Biol Theory (2014) 9:286–295
DOI 10.1007/s13752-014-0181-z

LONG ARTICLE

Is Cooperation a Maladaptive By-product of Social Learning? The Docility Hypothesis Reconsidered

Olivier Morin

Received: 29 July 2013 / Accepted: 29 April 2014 / Published online: 19 July 2014

Abstract The docility hypothesis holds that human social learning produces genuinely altruistic behaviors as a maladaptive by-product. This article examines five possible sources of such altruistic mistakes. The first two mechanisms, the *smoke-detector principle* and the *cost-accuracy tradeoff*, are not specifically linked to social learning. Both predict that it may be adaptive for cooperators to allow some altruistic mistakes to happen, as long as those mistakes are rare and cost little. The other three mechanisms are specific to social learning: Through culture, individuals may come to adopt altruistic *norms selected at the group level*. Culture may provide people with cheap, accessible, but occasionally mediocre information that they are too reliant upon—a kind of *informational dumping*. Lastly, people may copy sources good to follow in one domain (like technology) but not in another (cooperation), thus committing *calibration errors*. I argue that those sources of errors are unlikely to lead to important amounts of altruism toward non-kin. Experimental evidence shows humans to be sufficiently skeptical, discriminative, and conscious of their own interest to avoid such altruistic mistakes in most cases. Docile altruism is unlikely to be an important aspect of human cooperation.

Keywords Altruism · Bounded rationality · Cultural group selection · Cultural transmission · Social learning

The docility hypothesis (DH henceforth) proposes that human social learning can sustain a substantial range of biological maladaptations—behaviors that result in a net loss of biological fitness, and thus flaunt the logic of natural selection. In particular, the presence of culture in our species would allow the diffusion of biologically altruistic behaviors—a maladaptive by-product of the generally adaptive practice of learning from others. Various authors have championed this as a likely explanation for the peculiarities of human cooperation. Simon (1990, 1993) was the first to use the word "docility" in this connection, but the project of explaining human cooperation as a by-product of cultural learning is evident in some of Boyd and Richerson's (1985) earliest writings.[1] The view that social learning and cultural transmission have a lot to do with human altruism is also popular among experimental economists (Fehr and Fischbacher 2003). Simon's DH is at the heart of gene-culture coevolutionary explanations of human cooperation (see Gintis' 2003) modeling of Simon's hypothesis; see also Knudsen 2003).

Its proponents typically believe that docile altruism is not merely a matter of theoretical interest, but a crucial dimension of human cooperation.

> [T]he… possibility for a society to cultivate and exploit altruism has very strong implications for social theory, including economics and the theories of political institutions and other organizations. (Simon 1990, p. 1668)
>
> [A]ltruistic norms can hitchhike on personally fitness-enhancing norms. Were this not the case, human

O. Morin (✉)
KLI, Klosterneuburg, Austria
e-mail: olivier@cognitionandculture.net

[1] Although Boyd and Richerson's theories since this book cannot be reduced to the DH.

> society as we know it would not exist. (Gintis 2003, p. 417)

This article will argue that the amount of altruism produced as a maladaptive by-product of social learning does not warrant such assertions.

Following defenders of the DH (such as Simon 1990), I take the word "altruism" in its strict biological sense (Hamilton 1964). A behavior is altruistic if it results in a net inclusive fitness loss for the actor, and a net fitness gain for the recipient. This implies the actor's sacrifice is not compensated by reciprocity, an increase in reputation, the avoidance of punishment, or any other fitness benefit, direct or indirect. The word "cooperation" will be used in a broader sense, to cover both altruism and mutually beneficial forms of helping.

This definition of altruism might seem unduly restrictive, since the word is sometimes used as a mere synonym for cooperation, or even as a label for mutually beneficial cooperation (as in Trivers' (1971) reciprocal altruism). What is called here "genuine biological altruism toward non-kin" is the very opposite of Trivers' reciprocal altruism. The traditional acceptation is useful. It emphasizes an important divide between two ways of explaining cooperative behaviors.

Many approaches to the evolution of cooperation can be called mutualistic (following Baumard et al. 2013): they start from the fact that cooperative strategies have to boost the cooperator's fitness if they are to evolve. In the case of cooperation with non-kin, the reward typically takes the form of advantages linked to reciprocity, reputation, or the avoidance of retaliation. This is not to say that we are systematically selfish, or can calculate the costs and benefits of cooperation. Rather, the psychology of human cooperation is expected to consist in a set of imperfect, and often mistaken heuristics, that evolved to respond to cues indicating kinship, reputation, reciprocity, or retaliation. In the case of cooperation with non-kin, those mechanisms roughly approximate a rational response to incentives. In this view, we do not eschew all the risks associated with cooperation. On the whole, we deal with them in a roughly adaptive way, though we occasionally make uncompensated sacrifices to others.

The DH suggests a different explanatory strategy. It would explain the rise of cooperation without appealing to adaptive mechanisms like reciprocity, reputation, or punishment. In this view, cooperation evolves as a maladaptive side effect of social learning, and human cooperative behaviors do not have to obey adaptive constraints. They do not even need incentives. Genuine biological altruism is not only possible, but widespread. This strategy is attractive. There are problems that reciprocity, reputation, and retaliation do not solve perfectly. Incentives-based cooperation is unstable when reciprocators (or retaliators) are too rare, and when reputation systems are noisy, or underdeveloped. Those problems disappear if cooperation need not depend on incentives: it simply takes place because cooperators wish to sacrifice for others.

