

Russell Hardin

The Genetics of Cooperation

Abstract: Binmore analyzes the genetic basis of cooperation. Much of the literature doing this supposes that we must explain directly the cooperative tendency, whether by individual or group selection. A more effective way to go is to find something more general and likely more deeply embedded in personal traits that enables and even enhances cooperation. Hume, with whom Binmore claims affinities, long ago proposed a psychological phenomenon now called mirroring, which induces good relations through shared sentiments in a way that is essentially hard-wired. Mirroring indirectly contributes to cooperativeness. There may be other similarly indirect ways to account for human cooperativeness.

0. Introduction

Ken Binmore's *Natural Justice* is a book-length exegesis of a quip by J. B. S. Haldane in his response to the possibility that altruism is genetically based. Haldane said, Yes, I would lay down my life for two of my brothers or eight of my cousins. Why? He shares one-half of his genes with each brother and one-eighth with each cousin. To lay down his life for only one of his brothers would be to expend twice as many of his genes as he would save, and that would be evolutionarily wasteful. What has Binmore added to this quip? The big question here is how the genetic concern not to waste one's life while defending too few of one's genes is translated into actions in relevant contexts. Much of Binmore's text is given over to making and explaining this translation.

Consider an alternative, social, account that does not rely on the genetics of altruism. When the Israeli Phinehas zealously drove his spear through Zimri and his Midianitish woman in their tent (Numbers 25:6–8), he was plausibly acting on behalf of a tribal group whose life-or-death fate depended on strong internal loyalty. He seems even to have understood his own action this way. His kind of action may well have benefited the group of the Israelites. Supposing that he was at some risk when he did this, then Phinehas was altruistic to the Israelites. Out-group viciousness can be in-group altruism. Which was subject to the greater selection pressure? It seems extremely unlikely that actions such as that of Phinehas are genetically determined. They are more likely socially determined. If by chance some infants of the Israelis had been switched for some infants of the Midianites, those 'false' Israeli infants would have grown up to be culturally Israeli and their tribe-specific behavior would have been driven by their culture, not by their genes.

If I have a genetic attachment to certain kinds of food, there is no need for calculation to decide whether to eat such food when it becomes available. My genetically determined altruism seems far more likely to act only indirectly. The claim that we select genetically on our close relatives is partly analogous to the argument of David Gauthier (1986) that we are able to spot cooperators and therefore that we can choose to deal with those who are likely to bring benefits rather than losses to us. Just as it is hard to spot who is a cooperator, so too is it hard to know how we could recognize the level of closeness of our genetic relatives, or at least our siblings in an era before understanding of how pregnancy happens and an accounting of levels of relationship could be done. In that era, we generally might not have known who our father was, and in any case our genes would get little benefit from our protecting our parents because, by the time we might do that, they are likely to be past parenting anyone else with our genes.

If the genetic protection argument works, it must depend on some way of recognizing siblings, as might be possible in any species that has a nuclear familial structure. It would be that familial structure that might be genetically selected and that, once selected, might reproduce itself into future generations. Group selection probably makes great sense in *cultural or sociological selection*—as in the case of Phinehas. Such cultural selection seems likely to dominate over genetic selections at the group level. Indeed, we now know that many species of birds have eggs from more than one father in their nests, contrary to former beliefs that the birds were monogamous. It seems extremely unlikely that, for those species in which the fathers play a role in the lives of the chicks, the fathers have any genetic bias for their own chicks.

Clearly, the simplistic argument from shared genes cannot tell us much if it is too hard for individuals to act from their supposedly genetically selected urges. We need to know how large a social group tended to be in some earlier stage of evolution in order to know what was the impact on the genes of the next generation of any of my single actions. We might get further in such an analysis if we suppose that the survival of my small group (perhaps 50 to 200 members) depends on my helping in defense against saber tooth tigers, in capturing or killing mastodons, or in defense against other groups, perhaps especially of Neanderthals. This would be an argument from group selection.

