$$

From Eq. (3) we can see that the quantity $-\frac{1}{N-1} b - c$ is always negative. Thus no matter how much benefit is yielded for the cost paid by an altruistic individual, the number of altruistic offspring produced by an altruistic individual is on average lower than the corresponding number of selfish offspring produced. As a result, the frequency of altruistic individuals declines over generations and ultimately reaches zero.

This kind of reasoning has led Sober and Wilson to claim that altruism does not evolve in this setup because selfish individuals are favored by natural selection (1998, p. 21). Yet, before concluding that natural selection is responsible for this evolution, we need to ensure that the setup of the model satisfies all the requirements established in the sections above for natural selection to be possible. So far we have evidence that the reproductive output of selfish and altruistic individuals is different. But is this difference due to differences in intrinsic-invariable properties or differences in extrinsic or in intrinsic-variable properties which are themselves the result of differences in intrinsic-invariable properties as required by our formulation of natural selection?

It is easy to demonstrate that altruistic individuals have on average a difference in the extrinsic property "number of altruistic potential interactors" assuming every individual has the same probability to be chosen as an interactor. In fact, if each individual interacts randomly with other individuals of the population and there is a limited number of individuals in the population, altruistic individuals interact on average with one less altruistic individual than selfish ones because they cannot interact with themselves. Now, whether this difference can be attributed to a difference in intrinsic-invariable property between altruistic and selfish individuals in this setup and thus be legitimately associated with natural selection is too early to answer. I will come back to this question in the next section. At least at this stage, we must take into account the possibility that it is not and that Sober and Wilson's setup simply does not allow us to tell whether the evolutionary fate of the population results from the work of natural selection (differences in reproductive output due to differences in intrinsic-invariable properties) or results from the work of other processes (differences in reproductive output due to differences in extrinsic and/or intrinsic-variable properties). One way to avoid this difficulty is to assume a population of infinite size. When size is infinite the following assumption can be made:

$$\frac{p_A N - 1}{N - 1} \approx p_A \tag{4}$$

In this case one can consider that, on average, an altruistic individual has the same extrinsic property "number of altruistic potential interactors" as a selfish one. As a

[5] This is inspired from Sober and Wilson (1998, pp. 19–21)

result any evolution observed will be the work of natural selection (assuming types are not variable). In such a case one can assume that O is a good proxy for fitness (W). Thus, we have:

$$O_A = W_A \tag{5}$$

$$O_S = W_S \tag{6}$$

From now on, I will assume a population of infinite size. Plugging (4), (5), and (6) into (3) we find:

$$W_A - W_S = -c \tag{7}$$

Equation (7) demonstrates that when the population size is infinite, the difference in reproductive output between A and S is on average $-c$, which is always a negative quantity. Since the probability for an altruist to interact with another altruist is the same as the probability for a selfish individual to interact with an altruist, this means that in this setup selfishness evolves because of some difference in intrinsic-invariable property between the two types and not some difference in extrinsic property. Thus, Sober and Wilson's claim that selfishness evolves by natural selection is now fully justified.

With the result from Eq. (7) at hand we can now ask under which conditions altruism could evolve and then assess whether it can satisfy our necessary conditions for natural selection. The reproductive output of the individuals in our population depends on four parameters: X, b, c, and the probability of the focal individual to interact with an altruist, which in our setup is p_A. Because X and b have the same values in altruistic and selfish individuals, changing their values would have no consequences on the evolutionary fate of the population. Modifying the value of c, however, could change the evolutionary fate of the population and lead altruistic individuals to invade the population. For that to happen we would need to assign a negative value to c. Yet, by doing so an altruistic individual would by definition not be altruistic anymore. This solution is thus not desirable. The last possibility is to modify the probability for each type to interact with altruistic individuals in the population, so that it becomes superior to p_A for altruistic individuals in a way that leads the benefits they receive to outweigh the cost c they have paid. Let us call p_{AA} the probability for an altruist to interact with another altruist and p_{SA} the probability for a selfish individual to interact with an altruist so that, assuming infinite population size, we have:

$$O_A = X + p_{AA}b - c \tag{8}$$

$$O_S = X + p_{SA}b \tag{9}$$

For altruism to increase in frequency between two generations in the population, the following inequality must be true:

$$O_A > O_S$$

$$X + p_{AA}b - c > X + p_{SA}b$$

Simplifying, we find:

$$(p_{AA} - p_{SA})\, b > c$$

which we can rewrite as:

$$Rb > c \tag{10}$$

with $R = p_{AA} - p_{SA}$.