Using genuine altruism toward non-kin as the prototype of human cooperation comes at a price, which can be captured in one word: *maladaptation*. Biologically altruistic behaviors toward non-kin are likely to decrease the altruist's individual fitness. Uncompensated sacrifices can be quite thoughtful, deliberate, and, of course, unobjectionable; still, from a purely adaptationist point of view they are "mistaken" in that they do not reward the altruist. The DH could be viewed as consisting of two main claims:

(1) Altruistic "mistakes" are the chief cause of cooperation in our species: cooperative behaviors do not have to be adaptive at the individual level.
(2) Culture causes biologically altruistic "mistakes" to be much more common in our species.

A full rebuttal of claim (1) is beyond the scope of this article; one would need to provide a complete mutualistic account of human cooperation. Several hypotheses have been proposed (see for instance Baumard et al. 2013). This article focuses on claim (2): is culture responsible for an increase of altruistic mistakes in our species? It tackles it from two angles.

First, in the following section, I look at two main causes of altruistic mistakes that do not require social learning. These are:

The Smoke-Detector Effect Altruistic mistakes may be much less costly than selfish mistakes. This may happen because the benefits of reciprocity, of increasing one's reputation or avoiding punishment, are not entirely predictable. Therefore, it could be adaptive to err on the safe side, betting on good deeds not going unrewarded.

The Cost/Accuracy Tradeoff Investing time and energy in a cognitive task may not always be a worthwhile effort. Therefore, even if we could calculate the fitness costs and benefits of every decision (which we clearly cannot), it may not always be in our best interest to try.

The mistakes occasioned by these mechanisms are, I will argue, by-products of individually adaptive strategies for cooperation, and do not provide an independent solution to the puzzle of cooperation.

The next main section aims to show that the emergence of culture in our species did not change this. It considers three sources of altruistic mistakes which are a direct consequence of social learning:

A Cultural Selection for Altruistic Norms Competition between organizations will be easier on those which promote cooperation better. One possible consequence of this is the rise of norms prescribing biologically altruistic

behaviors. The presence of such norms in cultural repertoires would imply that social learners get more exposure to altruistic behaviors than they would if they relied on individual learning.

Informational Dumping The wide availability of cheap social information could make us less likely to use our own cognitive resources, even when those personal resources would yield better advice. Docile individuals would prefer the cheap, low-quality information they get from others, with maladaptive consequences.

Calibration Errors Social learning could induce numerous mistakes in the domain of cooperation, because we would be misled by its benefits in other areas, such as technology. We would copy maladaptive altruism from people who proved to be good models in other domains.

Those three mechanisms, I will argue, are not likely to work as expected by defenders of the DH. The last main section before the conclusion will examine some famous experimental examples of imitative altruism, to show how this literature can be reinterpreted in light of these considerations.

General Failures of Adaptive Cooperation

The Smoke Detector Principle

Many students of cooperation have remarked that when cooperation is adaptive most of the time, it makes sense to tolerate a certain rate of generous mistakes: giving goods and services even when we have no reason to think we might get a proportionate benefit. Why? Because one cannot entirely rule out the possibility that cooperation might be beneficial. You may think nobody is looking at you when you refrain from littering the streets, but you never know for sure. You might be escaping punishment, gaining reputation, or attracting reciprocity. Cognition for cooperation, in this view, is like a smoke detector: it is tuned to sound many false alarms, since those are less costly than failures to sound the alarm when it matters. Smoke detectors tend to overestimate the benefits of cooperation—that are, arguably, both important and hard to foresee (see, e.g., Delton et al. 2011; Baumard et al. 2013). This mechanism can be seen as an adaptive reaction to the unpredictable benefits of cooperation. It would not arise if those benefits were absent or scarce—which would be the case if maladaptive altruism were the standard form of cooperation. Thus, the smoke-detector hypothesis only makes sense in a mutualistic perspective.

Bounded Rationality and the Cost/Accuracy Tradeoff

Simon's DH was inspired by his theory of bounded rationality (Simon 1972), in which two sources of cognitive limitation, external and internal, are distinguished. In the case of fitness, the external sources of maladaptation are simply all the unpredictable events that will modify our fitness in the future. Only a supernaturally prescient agent could take those into account. Internal mistakes, on the other hand, are due to an agent's own cognitive limitations. In Simon's view, docility was fueled primarily by "internal" mistakes: docile altruism was a consequence of the fact that some fitness-related problems are too difficult for us to tackle.

According to bounded rationality, internal mistakes result from a tradeoff between computational costs and accuracy. To think is to spend some effort on computation, in order to make more accurate decisions. Given the limits of our intellectual capacities, there has to be a point where the benefits of computation no longer balance its costs. Investing more cognitive effort might improve our decisions, but the improvement would not be cost-effective. Prohibitive cognitive costs would thus have us accept inevitable mistakes.

What would cooperation look like if based on such mistakes? We might get a rough answer from what we know about misallocation of parental care. Breeding the wrong offspring is a typical example of maladaptive altruism. Recognizing one's own kin is not a trivial feat: computational costs will likely be high (all the more so since young non-kin have an incentive to deceive strangers into caring for them), and accuracy cannot be perfect. Even so, most animals perform remarkably well on this difficult task, thanks to cheap heuristics.