What we need is some deeper genetic selection that leads us all to be cooperative in general, although even then our cooperativeness will be enhanced by sociological devices, such as norms and rules. If we can find a deeper, more general trait that is genetically selected, we can avoid the just-so stories implicit in Haldane's arithmetic. Most of the extant just-so stories on the genetics of intra-species altruism are about how a single individual risks harm by warning others of an impending danger. A favorite example in this game is warnings calls by birds. A bird sees or senses a fox sneaking up on the flock and it calls special attention to itself by calling urgently and loudly to scatter the flock to escape the fox. That bird supposedly is acting altruistically. Against this charming tale, we must note that the bird that shrieks a warning presumably knows where the fox is, which the other members of the flock do not. The others scatter in all

directions, including in the direction of the fox. The ‘warning’ bird knows to fly away from the fox. This form of the story makes warning calls a matter of pure self interest. One bird can secure its own safety by compromising the safety of the others, who are its genetic relatives.

We need a more general account because it seems unlikely that the genetic story that Haldane spoofs can be heavily and selectively at work in our own lives in conditions in which the bulk of our dealings—often cooperative and even altruistic—may be with people to whom we are not closely related. We may not develop a gene directly for cooperation with our nearest genetic relatives, but we may develop genetic capacities that indirectly enable us to cooperate.

1. Naturalism

I am generally sympathetic to the naturalist turn that Binmore takes. He does not focus on what are the ‘right’ principles of justice but on what principles we naturalistically develop. This was David Hume’s view. His account of morality is naturalistic in the sense that Hume explains why we have the morals we have (Hardin forthcoming, chapter 1). He does not present a normative but only an explanatory theory. There is, of course, not unanimous agreement on this claim, but it is easily defended by direct statements of Hume. Consider two among many claims from those who find Hume to be a naturalist and not a moralist. Hume “has taken a sheerly naturalistic view of morals” (Kemp Smith 1941, 563). “Hume’s ethics is a twofold psychological enquiry into human nature. These enquiries are not the same as those which philosophers undertake in what are now called Normative Ethics and Meta-Ethics.” (Penelhum 1975, 131)

It is of interest that Hume’s explanation is not genetic but social. Of course, genetic theory was not available to him. He thinks we can be generous and altruistic, but the strength of our altruism declines sharply as we deal with those increasingly far from us, as in Haldane’s quip. Hume roughly agrees with Haldane, although Hume’s claim is founded in social preferences rather than genetic survival. He says, “A man naturally loves his children better than his nephews, his nephews better than his cousins, his cousins better than strangers, where every thing else is equal.” (T3.2.1.18, SBN 483–4) The “every thing else” might be the intensity of the relationships and the time spent in them, which would be a social rather than genetic artifact. Hume is the Haldane of beneficence.

Hume does, however, arguably present a psychological theory of why we have strong sympathy for others, and that theory is about how our brain works the way it seems to work. From this view, one can speculate (this entire enterprise is based primarily on speculation) that some motor at the level of individual survival explains the cooperativeness of humans. There might be many candidates for such an explanation. I will propose one, not with a claim that it is the one, but only that it seems at least as likely to work as any current explanation from direct genetic selection of genes for cooperation or altruism. Hume labels the phenomenon he analyzes as ‘mirroring’. I mirror your emotions even to the point of sharing them to some extent.

2. Mirroring

We now know we are hard-wired with many mirror neurons that do remarkable things for us. If you smile at me, your smile pushes me to smile. This is not merely an act of acknowledgement. It is a matter of my brain mirroring what it sees. It is likely that we have survived into our present condition because our ancient forebears had relationships with parents and others that were positive and helpful to them. For example, many people have noticed the phenomenon of the reciprocal smile. As the Latin poet Horace says, “As men’s faces smile with those who smile, so they weep with those who weep.” And as we all too often experience, our faces yawn with those who yawn. Hume quotes Horace’s remark in his own account.¹

It also seems clear that Hume grasped the fact of this motor without having a clear account of why it works as it does (Hardin forthcoming, chapter 2). Sadly, virtually no one picked up on Hume’s insights and his proto-theory of mirroring was not developed. It was independently discovered by a German psychologist around 1903, and finally explored in depth only very recently, partly in response to the realization that chimps and other primates seem to mirror sensations of humans and partly because it is now possible to gain some entrée to the phenomenon with ‘functional magnetic resonance imaging’ (fMRI).

What might stand behind the phenomenon that Hume recognizes and uses to ground his claims for sympathy, but that he does not explain? There was of course no need for him to explain; he could observe the phenomenon and could start from it to explain various outcomes or results of the fact of mirroring (EPM5.17n19, SBN 219–220n). There are now fMRI studies of the brain’s reaction to others’ sensations that corroborate Horace’s and Hume’s observed facts and that, in a sense, seem to show the phenomenon of *seeing* another’s emotions at work.

