

The Baltic International Yearbook of
Cognition, Logic and Communication

October 2012
pages 1-19

Volume 7: *Morality and the Cognitive Sciences*
DOI: 10.4148/biyclc.v7i0.1778

FRANCESCO ORSI
Tartu University

MORAL JUDGMENT, SENSITIVITY TO REASONS, AND THE MULTI-SYSTEM VIEW

ABSTRACT: In this paper I attempt a critical examination of the multi-system or dual-process view of moral judgment. This view aims to provide a psychological explanation of moral sensitivity, and in particular an explanation of conflicting moral sensitivities in dilemma cases such as the crying baby scenario. I argue that proponents of the multi-system view owe us a satisfactory account of the mechanisms underlying “consequentialist” responses to such scenarios. For one thing, the “cognitive” processes involved in consequentialist reasoning only seem to play a subserving role with respect to the final judgment (providing non-moral inputs to judgment, or exerting additional strength to override the immediate “deontological” response). In this sense, Greene and colleagues fail to identify a peculiar system of moral judgment specularly opposed to the affective “deontological” one. For another, Greene and colleagues’ work on the emotion-cognition dichotomy and the distinction between alarm-bell and currency emotions, though promising, still falls short of providing an adequate and consistent picture of the psychological mechanisms underlying “cognitive” evaluations and verdicts in dilemma scenarios. It is suggested that alongside further experimental work, proponents of this view should pay more attention to the conceptual underpinnings of their distinctions.

1. INTRODUCTION

In this paper I aim to critically examine the contribution that a certain psychological account of moral judgment, the multi-system view, can offer to the understanding of moral judgment. In the first part I introduce the notion of sensitivity to reasons, and briefly argue that it is a central component of our conception of moral judgment, a component that any theory of moral judgment should explain. In the second part, I introduce the multi-system view as developed in particular by F. Cushman and J. Greene in a number of papers. I will argue that such a view suffers from a problem of functional asymmetry among systems.

2. MORAL JUDGMENT AND SENSITIVITY TO REASONS

Sensitivity to moral reasons is an ineliminable component of moral judgment. In order to decide the moral status of an action, we need to ascertain the facts of the matter, and decide if they are morally relevant, that is, if they provide moral reasons for or against an action. This is because the moral status of an action depends on the moral reasons for or against the action. When the balance of moral reasons is against performing an action, then the action is morally wrong, or ought not to be done. When the balance of moral reasons favours performing an action, then the action is morally good and possibly obligatory. In sum, in order to judge an action, we need to be sensitive to the moral reasons. Sensitivity to reasons is thus the capacity to recognize normative facts, that is, to recognize certain facts as moral reasons for or against an action. The exercise of such capacity may go wrong in many ways of course: e.g. I may wrongly think that the fact that somebody has insulted me is a good moral reason to cut their throat. Or sometimes we may fail to exercise this sensitivity when we should: e.g. knowing well that she is in pain, I may fail to recognize my victim’s pain as a reason to stop torturing her.

Sensitivity to reasons is present both in moral action and in moral judgment in actual or hypothetical situations. Consider the crying baby case, much discussed in the psychological literature (and which I will return to below):

Enemy soldiers have taken over your village. They have

orders to kill all remaining civilians. You and some of your townspeople have sought refuge in the cellar of a large house. Outside, you hear the voices of soldiers who have come to search the house for valuables. Your baby begins to cry loudly. You cover his mouth to block the sound. If you remove your hand from his mouth, his crying will summon the attention of the soldiers who will kill you, your child, and the others hiding out in the cellar. To save yourself and the others, you must smother your child to death. Is it appropriate for you to smother your child in order to save yourself and the other townspeople?

Here, in order to decide what we ought to do, we must identify the reasons for or against alternative actions, weigh them, and decide which action is favoured by the balance of reasons. Whatever we conclude (it is permissible/obligatory/wrong to smother our child), we will have displayed our sensitivity to the reasons in the situation. Minimally, we will have recognized both the fact that we would be killing our child as a reason not to smother her, and the fact that if we let her cry, everybody will be killed as a reason to smother the poor child.

Sensitivity to reasons is a central component of moral judgment, for a subject who makes a moral judgment must be disposed to justify it with moral reasons. It is part of what makes a moral judgment something else than a mere subjective preference that we are ready to give reasons for it. If we think that an action is wrong, we will think that something about the action (or about the agent, or about the consequences) makes it wrong, and for this reason we judge it to be wrong.