Davis and Todd (1999) show, for instance, that birds feeding their brood may minimize their mistakes by using simple decision rules. In humans, Silk (1980) shows that in those cultures where adoption is most prevalent (up to one-third of children), it mostly concerns close kin, with a tight match between relatedness and the likelihood of adoption (individuals with a relatedness coefficient of .25 with their adoptive parents, like nephews or grandsons, account for almost 60 % of all adoptions). If we consider Silk's data on Oceanian cultures, and look at the cases of adoption which cannot be accounted for by a relatedness coefficient of at least .0625 (corresponding to the relatedness between you and your first cousin once removed), we can see that those represent a little less than 10 % of the total of adoption cases (not taking into account the cases for which relatedness is unknown, or information is unavailable)—which (given that adoptions amount to a third of children at most) puts the total proportion of children raised by genuine non-kin around 3 % or below. Concerning that 10 % of cases, however, Joan Silk stresses that most adoptions benefitted foster parents in some way: they may have received support in old age, an increase in reputation, a link with a rich and powerful family, all of which could translate into all sorts of social and material benefits.

Thus it seems that many animals either pay high cognitive costs to detect their own offspring, or find cheap ways around the problem. One might reply that kin recognition is an evolutionary ancient problem, whilst many cooperation problems are too recent for evolution to have equipped us with built-in heuristics. There is, however, no reason to think that the power of heuristic thinking should depend on hardwired modules. Individual learning matters. It can invent smart heuristics in ontogenic time. Gambetta and Hamill's (2005) study of taxi drivers in Belfast and New York presents a set of heuristics designed in a specific professional context, and effectively solving an intricate and multidimensional cooperation problem with limited cues. Taxi drivers use superficial cues, combined with their intuitive grasp of their city's sociology, to guess which would-be passengers would make good clients. A task that would look mind-boggling in the abstract—how to predict the outcome of a risky commercial interaction with a perfect stranger—appears to be manageable at a relatively low cost.

Such heuristics as used by taxi drivers suggest that Simon's tradeoff can be met without sacrificing a great deal of accuracy. Gigerenzer and his colleagues famously claim that cheap and simple heuristics sometimes perform almost as well as complex and expensive ones. Simon's tradeoff still rules in theory (if it did not exist, simple heuristics would not be needed at all), but in practice, we seldom need to choose between big efforts and big mistakes: "Models of inference do not have to forsake accuracy for simplicity. The mind can have it both ways" (Gigerenzer and Goldstein 1996).

Not to put too fine a point on it, Gigerenzer and Goldstein's heuristics are free lunches. Heuristic users do not choose the cheap-but-inaccurate option in Simon's cost/accuracy tradeoff. For all practical purposes, they *escape* the tradeoff.

Is Social Learning a "Simple Heuristic that Makes Us Smart"?

To someone familiar with the bounded rationality debate this will sound obvious. What these well-known views imply about social learning is not so clear, however.

On the one hand, most proponents of the DH (including, of course, Simon) agree that social learning is a cheap and adaptive shortcut across complex problems—one more "good trick" in the repertoire of human adaptations. Richerson and Boyd (2005, p. 120), using Gigerenzer's famous slogan, call social learning a "simple heuristic that makes us smart."More generally, in all models where altruism evolves because of social learning, it evolves as a by-product of a hugely beneficial adaptive strategy (André and Morin 2011).

In spite of this, defenders of the DH insist—in the same breath—that social learning is not a free lunch, after all. A few pages after endorsing Gigerenzer's view, Richerson and Boyd argue that using social learning comes at a price. "Culture," their slogan goes, "is built for speed, not for comfort" (Richerson and Boyd 2005, p. 187). In many cases, they argue, relying on social learning amounts to choosing the fast-and-inaccurate option in a speed/accuracy tradeoff. This will cause egregious mistakes that individual learners would not commit. To cite one of their favorite examples, social learning will push us to become kamikaze killers if that is what everyone else does (p. 204ff.). Clearly, in this view, social learning is not always a simple heuristic that makes us smart. It is a simplistic trick that makes us dumb—at least in some domains, at least with altruism.

What kind of a heuristic is social learning? Is it a heuristic *à la* Gigerenzer—in which case it would escape the cost/accuracy tradeoff? Or is it simply a choice for speed and economy over accuracy—in which case its maladaptive potential would be huge?

Does Social Learning Contribute to the Importance of Altruism?

As we saw, everyone seems to agree that social learning must enhance our capacity to make adaptive decisions, overall and all things considered. As far as cooperation is concerned, however, things could be different. It could be the case that social learning is a simple heuristic that makes human individuals smart when they learn to feed, to talk, to hunt, or to swim—but when it comes to cooperation, it would betray the individual and impose altruistic mistakes on him or her. The benefits of relying on culture in other domains might be so high that individuals would put up with the heavy maladaptations that ensue. As Simon put it, altruism is society's tax on the benefits of cultural transmission. In this scenario, social learning for cooperation does not make us ever so slightly smarter in that domain. Two mechanisms correspond to this view: *informational dumping* and *calibration errors*.

There is another possibility. Social learning may promote altruism without imposing any additional cost on us. This may happen if cultural evolution eliminates norms of cooperation that are maladaptive, but not altruistic, to replace them with altruistic norms. Thus the prevalence of altruism may be increased without adding anything to the burden of maladaptation. Gintis' (2003) version of the DH is based on such a mechanism, and on *cultural selection for altruism.* We start with it.

Would Cultural Selection Favor Altruism Over Other Forms of Cooperation?

That there would be cultural selection regarding norms of cooperation is hard to doubt: a look at any piece of

institutional history tells us of the many institutional arrangements that were rejected, faded away, or died out. If cultural selection makes altruistic norms more prevalent in cultural repertoires, this will increase the odds that social learners exposed to such repertoires will behave altruistically, compared to non-social learners.

