### Indignation: Psychology, Politics, Law

The Harvard community has made this article openly available. **Please share** how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Published Version</td>
<td><a href="http://chicagounbound.uchicago.edu/law_and_economics/263/">http://chicagounbound.uchicago.edu/law_and_economics/263/</a></td>
</tr>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:12921739">http://nrs.harvard.edu/urn-3:HUL.InstRepos:12921739</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP</a></td>
</tr>
</tbody>
</table>
2007

Indignation: Psychology, Politics, Law

Cass R. Sunstein

Daniel Kahneman

Follow this and additional works at: http://chicagounbound.uchicago.edu/law_and_economics

Part of the Law Commons

Recommended Citation


This Working Paper is brought to you for free and open access by the Coase-Sandor Institute for Law and Economics at Chicago Unbound. It has been accepted for inclusion in Coase-Sandor Working Paper Series in Law and Economics by an authorized administrator of Chicago Unbound. For more information, please contact unbound@law.uchicago.edu.
INDIGNATION: PSYCHOLOGY, POLITICS, LAW

Daniel Kahneman and Cass R. Sunstein

THE LAW SCHOOL
THE UNIVERSITY OF CHICAGO

July 2007

Indignation: Psychology, Politics, Law*

Daniel Kahneman
Princeton University

Cass R. Sunstein
University of Chicago

Abstract

Moral intuitions operate in much the same way as other intuitions do; what makes the moral domain distinctive is its foundations in the emotions, beliefs, and response tendencies that define indignation. The intuitive system of cognition, System I, is typically responsible for indignation; the more reflective system, System II, may or may not provide an override. Moral dumbfounding and moral numbness are often a product of moral intuitions that people are unable to justify. An understanding of indignation helps to explain the operation of the many phenomena of interest to law and politics: the outrage heuristic, the centrality of harm, the role of reference states, moral framing, and the act omission distinction. Because of the operation of indignation, it is extremely difficult for people to achieve coherence in their moral intuitions. Legal and political institutions usually aspire to be deliberative, and to pay close attention to System II; but even in deliberative institutions, System I can make some compelling demands.

There have been profound changes in the psychological analysis of moral sentiments over the last few decades, from a conception of morality as a system of abstract rules that can be understood and internalized (Kohlberg, 1969), to a view that emphasizes moral emotions and moral intuitions that are not anchored in reasons (e.g., Greene, 2004; Haidt, 2001; Rozin, 1999). In this brief essay we sketch an analysis of moral intuitions that builds on the new work, and relates it to a general approach to the study of intuitive thought. We suppose that the same cognitive machinery generates attitudes, judgments, beliefs and actions in moral domains and in other domains, and that the moral domain is distinctive because it involves a special attitude: the complex of emotion, beliefs and response tendencies that define indignation.

Rozin, Lowery and Imada (1999) suggested that there are three variants of indignation: anger, disgust and contempt. The first two, but perhaps not the third, are strongly associated with altruistic punishment. We are mainly concerned here with the

---

*Rozin, Lowery and Imada (1999) suggested that there are three variants of indignation: anger, disgust and contempt. The first two, but perhaps not the third, are strongly associated with altruistic punishment. We are mainly concerned here with the

1This paper was published under the title, “Cognitive Psychology of Moral Intuitions,” in Neurobiology of Human Values (Jean-Pierre P. Changeux et al. eds. 2005).
variant of indignation that involves anger. For an example, imagine that you see a bully beat up a weakling without any provocation. You will respond with indignation. Like other attitudes, indignation has three related constituents:

(1) an **emotion**—which, confusingly, is also named indignation
(2) a set of **beliefs and judgments** about the reprehensible action
(3) a **response tendency**, to administer punishment to the guilty actor, even at some cost to oneself (economists call this a readiness to provide altruistic punishment, because it occurs even in the absence of any expectation of further interaction with the target of punishment).

Like other intentional states, indignation can be explained in two quite different ways: by referring to reasons, or to psychological causes. As you see the bully assaulting his victim, you are likely to be aware of a reason for your emotion: the action violates an accepted (and in your view justified) social rule that prohibits unprovoked aggression. The categorization of the action provides a reason for indignation, a reason that the observer expects other objective observers to endorse. Classical analyses of moral development were much concerned with people’s ability to marshal reasons for their judgments; the reasons were often understood as causing those judgments.