The fMRI studies do not do much more than Hume already did—they establish that mirroring happens, although they are more definitive than Hume’s singular testimony. The part of the brain that perceives a smile is evidently the part that engineers a smile of our own, so that Horace’s observation may be a biologically hard-wired fact of our brains. Smiles evoke smiles. The evolution of this feature of our brains might be explained by the benefits of smiling in gaining the good graces of others, especially when we are too young to survive on our own.² Smiling may enable humans to enjoy very long periods of infancy, childhood, and adolescence so that we can develop extraordinary abilities that set us apart from other animals.

Hume further notes “we may remark, that the minds of men are mirrors to one another, not only because they reflect each other’s emotions, but also because those rays of passions, sentiments and opinions may be often reverberated, and

¹ Quoted in Latin in EPM5.18n1, SBN 220n.

² There are reputedly recent studies that suggest other connections. Those who yawn when another yawns seem to score higher on empathy tests than those who do not mirror the yawns of others (Henry Fountain, “Tarzan, Cheetah and the Contagious Yawn”, *New York Times*, 24 August 2004, F1).

may decay away by insensible degrees” (T2.2.5.21, SBN 365; see also T3.3.1.7, SBN 576 and EPM5.18, SBN 179) (see further, Penelhum 1993, 143). What in twentieth century philosophy was the problem of other minds (how can we know another’s mind?) is assumed away in limited part by Hume. For this too there may now be neurophysiological evidence from fMRI studies. Sympathy, these studies suggest, is a form of direct, non-verbal communication and the evocation of relevant feelings. Mirroring opens our black-box minds to others that they may feel what we feel.

It is psychological mirroring that leads me to like or dislike something that is done to you, by letting me sense what you enjoy or suffer. Contemporary neurophysiological findings seem to turn this trick and therefore seem to strengthen Hume’s claims for moral psychology, although the mechanism is not yet clear. Those readers who have had difficulty accepting this part of Hume’s argument might soon find it easy to accept. Hume appears to be right on the psychology here. The only question that might remain for some is that of his general claims for morality psychologized. Do we have moral reactions (approbation or disapprobation) to the feelings we get from mirroring? Those would be moral reactions on behalf of another. That is to say, the important and very difficult trick Hume needs to complete his explanatory theory is to evoke *my* sentiments—that is, a moral judgment—in response to actions that affect *your* interests.

From the fMRI data it appears possible that these two phenomena—sympathy and moral sentiments—are at least partially run together in our brains (see also, Árdal 1966, 47n.) Hence, Hume’s theory is complete but in a way that he apparently did not see. The knowledge and the feeling, the sympathy and the sentiments, may come in a single package. There is no mediating interpretation that our brains have to make. A nearly brand new baby smiles back at our smile. It is implausible to suppose that the baby is interpreting our kindness or good will in its first days of life; it is reacting from an apparently hard-wired capacity. Empathy seems to ‘mirror’ another person’s emotional responses in one’s own brain.³ Happily, “mirror” is Hume’s word and also the terminology of contemporary neurophysiological science (T2.2.5.21, SBN 365). Mimicry, which has long been noted and which is detailed by Darwin, has usually been explained as a two-step process. Our perceptions of, say, a smile stimulate thoughts, which guide our behavioral response: smiling back. Studies of brain activity, as measured by fMRI brain scans, suggest that the whole reaction is immediate in a single step, not mediated by thought. The part of my brain that recognizes a smile also forces or stimulates my own smile and my own feeling of pleasure.⁴ Seeing your smile triggers mine. We are to a degree hard-wired to each other.

At a very young age, Hume seems to have grasped the nature of this phenomenon to a sufficient degree as to make it the foundation of his moral psychology. He does not attempt an explanation of the phenomenon but merely starts from it to explain morality as a matter of fellow feeling. In fact, of course, he had no way to prove his assertion of the nature of this psychological trick other

³ The German psychologist, Theodore Lipps, coined the German term for empathy in 1903, and he described the phenomenon of mirroring; Bower 2003, 330.

⁴ See various contributions to Meltzoff and Prinz 2002. Also see Miller 2005.

than to elicit our agreement that we too have the experience he describes. The technology of fMRI now seems to give us some entrée to the phenomenon.