J. Haidt's research may seem to pose two challenges to the importance of sensitivity to reasons. One is the idea that 'justifying a judgment with reasons' is often just the result of post hoc rationalizations (Haidt 2001). The suggestion here is that in expressing moral judgments we voice primarily our preferences or prejudices, and only later, if pressured, do we come up with some rationally acceptable justification.

However, nobody denies that reasons for judgments might not consciously or explicitly guide our judgments. Instead, the very pressure to say something in defence of our judgment shows that we know we would not be making a moral judgment otherwise, but only voicing a

preference or prejudice of ours. If we want our judgment to be a moral judgment, there is nothing wrong with trying to find our reasons post hoc. In fact, engaging in the more or less conscious task of looking for our reasons, regardless of success, is enough to show that our judgment is a moral one.

The second apparent challenge is the persistence of certain judgments despite a recognized inability to give reasons for them. Haidt's example is the widespread judgment that an incestuous act between brother and sister would be wrong, despite the absence of any physical or emotional harm or other negative consequences for all involved (Haidt 2001).

Two remarks are in place. First, people who stick to their judgments though unable to justify them might nonetheless still be disposed to provide a reason, if they could find it. That is, they can acknowledge that a reason is needed—even when all they are able to come up with is the irrelevant self-report that 'it just feels wrong'. Second, Haidt shows little charity to an obvious alternative: sexual intercourse between close relatives may be perceived as morally wrong on its own right, and not for further reasons. Those who stick to their judgment may perceive incest as a morally wrong type of act (like killing the innocent), thus needing no *further* reasons to judge an instance of it as wrong.

Therefore, it seems quite plausible that an essential feature of moral judgments is the subject's disposition to give reasons for them, thus exercising their sensitivity to (moral) reasons. In the next sections I argue that a prominent psychological approach to moral judgment, the multi-system or dual-process view, cannot account for this aspect of moral judgment.

3. THE MULTI-SYSTEM VIEW

A number of authors, such as J. Greene and F. Cushman, have proposed that the best explanation of certain patterns of moral judgment in response to hypothetical scenarios is that two, or more, dissociable psychological mechanisms are at work when engaging in moral decisions. I concentrate here on the multi-system view as the best explanation for differences in judgments concerning the crying baby scenario

illustrated above, leaving aside its application to the well known trolley problem cases (Greene et al. 2001).

When presented with the crying baby scenario, people divide over what they ought to do (Greene et al. 2004). Some judge it permissible to smother the baby, some judge it impermissible. Where does the difference lie? Data from fMRI has been thought to help provide an answer. Some patterns of brain activation and reaction times are common to both categories of subject:¹

- 1) In both categories of subjects, brain regions associated with emotion and social cognition show increased activation. Those areas (the posterior cingulate cortex, the medial prefrontal cortex, and the amygdala) are also active in other cases of potential personal harm, like the “footbridge” case, or an infanticide case where a mother simply kills her newborn baby for no apparent reason;
- 2) Both categories of subjects have high reaction times, compared to judgments on cases (like the infanticide just mentioned) where the judgment is relatively quick. The high reaction times evidence that both categories of subject are engaged in “abstract reasoning” of the kind that is required by a cost-benefit analysis. Both subjects seem to be weighing the costs and benefits of either outcome, irrespective of their eventual judgment (Greene et al. 2004, p. 396);
- 3) Both categories of subjects show increased activity in brain regions (anterior cingulate cortex) associated with the detection of cognitive conflict.

It is natural to interpret this cognitive conflict as arising from an opposition between the neuropsychological processes alluded to in (1) and (2). Presumably, the emotional reaction in (1) is negative—prohibiting smothering the baby—whereas the result of the cost-benefit analysis is positive—favouring smothering the baby. The crucial difference between the two categories of subjects seems to lie in the differential activation of the dorsolateral prefrontal cortex and the inferior parietal cortex. Subjects judging that smothering the baby is permissible show higher activation of these areas than subjects judging that it is not permissible. Now, these brain regions are associated with a kind of function which Greene et al. (2004) term “cognitive” (or “regulative”)

“control” (presumably following Miller & Cohen 2001). In (Greene et al. 2004, p. 396), these regions are said to play an important role in the regulation of potentially counterproductive emotions in many contexts: social decision-making, placebo effects, the evaluation of trade-offs between future and immediate rewards. In (Greene 2007, p.46) the associated functions are the following: executive control, complex planning, deductive and inductive reasoning, taking the long view in economic decision making. Finally, Cushman & Greene (in press) also mention “thinking guided by explicit rules” (p. 2).

