



# Dupoux and Jacob's moral instincts: throwing out the baby, the bathwater and the bathtub

Susan Dwyer<sup>1</sup> and Marc D. Hauser<sup>2</sup>

<sup>1</sup> Department of Philosophy, UMBC, 1000 Hilltop Cir., Baltimore, MD 21250, USA

<sup>2</sup> Departments of Psychology, Biological Anthropology, and Organismic & Evolutionary Biology, Harvard University, 33 Kirkland St., Cambridge, MA 02138, USA

We agree with Dupoux and Jacob (DJ) [1] that human moral capacities are, in ways yet to be understood, grounded in our biology. However, we disagree with DJ's outright rejection of the linguistic analogy (LA). Fully cognizant that our present understanding of the moral faculty is no better, and perhaps is worse, than the state of affairs in linguistics circa 1950, we believe that it has opened up new and exciting research questions for scientists and moral philosophers alike, with empirical work emerging apace [2–4].

DJ claim to acknowledge that LA 'usefully' organizes inquiry into morality around the five central questions familiar from the study of language [5] and concede that recent studies of moral judgment [2,5–7] lend *prima facie* support to the value of LA. Yet they insist that the plausibility of LA depends on whether 'the mechanisms underlying moral judgments make use of moral information encoded in a dedicated moral grammar (374)'. DJ's attempted immolation of this straw person reveals not only their misunderstanding of the hypothesized moral faculty (MF), but also of the language faculty (LF).

Here, we summarize some misconceptions:

- LF is not agreed to be, and perhaps is not, a module in the Fodorian sense [8].
- It is plainly question begging to assert, 'metacognitive processes of justification operating on explicit moral beliefs are fully part of the moral faculty (377)'. This claim is tantamount to saying that rhetoric is a proper object of scientific research. Hence, LA is not undermined, even if MF is not a module in the Fodorian sense.
- There is evidence that humans parse the world in ways strongly suggestive of a grammar of action [3,5,6]. In addition, just as the meanings of complex expressions can be fully determined by the meanings of their parts and the way in which they get put together without all those parts being fully recoverable through simple reversibility, failure of full reversibility from moral valences to structural descriptions of actions is not evidence for the absence of compositionality.
- The fact that part of our moral psychology depends on explicit beliefs is not counterevidence against an intuitive, unconscious component, viz. a moral grammar. Part of our language psychology also depends on

explicit beliefs about what we will say in response to criticism or praise in certain contexts. It is the abstract structure of these statements, as opposed to their content, that carries the signature of the language faculty; similarly, moral judgments might also carry the signature of the moral faculty. We simply have to explore this possibility, not reject it.

Two further points need emphasis. First, advocates of LA, including Hauser [6], do not discount the relevance of emotion to moral judgment, and they certainly do not dole out questions about emotion to the domain of moral performance to make moral judgment seem more like linguistic judgment, as DJ insinuate. Rather, once intuitive moral judgment becomes an object of serious empirical inquiry, the causal necessity of emotion in the generation of moral judgment *tout court* cannot be assumed. What is necessary is to ask when and what kinds of emotion have a role, and what other mechanisms might precede the emotions, having a crucial role in triggering them. Recent studies of patients with ventromedial prefrontal cortex damage show that they have flattened social emotions [9,10] but also show that some moral judgments are nonetheless completely preserved, whereas others are not [11]. This work, motivated by LA, enables a much more careful dissection of the moral sphere, pushing on the computations that precede emotions, and also on aspects of moral decision making that can be computed either in the absence of emotional input or before it.

Second, the principles and parameters (PandP) approach to accounting for linguistic diversity is one of several theories of how a universal competence is variably manifest. Hauser [6] and Dwyer [12] invoked it because it provides a heuristically useful way to think about moral diversity. It is an empirical question how much moral diversity exists, and the identification of substantive differences that require explanation will depend on what the units of comparison are (e.g. individual intuitive judgments, reflective judgments, individual behaviors, cultural norms or cultural practices) and, in particular, on the development of a uniform method to elicit moral judgments cross-culturally. The one study that DJ cite to undermine PandP as a model for accounting for moral diversity – by Shweder *et al.*, on the comparative work on sleeping arrangements in a large city in Illinois, USA, and a small rural town in India [13] – is ill chosen because of flawed methods: the sample size was too small and the study could not distinguish between the relative contributions

Corresponding author: Dwyer, S. (dwyer@umbc.edu).  
Available online xxxxxx.

to subjects' moral judgments of cultural differences, on the one hand, and differences due to living in a big city versus a rural community, on the other. More importantly, incest avoidance was identified in both subject groups as a crucial feature in judging the appropriateness of sleeping arrangements. That other differing features are deemed relevant by Indians (female chastity) and North Americans (parents' privacy) does not show that there cannot be something akin to parametric variation in general moral principles.

The primary aim of LA is to account for the descriptive principles that underpin our moral judgments; equally important are the developmental processes that enable this capacity. No advocate of LA claims that morality is, in every respect, similar to language, a contention that is clearly incoherent, given their transparent functional differences. Rather, LA takes advantage of the questions raised in the Chomskyan tradition of linguistics to ask similar questions about morality and, thus, to understand how this domain of knowledge works. It is utterly surprising that basic questions of domain-specific mechanisms, poverty of the stimulus, crucial periods of acquisition, selective neural breakdown and a competence–performance distinction are not only unanswered, but virtually never asked. We hope that others interested in the nature of our moral psychology will pursue these issues, recognize that they in no way compromise research on the emotions, or on the actions we take and the moral justifications we offer for them. Let the baby have her bath!

## References

- 1 Dupoux, E. and Jacob, P. (2007) Universal moral grammar: a critical appraisal. *Trends Cogn. Sci.* 11, 373–378
- 2 Hauser, M. *et al.* (2007) A dissociation between moral judgment and justification. *Mind Lang.* 22, 1–21
- 3 Cushman, F. *et al.* (2006) The role of conscious reasoning and intuition in moral judgments: testing three principles of harm. *Psychol. Sci.* 17, 1082–1089
- 4 Young, L. *et al.* (2006) Does emotion mediate the relationship between an action's moral status and its intentional status? Neuropsychological evidence. *J. Cogn. Cult.* 6, 291–304
- 5 Mikhail, J. (2007) Universal moral grammar: theory, evidence and the future. *Trends Cogn. Sci.* 11, 143–152
- 6 Hauser, M.D. (2006) *Moral Minds: How Nature Designed our Universal Sense of Right and Wrong*, Ecco Press
- 7 Haidt, J. (2001) The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychol. Rev.* 108, 814–834
- 8 Collins, J. (2004) Faculty disputes. *Mind Lang.* 19, 503–533
- 9 Damasio, A.R. *et al.* (1990) Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behav. Brain Res.* 41, 81–94
- 10 Beer, J.S. *et al.* (2003) The regulatory function of self-conscious emotion: insights from patients with orbitofrontal damage. *J. Pers. Soc. Psychol.* 85, 594–604
- 11 Koenigs, M. *et al.* (2007) Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature* 446, 908–911
- 12 Dwyer, S. (2006) How good is the linguistic analogy? In *The Innate Mind: Culture and Cognition* (Carruthers, P. *et al.*, eds), pp. 237–256, Oxford University Press
- 13 Shweder, R.A. *et al.* (1995) Who sleeps by whom revisited: a method for extracting the moral goods implicit in practice. *New Dir. Child Dev.* 67, 21–39