Gintis (2003) argues that this need not introduce any additional maladaptation. Why? Because cultural selection could replace non-altruistic maladaptive norms with altruistic ones. In other words, it would select against merely detrimental maladaptive behaviors with one hand, while positively selecting altruistic maladaptive behaviors with the other. Cultural group selection could increase the proportion of altruistic behaviors in the maladaptive part of the cultural repertoire, relative to merely detrimental behaviors. Thus, in theory, cultural altruism might be adaptively neutral overall. It would take the garbage of merely maladaptive mistakes and turn it into altruistic gold.

Whether it does so depends on a tricky combination of factors. In Gintis' view, cultural selection must replace simply maladaptive moves with altruistic (and maladaptive) moves—but it must destroy at least as much maladaptation as it creates. This suggests that cultural group selection for altruism requires a very specific balance of forces. Even if we grant that this complex equilibrium obtains, there are additional conditions that must be satisfied for altruism to become pervasive.

Condition 1: Cultural Selection Must Favor Altruistic Norms Specifically, Rather Than Norms of Cooperation in General

A society can only implement so many norms of cooperation, and altruistic norms are not the only option. Other forms of cooperation (where cooperators are rewarded) can also produce shared benefits. Altruistic norms do not necessarily outcompete others. We are used to thinking of altruism as the most valuable form of cooperation, because of the outstanding moral virtues that it requires. Yet some forms of cooperation may benefit society more than altruism. Indirect and direct reciprocity, in particular, arguably produce better outcomes for entire groups in some cases. Non-zero-sum games are such a case: when the coordinated efforts of each participant can increase the total gain to be shared, reciprocity gives everyone an incentive to join. Such was the fundamental insight of early economic theory: charity, like theft, merely transfers goods from one person to another, while trade spurs everyone to produce more. Selection at the group level can be sensitive to those benefits, just as it takes the benefits of altruism into account. Thus, in Friedrich Hayek's version of cultural group selection (1988), market institutions are favored over forms of cooperation based on uncompensated transfers, because the former are able to produce common goods that are inaccessible to the latter.

Condition 2: Individuals Must Not Select Against Altruistic Norms

It may be useful to distinguish two forces acting upon cultural selection: demographic dynamics and individual choices. Demographic forces guide cultural selection when some norms outlast others because some groups drive other groups to extinction (Darwin 1871/2004; Hayek 1988; Sober and Wilson 1999). Individual choices need not figure in the process: norms could be selected at the group level even if individuals had no choice at all—if they followed whatever norms prevailed in their society.

Other forces drive cultural selection, besides demographic selection at the group level. Individuals and their choices also weigh on cultural selection. They can, for instance, vote with their feet against a culture where they would be asked to sacrifice themselves without getting anything in return. Somewhat confusingly, such individual choices play a very important role in many models of "cultural group selection." Thus, Boyd and Richerson (2009) argue that mutually beneficial norms of cooperation may evolve if people tend to join the groups where cooperation is most beneficial for them. This leads to the growth of groups where mutually beneficial norms of cooperation circulate, and to the gradual attrition of others. Even though they describe the process as a cultural group selection phenomenon, the growth of cooperative groups is nothing but a direct result of individual choices. It is no coincidence, I think, that the resulting form of cooperation happens to be mutualistic, not altruistic.

Historically, however, the cultural group selection hypothesis has been used to shortcut individual choices. In Simon's DH, cultural evolution is led by demographic success alone. Individual choices do not figure.

> A society that instilled such behaviors [i.e., altruistic behaviors] in its docile members would grow more rapidly than one that did not; hence such behaviors would become, by evolution at the social level, a part of the repertory of proper behavior of successful societies. Societies that did not develop such a repertory… would ultimately disappear. (Simon 1990, p. 1667)

We may not have any preference for altruism; it will spread nonetheless, as long as we do not actively reject it, because altruistic societies will survive. This strategy of bracketing individual choices away is not specific to defenders of the DH. In Hayek's theory of cultural group selection, the rise of market-based cooperation happened in spite of

individuals. Human psychology, Hayek thought, is fundamentally altruistic—as a consequence, we have an instinctive aversion to market principles. They would not have prevailed if market-based societies had not demographically outperformed other societies. In both Simon and Hayek, cultural group selection acts independently of individual preferences, and sometimes in spite of them.

This strategy has one obvious drawback: how do we make sure that individual selection does not disturb group-level selection? The problem is particularly keen for the DH. So far, we are trying to maintain the assumption that social learning makes us smart. This means individuals are willing and able to screen off altruistic behaviors, and social learning does not prevent them from doing so, quite the contrary.

Do these assumptions hold in models where altruism evolves because of culture? I do not think so. Consider two models often cited to argue that biologically altruistic punishment evolves by cultural transmission: Guzmán et al. (2007), and Henrich et al. (2001). Both papers model a population where norms of cooperation and norms of punishment are applied by some agents, agents that others can copy. The norm of cooperation is a norm of mutual help: whoever applies it benefits in the long run. The norm of punishment, in those two models, is genuinely altruistic: agents punish other agents at a cost to themselves, without getting anything in return. Now, some agents are predisposed to copy others; some agents are not. Those who copy reproduce the norm of mutual help. As a result, they fare better than non-copiers, because mutualistic cooperation is beneficial. However, there is a catch: the agents who copy the (beneficial) norm of mutualistic cooperation *must* also copy the (detrimental) norm of punishment. Selectivity is not an option.