These lists are clearly a mixed bag. However, in light of previous evidence that such “cognitive” regions are associated with utilitarian kinds of judgments in cases such as the “switch” version of the trolley (Greene et al. 2001), Greene and colleagues suggest that the function of the increased activation of these regions in subjects judging it permissible to smother the baby is to respond to the cognitive conflict by overriding the strong, negative emotional reaction to the prospect of smothering one’s baby, and allowing the evaluation based on the cost-benefit analysis (very roughly: many lives are worth more than one, or an outcome where one ends up killed is better than an outcome where everybody ends up killed) to decide the conflict. By contrast, in subjects judging it impermissible to smother the baby, the “cognitive control” mechanisms are not strong enough to override the negative emotional reaction.

Based on these and further neuroscientific data,² Greene and colleagues argue that different systems are in play in these moral judgments (Greene 2007; Cushman & Young 2009; Cushman et al. 2010; Cushman & Greene in press). When a situation presents the possibility of personally harming someone to save others, an emotional or affective system reacts to the input with a strong negative response. At the same time, abstract reasoning systems underlie the realization that by harming one we can save others—some form of cost-benefit analysis is performed here. The perceived conflict between the alternatives is the result of the clash of these systems. Eventually, one’s final judgment will depend on the work of mechanisms of “cognitive control”. Importantly, since both the affective and the ‘cold’ systems are demanding our attention, “no matter what you do, part of you is going to be dissatisfied” with the eventual moral decision (Cushman &

Young 2009, p. 19). As Cushman and Young put it, “the moral mind is a constellation of distinct cognitive processes that can operate independently, often interact, and sometimes compete” (Cushman & Young 2009, p. 11).

Now, in proposing the competition between these mechanisms as explaining the psychology of moral judgment in conflict cases, Cushman and Young encourage the idea that a multi-system view has all the features needed to explain the sensitivity to reasons of moral judgment:

First, the mechanisms can each be characterized by distinct axioms. Second, the mechanisms produce opposing judgments in an identical category, such as a normative demand on action or a judgment of responsibility. Third, the demands on and judgments of behavior produced by each mechanism are inherently non-negotiable—these do not represent mere votes in favor of a particular conclusion but rather require a particular conclusion. (Cushman & Young 2009, p. 17, see also 15–6, and Cushman & Greene in press).

The authors credit the mechanisms with a rich conceptual articulation that mirrors the articulation of an exercise of sensitivity to moral reasons.³ The systems process the factual inputs—representations of possible actions, harms, alternative outcomes—by causally responding to them in an ‘affective’ or a ‘cold’ fashion, which includes the possibility of a conflict, understood as a conflict of purely psychological impulses.

According to Cushman and Young, each system is better described as assigning an evaluation to the inputs in accordance with its own axioms, and generating its own final moral verdict (a non-negotiable demand). In the crying baby case, the affective system applies the axiom “It is prohibited to do harm to an individual as a means to an end”—or a more specific one—and generates the verdict that it would be absolutely wrong to smother the baby. In other words, the affective system takes the prospect of killing the baby as a means to an end as a sufficient and decisive reason not to smother the baby. On the other hand, the ‘cold’ system, recognizing that it would be worse to let all be killed than to kill one, applies some consequentialist axiom such as “It is required to minimize overall harm”, and demands that

one’s child be sacrificed. The ‘cold’ system thus displays a sensitivity to other facts of the scenario as decisive reasons to smother the baby (namely the fact that everyone would be killed otherwise). Moreover, the ensuing dilemma is represented as a full-blown normative conflict between contradictory verdicts rather than merely a battle of opposing impulses.

4. CRITICISM: FUNCTIONAL ASYMMETRY

I present here a criticism of the multi-system view understood as an attempt to capture the psychological structure of sensitivity to reasons in moral judgment. This is not meant to be a decisive objection to any multi-system account, but rather as a difficulty that a more careful analysis and formulation of the view might in the future overcome.

I call the problem a “functional asymmetry” between mechanisms. The structure of the objection is the following. Systems are functionally symmetrical when they “do the same kind of job”. In this case, the job is to be sensitive to moral reasons: to process *independently given* inputs (representations of the facts of the matter) and yield a moral verdict based on them. If the systems are not functionally symmetrical in this sense, then they cannot be opposed to one another in such a way as to give rise to a moral conflict, i.e. a perceived conflict between moral reasons. If they cannot be opposed in such a way, then they do not represent competing sensitivities to moral reasons. But, as I will argue, the systems have not been shown to be functionally symmetrical. Therefore they do not represent competing sensitivities to moral reasons.