Inequality (10) is a condition for altruism to evolve put under a similar form as Hamilton's rule. In Hamilton's rule, although R is usually interpreted as a coefficient of relatedness (r), the demonstration provided here shows that relatedness is not fundamental for altruism to evolve.[6] What is fundamental is that individuals of both types interact differentially with other individuals and more precisely that altruistic individuals interact sufficiently more with individuals of their type to outweigh the cost (in terms of indirect offspring produced) they paid for helping others; R is thus a measure of positive assortment (Fletcher and Doebeli 2009; Bourke 2011). A concise but thorough treatment of R as a measure of positive assortment, using a covariance approach, can be found in Queller (1985). For a dynamically richer approach see Van Baalen and Rand (1998). The generalization of Hamilton's rule I propose with inequality (10) is very similar to the one proposed by Fletcher and Doebeli (2009), although they arrive at it via a different road. Inequality (10) is thus consistent with classical evolutionary theory. Here, I have just applied it to a particular setup.

The Evolution of Altruism by Natural Selection

With the conditions for the evolution of altruism in terms of assortment specified in inequality (10) we can now predict when it will increase in frequency over two generations in a population of individuals. Yet one legitimate question to ask is whether this evolution is necessarily the result of natural selection. In this section, I argue that altruism can evolve systematically in a population but that it is not systematically the result of natural selection as per the necessary condition for natural selection specified in the first section.

Remember that one necessary condition for evolution by natural selection to occur is to have either *direct* or *indirect* (that is, via extrinsic and/or intrinsic-variable properties) differences in intrinsic-invariable properties that lead to differences in reproductive output between the individuals

[6] This is also a point made in the literature on reciprocal altruism (Trivers 1971; Axelrod 1984).

of a population. We have seen that when individuals interact randomly in a population of infinite size, altruism cannot evolve by natural selection since the differences in intrinsic-invariable properties between selfish and altruistic individuals (their type) lead selfish individuals to increase in frequency between generations. Finally, we have seen that for altruism to evolve, altruistic individuals must interact preferentially with other altruistic individuals. Yet, if we suppose a population in which altruism evolves–that is, inequality (10) is verified–two general explanations, in relation to our necessary condition for natural selection, are consistent with it. Under the first one, altruism evolves because altruistic individuals happen to interact preferentially with other altruistic individuals for reasons that depend ultimately on some difference in extrinsic (or intrinsic-variable that ultimately depend on extrinsic) properties of the types in a given particular environmental setup. Under the second one, altruism evolves because altruistic individuals interact preferentially with other altruistic individuals for reasons that depend ultimately (that is, directly or indirectly) on some difference in intrinsic-invariable properties of the types in a given particular environmental setup. Let us recall that an example of extrinsic property is location and an example of intrinsic-invariable property for a biological organism is having a particular sequence of DNA.

To illustrate this point, let us take the same setup as in the previous section of a population made up of two types: altruist A and selfish S and in which individuals interact two by two. If no other parameter is specified, individuals interact perfectly randomly, and the probability of interaction with an altruistic individual depends on the number of altruistic individuals in the population N_A for both types. Let us suppose further that for an individual (whether A or S) the probability of interaction with an altruistic individual depends on two causal factors.[7] The first one is an intrinsic-invariable bias β for altruistic individuals in choosing an altruistic partner over a selfish one, with $\beta \geq 1$.[8] This leads an individual A to choose a partner as if it was seeing $N_A\beta$ altruistic individuals with which it can interact, when an individual S in the same conditions sees only N_A altruistic individuals. For the second factor we suppose that the focal individual can only interact with its neighbors (with the same probability in the absence of intrinsic bias). The second factor is thus the proportion of altruistic individuals in the neighborhood $\frac{N_A}{N_T}$ with N_T being the total number of neighbors for the focal individual, which is an extrinsic property of the focal individual. For simplicity, we also suppose that in our setup there is no difference in intrinsic-variable properties between the two types.

Taking into account these two factors we can now write the probabilities of interactions of an individual A and an individual S with an individual A as:

$$p_A = \frac{N_{AA}\beta}{N_{AA}\beta + N_T - N_{AA}} = \frac{N_{AA}\beta}{N_{AA}(\beta - 1) + N_T} \text{ and } p_S = \frac{N_{SA}}{N_T}$$

with N_{AA} and N_{SA} being the number of altruistic individuals an altruistic and a selfish individual respectively can interact with in the neighborhood.

