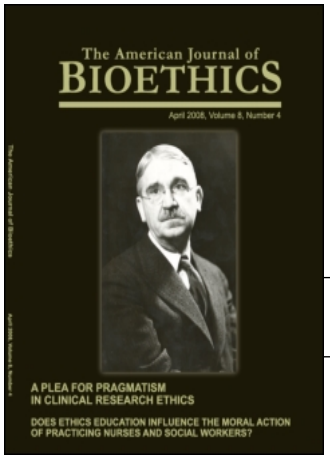

The American Journal of Bioethics
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713606739

Wired for Autonomy

Antonio Casado da Rocha [a]; Alvaro Moreno Bergareche [a]
[a] University of the Basque Country,

First Published on: 01 May 2008

To cite this Article: da Rocha, Antonio Casado and Bergareche, Alvaro Moreno (2008) 'Wired for Autonomy', The American Journal of Bioethics, 8:5, 23 — 25

To link to this article: DOI: 10.1080/15265160802180042
URL: http://dx.doi.org/10.1080/15265160802180042

particularly reciprocal altruism. The point I would like to stress is that *moral imagination*, even if this term is understood in neurophysical terms, would seem to require more than the individual distinguishing himself from others.

In the authors' account of the blurring process, the process ends with M deciding not to act if the consequences of the act are bad for M. This raises two questions. Does M makes this decision because he "identifies" himself with the original recipient, N? How does this relate to reciprocal altruism? If M "identifies" in some sense with the original recipient then there must be something about the abstract, intermediate image in which M recognizes himself; for, if this were not the case, it is difficult to understand why M would subsequently be lead to consider the consequences of the act to himself. In other words, either the blurring is not complete, or we need a "de-blurring" stage. The bigger problem, though, is that we do not seem to have made much progress to explaining why or how M would act reciprocally. One could claim that M "recognizes" that, on account of the blurring, the envisioned harm could also apply reciprocally to him, and hence he should not act in this way towards N. But this assumes exactly what we are trying to explain, namely, reciprocal altruism. M's identification with the potential harm to N and its resultant moral effect makes sense only assuming something like reciprocal altruism is already in place.

The authors claim that there account is "parsimonious" but one can imagine an equally or even more economical model that omits the blurring step. This article, like many others, conceives of moral behavior in terms of self-interest: the agent is motivated by self-interest and thus we need some explanation as to why and how the agent would behave in an altruistic fashion. But could we not also conceive of ourselves as essentially other-directed? Emotions such as compassion, sympathy, or empathy do not have a clear first-personal sense: a person cannot feel compassionate or empathetic towards herself; and although it is meaningful to say that a person can have sympathy for herself I would contend that this sentiment is not equivalent to the sympathy we feel towards others. If this is plausible, then perhaps some types of moral judgment and behavior do not require the person to envision her actions toward others in light of their actions towards her.

REFERENCES

Glannon, D. 2005. Neurobiology, neuroimaging, and free will. *Midwest Studies in Philosophy* 29: 68–82.

Greene, J. D., Nystrom, L. E., Engell, A. D., et al. 2004. The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44: 389–400.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., et al. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* 293: 2105–2108.

Pfaff, D., Kavaliers, M., and Choleris, E. 2008. Mechanisms underlying an ability to behave ethically. *American Journal of Bioethics—Neuroscience* 8(5): 10–19.

Tancredi, L. 2005. *Hardwired behavior: What neuroscience reveals about morality*. Cambridge, MA: Cambridge University Press.

Wired for Autonomy

Antonio Casado da Rocha, University of the Basque Country
Alvaro Moreno Bergareche, University of the Basque Country

The target article by Donald Pfaff and his colleagues (2008) postulates a mechanism operating in the central nervous system to produce what they call in their title *Ethical Behavior*, and which is consistently described in the body of the article as beneficial or *altruistic*. This emphasis on the *ethical* might be misleading, at least in an academic or philosophical setting, because ethics is generally considered more a domain than a specific content, no matter how altruistic. For example, Aristotle saw ethics as the domain of deliberation and choice; and, for Kant, the morality of an action is not a function of the agent's intentions, but a function of the exercise of their normative self-government or autonomy.

Being members of a research group focusing on the philosophy of biology, we are sympathetic to moderate naturalistic or empirical approaches to bioethics, in which facts and values are construed as a continuum rather than as a dichotomy. However, we do not base the "ability to behave ethically" on an inclination towards doing good (or avoiding harm) to others, but rather on *self-regulatory capacities for acting autonomously*, which for contemporary naturalism is a key aspect of life and cognition.

