In another study, Pizarro et al. (2003) found that participants discounted blame for intentional actions that were not carried out quite as intended (i.e., acts that lacked "intentions-in-action"; Searle 1983). For example, when a murderer tripped and accidentally stabbed his victim in the process of attempting to kill him, he was perceived as less blameworthy. Interestingly, when asked to give their most rational response, participants judged acts that did and did not possess intention-in-action to be equally blameworthy. This suggests that, at least for some, discounting blame for acts that lacked intention-in-action was subjectively irrational.

In another example, Tetlock et al. (under review) examined conservative and liberal managers' reactions to a hypothetical employee error (failure to mail a package on time) with either mild or severe consequences. Both conservative and liberal managers judged the employee more harshly when the consequences of the error were severe (this has been referred to as an "outcome bias" and "moral luck"; Baron & Hershey 1988). Liberals viewed this outcome bias as an error, and reduced their recommended punishment in the severe consequences case when asked to consider how they would have reacted had the consequences been mild. In contrast, conservatives saw it as perfectly appropriate to determine the employee's punishment based on the consequences of his or her actions.

Liberals and conservatives also disagree regarding whether certain socialized intuitions are rational. Ingenious studies by Jonathan Haidt and his colleagues demonstrate that most people find it intuitively wrong to wash one's toilet with the American flag, eat one's recently expired pet, or masturbate into a dead chicken (Haidt 2001; Haidt et al. 2003). When asked to make the most rational judgment possible, liberals appear to correct for their intuitions – reducing blame for eating Fido, for example (Uhlmann et al., in preparation; see also Haidt & Hersh 2001). In contrast, conservatives provide essentially the same judgments when asked to respond rationally versus intuitively. For liberals, the judgments identified by Haidt exert a subjectively irrational influence on their judgments. But for conservatives, who place a high priority on traditional values, such judgments may seem perfectly well-grounded.

If people are indeed exhibiting "absurd moral judgments" (target article, Abstract), we suggest that this is not because heuristics lead individuals' moral judgments to diverge from some objective standard of morality (such as weak consequentialism), but because these judgments would be deemed irrational by the participant himself upon reflection. Perhaps this sense of the term "error" may be the best way to avoid the morass of subjectivity inherent in studying the moral judgments of other people, and may also keep researchers from hurling insults at each other's normative theories of choice.

ACKNOWLEDGMENT
We thank Andy Poehlman for his comments on an earlier draft of this article.

Cognitive heuristics and deontological rules

Ilana Ritov

School of Education, Hebrew University, 91905 Jerusalem, Israel.
msiritov@mscc.huji.ac.il

Abstract: Preferences for options that do not secure optimal outcomes, like the ones catalogued by Sunstein, derive from two sources: cognitive heuristics and deontological rules. Although rules may stem from automatic affective reactions, they are deliberately maintained. Because strongly held convictions have important behavioral implications, it may be useful to regard cognitive heuristics and deontological rules as separate sources of nonconsequential judgment in the moral domain.

The idea of error-prone heuristics is especially controversial in the moral domain, as Sunstein notes, although examples of choices that violate consequential principles are abundant. Among those examples are the "punishment" of companies for cost–benefit analyses to determine their investment in safety, the betrayal aversion, the resistance to "tampering with nature," and the rejection of probability of detection as a normative factor in determining punitive damages. These choices have grave consequences for the lives and well-being of many people, and the contribution of this article in drawing attention to these problems is highly important.

To ascertain that those nonconsequential judgments result from the application of mental heuristics, it is necessary to address the question of what a heuristic is. The notion of a heuristic is not well defined in the psychological literature. As Sunstein notes, Tversky and Kahneman (1984) used the term heuristic to refer to a strategy that "relies on a natural assessment to produce an estimation or a prediction." These strategies take on the form of mental shortcuts, or general purpose rules, often applied without consciousness, in judgmental tasks requiring assessment of unknown values. More recently, the evolving research on dual process theories led to a broader view of the nature of heuristics. Heuristics have come to be equated with processes of System I. This system, also referred to as the experiential system, operates automatically and effortlessly, is oriented to concrete images, and responds affectively. By contrast, the rational system, or System II, operates consciously and effortfully, and is deliberate and reason-oriented (Epstein & Pacini 1999).