The view that has gained currency in recent years is quite different. In this view, indignation is like a fear of spiders. One does not fear spiders because they are dangerous—one just fears them. Because people tend to attribute their reactions to the objects that evoke these reactions, the feared spider is perceived as a dangerous spider. However, the perception of dangerousness is not the reason for the fear or even its cause; both the fear and the perception are symptoms of an uncontrolled reaction to spiders. Many people who are afraid of spiders know that their fear is objectively groundless and lacks a reason. Haidt (2001) described the equivalent state in the moral domain as “moral dumbfounding”: the experience of strong moral reactions for which no adequate reason comes to mind. Indignation, we suggest, is often not caused by reasons, and people can be dumbfounded when they are asked to explain why they are indignant. In fact some puzzling outcomes, in both politics and law, are a product of indignation that is intense and hard to justify (Sunstein, 2005). Moral dumbfounding find its mirror image in moral numbness, in which people are not indignant even though they have reason to be, and know they do.
The two-system model of the mind

Consider the expression “17 x 24 = ?”. For the great majority of people, the correct answer to the question will come to mind only if it is produced by a voluntary mental activity, which involves deliberate application of a rule, requires several steps of computation, storage and retrieval and takes a significant amount of time. For contrast consider the word “vomit.” For the great majority of people, disgust will come to mind in a completely involuntary process, which is produced very quickly by a process which is itself unconscious—one is aware only of its outcome. The two examples represent different families of cognitive processes.

The ancient idea that cognitive processes can be partitioned into two main families—traditionally called “intuition” and “reason”—is now widely embraced under the general label of dual-process theories (Chaiken and Trope, 1999; Sloman, 1996). Dual-process theories come in many forms, but all distinguish cognitive operations that are quick and associative from others that are slower, more reflective, and frequently more calculative (Gilbert, 1999). We adopt the generic labels ‘System 1’ and ‘System 2’ from Stanovich and West (2001). These terms may suggest the image of autonomous homunculi, and there is in fact evidence that the two systems correspond to different locations in the brain; but we do not suggest that the two systems are independent. We use “systems” as a label for collections of processes that are distinguished by their speed, their controllability, and the contents on which they operate (see Table 1).

Table 1. Two Cognitive Systems

<table>
<thead>
<tr>
<th>System 1 (Intuitive)</th>
<th>System 2 (Reflective)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatic</td>
<td>Controlled</td>
</tr>
<tr>
<td>Effortless</td>
<td>Effortful</td>
</tr>
<tr>
<td>Associative</td>
<td>Deductive</td>
</tr>
<tr>
<td>Rapid</td>
<td>Slow</td>
</tr>
<tr>
<td>Opaque Process</td>
<td>Self-Aware</td>
</tr>
<tr>
<td>Skilled</td>
<td>Rule-Following</td>
</tr>
</tbody>
</table>

Although System 1 is more primitive than System 2, it is not necessarily less capable. On the contrary, complex cognitive operations eventually migrate from System 2 to System 1 as proficiency and skill are acquired. A striking demonstration of the
intelligence of System 1 is the ability of chess masters to perceive the strength or weakness of chess positions instantly. For those experts, pattern matching has replaced effortful serial processing. Some people are widely taken to be moral experts as well, and it should be clear that pattern matching occurs in the moral domain; indignation is often a result.

In the particular dual-process model that we assume, System 1 quickly proposes intuitive answers to judgment problems as they arise, and System 2 monitors the quality of these proposals, which it may endorse, correct, or override. The judgments that are eventually expressed are called ‘intuitive’ if they retain the hypothesized initial proposal without much modification. The roles of the two systems in determining stated judgments depend on features of the task and of the individual, including the time available for deliberation (Finucane et al., 2000), the respondent’s mood (Isen, Nygren & Ashby, 1988; Bless et al., 1996), and intelligence (Stanovich & West, 2001). Without time for deliberation, for example, indignation can be extremely intense; when people have time to reflect, their reaction diminishes. And when System 1 is not indignant, and people are morally numb, deliberation can heighten moral concern and sometimes produce indignation (though this can take a great deal of heavy lifting on the part of System 2).

We assume that System 1 and System 2 can be concurrently active, that automatic and controlled cognitive operations compete for the control of overt responses, and that deliberate judgments are likely to remain anchored on initial impressions. Our views in these regards are similar to other dual-process models (Epstein, 1994; Gilbert, 1989, 2002; Hammond, 1996; Sloman, 1996).

The properties listed in Table 1 are shared by the system that produces intuitive thoughts and by the perceptual system. Intuitive judgments appear to occupy a position—perhaps corresponding to evolutionary history—between the automatic operations of perception and the deliberate operations of reasoning. Unlike perception, however, the operations of System 1 are not restricted to the processing of current stimulation. Like System 2, the operations of System 1 deal with stored concepts as well as with precepts, and can be evoked by language. This view of intuition suggests that the vast store of scientific knowledge available about perceptual phenomena can be a source of useful
hypotheses about the workings of intuition. The strategy of drawing on analogies from perception is applied in the following section.

A defining property of intuitive thoughts is that they come to mind spontaneously, like percepts. The technical term for the ease with which mental contents come to mind is *accessibility* (E. Tory Higgins, 1996). To understand intuition, we must understand why some thoughts are accessible and others are not.