Exactly *which* are the systems that are supposed to compete? Greene and colleagues unfortunately give less than a satisfactory or univocal answer. On the one hand, it seems that there is an affective system which responds, with a negative reaction, to an independently given input, namely representations such as the prospect of the child’s death, or more likely the physical action of smothering one’s own child. The affective system thus operates upon inputs that have already been “taken in” thanks to the work of other cognitive mechanisms (presumably perspective-taking, and perhaps simulated motor planning—see Cushman & Greene in press). This makes it a candidate apparatus

for sensitivity to reasons: sensitivity to reasons works upon facts which have already been cognitively digested by the subject.

However it is far from clear what in the opposing ‘cold’ system does a kind of job comparable to the affective reaction. When describing this separate system, Greene and colleagues mention increased activation in brain areas associated with basically two cognitive functions or processes: abstract reasoning, and all that goes under “cognitive control”. Let’s start with the latter. As pointed out, the “job” of cognitive control is to override the emotional response and let the ‘cold’ verdict win out. Therefore such a function could only logically and psychologically *follow* an already reached ‘cold’ verdict (kill the one to save the many). Cognitive control is part of the ‘cold’ system only to the extent that it is “recruited” by such a system to contrast the negative emotional reaction once the verdict has already been represented by some *other* mechanism. Even though the “utilitarian” verdict could not guide our judgment without a sufficiently strong cognitive control, the exercise of cognitive control is not the *locus* of the “utilitarian” verdict in the way that emotional processes are (or at least seem to be) both the locus of the “deontological” verdict *and* the driving force behind the eventual deontological judgment.

The remaining candidate process is “the abstract reasoning that constitutes a utilitarian analysis” (Greene et al. 2004, p. 396). In the specific crying baby case, the output of such process is the “abstract, ‘cognitive’ understanding that, in terms of lives saved/lost, one has nothing to lose (relative to the alternative) and much to gain by carrying out this horrific act” (ibid: 390). When pitted against the negative emotional reaction, in some people “this ‘cognitive,’ cost-benefit analysis wins out, and these people say ‘yes’” (Greene 2007, p. 46). Engaging in a cost-benefit analysis must surely precede the operation of cognitive control, and is somehow relevant to arriving at the judgment that it is permissible to smother the baby. However, it is not clear exactly *how* it is relevant. There seem to be two possibilities: (1) abstract reasoning as a calculation of the alternatives; (2) abstract reasoning as an evaluation of the alternatives.

If abstract reasoning just is the performance of calculation of the alternative outcomes, relative to lives saved/lost (smothering the baby: one life lost, many lives saved; not smothering the baby: no lives

saved), then abstract reasoning is engaged in *generating* some relevant inputs to the moral judgment, rather than in responding to independently given inputs. Abstract reasoning thus cannot be a mechanism of moral sensitivity symmetrically opposed to the affective one, but performs an evaluatively neutral preliminary function akin to that played by, say, simulated motor planning when we represent the prospect of physically smothering the child, while leaving it to some other process to evaluate the alternative outcomes and prescribe a certain course of action.

On the other hand, abstract reasoning could have a jointly calculative *and* evaluative/prescriptive function, generating both the relevant inputs in terms of lives saved/lost, and the *evaluative* output that smothering the baby is the *best* outcome in terms of lives saved—plus the prescriptive demand that this is what we morally *ought* to do. Such a mechanism would thus be symmetrically parallel to the affective process, issuing the opposed and conflicting verdict.

This idea is attractive and sensible—indeed, it seems to be the only sensible way to respond to the functional asymmetry objection as we set it up—but a few problems remain to be solved. The first one stems from the experiment design itself, and seems to have gone unnoticed so far. It is not obvious that the “abstract reasoning” process yields a *bona fide moral* evaluation. As it is described, in the crying baby case the agent’s life is as much at stake as the others’, and the research subjects are asked to identify with the agent. The cost to oneself of not smothering the baby is nothing less than being killed. Smothering the baby thus crucially means saving oneself. So, it might be that in issuing the judgment that it is appropriate or permissible or even obligatory to smother the baby to save the most people, the subjects are actually expressing the result of a prudential kind of cost-benefit analysis, which in this case happens to coincide with the consequentialist verdict. In the absence of further control experiments, it cannot be ruled out that the cognitive system might, in this case, simply be guided by a prudential bias towards saving one’s life rather than issue an impartial moral judgment based on some consequentialist principle. If the former is the case, then the abstract reasoning process is not producing a symmetrically opposed verdict “in an identical category”, as Cushman & Young require (2009, p. 17).