In the current broad view of heuristics, not only are estimates of quantities by rules of thumb seen as the products of heuristics, but any expression of preference derived through the experiential system is regarded as such, as well. Although the boundaries of the set have not been explicitly delineated, the most notable feature of a heuristic process that distinguishes it from the cognitive processes classified as reasoning or rational is its nondeliberative nature. Although the outcomes of a heuristic can be deliberately adopted by System II, judgment by heuristic is typically an intuitive and unintentional process (Kahneman & Frederick 2002). It is usually passive and preconscious.

Returning to the examples discussed by Sunstein in the present article, these can arguably be roughly classified into two kinds: the ones that reflect the use of general cognitive heuristics in judgments (applied in the moral domain), and others that deliberate application of rules. The clearest example of a non-deliberative heuristic is the outrage heuristic in punishment. Although people are certainly aware of their outrage, they are most likely not aware of using this emotional reaction as the primary, or even the sole determinant of the punishment they set.

The resistance to cloning, stemming from the conviction that one should not "play god," or "tamper with nature," is an example of the second kind. Although the belief itself may stem from an emotional reaction, it is explicitly adopted by the rational system. The principle is held consciously and deliberately. It is relatively abstract and context-general. Similarly, the rejection of the role of probability of detection in setting punitive damages is the result of deliberate processing, often by expert and sophisticated respondents (Sunstein et al. 2000). In both of these examples, as well as in other ones, the judgment is determined by a deontological rule.

Deontological rules are rules that concern actions rather than consequences. These rules are often associated with values that people think of as absolute, not to be traded off for anything else (Baron & Spranca 1997). These protected values, compared to values that are not absolute in this way, have various predicted properties, such as insensitivity to quantity: The amount of the harm done when they are violated does not matter as much as for other values. Furthermore, in judgments involving a deontological rule or a protected value, the participation of the actor is crucial, even when the consequences are the same. The tendency to punish companies that base their decisions on cost–benefit analysis, even if a high valuation is placed on human life, may reflect the agent relativity characteristic of the rule "do not trade human life for money."

https://doi.org/10.1017/S0140525X05400096 Published online by Cambridge University Press

As protected values are related to deontological rules against action ("do not play god," "do not tamper with nature," "do not cause death," etc.) they tend to amplify omission bias (Ritov & Baron 1999). If a person has a protected value against, for example, destroying species, this value seems to apply to action rather than inaction. That person might be unwilling to take an action that would cause the extinction of one species, in order to save five, even when the relative outcomes are fully spelled out. By contrast, another person, who cares just as much about preventing the extinction of species as the first person, but not as a protected value, would prefer that action be taken in order to achieve better consequences as a whole. Although the values people hold protected vary considerably, the basic finding of greater omission bias for protected values holds across a wide array of issues, ranging from endangered species, to withdrawal from occupied territories (for Israeli respondents). In all those cases, people holding protected values deliberately preferred omission, despite the fact that they knew explicitly that action would yield better consequences with respect to the specific problem.

The origin of deontological rules is the subject of much research. They may be the result of generalization from a range of problems. Deontological rules are undoubtedly closely linked with affect, but it remains an open question whether their impact is fully mediated by emotions. Even if espousing that deontological rules are primarily an expression of extreme affect, the judgmental process is different from other experiential processes in its explicit and deliberate nature. Until further research provides better understanding of those processes, it may be more useful to regard cognitive heuristics and deontological rules as separate sources of nonconsequential judgment in the moral domain.

Intuitions, heuristics, and utilitarianism

Peter Singer
University Center for Human Values, Princeton University, Princeton, NJ 08544; and Centre for Applied Philosophy and Public Ethics, University of Melbourne, Victoria 3010, Australia. **psinger@princeton.edu**

Abstract: A common objection to utilitarianism is that it clashes with our common moral intuitions. Understanding the role that heuristics play in moral judgments undermines this objection. It also indicates why we should not use John Rawls' model of reflective equilibrium as the basis for testing normative moral theories.