Some attributes are more accessible than others, both in perception and in judgment. Attributes that are routinely and automatically produced by the perceptual system or by System 1, without intention or effort, have been called *natural assessments* (Tversky and Kahneman, 1983). For example, experimental evidence shows that when a perceiver is exposed to a set of objects of the same general kind (e.g., a set of lines of different size), attributes of a prototypical member of the set (e.g., the average length of the lines) are computed effortlessly and automatically. Other attributes (e.g., the total length of the lines) are not accessible—they can only be assessed by a deliberate and quite laborious computation. Thus, average length is a natural assessment, but total length is not. Kahneman and Frederick (2002) compiled a partial list of these natural assessments. In addition to physical properties such as size, distance, and loudness, the list includes more abstract properties such as similarity, causal propensity, surprisingness, affective valence, and mood.

The evaluation of stimuli as good or bad is a particularly important natural assessment. The evidence, both behavioral (John A. Bargh, 1997; Robert B. Zajonc, 1998) and neurophysiological (e.g., Joseph E. LeDoux, 2000), is consistent with the idea that the assessment of whether objects are good (and should be approached) or bad (should be avoided) is carried out quickly and efficiently by specialized neural circuitry. A remarkable experiment reported by Bargh (1997) illustrates the speed of the evaluation process, and its direct link to approach and avoidance. Participants were shown a series of stimuli on a screen, and instructed to respond to each stimulus as soon as it appeared, by moving a lever that blanked the screen. The stimuli were affectively charged words, some positive (e.g., LOVE) and some aversive (e.g., VOMIT), but this feature was irrelevant to the participant’s task. Half the participants responded by pulling the lever toward themselves, half responded by pushing the lever away. Although the response was
initiated within a fraction of a second, well before the meaning of the stimulus was consciously registered, the emotional valence of the word had a substantial effect. Participants were relatively faster in pulling a lever toward themselves (approach) for positive words, and relatively faster pushing the lever away when the word was aversive. The tendencies to approach or avoid were evoked by an automatic process that was not under conscious voluntary control. Several psychologists have commented on the influence of this primordial evaluative system (here included in System 1) on the attitudes and preferences that people adopt consciously and deliberately (Zajonc, 1998; Kahneman, Ritov and Schkade, 1999; Paul Slovic et al, 2002; Epstein, 2003). The implicit attitude test, designed to measure racial and other biases, finds that people show an automatic bias against African-Americans, older people, and others, even when they are unaware of it and wish to be unbiased (Nosek, Banaji, and Greenwald, 2002).

To complete this sketch of the operations of System 1, we introduce a process of attribute substitution that shapes many judgments and choices. The concept was introduced by Kahneman and Frederick (2002) as a basic mechanism to explain the basic results of heuristic judgment. They proposed that the reduction of complex tasks to simpler operations which characterizes such judgments is achieved by an operation in which an individual assesses “a specified target attribute of a judgment object by substituting another property of that object—the heuristic attribute—which comes more readily to mind” (p. 53). In the moral domain, as we show later, individuals charged with the task of determining the severity of a punishment appear to solve this difficult problem by consulting the intensity of their outrage.

Several of the processes we have discussed are involved in the explanation of a study reported by Miller and McFarland (1986), in which respondents determined the appropriate compensation for a man who was shot in the arm during the robbery of a grocery store. Some respondents were told that the robbery happened at the victim’s regular store. Other respondents were told that the victim was shot in a store that he visited for the first time, because his usual store happened to be closed that day. The two versions obviously differ in poignancy, because the counterfactual ‘undoing’ of an unusual event comes more easily to mind than the undoing of a normal occurrence. The difference of poignancy translated into a difference of $100,000 in the median award.
judged appropriate for the two cases. The participants in this experiment apparently answered the difficult question of appropriate question by mapping onto a scale of dollars their answer to a simple question: how much were they emotionally touched by the story.

It is unlikely that the respondents deliberately chose to provide this large compensation for poignancy. Indeed, when respondents were presented with both versions of the robbery story and asked whether a compensation board should make different awards in the two cases, 90% thought it should not. In the terms of the present discussion, the emotion-anchored process that produced the initial awards is dominated by System 1. The requirement to compare two questions evokes a much more complex activity, here attributed to System 2, which identifies the distinctive element that separates the two versions, and is unable to find any moral justification for different awards. This can be seen as an instance of “moral dumbfounding” (Haidt, 2001), in which a strong intuition exists that cannot be anchored in rules that the person consciously accepts.

**Outrage and Harm**

We have studied the operation of moral judgments in the domain of punitive damage awards (Kahneman, Schkade and Sunstein, 1998; Sunstein, Kahneman and Schkade, 1998). One of our hypotheses, couched in the language of the present treatment, was that the setting of such awards is mediated by an outrage heuristic.