With these two probabilities we can now calculate R for this setup. We have:

$$R = p_{AA} - p_{SA} = \frac{N_{AA}\beta}{N_{AA}(\beta - 1) + N_T} - \frac{N_{SA}}{N_T} \tag{11}$$

If we plug (11) into (10), we obtain the conditions for altruism to evolve in this setup that makes visible differences in intrinsic-invariable and extrinsic properties between the types. We obtain:

$$\left(\frac{N_{AA}\beta}{N_{AA}(\beta - 1) + N_T} - \frac{N_{SA}}{N_T}\right) b > c \tag{12}$$

To see the difference of conditions for the evolution of altruism due to natural selection and due to other evolutionary processes using inequality (12), let us first suppose that there is no difference in bias between the two types for choosing to interact with an altruistic individual. We thus have $\beta = 1$.

Starting from (12), this leads to the following condition for altruism to increase in frequency between two generations:

$$\left(\frac{N_{AA} - N_{SA}}{N_T}\right) b > c \tag{13}$$

(13) represents the conditions for altruism to increase in frequency between two generations when the *only* differences between types are differences in extrinsic properties, that is, in our setup, differences in the proportion of altruistic individuals in neighborhoods. It represents thus the conditions for altruism to evolve *purely* by processes different from natural selection. One scenario that could satisfy (13) and lead N_{AA} to be sufficiently superior to N_{SA} is that the population is viscous through limited dispersal, on top of being viscous through interaction limited to the neighborhood (as has been assumed so far). In such a scenario, because of limited dispersal, after a few generations, altruistic individuals are surrounded on average by more altruistic individuals than selfish individuals are, because their parents being altruistic will have produced altruistic offspring and dispersed them in the vicinity. The

[7] By "causal factors" I mean "difference makers" following Woodward's (2003) interventionist account of causation.

[8] Note that β, in the general case, could be inferior to 1 (but superior or equal to 0), in which case altruistic individuals would "avoid" interacting with other altruistic individuals. I consider here only the case in which altruistic individuals "seek" other altruistic individuals, hence why I assume $\beta \geq 1$.

same reasoning holds for selfish individuals being surrounded by more selfish individuals. Because inequality (13) being verified in this instance does not depend on differences in intrinsic-invariable properties between the two types, if two individuals, one of each type, have the same extrinsic properties, namely the property of being surrounded by the same number of neighbors of each type, then, on average, a selfish individual will be causally responsible (directly or indirectly) for the production of more selfish offspring than the altruistic individual will be of altruistic offspring.

Let us now consider the case in which there is no viscosity anymore in the dispersal of offspring: on average and at each generation altruistic and selfish individuals are surrounded by the same proportion of altruistic individuals in their neighborhood so that $N_{AA} = N_{SA} = N_A$. Under this condition, we suppose that the fact of being altruistic also has an effect on the probability of interacting more often with other altruistic individuals. This leads to the following condition for altruism to increase in frequency between two generations:

$$\left(\frac{N_A\beta}{N_A(\beta-1)+N_T}-\frac{N_A}{N_T}\right)b > c \tag{14}$$

which is equivalent to:

$$\beta(bN_A(N_T-N_A)-cN_TN_A) > bN_A(N_T-N_A) + cN_T(N_T-N_A) \tag{15}$$

(15) represents the condition for altruism to increase in frequency between two generations when the only differences between types are differences in intrinsic-invariable properties, that is, for altruism to evolve *purely* by natural selection. We can show that if this inequality is verified, then $\beta > 1$ (see Appendix), which corresponds to a situation of positive assortment. Contrary to the previous case, because the difference between the two types is intrinsic-invariable, if two individuals, one of each type, are placed, on average, under the same initial conditions or surrounded by the same number of altruistic individuals (that is, they have the same extrinsic properties), this will not prevent altruism from increasing in frequency over two generations.

It is thus possible to separate, in this setup, the two general conditions for altruism to evolve in our population outlined in the earlier section: one that amounts to some difference in extrinsic (and/or intrinsic-variable) properties of the type and that we can associate with evolutionary processes different from natural selection; the other that amounts to difference in intrinsic-invariable properties of the type and that we can associate with natural selection. Of course, if inequality (12) is verified and that $\beta > 1$ and $N_{AA} > N_{SA}$, this can lead to a mixed case of evolution of altruism by natural selection and other evolutionary processes.