A second problem can be appreciated by looking at the details and the more speculative parts of Greene's theory. This requires a separate treatment.

5. FROM EMOTION VS. COGNITION TO ALARM-BELL VS. CURRENCY EMOTIONS

Which mechanisms realize the evaluative and prescriptive function played by "utilitarian" abstract reasoning? The answer, we know, is "cognitive" mechanisms as opposed to affective or emotional mechanisms. We also know from the previous section that "cognitive" mechanisms involved in (1) cognitive control and (2) mere calculation play at most a supporting or subservient function with respect to the "utilitarian" judgment. What we are looking for is therefore another sort of "cognition". Greene and colleagues' work offers at this point two possible directions to address the question: i) an elaboration of the dichotomy emotion vs. cognition, ii) a distinction between two different kinds of emotions related to "deontological" and "consequentialist" moral judgment.

Greene has suggested that the emotion/cognition dichotomy, which seems to underlie the deontology/consequentialism divide, is better understood as a difference in degree rather than in kind. In particular, the distinction is in terms of the motivational impact of different representations:

[O]ne might render the emotion/cognition distinction in terms of a contrast between, on the one hand, representations that have direct motivational force and, on the other hand, representations that have no direct motivational force of their own, but that can be contingently connected to affective/emotional states that do have such force, thus producing behavior that is both flexible and goal directed (Greene et al. 2004, p. 397–8).

In his 2007 paper, Greene elaborates a bit further:

The rough idea is that "cognitive" representations are inherently neutral representations, ones that do not automatically trigger particular behavioral responses or dispo-

sitions, while "emotional" representations do have such automatic effects, and are therefore behaviorally valenced... Highly flexible behavior requires "cognitive" representations that can be easily mixed around and recombined as situational demands vary, and without pulling the agent in sixteen different behavioral directions at once. For example, sometimes you need to avoid cars, and other times you need to approach them. It is useful, then, if you can represent CAR in a behaviourally neutral or "cognitive" way, one that doesn't automatically presuppose a particular behavioral response. Stereotyped behavior, in contrast, doesn't require this sort of flexibility and therefore doesn't require "cognitive" representations, at least not to the same extent (Greene 2007, p. 40).

The idea thus would be that consequentialist judgment is characterized by "cognitive", i.e. motivationally neutral representations, which need external support (e.g. from mechanisms of cognitive control) to "impose themselves" and guide our decisions and behaviour. No such support is needed for the "emotional" representations characterizing deontological judgment, whose motivational impact is immediate.

However, this distinction does not help with our question, because it does not help us to identify the "cognitive" mechanism which underlies consequentialist judgment. On the basis of the distinction, we could say that a mechanism is cognitive (as opposed to affective) if the representations which it generates are cognitive, i.e. motivationally neutral. But what kind of representations does Greene have in mind? On the one hand, he might think of structured, explicitly morally or evaluatively valenced representations, such as "five lives saved are better than no lives saved" for the consequentialist judgment, and "smothering my baby is absolutely wrong" for the deontological judgment. One might then assume that the first type of representations are motivationally neutral whereas the second type are intrinsically motivating (immediately inducing me to reject the option of smothering the baby). However, this way of understanding the emotion/cognition dichotomy does not obviously distinguish separate *mechanisms*. For representations such as "five lives saved are better than no lives saved" are very likely to be intrinsically motivating in non-dilemmatic situations, when

there is no countervailing factor, inducing us to do what we can to save the five. Therefore, a mechanism cannot be said to be “cognitive” insofar as it generates representations with such a content, because such representations can well be “emotional”, i.e. intrinsically motivating.