Participants drawn from a jury roll in Texas were shown vignettes of legal cases in which a plaintiff had suffered a personal injury while using a product. For example, one of the scenarios concerned a child who had been burned when his pajamas caught fire as he was playing with matches. The pajamas were made of fabric that was not adequately fire-resistant, and the defendant firm had been aware of the problem. For some of the scenarios, alternative versions were constructed that differed in the severity of harm. In the high-harm version of the pajamas case, for example, the child was “severely burned over a significant portion of his body and required several weeks in hospital and months of physical therapy”. In the low-harm version “his hands and arms were badly burned and required professional medical treatment for several weeks”. Participants were told that the plaintiff had already been awarded compensatory damages.
One group of respondents indicated whether punitive damages were appropriate, and if so in what amount. Another group rated the outrageousness of the defendant’s behavior. In a subsequent re-analysis of this study, Kahneman and Frederick (2002) also obtained ratings of the severity of the harm suffered in each of the 14 vignettes. Lawsuits were not mentioned in these descriptions of harm. The same basic design was replicated twice, varying the size of the defendant firm.

The results supported the conclusion that assessments of punitive damages (the target attribute in this study) were mediated by an outrage heuristic. In the analysis offered by Kahneman and Frederick (2002), the outrage associated with each case was estimated by the product of the product of the average ratings of outrageousness and of harm. The correlations (over 14 vignettes) between the estimate of outrage and mean punitive damages were .90 in one of the firm-size conditions and .94 in the other.

The role of actual harm as a determinant of outrage in this experiment is of interest as a potential case of moral dumbfounding. The legally recognized distinction between murder and attempted murder is a salient example of the issue we raise. Consider the following scenarios:

1. A wishes B dead but does nothing about it
2. A tries to kill B and fails by chance
3. A tries to kill B and succeeds

It is difficult to justify a moral distinction between the last two cases. Indeed, it is safe to assume that people are asked to judge the outrageousness of the actions, there will be no difference. But punitive intent reflects the emotional intensity of the response to the event, and the emotion evidently depends on the harm that actually occurred. In the terms of the present analysis, the severity of punishment reflects the intensity of an emotional reaction in System 1. Punishments that are determined in this manner are expected to be crudely retributive, which is what we observe. Note that we are not claiming that it is impossible to defend the distinction, drawn by the criminal law, between murder and attempted murder. There may be good reasons for drawing that distinction. What we are arguing is that the distinction is not caused by those reasons; it is caused by the fact that moral intuitions, automatic and uncontrolled, are different in the two cases.
Indignation and the Definition of Harm

Indignation is evoked in an observer “by an agent who, intentionally and without provocation or adequate reason, causes a victim to suffer harm.” This highly abstract statement could be read as a rule that specifies appropriate reasons for indignation. Instead, we take the statement as an attempt to describe the rules that govern the elicitation of indignation in the observer’s System 1. As in many other situations, the use of a language of reason to describe processes that reason does not govern creates ambiguity and the risk of confusion. Indeed, every word in the statement changes its meaning across the two contexts—and the meaning that applies to the description must be a concept that System 1 is able to process.

An example is the notion of harm. Here again, we can offer a defining statement: “Harm is a loss relative an entitlement. An individual’s entitlements are governed by rules and expectations that are shared by the community.” This statement obviously covers many different kinds of harm, from physical injury to loss of property, and also to loss of reputation—depending on the nature of the entitlement that is violated. Here again, the statement sounds like a rule. It also appears to lead us nowhere, because now we must ask: what does System 1 know about entitlements? System 2 might, of course, have a theory of one or another kind, but that theory may or may not map onto moral intuitions. For System 1, the answer is that an entitlement is a socially endorsed normal state, also called a reference state, relative to which losses are defined. A reference state is an expectation that a valued state will be maintained. A valued state is defined by the fact that deviations from it produce positive or negative affect. Now we are making some progress, because System 1 (and the perceptual system) are quite capable of setting up expectations and in detecting deviations from them, and System 1 also produces emotional responses to changes.

The intuitive notion of entitlement was explored in a study of the moral attitudes to the behavior of firms in the market (Kahneman, Knetsch and Thaler, 1986). Residents of two Canadian cities evaluated the fairness of various unilateral actions that a firm might take to set the terms of its transactions with customers, employees or tenants. Table 2 presents examples that convey the general trends of the results, by attaching the labels Fair and Unfair to several sets of actions. Each set may be viewed as an experiment,
intended to isolate a critical factor in judgments of unfairness. Actions that are grouped together were always evaluated by different respondents.