In both models, free riders are avoided by construction: an agent simply cannot share in the mutually beneficial norm without paying the cost of altruistic punishment. (Whether this altruistic cost is ever paid is debatable—in Henrich and Boyd's model, altruistic punishment might never have to be resorted to.) Likewise, in Gintis' version of the DH, altruistic norms hitchhike on the ride of other transmitted norms, only because agents are too docile to select against altruistic norms (Gintis 2003, p. 414). In all those cases, selection by individuals is implicitly or explicitly kept from interfering with the rise of altruism.

In summary, cultural evolution at the group level may increase the level of altruism only if three demanding conditions are met: (1) a complex equilibrium between the elimination of maladaptation and the promotion of altruism; (2) altruistic norms that outcompete norms of cooperation based on reputation, reciprocity, or punishment; and (3) docile individuals, who do not care whether the norms they adopt will compel them to make costly sacrifices.

What happens if the conditions are not met? Then the DH might still be true, but it would need to take a new premise on board: it would need to assume that social learning turns us into dumb cooperators. Let us examine two mechanisms that would make this possible.

Does Informational Dumping Foster Docility?

Docile individuals are often described as unable or unwilling to use their own cognitive resources to evaluate the payoffs of various forms of cooperation, blindly following social models instead: "Docility will reduce the inclination to evaluate independently the contributions of behavior to fitness.… Hence the docile individual will necessarily also incur the cost, c, of altruism" (Simon 1990, p. 1667).

One possible cause of such docility might be called "informational dumping."The social world is an abundant source of cheap and easily accessible information about the cost and opportunities of cooperation. Defenders of docility (e.g., Gintis 2003, p. 414) often argue that such information is almost impossible to obtain individually. In such cases, it would make sense to choose the cheap-and-dirty alternative on the cost/accuracy tradeoff, and blindly rely on outside sources when it comes to opting for one form of cooperation over another. As happens with commercial dumping, culturally imposed norms would win over not because they are any better than individually chosen forms of cooperation, but simply because they are cheaper to implement.

There is, however, an obvious limit to the maladaptations that can be imposed in this way. Cheap and low-quality sources of information do not outcompete expensive but reliable sources all the time (otherwise there would be no market for information, as distinct from propaganda or advertisement). A boundedly rational learner should balance the cost of tapping a source of information with the reliability of that source. If the benefits of tapping a source are not worth the costs, then the source should never be tapped, even if it were the only available source. Thus the presence of a cheaper alternative can only make learners better off.

If, on the other hand, a source is worth paying the costs of tapping into it, we should not prefer a cheaper source, unless it offers comparable informational benefits. More specifically, the ratio setting informational benefits against access and computation costs should be better in the new source. Moreover, our readiness to appeal to costly computational procedures should vary in function of the interests at stake in a given decision: the higher the stakes, the more important the cognitive effort we should devote to it. As a result, we should prefer an inferior-but-cheaper source only when the stakes are low enough, the better

source is expensive enough, and the cheaper source is reliable enough.

As we shall see below, experimental data from social psychology confirm this pattern. Those three constraints (low stakes, a cheaper but not so unreliable source) strongly limit the effects of informational dumping on maladaptive behavior. We should not expect boundedly rational learners to trust clearly inferior sources when making important decisions.

How Important are Calibration Errors?

We can misevaluate a source of information in several ways: we can over- or underestimate a given source's competence (*misestimation*). I misestimate Mary if I think that she is a reliable bird expert, while in fact she doesn't know a cuckoo from a robin. We can also err by generalizing an evaluation of a source made in one domain, to a domain where the evaluation is no longer valid (*bad calibration*). My evaluation of Mary is not properly calibrated if I take her to be a reliable bird expert just because she is an outstanding botanist.

Misestimation in itself does not create any bias. Errors go both ways, underestimation being just as likely as overestimation. Our evaluations of various sources are imperfect and full of errors, and, while this can be a recipe for excessive docility, it can just as well make us overly unruly.

Bad calibration of an evaluation for a given source is more important. The bias it creates is well illustrated by an anecdote often cited by Boyd et al. (2011, p. 10922), an advertisement campaign where Michael Jordan sells underwear. Boyd and Richerson assume that the ad campaign had some impact: people followed Michael Jordan's implicit recommendation. If they did, they were victims of a bias. This bias consists in extending an attribution of competence in one area (basketball) to another area (one's choice of underwear). Their evaluation of Michael Jordan would not be properly *calibrated*: Michael Jordan's competence lies in the area of basketball, where we evaluate it correctly, not to an area where it should not be relevant. Such a bias would have maladaptive consequences.

Calibration errors are bound to happen, of course, but are they pervasive? Classic sociological work on "influentials" has shown that social influence is typically narrow and domain-specific. Two-thirds of the influentials studied by Katz and Lazarsfeld (1955) in their study of an Illinois village were "monomorphic." In other words, they influenced others only in their one domain of competence (see also Merton (1968, Chap. XII): "Patterns of influence in local and cosmopolitan influentials."). Even the Michael Jordan anecdote is misleading. We know very little about the impact of celebrity endorsement on actual purchases. In their meta-analysis of hundreds of studies on celebrity endorsement effects, Amos et al. (2008) cite only four studies based on data concerning actual purchases. None of those studies found a significant effect of celebrity endorsement. In laboratory interviews, people do claim that celebrity endorsement could make a difference on purchases. Yet, as Amos et al. make clear, that effect is usually observed when there is a tight link between the celebrity and the product—for instance, if Michael Jordan sold basketballs or basketball shoes. It is unlikely that Jordan ever helped sell underwear.