On the other hand, it is possible that Greene has in mind factual or descriptive representations, as the example of CAR suggests. CAR is represented in a neutral fashion, because it does not automatically trigger a response of attraction or repulsion. The idea would then be that the structured *factual* representations on which consequentialist judgment depends are motivationally neutral, whereas the factual representations on which deontological judgment depends are “emotional”. However, such an idea is dubious and does not help identify distinct mechanisms. First, consequentialist and deontological judgments operate on some shared factual representations in the crying baby case, e.g. the representation “me smothering my baby by hand”. This same representation is at once emotional and cognitive: it is emotional insofar as it triggers an automatic response of repulsion, and cognitive insofar as it is factored in as a “cost”, but not rejected outright in the consequentialist evaluation. Moreover, the factual representation “more lives saved than killed” which drives the consequentialist judgment is not “inherently neutral”: it surely must automatically trigger some positive response, at least when considered in abstraction from the means of saving the lives. So representations are not cognitive or emotional on the basis of their content (be it evaluative or factual): the same representation can be emotional and cognitive at the same time or at different times given appropriate circumstances. But then the cognitive or emotional nature of representations cannot be a good guide to the cognitive or emotional nature of the mechanisms. In particular, one cannot say that the mechanism underlying consequentialist judgment is cognitive because it generates (or operates upon) cognitive representations in this sense.

The second theoretical option intends to go beyond the emotion/cognition dichotomy, and starts from a difference between two kinds of emotions related to “deontological” and “consequentialist” moral judgment. Greene and colleagues have recently developed a somewhat speculative distinction between two types of emotional processing: ‘alarm-bell’-like and ‘currency’-like (Greene

2007; Cushman et al. 2010). According to their hypothesis, alarm-bell emotions—immediate, vivid and motivationally demanding—underlie deontological judgments in cases such as the crying baby, whereas currency emotions—being mediated by reasoning, more subtle and “provisional”—underlie consequentialist reasoning and judgment. Given that both verdicts are thus ultimately based on emotional responses of some kind, Greene asserts his multi-system or dual-process view to be “sympathetic to Hume’s claim that all moral judgment (including consequentialist judgment) must have some emotional component” (Greene 2007, p. 41).

Here is how Cushman et al. understand the distinction and point to possible neural localizations of the respective processes:

The core idea is that alarm-bell emotions are designed to circumvent reasoning, providing absolute demands and constraints on behavior, while currency emotions are designed to participate in the process of practical reasoning, providing negotiable motivations for and against different behaviors. For example, the amygdala, which has been implicated in responses to personal moral dilemmas, reliably responds to threatening visual stimuli such as snakes and faces of out-group members...Thus the amygdala is a good candidate for a region that is critical for supporting at least some alarm-bell emotions. In contrast, Knutson and colleagues...have identified a set of meso-limbic brain regions that appear to represent expected monetary value in a more graded fashion, with distinct regions tracking a stimulus’s reward magnitude, reward probability, and expected value. These regions, in a rather transparent way, support currency-like representations (Cushman et al. 2010, p. 62).

In his 2007 paper, Greene captures the difference as follows. The cost-benefit analysis that characterizes consequentialist judgment is

a weighing process and not an “alarm” process. The sorts of emotions hypothesized to be involved here say, “Such-and-such matters this much. Factor it in.” In contrast, the emotions hypothesized to drive deontological judgment are far less subtle. They are...alarm signals that issue simple

commands: “Don’t do it!” or “Must do it!” While such commands can be overridden, they are designed to dominate the decision rather than merely influence it (Greene 2007, p. 65).

The “cognitive” mechanism that underlies consequentialist judgment operates via currency emotions: the process of weighing up requires that each alternative only offer a *prima facie* or *pro tanto* reason for action and does not dominate over the others, psychologically (motivationally) and normatively speaking. Here is a non-moral example:

The desire for ice cream on a hot summer day is an example of a currency emotion: it supplies a reason to pursue the Good Humor truck, but this reason can be traded off against others, such as maintaining a slim poolside profile. Currency-like emotions function by adding a limited measure of motivational weight to a behavioral alternative, where this weighting is designed to be integrated with other weightings in order to produce a response. Such emotional weightings say, “Add a few points to option A” or “Subtract a few points from Option B,” rather than issuing resolute commands (Cushman et al. 2010, p. 63).

However this is not the end of the story. Currency responses assign a partial evaluation to the different alternatives (guided by some criterion or other, such as monetary gain, personal well-being, or lives saved/lost). But the weighing up process, in the crying baby case, results in a precise verdict: it is acceptable (or even obligatory) to smother the baby. Moreover, this is a resolute, non-negotiable demand, if it is to conflict with the analogously non-negotiable demand not to smother the baby, issued by the “alarm-bell” affective system. So the question for a multi-system view is: how can a non-negotiable demand result from partial, negotiable valuations?