### Table 2

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A1—Unfair</strong></td>
<td>A hardware store raises the price of snow shovels during a spring blizzard.</td>
</tr>
<tr>
<td><strong>A2—Fair</strong></td>
<td>A store maintains the price of a good when its costs drop, thus increasing its profits.</td>
</tr>
<tr>
<td><strong>B1—Unfair</strong></td>
<td>An employer who is doing well cuts wages by 5%, because many unemployed workers would be willing to work at the lower rate.</td>
</tr>
<tr>
<td><strong>B2—Fair</strong></td>
<td>An employer who is doing poorly cuts wages by 5%, to avoid or diminish losses.</td>
</tr>
<tr>
<td><strong>C1—Fair</strong></td>
<td>The worker replacing an employee who quit is paid less for doing the same job.</td>
</tr>
<tr>
<td><strong>C2—Fair</strong></td>
<td>An employer changes his line of business (from painting to landscaping), retains the same workers but reduces their wage.</td>
</tr>
<tr>
<td><strong>D1—Unfair</strong></td>
<td>A landlord raises the rent of a sitting tenant after learning that the tenant has found a good job in the area and will not want to move.</td>
</tr>
<tr>
<td><strong>D2—Fair</strong></td>
<td>Facing increasing costs, a landlord raises the rent of a sitting tenant who lives on a fixed income and may be forced to move.</td>
</tr>
<tr>
<td><strong>E1—Unfair</strong></td>
<td>A grocery store with a large stock of peanut butter raises its price immediately when it learns that wholesale prices have risen.</td>
</tr>
<tr>
<td><strong>E2—Fair</strong></td>
<td>A landlord who rents apartments in two identical buildings charges higher rent for one of them, because a more costly foundation was required for that building.</td>
</tr>
<tr>
<td><strong>F1—Unfair</strong></td>
<td>An employer cuts wages by 7% in a period of high unemployment and no inflation.</td>
</tr>
<tr>
<td><strong>F2—Fair</strong></td>
<td>An employer raises wages by 5% in a period of high unemployment and 12% inflation.</td>
</tr>
</tbody>
</table>

**Unfair**—A car dealer raises the price of a popular model above list price when a shortage develops.  
**Fair**—A car dealer eliminates a discount on a popular car model.

A surprisingly simple model, growing out of a conception of what counts as harm, adequately summarizes those intuitive judgments. In this model the relations between a firm and its transactors (customers or employees) are regulated by a reference transaction, which specifies a reference profit for the firm and reference terms (price, wage, rent) for the transactors. The firm has the power to impose a change of these terms
by posting new prices or by setting new wages; it is also free to maintain the reference terms of trade as circumstances change. Actions that abuse this power will be called unfair.

The reference transaction specifies what the transactors “have a right to expect” in their dealings with the firm. According to widespread moral intuitions, both sides have an informal entitlement to the terms of that transaction. The reference terms often correspond to the status quo. Indeed, the linking of entitlements to the reigning status quo is a powerful mechanism that transforms is into ought. However, states other than the status quo can be designated as the relevant reference. For example, a firm that labels a price as discounted or a part of the worker’s income as a bonus, reserves the right to limit its transactors’ entitlement to the ‘regular’ prices and wages.

A few critical distinctions are made. The first is between actions that the firm takes to protect its profits from an exogenous threat and actions that it initiates to increase its profit at someone’s expense. The principal constraint imposed by the intuitive principles of fairness is that a firm cannot fairly achieve a gain by exploiting its power to impose a loss on its transactors. This is the sense in which transactors have an entitlement to the reference terms. However, the entitlement is not unconditional. The firm is also entitled to its reference profit—the profit it achieves in the reference transaction—and this entitlement trumps the transactors’ entitlement. When its reference profit is threatened, it is not unfair for a firm to protect itself by passing on losses to transactors; the rules of fairness do not require the firm to absorb any part of an exogenously imposed loss. A second distinction is between losses and failures to gain. The rules of fairness do not require the firm to share with its transactors any gains that it obtains, whether by good fortune or by its efforts. For example, a large majority of respondents in a subsequent study believed that a firm has no obligation to pass on any reduction of wholesale prices to its customers. Of course, a firm that shares its gains with its customers or employees will be admired. But a firm that fails to do so will not evoke indignation.

As we might expect of judgments that arise from intuitions rather than from a rule, the moral judgments in Table 1 are specified at a strikingly low level of abstraction. For example, the reference price for a can of peanut butter appears to be linked to the price at which it was bought, rather than to the price at which it will be replaced (E2), and
the firm’s obligation not to cut wages evidently protects only the nominal wage rather than the workers’ real income (F1 vs. F2). The entitlement to the reference terms is also personal: it protects the current tenant or employee, but not their replacements.

It seems safe to predict that the factors that determine the fairness of market transactions will also prove to be critical in the broader context of social justice. In law as well as morality, the issue of affirmative action provides an instructive example. Hardly anyone will think it fair to achieve the goals of affirmative action by a policy in which current employees would lose their jobs to make room for replacements that would improve the representativeness of the work force. In fact, the Supreme Court of the United States has vindicated this intuition, suggesting that affirmative action programs are unacceptable when they impose losses on existing workers (Wygant v. Jackson Bd. of Educ., 476 US 267, 1986). On the other hand, a policy that discriminates among new hires is much more acceptable, because it appears to create only gainers and non-gainers, but no losers. As this example illustrates, the distinction between actual losses and opportunities foregone, however fragile after reflection, is psychologically and morally real. Indeed, the distinction appears to be categorical: the coefficient of loss aversion is much higher in the moral domain than in the domain of individual choice. System 2, of course, raises serious questions about the distinction between actual losses and opportunities foregone; but it is often silenced in the face of moral indignation.