At one point Cass Sunstein suggests that his assertion that heuristics play an important role in our moral judgments does not really favor one side or the other in the debate between utilitarians and deontologists:

> If moral heuristics are in fact pervasive, then people with diverse foundational commitments should be able to agree, not that their own preferred theories are wrong, but that they are often applied in a way that reflects the use of heuristics. Utilitarians ought to be able to identify heuristics for the maximization of utility; deontologists should be able to point to heuristics for the proper discharge of moral responsibilities; those uncommitted to any large-scale theory should be able to specify heuristics for their own more modest normative commitments. (target article, sect. 1, para. 4)

This seems to me to lean too far towards normative neutrality. Seen against the background of a long-running debate in normative ethics, Sunstein's illuminating essay gives support to utilitarians, and not to deontologists.

A major theme in normative ethics for the past two centuries has been the debate between those who support a utilitarian, or more broadly consequentialist, normative ethical theory and those who ground their normative ethics on our common moral judgments or intuitions. In this debate, the standard strategy employed by deontologists has been to present examples intended to show that the dictates of utilitarianism clash with moral intuitions that we all share – and that fit with deontological views of ethics. A famous literary instance occurs in Dostoyevsky's *The Karamazov Brothers*. Ivan challenges Alyosha to say whether he would consent to build a world in which people were happy and at peace, if this ideal world could be achieved only by torturing "that same little child beating her chest with her little fists." Alyosha says that he would not consent to build such a world on those terms (Dostoyevsky 1879). Hastings Rashdall purported to refute hedonistic utilitarianism by arguing that it cannot explain the value of sexual purity (Rashdall 1907, p. 197). H. J. McCloskey, writing at a time when lynchings in the American South were still a possibility, thought it a decisive objection to utilitarianism that the theory might direct a sheriff to frame an innocent man in order to prevent a white mob from lynching half a dozen innocents in revenge for a rape (McCloskey 1957). Bernard Williams invited utilitarians to ponder a similar example, of a botanist who wanders into a village in the jungle where twenty innocent people are about to be shot. He is told that nineteen of them will be spared, if only he will himself shoot the twentieth (Williams 1973).

Initially, the use of such examples to appeal to our common moral intuitions against consequentialist theories was an ad hoc device lacking meta-ethical foundations. It was simply a way of saying: "If Theory U is true, then in situation X you should do Y. But we know that it would always be wrong to do Y, therefore U cannot be true." This is an effective argument against U, as long as the judgment that it would always be wrong to do Y is not challenged. But the argument does nothing to establish that it is always wrong to do Y, nor what a sounder theory than U would be like.

John Rawls took the crucial step towards fusing this argument with an ethical methodology when he argued that the test of a sound moral theory is that it can achieve a "reflective equilibrium" with our considered moral judgments. By "reflective equilibrium" Rawls meant that, where there is no inherently plausible theory that perfectly matches our initial moral judgments, we should modify either the theory, or the judgments, until we have an equilibrium between the two.

The model here is the testing of a scientific theory. In science, we generally accept the theory that best fits the data, but sometimes, if the theory is inherently plausible and fits some of the data, we may be prepared to accept it despite its failure to fit all the data. We assume, perhaps, that the outlying data are erroneous, or that there are undiscovered factors at work in that particular situation. In the case of a normative theory of ethics, Rawls assumes that the raw data are our prior moral judgments. We try to match them with a plausible theory, but if we cannot, we reject some of the judgments, and modify the theory so that it matches others. Eventually the plausibility of the theory and of the surviving judgments reach an equilibrium, and we then have the best possible theory. In this view, the acceptability of a moral theory is not determined by the internal coherence and plausibility of the theory itself, but rather, to a significant extent, by its agreement with those of our prior moral judgments that we are unwilling to revise or abandon. In *A Theory of Justice,* Rawls uses this model to justify tinkering with his original idea of a choice arising from a hypothetical contract, until he is able to produce results that are not too much at odds with our ordinary ideas of justice (Rawls 1951; 1971, p. 48).

The model of reflective equilibrium has always struck me as dubious. The analogy between the role of a normative moral theory and a scientific theory is fundamentally misconceived (Singer 1974). Our common moral intuitions are not "data" in the sense that a series of measurements of the positions of electrons may be data that any credible scientific theory must explain. A scientific theory seeks to explain the existence of data that are about a world "out there" that we are trying to explain. Granted, the data may have been affected by errors in measurement or interpretation, but unless we can give some account of what the errors might have been, it is not up to us to choose or reject the observations. A normative ethical theory, however, is not trying to explain our common moral intuitions. It might reject all of them, and still be su-

https://doi.org/10.1017/S0140525X05400096 Published online by Cambridge University Press