The problem is, again, one of functional symmetry. The distinction between currency and alarm-bell emotions may be psychologically real and important, but does not pick out the two opposing moral sensitivities supposedly at work in the crying baby dilemma. Alarm-bell emotions do issue one kind of final verdict (“deontological”). By contrast, currency emotions, to the extent that each of them assigns some

partial valence to a different outcome or aspects thereof (smothering the baby: *prima facie* bad; saving five lives: *prima facie* good; etc.), still appear to be mechanisms subserving the cost-benefit analysis required by the consequentialist judgment, rather than being the locus of such judgment. Currency emotions might be psychologically necessary to carry out a sensible cost-benefit analysis, but—as described—are not sufficient to generate an overall verdict.

In search of an answer to the problem, one can consider Cushman et al.’s hypothesis that there might be some currency affective “premises” which are *general* in content and guide the particular currency valuations towards particular overall valuations:⁴

[H]arm is bad, regardless of who experiences it. Benefits are good, regardless of who experiences them. More harm is worse than less harm. More benefits are better than fewer benefits. Small harms can be outweighed by large benefits. Small benefits can be outweighed by large harms (Cushman et al. 2010, p. 65).

On the basis of such affective premises, we rationally construct consequentialist practical principles telling us to, e.g., minimize harms. And one might think that this *non-negotiable* principle drives the non-negotiable response which conflicts with the “deontological” response. So, at last, we can perhaps establish functional symmetry by seeing the “consequentialist” sensitivity as the result of a mixed affective-cognitive process which stands in opposition to the purely affective alarm-bell-like “deontological” sensitivity.

Cushman et al. themselves admit that the idea is speculative. But there seems to be an important conceptual obstacle to the proposal. First we were told that the defining trait of currency-like emotions is that they “provide negotiable motivations for and against different behaviors”. Applied to the crying baby case, this means that the currency affective system regards the prospective death of the baby as bad, but as a negotiable harm nonetheless (a negotiable reason not to kill the baby). However, according to the proposal in the previous paragraph, there are general currency-like affective “premises” which guide consequentialist judgment. Since the latter eventually issues non-negotiable verdicts, the non-negotiability would then be somehow derived or inherited from the affective premises on which the judgment is based.

For instance, the premise “more harm is worse than less harm” seems to imply a *non-negotiable* reason to choose less harm over more harm. Applied to the crying baby case, this premise will determine a non-negotiable reason to choose killing the baby over not killing the baby. What we have here then is a *currency-like* emotion which provides a *non-negotiable* motivation for a certain behaviour. But this contradicts the defining trait of currency-like emotions. So, more work seems required to get clear on the nature of currency-like emotions.

The alternative is that the set of affective premises which guide consequentialist judgment are constituted by distinctive emotions which are both *general* in the content they underwrite and *alarm-bell-like*. This proposal would fit better with the negotiable/non-negotiable distinction, since the guiding alarm-bell emotions, *qua* alarm-bell-like, would accordingly issue the non-negotiable verdicts of consequentialism.

However, there seem to be two problems with this suggestion. One is that the “premises” which guide consequentialist judgment—harm is bad, benefits are good, more harm is worse than less harm, etc.—cannot all be constituted by alarm-bell emotions. In particular, if “harm is bad” were systematically underwritten by an alarm-bell emotion, our motivation to avoid harm would be non-negotiable. Consequentialist thinking, characterized as it is by the possibility of balancing harms against benefits, thus could not even get off the ground.⁵ The second problem is that to construe *general* evaluative attitudes as alarm-bell-like seems to stretch the concept of an alarm-bell reaction too far. For all that Cushman et al. tell us, alarm-bell aversions seem to typically arise in response to *particular*, concrete threatening stimuli (e.g. snakes and faces of out-group members are mentioned above, to which one might add the specific imagined action of smothering one’s baby).

The overall conclusion of this section thus is that neither the emotional vs. cognitive representation dichotomy, nor the alarm-bell vs. currency emotions distinction satisfactorily carve out the psychological processes responsible for the deontological and the consequentialist moral judgments in the crying baby case. Consequentialist thinking is characterized by both negotiable reasons, such as a *prima facie* reason to avoid harm, and non-negotiable demands, such as the requirement to minimize harm, even when this involves causing some harm.

While “cognitive” (i.e. not immediately motivating) representations and currency-like emotions might reasonably underwrite and assist in the process of weighing up harms and benefits, pro and con reasons, it is still far from clear which representations and emotional processes (if emotional at all) guide and underlie the realization that smothering the baby is the best course of action, given the circumstances, and the non-negotiable injunction (or at least permission) to proceed with such an action.