Moral Framing

A framing effect is said to occur when two extensionally equivalent statements evoke different judgments or preferences when presented singly, yet appear transparently equivalent when shown together (Tversky and Kahneman, 1981). Framing effects arise because statements that are extensionally equivalent may nevertheless evoke different associations and different emotional responses: A cold cut described as 90% fat-free is more attractive than if it is described as 10% fat, and more likely to be purchased. Framing effects are a manifestation of the associative and emotional processes of System 1. There have been several demonstrations of framing effects in the domain of moral judgment.
Thomas Schelling has offered a compelling example, one that illustrates the difficulties that may arise when outcomes are evaluated as changes relative to an arbitrary reference level. Schelling reports asking his students to evaluate a tax policy that would allow a larger child exemption to the rich than to the poor. Many listeners will immediately respond as Schelling’s students do, that this proposal is outrageous. The intuition that leads to this quick conclusion is a simple and attractive rule: no social allocation of gains or losses should be more favorable to the rich than to the poor. Schelling next points out to his students that the default case in the standard tax table is a childless family, with special adjustments for families with children. Of course, the existing tax schedule could be rewritten with a family with two children as the default case. Childless families would then pay a surcharge. Should this surcharge be as large for the poor as for the rich? Of course not.

Schelling’s example is a framing effect, in which an inconsequential aspect of the statement of the problem appears to control moral judgments. The intuitions that are evoked by formulating the problem in terms of exemptions and surcharges are incoherent. Neither intuition survives when the inconsistency is pointed out. It soon becomes apparent that the only formulation of the tax problem that avoids arbitrariness is a complete table that determines the after-tax income of families for every level of pre-tax income and number of children. This description in terms of final states is superior, precisely because it avoids powerful intuitions that can be manipulated by the framing of the problem. Unfortunately, there are no clear moral intuitions left to resolve the problem when it is properly framed. In this case as well as in others the only potent moral intuitions apply to changes and to differences, not to states.

From money we turn to legal questions involving the valuation of injuries to health, and to an experiment in which the same difference between two states of health was caused to be coded either as a loss or as a gain. The experiment was concerned with lay assessments of appropriate monetary compensation for the pain and suffering associated with personal injuries, such as “losing mobility in one knee for four years” (McCaffery, Kahneman and Spitzer, 1994). Separate samples of respondents were given different “jury instructions” describing the thought experiment they should conduct to determine fair compensation. One of the instructions suggested a positive choice between
two desirable options. The respondents were instructed to imagine that the victim had very recently suffered the injury and was now offered a choice between a complete and immediate cure and an amount of money. Fair compensation was to be set at the highest amount for which the victim would still prefer the cure. In contrast, the selling instruction required the respondent to assume that the victim considered an ex ante proposition to accept the injury in conjunction with a payment of money. Fair compensation was to be set at the lowest payment for which the victim would have accepted the offer.

The difference between health and injury is coded as a gain in the former case and as a loss in the latter—this is the pattern of an endowment effect. In terms of final states, of course, the two versions of the problem are not distinguishable. As expected, the average judgment of fair compensation was about twice as high with the selling than with the choice instruction. This is also a framing effect, but unlike Schelling’s example the people who are susceptible to this effect rarely acknowledge its existence. When the participants in each experimental condition were shown the instruction given to the other group, they thought both instructions were fair and did not notice that they were likely to evoke discrepant responses. The legal system typically uses a version of the buying instruction, but without a great deal of reflective thinking about why that instruction should be preferred; and creative lawyers are able to frame the problem so as to ensure that a selling instruction comes before the jury, in a way that produces predictably higher dollar awards.

Moral framing has been demonstrated in the important context of obligations to future generations (see Frederick 2003), a much-disputed question of morality, politics, and law (Revesz 1999). To say the least, the appropriate discount rate for those yet to be born is not a question that most people have pondered, and hence their judgments are highly susceptible to different frames. From a series of surveys, Maureen Cropper and her coauthors (1994) suggest that people are indifferent between saving one life today and saving 45 lives in 100 years. They make this suggestion on the basis of questions asking people whether they would choose a program that saves “100 lives now” or a program that saves a substantially larger number “100 years from now.” It is possible, however, that people’s responses depend on uncertainty about whether people in the future will otherwise die (perhaps technological improvements will save them?); and
other ways of framing the same problem yield radically different results (Frederick 2003). For example, most people consider “equally bad” a single death from pollution next year and a single death from pollution in 100 years. This finding implies no preference for members of the current generation. The simplest conclusion is that people’s moral intuitions about obligations to future generations are very much a product of framing effects (for a similar result, see Baron 2000).