The Experimental Evidence for Smart Social Learning Versus Docile Altruism

When it Matters, We Base Our Decisions on Accurate, Well-Calibrated Evaluations of Social Sources

Abundant experimental evidence shows that, when it matters to them, human adults calibrate their trust in a careful way, and mostly for their own benefit. An example is provided by the many replications of Asch's famous conformity experiment. As is well known, a substantial minority of people will systematically endorse the false opinion of a majority (Asch 1955). One should note, however, that in most versions of the experiment, imitating the majority entails no cost at all (on the contrary, pleasing others may be counted as a benefit). What happens when rewards are introduced?

In a modified version of Asch's paradigm, Baron et al. (1996) asked subjects to recognize, in a lineup, an individual they had previously seen on a picture. They varied both the amount of information available to the subjects (by changing the time of exposure to pictures), and the importance of the task (by introducing monetary incentives). Subjects blindly imitated a misleading confederate when the stakes were not high, or when their own personal information was unreliable (i.e., when the task was difficult). They trusted their own judgments otherwise, i.e., in the condition where the stakes were high and the task was easy. It should be noted that when the task is difficult and the motivation is high, subjects are much more likely to imitate the confederates than in any other condition. This makes perfect sense, since they have every reason to trust the unanimous confederates and few reasons to trust their own dubious perception. Of course, the experiment is rigged, so that trusting the confederates is always a losing strategy. Even so, mistakes in this condition are not more frequent than in the control condition (where subjects pass the difficult task alone). In other words, subjects followed the misleading confederates mostly when their own judgment would have been wrong as well. Their use of social

information was flexible, indeed close to optimal given the constraints.

This result mirrors some interesting data obtained in the psychology of persuasion. People let others' arguments inform their own decisions to the extent that better and more direct sources of information are absent or mediocre (Conway and Schaller 2005). We endorse other people's opinion when the issue at stake is perceived as irrelevant, or when a lack of information makes it rational to do so, but not otherwise. This applies to the influence of prestige (Rhine and Severance 1970; Petty et al. 1981; Axsom et al. 1987; Petty and Wegener 1998). This also curbs the influence of conformity (Axsom et al. 1987; Mackie et al. 1990). We trust others' information conditionally on their competence and benevolence (Pasquini et al. 2007; Mascaro and Sperber 2009).

The Experimental Evidence for Costly Imitative Altruism

The DH predicts that altruistic behaviors should be much more prevalent in a species of social learners, and should overwhelmingly be a product of imitation. Numerous results coming from social psychology are thought to support the view that humans routinely reproduce acts of generosity carrying an important cost, for reasons having nothing to do with punishment, reciprocity, or reputation.

Imitative altruism in children

The best evidence of costly imitation probably comes from studies showing that children are more likely to be generous (or violent) when they have witnessed a model behaving generously (or violently) (Bandura 1963; Bryan 1971). In Bryan's "jar studies" in particular, children are made to win a small reward in chips (which may be exchanged for real toys), and then told they may give a part of it away to a child in need by placing it in a jar. The setup resembles a dictator game, and just like in the dictator game, children are quite likely to show some generosity, with or without imitation. However, when the experimenter sets the example by giving away her own chips, children are more generous. The effect, however, is weak, and the authors argue that it can be explained away by a simple disinhibition effect:

> The effect of the generous model is hardly a strong one …. A hypothesis that appears reasonable concerns the disinhibition of behavior.… For many children, set as they are in the novel contexts of both a laboratory and a helping situation, the witnessing of a "novel" behavior without reprimand would subsequently increase the likelihood of such behavior. (Bryan and Walbek 1970, pp. 346–347)

In other words, the model shows giving is permitted and has no unpleasant consequences. She may also reinforce the salience of the altruistic action by attracting the child's attention to it (the same could be said of adult studies on model-induced giving; Bryan and Test 1967; Rushton and Campbell 1977[2]).

Most other effects attributed to costly imitation in children (like "overimitation" phenomena observed in three- to five-year-olds; Lyons et al. 2011) are weak and heavily context-dependent. The costs they involve are trivial compared to the costs of any important cooperation episode among adults.

Economic Games

In most experimental settings, imitating others entails only the smallest costs. A child who hits a plastic doll believing he is not observed suffers no obvious and important cost (Bandura 1963). Likewise, passersby who imitate a demonstrator looking above her own head (Milgram et al. 1969), students covering a computer's keyboard (Coultas 2004), subjects reproducing an experimenter's posture or gestures (Dijksterhuis and Bargh 2001) do not risk or gain anything valuable. The situation is different in economic games, where each move entails relatively important risks and gains. Still, research on social learning in economic games suggests that subjects rely on imitation (if ever) to the extent that it seems likely to improve their own payoff. (McElreath et al. 2005, 2008; Efferson et al. 2007, 2008). Such examples do not support the view that imitation is, on average, more likely than individual decisionmaking to yield maladaptive decisions regarding cooperation.

Some authors (most clearly Gintis et al. 2003) have taken the existence of cross-cultural variation in economic games (Henrich et al. 2005) as evidence that culturally transmitted norms can cause altruistic behavior. There are many reasons to resist this interpretation.

First of all, there is no consensus on the motivations underlying altruistic giving in, for instance, the dictator game. It is increasingly clear that (guarantees of anonymity notwithstanding) some implicit concern for one's reputation still moves most players (Dana et al. 2007). Framing effects and experimenter demands are also increasingly suspected to underlie altruistic giving (as argued by

[2] It is debatable whether Rushton and Campbell's (1977) study shows what it says in the abstract. In their experiment, 27 subjects were exposed to a model who agreed to give her blood, while eight subjects were not. Eighteen out of the 27 subjects, and two out of the eight agreed to give blood in the future. Weeks later, nine subjects out of the 27, and none out of the eight, actually agreed to give blood. Contrary to what the authors claim, neither comparison is significant on a Fisher exact test.