6. CONCLUSION

In this paper I have attempted a critical examination of the multi-system or dual-process view of moral judgment. This view aims to provide a psychological explanation of moral sensitivity, and in particular an explanation of conflicting moral sensitivities in dilemma cases such as the crying baby scenario. I have argued that proponents of the multi-system view still owe us a satisfactory account of the mechanisms underlying “consequentialist” responses to such scenarios. For one thing, the “cognitive”, as opposed to “emotional”, processes involved in consequentialist reasoning only seem to play a subserving role with respect to the final judgment (providing non-moral inputs to judgment, or exerting additional strength to override the immediate “deontological” response). In this sense, Greene and colleagues fail to identify a peculiar system of moral judgment specularly opposed to the affective “deontological” one. For another, Greene and colleagues’ work on the emotion-cognition dichotomy and the distinction between alarm-bell and currency emotions, though promising, still falls short of providing an adequate and consistent picture of the psychological mechanisms underlying “cognitive” evaluations and verdicts in dilemma scenarios. It is suggested that alongside further experimental work, proponents of this view should pay more attention to the conceptual underpinnings of their distinctions.

Notes

¹Or at least to all subjects whose response was analyzed and not discarded as abnormal or displaying “unbalanced factors” (Greene et al. 2004, p. 398).

²Such as (1) evidence from studies on subjects with damage to the ventro medial prefrontal cortex (an area associated with emotions), who seem to be more disposed than healthy subjects to endorse harmful behaviour in order to promote the greater good; (2) evidence from subjects under “cognitive load”: “consequentialist” judgments seem to take longer to make than “deontological” ones, suggesting that cognitive tasks interfere with consequentialist reasoning but not with deontological reasoning (Greene et al. 2008).

³Or at least a form of sensitivity to reasons that proceeds from axioms and directly issues verdicts rather than prima facie judgments.

⁴Cushman et al. consider this as a hypothesis regarding the *origin* of consequentialist thinking. My aim is to understand whether the hypothesis might work as an explanation of the psychological mechanism underlying consequentialist responses to the crying baby case.

⁵Of course it might be instead plausible to think that the qualified premises on which consequentialism is built (“harm is *pro tanto* bad”), and their attached “currency” emotions, are historically the result of modifications to original alarm-bell responses to harm. See Cushman et al. (2010) and Nichols (2004).

References

- Cushman, F. A. & Greene, J. D. in press. ‘Finding faults: How moral dilemmas illuminate cognitive structure’. In J. Decety & J. Cacioppo (eds.) ‘The Handbook of Social Neuroscience’, 269–279. New York: Oxford University Press.
- Cushman, F. A. & Young, A. W. 2009. ‘The psychology of dilemmas and the philosophy of morality’. *Ethical Theory and Moral Practice* 12, no. 1: 9–24.
- Cushman, F. A., Young, L. & Greene, J. 2010. ‘Multi-system Moral Psychology’. In J. Doris et al. (ed.) ‘The Oxford Handbook of Moral Psychology’, 46–69. Oxford: Oxford University Press.
- Greene, J. 2007. ‘The secret joke of Kant’s soul’. In W. Sinnott-Armstrong (ed.) ‘Moral Psychology (Vol. 3)’, 35–117. Cambridge, MA: MIT Press.
- Greene, J., Sommerville, R. B., Nystrom, L. E., Darley, J. M. & Cohen, J. D. 2001. ‘An fMRI investigation of emotional engagement in moral judgment’. *Science* 293: 2105–2108.
- Greene, J., Nystrom, L. E., Engell, A. D., Darley, J. M. & Cohen, J. D. 2004. ‘The neural bases of cognitive conflict and control in moral judgment’. *Neuron* 44: 389–400.
- Greene, J., Morelli, S. A., Lowenberg, K., Nystrom, L. E. & Cohen, J. D. 2008. ‘Cognitive load selectively interferes with utilitarian moral judgment’. *Cognition* 107: 1144–1154.
- Haidt, J. 2001. ‘The emotional dog and its rational tail: A social intuitionist approach to moral judgment’. *Psychological Review* 108: 814–834.
- Miller, E.K. & Cohen, J.D. 2001. ‘An integrative theory of prefrontal cortex function’. *Annual review of neuroscience* 24, no. 1: 167–202.
- Nichols, S. 2004. *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford: Oxford University Press.