The same point holds for the question whether government should consider not only the number of “lives” but also the number of “life years” saved by regulatory interventions. If the government focuses on life-years, a program that saves children will be worth far more attention that a similar program that saves senior citizens. Is this immoral? People’s intuitions depend on how the question is framed (see Sunstein 2004). People will predictably reject an approach that would count every old person as “worth less” than what every young person is worth. But if people are asked whether they would favor a policy that saves 105 old people or 100 young people, many will favor the latter, in a way that suggests a willingness to pay considerable attention to the number of life-years at stake.

Framing effects present a large difficulty for the achievement of coherent judgments and preferences. The normal process of comprehension takes a given message to a state of the world, but the correspondence of messages and states is not one-to-one. Ambiguity arises when a single message is compatible with multiple states of the world. Framing effects arise when a single state of the world may be described in multiple ways, and when a relevant response is description-dependent. Thus, the avoidance of framing effects requires a search through the set of descriptions that are extensionally equivalent to the original message. Unfortunately, the human mind is not equipped to solve this problem. People are therefore given no warning, when confronted with a particular version of Schelling’s tax puzzle that there is an alternative version of the same problem to which they would have responded differently. In the context of punitive damage awards and valuation of environmental amenities, we have found that incoherence in moral judgments is predictable (Sunstein, Kahneman, Schkade, and Ritov, 2003). We leave to others the possible implications of these observations for the notion of reflective equilibrium.
Agency, Omission and Brutal Commission

To say the least, there has been much discussion of whether and why the distinction between acts and omissions might matter for morality, law, and policy. In one case, for example, a patient might ask a doctor not to provide life-sustaining equipment, thus ensuring the patient’s death. In another case, a patient might ask a doctor to inject a substance that will immediately end the patient’s life. Many people seem to have a strong moral intuition that the failure to provide life-sustaining equipment, and even the withdrawal of such equipment, is acceptable and legitimate—but that the injection is morally abhorrent. And indeed American constitutional law reflects judgments to exactly this effect: People have a constitutional right to withdraw equipment that is necessary to keep them alive, but they have no constitutional right to physician-assisted suicide (see Washington v. Glucksberg 1997, pp. 724–25). But what is the morally relevant difference?

It is worth considering the possibility that the act-omission distinction is rooted in System I, and is in some cases very hard to defend in principle. The moral puzzles arise when life, or a clever interlocutor, comes up with a case in which there is no morally relevant distinction between acts and omissions, but when moral intuitions strongly suggest that there must be such a difference. As an example, consider the vaccination question; many people show an omission bias, favoring inaction over statistically preferable action (Baron and Ritov 1993). The persistent acceptance of withdrawal of life-saving equipment, alongside persistent doubts about euthanasia, may be another demonstration of the point.

Compare the dispute over two well-known problems in moral philosophy (see Thomson 1986, pp. 94–116). These problems do not involve the act-omission distinction; no omission is involved. But the problems implicate closely related concerns. The first, called the trolley problem, asks people to suppose that a runaway trolley is headed for five people, who will be killed if the trolley continues on its current course. The question is whether you would throw a switch that would move the trolley onto another set of tracks, killing one person rather than five. Most people would throw the switch. The second, called the footbridge problem, is the same as that just given, but with one difference: the only way to save the five is to throw a stranger, now on a footbridge that
spans the tracks, into the path of the trolley, killing that stranger but preventing the trolley from reaching the others. Most people will not kill the stranger; in fact they are indignant at the suggestion that they ought to do so. But what is the difference between the two cases, if any? A great deal of philosophical work has been done on this question, much of it trying to suggest that our firm intuitions can indeed be defended in principle.

Without engaging these arguments, let us suggest the possibility of a simpler answer. As a matter of principle, there may or may not be a difference between the two cases. But people’s different reactions are based on automatic moral intuitions that condemn the throwing of the stranger but support the throwing of the switch. As a matter of intuition, it is worse to throw a human being in the path of a trolley than to throw a switch that (indirectly?) leads to a death. People also struggle heroically, and by reference to System 2, to rescue their intuitions and to establish that the two cases are genuinely different in principle. whether or not this is so, But system 1, and indignation about brutal acts of commission, are responsible for the underlying intuitions.

Consider a suggestive experiment designed to see how the human brain responds to the two problems (Greene et al. 2001). The authors do not attempt to answer the moral questions in principle, but they find “that there are systematic variations in the engagement of emotions in moral judgment,” and that brain areas associated with emotion are far more active in contemplating the footbridge problem than in contemplating the trolley problem. An implication of the authors’ finding is that human brains are hard-wired to distinguish between bringing about a death “up close and personal” and doing so at a distance. It follows that acts, especially brutal acts, would be far more likely to produce reactions from the brain areas associated with emotions than omissions that cause identical harms. Compare the case of fear, where an identifiable region of the brain makes helpfully immediate but not entirely reliable judgments (Ledoux 1998), in a way that suggests a possible physical location for some of the operations of System I. So too, we think, in the context of morality, politics, and law (Greene & Haidt 2002). A clear implication involves moral numbness: Many omissions do not trigger indignation on the part of System 1, but might well be subject to moral criticism from the standpoint of System 2, if only it can become sufficiently active.
Conclusion