Incidentally, mirroring seems to be very weak in those with autism (Miller 2005, 947). Hence, David Owen is apparently right in saying that the renowned autistic woman, Temple Grandin, is lacking in Humean sympathy.⁵ Her lack is organic. Her reason functions very well so that she is able to see herself as like, in her words, an anthropologist on Mars, where she would have little in common with and no sympathy for others—this is the condition she is in *on earth*.

Sympathy in the fMRI studies appears to be a form of direct, non-verbal communication and the evocation of relevant feelings. In another context, Hume dismissively says of the possibility of an innate sense of rules of property, “We may as well expect to discover, in the body, new senses, which had before escaped the observation of all mankind.” (SBN EPM3.40) In actual fact, he may well have discovered, along with some others, including Horace and recent psychologists, what we might come to call a sense: the sense of sympathy. It is a sense that is much more acute in Hume than in most people but that is clearly evident in large numbers of people, including new born babies. It may be as hard-wired as the sense of taste or smell.

Hume says our sympathy for those on a ship sinking off shore will be greatly heightened if they are close enough for us to see their faces and their frightened responses. He does not explain this fully but only says that contiguity makes their suffering clearer to us (T3.3.2.5, SBN 594–5; also see T2.1.11.6 and 8, SBN 318 and 320). The fMRI studies suggest that the issue is not that we have to see their expressions in order to understand their emotions; our reason is adequate for such understanding. The issue is that *we have to see their expressions in order to trigger the mirroring of our own similar emotions*. This is a phenomenon that is not mediated by thought or reason, and perhaps it cannot be replaced by thought or reason when the actual visions are not available.

Suppose we accept this entire account of our moral sentiments and of their apparent mirroring. If they are merely a fact of our psychology, should they determine our morality? Yes, in Hume’s functional way. That is, our sentiments about others evoke responses from us that are responses to the utility, pleasure, or pain of those others. What typically brings pleasure to others is their own benefit, which is *good for them*. We cannot go further to say it is good per se unless we go so far as to say that something like utilitarianism is the right moral theory. Hume does not make this claim, but in his analysis of the motivating force of mirrored reactions he does imply that he and we are psychologically utilitarian. One of the things we can tell about another through mirroring is how something affects their welfare, pleasure, or pain. This fact is important if we are psychologically utilitarian—and mirroring virtually makes us be, as though evolution has produced utilitarianism as our moral response. Psychological utilitarianism connects observation to judgment. These facts do not make utilitarianism the true moral theory; they merely characterize our psychology as moralized through mirroring. This psychology gives us a science of moral

⁵ Owen 1994, 195. For more on Temple Grandin, see Sacks 1993/1994.

beliefs and approbations; it cannot additionally justify those approbations or make them right.

Mirroring is a major discovery for Hume despite the fact that seemingly all people experience it, so that it might well have been a matter of widespread common knowledge. It apparently remains unconscious and inarticulate to most people even while it often regulates their emotions and behavior. Hume is sufficiently perceptive that, once he has noticed the phenomenon, he finds mirroring to be a fundamental part of the psychology of sympathy and therefore a fundamental part of distinctively moral psychology. Mirroring makes Hume's theory psychologically richer than any of the then contemporary moral sense and sentiments theories, which are inherently psychological in their foundations. Their proponents are generally content to stop their inquiries at the point of asserting that we just do know right from wrong, that reason can determine these, or that god has given us such knowledge. Hume has empirically observed—and supposes we can all observe—the phenomenon of mirroring and sympathy.

There is a further trick that might still be difficult: evoking an emotional response of a similar kind in response to the interests of society. Identification with the interests *of society* must be very weak psychologically and it is grounded in reason more than in sympathy. The neurophysiological studies probably cannot address such an abstract phenomenon as responding psychologically to the interests of society. Seeing the interests of society forwarded or abused is not comparable to the visual cue of a smile or frown.

3. Back to Genes

Now let us turn full circle back to genetic evolution. Trying to give an account of the evolution of the phenomenon of mirroring should be much simpler than giving an account of Haldane's arithmetical altruism. It must generally be true that human babies are at grievous risk of not surviving unless adults sympathize enough to protect them through their earliest years. The length of the period of mothering and care-giving more generally is extreme for humans. The few cases of wild children that have been discovered and studied are woeful (see for example, Lane 1976). After some ethological stage, perhaps roughly at eleven years, a child no longer has the capacity to master language.