Bioethics needs a scientific explanation of the complex system in which autonomous deliberation and choice is rooted. It would be overly simplistic to oppose the position of the target article by claiming, as so-called humanists

Address correspondence to Antonio Casado da Rocha, University of the Basque Country, Filosofia de los Valores y Antropologia Social (UPV/EHU) Tolosa etorbidea, 70. 20018 San Sebastian, Spain. E-mail: antonio.casado@ehu.es

sometimes do, that mechanistic approaches are reductionist or determinist and therefore cannot explain the distinctive ability of humans to act for reasons of their choice. An adequate account of moral agency should move beyond the common conception of mechanisms as purely reactive systems, and understand our capacity to be free, autonomous, and responsible as arising "not in spite of but in virtue of the kind of mechanisms that constitute us" (Bechtel 2008, 240). Mechanisms need to be explored in order to explain ethical behavior and thus contribute to better decision-making; that is precisely the mission of neuroethics, defined as the examination of morality "informed by our understanding of underlying brain mechanisms" (Gazzaniga 2005, xiv–xv).

Forms of reciprocal altruism are ubiquitous amongst non-human primates, but they do not constitute examples of ethical behavior in the sense that they are not results of deliberation and choice. They might be part of what Frans de Waal has called the "building blocks" of morality—the capacity for empathy, a tendency towards reciprocity, a sense of fairness, and the ability to harmonize relationships—but ethical behavior *sensu stricto* emerges at a higher level—that of moral autonomy, which, of course, is based on biological and cognitive levels of autonomy (see Moreno et al. 2008, 313). Arguably, there is a strong difference between the kind of normativity of an animal for which awareness covers only a short present time span (a conscious scene built on its own history of value-dependent answers) and that of humans, who are able to foresee the long-term consequences of our actions. Human agency is intrinsically social, embedded in culturally elaborated norms and habits; the more complex the social structure, the greater the autonomy of the individual, which in turn prompts the need for further norms. The capacity for moral autonomy appears at this level of self-regulated agency.

As explained by Christine Korsgaard (2006, 113) in a commentary to de Waal, the capacity for moral autonomy can be empirically explained—there is nothing unnatural about it—but nevertheless requires a certain form of self-consciousness, which does not arise without specifically human abilities for language and a sense of self. Using her own words to restate the explanation offered by Pfaff and colleagues (2008), M may be conscious of the object of his fear (the negative consequences of his act for N) or desire (the positive consequences thereof), and M may be conscious of this object as fearful or desirable, and thus something to be avoided or sought. This is the ground "underlying" M's action, and as such it can be present both in human and non-human primates. But Korsgaard (2006) adds that, when acting morally, humans are conscious of something else: we are conscious *that* we fear or desire something, and *that* we are inclined to act in a certain way as a result. As she puts it, we are conscious of the ground *as a ground*.

If we view the example provided by Pfaff and colleagues (2008) in this way, when M refrains from knifing N *for moral reasons*, M does not just think about the object that he fears or even about its fearfulness, but rather about his fears and desires themselves as a ground for his action. The underlying mechanisms posited in the article do provide an explanation of our inclinations, but inclinations are the rawest material for ethics: it is not our inclinations, but rather what we do with them, that counts in ethics as a normative activity. One positive aspect of the kind of work carried out by Damasio, Pfaff (2008), and others, is that it suggests why it would be a mistake to follow the Kantian requirement to exclude emotion and feelings from the decision-making process. Another positive aspect is that it emphasizes the crucial role played by an individual's "social history" in their behavior. Habits are attempts to modulate our inclinations, emotions, and feelings, in order to transform them into virtues. We cannot choose our inclinations, but to a certain extent we choose our habits (if only consequently to be driven by them) by means of education and other socially enforced constraints on our behavior.

There is yet another sense in which the work by Pfaff and colleagues (2008) contributes positively to bioethics. Although the article is not intended to solve disputes among philosophers, it provides new insights to old debates concerning the role of shared fear as a basis for social behavior and of ignorance or loss of social information as a basis for fairness. Elucidating the biological mechanisms underlying prosocial behaviors and reciprocally beneficial responses may help us favor one moral theory over another, simply because one of them fits better with the best description provided by neuroscience of the mechanisms underlying human motivation.

To start with Thomas Hobbes—the most complex philosopher of fear, and one of the first to provide a mechanicist account of human behavior—it is worth remembering how in *Leviathan* he explains that equality between humans is based on the fact that "the weakest has strength enough to kill the strongest" ([1651] 1998, 82), thus making it impossible for anyone to live without fear. Shared fear is at the heart of his "fundamental law of nature" to seek peace. Hobbes is thinking in terms of the vulnerability caused by social dynamics, but his argument fits well with the mechanistic explanation offered by Pfaff and colleagues (2008) when they propose several ways in which the perceived difference between M and N can be reduced; this "blurring" of personal difference would make it easier for M to have a sense of shared fate with N, a sense of their basic equality when facing fear.