Let us remark that in cases in which the types are strictly identical except on one gene or more generally one intrinsic-invariable property that leads to some difference in reproductive output, the only way to satisfy inequality (12) and have *some* evolution by natural selection of altruism, is for this property to have a pleiotropic effect, as in cases of green-beard effect (for more details on this effect see Hamilton 1975; Dawkins 1979; Gardner and West 2010). Although the green-beard effect has been rarely found in nature (Keller and Ross 1998), it seems that it represents the only form of evolution of altruism by natural selection with the same cause being responsible for altruism and positive assortment. While cases in which two (or more) intrinsic-invariable properties or genes (with at least one responsible for the social behavior, and at least another one for the assortative behavior) are transmitted altogether over generations in the population can *theoretically* lead to the evolution of altruism by natural selection, we should expect them to be rare in real populations due to genetic recombination or mutation of one of the two properties. In fact, any allele regulating in a particular way the choice of a social partner should be found in equal proportions in the altruist and the selfish type unless it is always transmitted together with the gene regulating the actual social behavior (selfish or altruist) and that mutations are deleterious. This is because it is advantageous for both types to choose altruistic individuals over selfish individuals as social partners (Nunney 1985). We thus expect β to be similar for the altruistic and the selfish type.

Conclusion

The main point of this article has been to show that following consistent concepts of altruism and natural selection, the evolution of altruism can be considered as the result of natural selection only under a limited number of cases. When individuals interact preferentially with altruistic individuals for reasons that do not depend on intrinsic-invariable properties, the resulting evolution should be attributed to (an)other process(es) than natural selection. To be clear, I do not want to undermine the role of Hamilton's rule in evolutionary theory nor all the theoretical and empirical research that has stemmed from it. Rather, I have shown that when one uses consistent terms, at least two distinct evolutionary processes satisfy the conditions of this rule, only one of which should be understood as natural selection if the distinction between differences in reproductive output due to extrinsic and intrinsic-variable/

intrinsic-invariable properties holds. It should be noted that a possible objection to my approach is that natural selection has several meanings and that it is legitimate to call "evolution by natural selection," at least in some cases, some evolutionary change resulting from differences in extrinsic or in intrinsic-variable properties. Although I accept this as a possibility, I have pulled apart two fundamentally different processes that can lead to the evolution of altruism. If they both are considered by some to be cases of evolution by natural selection, one will have to admit that natural selection is not a unified causal process and that the conceptual distinction made here is still a valuable one.

Acknowledgments I am thankful to Patrick Forber, Arnaud Pocheville, and two anonymous reviewers for their comments on earlier versions of this paper. I am particularly thankful to Arnaud Pocheville for his extensive help on the most technical parts of the paper and for his thorough proofreading. This research was supported under Australian Research Council's Discovery Projects funding scheme (Projects DP0878650 and DP150102875).

Appendix

Let us start from (15):

$$\beta(bN_A(N_T - N_A) - cN_TN_A) > bN_A(N_T - N_A) + cN_T(N_T - N_A)$$

To be meaningful, we assume $\beta \geq 0$. Let us call the term in brackets on the left-hand side of (15) $P = bN_A(N_T - N_A) - cN_TN_A$. Let's call Q the right hand side.

If $P > 0$ we have:

$$\beta > Q/P$$

We know that $cN_T(N_T - N_A) > 0 > -cN_TN_A$, thus $Q > P$ which means that $\beta > 1$.

If $P = 0$ then $0 > Q$, which is impossible.

If $P < 0$ then either $0 > Q$, which is impossible, or $\beta < 0$, which is impossible.

Thus, if (15) holds, $\beta > 1$.

Then $P > 0$ implies $N_A > 0$ and $b > c\frac{N_T}{N_T - N_A}$, in the case where $N_T \neq N_A$. This is an interesting constraint bearing on b which is consistent with the hypotheses classically made in models on the evolution of altruism, namely that the benefit received by the focal altruistic individual is larger than the cost it pays.

Thus if (15) holds, $\beta > 1$ (which satisfies the intuition), $N_A > 0$ (which is expected), and $b > c\frac{N_T}{N_T - N_A}$, if $N_T \neq N_A$.