The development of the capacity for mirroring probably preceded language and, indeed, likely helped in the development of language. Other animals seem to mirror each other's emotions. In fact, the recent efflorescence of work on mirroring was sparked by a researcher's realization that a chimp in a cage was envying his lunch. The extensiveness of the capacity for and incidence of mirroring might well explain the human smile. There are specifically dedicated mirror neurons for the recognition of smiles.

4. Concluding Remarks

Part of the appeal of the account of cooperativeness as backed by the emotional ties communicated and even created by mirroring is that it does not need complicated arithmetical work on the part of the individuals to act on it. Mirroring is a hard-wired structure that apparently has a large payoff in cooperativeness but without making significant demands on the actors involved. It requires little knowledge about others with whom one interacts. The individual actors probably do not even realize what is happening in their own minds most of the time, but merely respond to others in positive ways. Were we more alert to what is happening, we would long ago have recognized the phenomenon without much difficulty, so that Horace and Hume would not seem so brilliant.

Another part of its appeal is that it does not require the specific genetic structure of the genetic accounts of Binmore and many others. Mirroring is a very generalized phenomenon that simultaneously benefits us and those with whom we interact as we experience it. This generality makes its development seem natural and easily explained, even though the chief benefits of mirroring are to us in our face-to-face interactions with small numbers of other people.

Mirroring also has further payoffs for us if it allows and even prompts us collectively to develop language, norms, and mutual understandings that all enhance our possibilities for cooperation. It makes us cooperative. Those who cannot mirror, such as highly autistic children and the highly skilled veterinarian Temple Grandin, lose in many ways. They may not suffer painful feelings of loss because they cannot empathize enough with others even to recognize what or even that they lose. They might be reduced to a kind of deductive assessment of the value to them of cooperating with others. In this, they may be more like Haldane's calculating kinsman than are most of us.

Oliver Sacks (1993/94, 112) says of Grandin's growing up that, even after she learned language, "Something was going on between the other kids, something swift, subtle, constantly changing—an exchange of meanings, a negotiation, a swiftness of understanding so remarkable that she wondered if they were all telepathic." The other children were mirroring and communicating—probably—below the level of their own awareness. They did not need to keep straight who was whose sibling (at one remove genetically) or who was a virtually complete stranger (at perhaps ten or more removes, for whom Haldane would have to lay down his life for a thousand others). Their capacity was generalized, not specific to genes shared with only certain others. If we want genetic explanations of social behavior, we must generally build our accounts on such generalized capacities that do not require precise differentiations of people in our hopes for cooperativeness.

Bibliography

- Árdal, P. (1966), *Passion and Value in Hume's Treatise*, Edinburgh
- Binmore, K. (2005), *Natural Justice*, Oxford-New York
- Bower, B. (2003), Repeat After Me: Imitation is the Sincerest Form of Perception, in: *Science News 163 (May 24)*, 330–32
- Gauthier, D. (1986), *Morals by Agreement*, Oxford
- Hardin, R. (forthcoming), *Hume's Moral and Political Philosophy*, Oxford
- Hume, D. ([1739-40] 2000), *A Treatise of Human Nature*, Oxford, ed. D. F. Norton and M. J. Norton (cited as T with book, chapter, section, and paragraph numbers in order)
- Kemp Smith, N. (1941), *The Philosophy of David Hume: A Critical Study of its Origins and Central Doctrines*, London
- Lane, H. (1976), *The Wild Boy of Aveyron*, Cambridge/MA; reprinted New York 1977
- Meltzoff, A. N./W. Prinz (eds.) (2002), *The Imitative Mind*, Cambridge
- Miller, G. (2005), Reflecting on Another's Mind, in: *Science 13, May*, 945–7
- Owen, D. (1994), Reason, Reflection, and *Reductios*, in: *Hume Studies 20.2*, 195–210
- Penelhum, T. (1975), *Hume*, New York
- (1993), Hume's Moral Psychology, in: D. F. Norton (ed.), *The Cambridge Companion to Hume*, Cambridge, 117–47
- Sacks, O. (1993/1994), An Anthropologist on Mars, in: *New Yorker (27 December/3 January)*, 106–125