Moral intuitions operate in much the same way as other intuitions do; what makes the moral domain is distinctive is its foundations in the emotions, beliefs, and response tendencies that define indignation. System I is typically responsible for indignation; System II may or may not provide an override. Moral dumbfounding and moral numbness are often a product of moral intuitions that people are unable to justify. We have argued that an understanding of indignation helps to explain the operation of the outrage heuristic, the centrality of harm, the role of reference states, moral framing, and the act-omission distinction. We have also suggested that because of how indignation operates, it is extremely difficult for people to achieve coherence in their moral intuitions.

The intuitions described here play an important role in multiple domains, including families, workplaces, sporting events, and religious organizations. But as many of our examples suggest, they also influence the decisions of legal and political institutions. Such institutions usually aspire to be deliberative, and to pay close attention to System II; but even in deliberative institutions, System I can make some compelling demands.
References


Readers with comments should address them to:

Professor Cass Sunstein
University of Chicago Law School
1111 East 60th Street
Chicago, IL  60637
csunstei@midway.uchicago.edu
For a listing of papers 1–200 please go to Working Papers at http://www.law.uchicago.edu/Lawecon/index.html

201. Douglas G. Baird and Robert K. Rasmussen, Chapter 11 at Twilight (October 2003)
205. Lior Jacob Strahilevitz, The Right to Destroy (January 2004)
208. Richard A. Epstein, Disparities and Discrimination in Health Care Coverage; A Critique of the Institute of Medicine Study (March 2004)
209. Richard A. Epstein and Bruce N. Kuhlik, Navigating the Anticommons for Pharmaceutical Patents: Steady the Course on Hatch-Waxman (March 2004)
213. Luis Garicano and Thomas N. Hubbard, Specialization, Firms, and Markets: The Division of Labor within and between Law Firms (April 2004)
216. Alan O. Sykes, The Economics of Public International Law (July 2004)
225. Christine Jolls and Cass R. Sunstein, Debiasing through Law (September 2004)
228. Kenneth W. Dam, Cordell Hull, the Reciprocal Trade Agreement Act, and the WTO (October 2004)
230. Lior Jacob Strahilevitz, A Social Networks Theory of Privacy (December 2004)
231. Cass R. Sunstein, Minimalism at War (December 2004)
238. Randal C. Picker, Copyright and the DMCA: Market Locks and Technological Contracts (March 2005)
239. Cass R. Sunstein and Adrian Vermeule, Is Capital Punishment Morally Required? The Relevance of Life-Life Tradeoffs (March 2005)
240. Alan O. Sykes, Trade Remedy Laws (March 2005)
250. Lior Jacob Strahilevitz, Exclusionary Amenities in Residential Communities (July 2005)
255. David A. Weisbach, Pareto Intergenerational Discounting (August 2005)
257. Adrian Vermeule, Absolute Voting Rules (August 2005)
258. Eric Posner and Adrian Vermeule, Emergencies and Democratic Failure (August 2005)
260. Adrian Vermeule, Reparations as Rough Justice (September 2005)
262. Adrian Vermeule, Political Constraints on Supreme Court Reform (October 2005)
264. Lior Jacob Strahilevitz, Information Asymmetries and the Rights to Exclude (November 2005)
265. Cass R. Sunstein, Fast, Frugal, and (Sometimes) Wrong (November 2005)
266. Robert Cooter and Ariel Porat, Total Liability for Excessive Harm (November 2005)
273. Cass R. Sunstein, Burkean Minimalism (January 2006)
274. Cass R. Sunstein, Misfearing: A Reply (January 2006)
275. Kenneth W. Dam, China as a Test Case: Is the Rule of Law Essential for Economic Growth (January 2006, revised October 2006)
316. Ariel Porat, Offsetting Risks (November 2006)
324. Wayne Hsiung and Cass R. Sunstein, Climate Change and Animals (January 2007)
332. Cass R. Sunstein, If People Would Be Outraged by Their Rulings, Should Judges Care? (February 2007)
334. Eugene Kontorovich, Inefficient Customs in International Law (March 2007)
335. Bernard E. Harcourt, From the Asylum to the Prison: Rethinking the Incarceration Revolution. Part II: State Level Analysis (March 2007)
337. Adam B. Cox and Thomas J. Miles, Judging the Voting Rights Act (March 2007)
338. M. Todd Henderson, Deconstructing Duff & Phelps (March 2007)
341. Anup Malani, Valuing Laws as Local Amenities (June 2007)
342. David A. Weisbach, What Does Happiness Research Tell Us about Taxation? (June 2007)
344. Christopher R. Berry and Jacob E. Gersen, The Fiscal Consequences of Electoral Institutions (June 2007)