References

Abrams M (2009) The unity of fitness. Philos Sci 76:750–761
Axelrod R (1984) The evolution of cooperation. Basic Books, New York
Bouchard F (2008) Causal processes, fitness, and the differential persistence of lineages. Philos Sci 75:560–570
Bouchard F, Rosenberg A (2004) Fitness, probability and the principles of natural selection. Br J Philos Sci 55:693–712
Bourke AF (2011) Principles of social evolution. Oxford University Press, Oxford
Bourrat P (2014) Reconceptualising evolution by natural selection. Dissertation, University of Sydney
Bourrat P (2015a) Levels of selection are artefacts of different fitness temporal measures. Ratio 28:40–50. doi:10.1111/rati.12053
Bourrat P (2015b) Levels, time and fitness in evolutionary transitions in individuality. Philos Theory Biol. doi:10.3998/ptb.6959004.0007.001
Brandon RN (1990) Adaptation and environment. Princeton University Press, Princeton
Brandon RN (2014) Natural selection. Stanford Encyclopedia of Philosophy. http://www.plato.stanford.edu/archives/spr2014/entries/natural-selection/. Accessed Jan 2015
Dawkins R (1976) The selfish gene. Oxford University Press, Oxford
Dawkins R (1979) Twelve misunderstandings of kin selection. Z für Tierpsychol 51:184–200
Dugatkin LA, Reeve HK (1994) Behavioral ecology and levels of selection: dissolving the group selection controversy. Adv Study Behav 23:101–133
Fletcher JA, Doebeli M (2009) A simple and general explanation for the evolution of altruism. Proc R Soc B 276:13–19
Gardner A, Foster KR (2008) The evolution and ecology of cooperation–history and concepts. In: Korb J, Heinze J (eds) Ecology of social evolution. Springer, Heidelberg, pp 1–36
Gardner A, West SA (2010) Greenbeards. Evolution 64:25–38
Godfrey-Smith P (2009) Darwinian populations and natural selection. Oxford University Press, New York
Grafen A (1984) Natural selection, kin selection and group selection. In: Krebs JR, Davies NB (eds) Behavioural ecology: an evolutionary approach. Blackwell, Oxford, pp 62–84
Haig D (2012) The strategic gene. Biol Philos 27:461–479
Hamilton WD (1963) The evolution of altruistic behavior. Am Nat 97:354–356
Hamilton WD (1964a) The genetical evolution of social behaviour. I. J Theor Biol 7:1–16
Hamilton WD (1964b) The genetical evolution of social behaviour. II. J Theor Biol 7:17–52
Hamilton WD (1975) Innate social aptitudes of man: an approach from evolutionary genetics. In: Fox R (ed) Biosocial anthropology. Malaby Press, London, pp 133–153
Keller L, Ross KG (1998) Selfish genes: a green beard in the red fire ant. Nature 394:573–575
Kerr B, Godfrey-Smith P (2002) Individualist and multi-level perspectives on selection in structured populations. Biol Philos 17:477–517
Kerr B, Godfrey-Smith P, Feldman MW (2004) What is altruism? Trends Ecol Evol 19:135–140
Mills SK, Beatty JH (1979) The propensity interpretation of fitness. Philos Sci 46:263–286
Nunney L (1985) Group selection, altruism, and structured-deme models. Am Nat 126:212–230

Okasha S (2006) Evolution and the levels of selection. Oxford University Press, New York
Okasha S (2015) The relation between kin and multilevel selection: an approach using causal graphs. Br J Philos Sci. doi:10.1093/bjps/axu047
Pocheville A (2010) What niche construction is (not). La niche écologique: concepts, modèles, applications. Ecole Normale Supérieure, Paris
Queller DC (1985) Kinship, reciprocity and synergism in the evolution of social behaviour. Nature 318:366–367
Rosas A (2010) Beyond inclusive fitness? On a simple and general explanation for the evolution of altruism. Philos Theory Biol. doi:10.3998/ptb.6959004.0002.0004
Rousset F (2004) Genetic structure and selection in subdivided populations (MPB-40). Princeton University Press, Princeton
Sober E (1984) The nature of selection. MIT Press, Cambridge
Sober E, Wilson DS (1998) Unto others: the evolution and psychology of unselfish behavior. Harvard University Press, Cambridge
Sterelny K (1996) The return of the group. Philos Sci 63:562–584
Sterelny K, Griffiths PE (1999) Sex and death: an introduction to philosophy of biology. University of Chicago press, Chicago
Taylor PD, Frank SA (1996) How to make a kin selection model. J Theor Biol 180:27–37
Trivers RL (1971) The evolution of reciprocal altruism. Q Rev Biol 46:35–57
Van Baalen M, Rand DA (1998) The unit of selection in viscous populations and the evolution of altruism. J Theor Biol 193:631–648
Wenseleers T, Gardner A, Foster KR (2010) Social evolution theory: a review of methods and approaches. In: Szekely T, Moore AJ, Komdeur J (eds) Social behaviour: genes, ecology and evolution. Cambridge University Press, Cambridge, pp 132–158
West SA, Griffin AS, Gardner A (2007) Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. J Evol Biol 20:415–432
West SA, Griffin AS, Gardner A (2008) Social semantics: how useful has group selection been? J Evol Biol 21:374–385
Wilson DS (1980) The natural selection of populations and communities. Benjamin/Cummings, Menlo Park
Wilson DS, Wilson EO (2007) Rethinking the theoretical foundation of sociobiology. Q Rev Biol 82:327–348
Woodward J (2003) Making things happen: a theory of causal explanation. Oxford University Press, New York