In contrast, the mechanism described by Pfaff and colleagues (2008) bears a striking resemblance to John Rawls' *veil of ignorance*, an heuristic device designed to reflect the kind of society in which we would be "treated equally as moral persons" (1999, 122). In Rawls' theory, loss of information is used as a guarantee of fairness, serving to push rational agents toward principles of justice that would be reasonably accepted by all and represent "a genuine reconciliation of interests" (1999, 122). Empirical approaches to ethics often stress the central role of sympathy in the theories offered by David Hume, and Adam Smith, but dismiss Hobbes ([1651] 1998) and Rawls (1999) because their

contractualism creates an illusion of human society as the result of a voluntary arrangement between free and equal agents (after all, as primatologists know, there was never a single point at which we became social). However, the fact that there is a biological basis to the idea that the loss of social information plays a role in treating others as we want to be treated ourselves somehow makes Rawls use of his device more convincing.

The fact that it is physiologically feasible to explain reciprocal altruism makes it easier to morally demand it: after all, *ought* implies *can*. But the fact that one *is* wired for reciprocity does not imply that one *ought* to act reciprocally (Racine 2008, 98)—after all, behavior following the Golden Rule can reasonably be demanded, but always depending on the interaction between ethical reasoning and context, as described in the Rawlsian method of reflective equilibrium applied by Beauchamp and Childress (2001, 397–401) to the field of bioethics.

In conclusion, the importance of this kind of research is not to be belittled, but it should be made clear from the start that neuroscience "will never find the brain correlate of responsibility" (Gazzaniga 2005, 101). We might be "wired for reciprocity", as Pfaff and colleagues (2008) argue, but we are also biologically and socially *wired for autonomy*.

REFERENCES

Beauchamp, T., and Childress, J. 2001. *Principles of biomedical ethics*, 5th ed. New York, NY: Oxford University Press.

Bechtel, W. 2008. *Mental mechanisms. Philosophical perspectives on cognitive neuroscience*. New York, NY: Routledge.

Gazzaniga, M. S. 2005. *The ethical brain.* New York, NY: The Dana Foundation.

Hobbes, T. [1651] 1998. *Leviathan*. Oxford, UK: Oxford University Press.

Korsgaard, C. M. 2006. Morality and the distinctiveness of human action. In *Primates and philosophers*, ed. F. B. M. de Waal. Princeton, NJ: Princeton University Press, 98–119.

Moreno, A., Etxeberria, A., and Umerez, J. 2008. The autonomy of biological individuals and artificial models. *BioSystems* 91(2): 309–319.

Pfaff, D., Kavaliers, M., and Choleris, E. 2008. Mechanisms underlying an ability to behave ethically. *American Journal of Bioethics—Neuroscience* 8(5): 10–19.

Racine, E. 2008. Which naturalism for bioethics? A defense of moderate (pragmatic) naturalism. *Bioethics* 22(2): 92–100.

Rawls, J. 1999. *A theory of justice*. Oxford, UK: Oxford University Press.

Ethical Decision-Making as Enlightened Behavior

Amaris Keiser, Mount Sinai School of Medicine
Eric Gehrie, Mount Sinai School of Medicine

Pfaff and colleagues (2008) argue that "blurring of identity" (10) is "the crucial step" (10) that must occur in order for an individual to act in accordance with the "Golden Rule" (10). They contend that "instead of seeing the consequences of his act solely for the other individual" (10), the ethical person "*loses the difference*" (original emphasis, 10) between himself and the other individual who is affected by his actions. If the ethical person finds that his proposed action is good for both himself and the other individual, then he does it; if the proposed action would be hurtful, he abstains.

This theory has a certain Rawlsian quality to it, in that it suggests that ethical or just decisions are best made in a context where decision-makers are blind to certain personal details. Through the "blurring of identities", the authors create a construct in which an individual involuntarily feels for himself the consequences of his actions for another. This motivates the individual to make decisions that are good for both himself and the recipient of his actions, and thus ethical behaviors are encouraged. Similarly, by virtue of the "veil of ignorance," individuals in the state of nature do not know any details about themselves, and are thus motivated to make choices that are fair and equitable for all. However, whereas Rawls considered the original position and the veil of ignorance to be hypothetical considerations meant to aid individuals in fulfilling the task of "define[ing] the principles of justice" (Rawls 1971, 19), Pfaff and colleagues (2008) suggest that humans instinctively create a practical, real-world veil of ignorance, resulting in the creation of a mechanism for making ethical decisions. The authors further argue that the veil is created through a temporary loss of information at the time of the decision-making. Although we do understand the appeal of this theory, we find it problematic due to residual ambiguity regarding the precise mechanism employed, as well as the fact that it suggests the existence of a set of human behaviors that are not typically observed.

Address correspondence to Amaris Keiser, Mount Sinai School of Medicine, New York, NY 10029. E-mail: amaris.keiser@mssm.edu