Winking and Mizer (2013), who find no donation at all in a dictator game played in an ecological setting).

Second, the existence of variations from one geographical setting to another is not enough to prove that culturally transmitted norms are the cause. Important differences also exist between neighboring communities with no obvious cultural differences: adjacent Tsimane villages (Gurven et al. 2008), different neighborhoods of Newcastle-upon-Tyne (Nettle et al. 2011), communities of horticulturalist–foragers less than 100 km apart (Lamba and Mace 2011). Quantitatively speaking, those intra-cultural differences are quite comparable to inter-cultural differences (Lamba and Mace 2012).

Conclusion

Both cultural group selection for altruism and maladaptive mistakes of social learning occupy a central place in current work on the evolution of cooperation. If the present account is on the right track, a change of focus may be in order. Proponents of the DH are right when they note the existence of altruistic mistakes resulting from social learning, but this article found no reason to believe those mistakes should be numerous or important.

Altruistic mistakes may spring from two kinds of sources: some have to do with cooperation in general, and some are specifically linked to social learning. The benefits of cooperation are both high and unpredictable, so much so that it makes sense to be generous even when rewards seem unlikely. Generous behaviors do happen as a result, but this is because they tend to be adaptive. The tradeoff between cost and accuracy that decision makers face can often be solved in a simple and smart way with cheap strategies that do not occasion many mistakes. Social learning is one of those smart strategies. Yet the DH suggests that it is also an important source of altruistic mistakes.

Cultural selection at the group level may favor altruistic norms, but it may just as well favor other forms of cooperation, like mutualistic norms of fairness that benefit everyone, including cooperators (Baumard et al. 2013). Furthermore, cultural selection is also driven by individual choices, and individuals are likely to prefer mutualistic norms (which entail no detrimental sacrifice while still permitting cooperation), thus standing in the way of altruism.

Social information may also play havoc with our capacity to make adaptive decisions. It could flood us with cheap and inaccurate information that would prove thoroughly unhelpful in the area of cooperation (though it could be precious elsewhere). I doubt, however, that such informational dumping could blind us to the costs of altruism. The experimental literature shows humans to be discriminating and flexible imitators.

Thus, the maladaptive side effects of social learning are unlikely to have been important forces in the evolution of cooperation. The most promising approaches of this phenomenon appear to be mutualistic: they start from the fitness benefits accruing to individual cooperators. These may consist in an escape from punishment, an increase in reputation, or some reciprocation (direct or indirect). Those mechanisms are already at the core of most accounts of the evolution of cooperation. In spite of this, the credit for maintaining human societies in existence is often given to docility, altruism, and cultural group selection—in other words, to the benefits of blindly trusting a culture engaged in a war of norms with other cultures.

Acknowledgments This paper was written while the author was a postdoctoral research fellow at the Cognitive Science Department of the Central European University in Budapest. Jean-Baptiste André contributed many important remarks, comments, and reflections on earlier versions of this paper. I thank Dan Sperber, Francesca Giardini, Olivier Mascaro, and Christophe Heintz for helpful comments.

References

Amos C, Holmes G, Strutton D (2008) Exploring the relationship between celebrity endorser effects and advertising effectiveness—a quantitative synthesis of effect size. Int J Advert 27:209–234

André J-B, Morin O (2011) Questioning the cultural evolution of altruism. J Evol Biol 24:2531–2542

Asch S (1955) Opinions and social pressure. Sci Am 193(5):31–35

Axsom D, Yates S, Chaiken S (1987) Audience response as a heuristic cue in persuasion. J Pers Soc Psychol 53:30–40

Bandura A (1963) The role of imitation in personality development. J Nurs Educ 18:1–9

Baron R, Vandello J, Brunsman B (1996) The forgotten variable in conformity research: impact of task importance on social influence. J Pers Soc Psychol 71:915–927

Baumard N, Andre J-B, Sperber D (2013) A mutualistic approach to morality. Behav Brain Sci 36:59–122

Boyd R, Richerson PJ (1985) Culture and the evolutionary process. University of Chicago Press, Chicago

Boyd R, Richerson PJ (2009) Voting with your feet: payoff-biased migration and the evolution of group-beneficial behavior. J Theor Biol 257:331–339

Boyd R, Richerson PJ, Henrich J (2011) The cultural niche: why social learning is essential for human adaptation. Proc Natl Acad Sci USA 108:10918–10925

Bryan J (1971) Model affect and children's imitative altruism. Child Dev 42:2061–2065

Bryan J, Test MA (1967) Models and helping: naturalistic studies in helping behavior. J Pers Soc Psychol 6:400–407

Bryan J, Walbek N (1970) Practicing and preaching generosity: children's actions and reactions. Child Dev 41:329–353

Conway LG, Schaller M (2005) When authorities' commands backfire: attributions about consensus and effects on deviant decision-making. J Pers Soc Psych 89:311–326

Coultas JC (2004) When in Rome…An evolutionary perspective on conformity. Group Process Intergroup Relat 7:317–331

Dana R, Weber RA, Xi Kuang J (2007) Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. Econ Theory 33:67–68

Darwin C (1871/2004) The descent of man and selection in relation to sex. Penguin, London
Davis J, Todd P (1999) Parental investment by simple decision rules. In: Gigerenzer G, Todd P (eds) Simple heuristics that make us smart. Oxford University Press, Oxford
Delton AW, Krasnow MM, Cosmides L et al (2011) Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. Proc Natl Acad Sci USA 108:13335–13340
Dijksterhuis A, Bargh J (2001) The perception-behavior expressway: automatic effects of social perception on social behavior. Adv Exp Soc Psychol 33:1–40
Efferson C, Richerson PJ, McElreath R et al (2007) Learning, productivity, and noise: an experimental study of cultural transmission on the Bolivian altiplano. Evol Hum Behav 28:11–17
Efferson C, Lalive R, Richerson PJ et al (2008) Conformists and mavericks: the empirics of frequency-dependent cultural transmission. Evol Hum Behav 29:56–64
Fehr E, Fischbacher U (2003) The nature of human altruism. Nature 425:784–791
Gambetta D, Hamill H (2005) Streetwise: how taxi drivers establish their customers' trustworthiness. Russell Sage Foundation, New York
Gigerenzer G, Goldstein DG (1996) Reasoning the fast and frugal way: models of bounded rationality. Psychol Rev 103:650–669
Gintis H (2003) The hitchhiker's guide to altruism: gene-culture coevolution, and the internalization of norms. J Theor Biol 220:407–418
Gintis H, Bowles S, Boyd R, Fehr E (2003) Explaining altruistic behavior in humans. Evol Hum Behav 24:153–172
Gurven M, Zanolini A, Schniter E (2008) Culture sometimes matters: intracultural variation in pro-social behavior among Tsimane Amerindians. J Econ Behav Org 67:587–607
Guzmán R, Rodriguez-Sickert C, Rowthorn R (2007) When in Rome, do as the romans do: the coevolution of altruistic punishment, conformist learning, and cooperation. Evol Hum Behav 28:112–117
Hamilton W (1964) The genetical evolution of social behavior (I&II). J Theor Biol 7:1–52
Hayek A (1988/1991) The fatal conceit. University of Chicago Press, Chicago
Henrich J, Boyd R, Bowles S et al (2001) In search of Homo economicus: behavioral experiments in 15 small-scale societies. Am Econ Rev 91:73–78
Henrich J, Boyd R, Bowles S et al (2005) Economic man in cross-cultural perspective: behavioral experiments in 15 small-scale societies. Behav Brain Sci 28:795–855
Katz E, Lazarsfeld P (1955) Personal influence. Free Press, Glencoe, IL
Knudsen T (2003) Simon's selection theory: why docility evolves to breed successful altruism. J Econ Psychol 24:229–244
Lamba S, Mace R (2011) Demography and ecology drive variation in cooperation across human populations. Proc Natl Acad Sci USA 108:14426–14430
Lamba S, Mace R (2012) Reply to Henrich et al.: behavioral variation needs to be quantified at multiple levels. Proc Natl Acad Sci USA 109:E34
Lyons D, Damrosch D, Lin J et al (2011) The scope and limits of over imitation in the transmission of artifact culture. Proc R Soc Lond B366:1158–1167
Mackie D, Worth L, Asuncion A (1990) Processing of persuasive in-group messages. J Pers Soc Psychol 58:812–822
Mascaro O, Sperber D (2009) The moral, epistemic and mindreading components of children's vigilance towards deception. Cognition 112:367–380
McElreath R, Lubell M, Richerson P et al (2005) Applying evolutionary models to the laboratory study of social learning. Evol Hum Behav 26:483–508
McElreath R, Bell A, Efferson C et al (2008) Beyond existence and aiming outside the laboratory: estimating frequency-dependent and payoff-biased social learning strategies. Phil Trans R Soc Lond B 363:3515–3528
Merton RK (1968) Social theory and social structure. Free Press, Glencoe, IL
Milgram S, Bickman L, Berkowitz L (1969) Note on the drawing power of crowds of different size. J Pers Soc Psychol 13:79–82
Nettle D, Colleony A, Cockerill M (2011) Variation in cooperative behavior within a single city. PLoS One 6(10):e26922
Pasquini E, Corriveau H, Koenig M et al (2007) Preschoolers monitor the relative accuracy of informants. Dev Psychol 43:1216–1226
Petty RE, Wegener DT (1998) Attitude change: multiple roles for persuasion variables. In: Lindzey G, Fiske S, Gilbert D (eds) The handbook of social psychology. McGraw-Hill, Columbus, OH, pp 323–390
Petty RE, Cacciopo J, Goldman R (1981) Personal involvement as a determinant of argument-based persuasion. J Pers Soc Psychol 41:847–855
Rhine R, Severance L (1970) Ego-involvement, discrepancy, source-credibility and attitude change. J Pers Soc Psych 16:175–190
Richerson PJ, Boyd R (2005) Not by genes alone: how culture transformed human evolution. University of Chicago Press, Chicago
Rushton JP, Campbell AC (1977) Modelling, vicarious reinforcement and extraversion on blood donation in adults: immediate and long-term effects. Eur J Soc Psychol 7:297–306
Silk J (1980) Adoption and kinship in Oceania. Am Anthropol 82:799–820
Simon HA (1972) Theories of bounded rationality. In: McGuire J, Radner R (eds) Decisions and organizations. North-Holland, Amsterdam, pp 161–176
Simon HA (1990) A mechanism for social selection and successful altruism. Science 250:1665–1668
Simon HA (1993) Altruism and economics. Am Econ Rev 83:156–161
Sober E, Wilson DS (1999) Unto others: the evolution and psychology of unselfish behavior. Harvard University Press, Cambridge, MA
Trivers R (1971) The evolution of reciprocal altruism. Q Rev Biol 46:35–57
Winking J, Mizer N (2013) Natural-field dictator game shows no altruistic giving. Evol Hum Behav 34:288–293