ISSN 1045-6333

HUMAN NATURE AND THE BEST CONSEQUENTIALIST MORAL SYSTEM

Louis Kaplow
Steven Shavell

Discussion Paper No. 349

02/2002

Harvard Law School
Cambridge, MA 02138

The Center for Law, Economics, and Business is supported by
a grant from the John M. Olin Foundation.

JEL Classes D11, D62, H00, K00

Human Nature and the Best Consequentialist Moral System

Louis Kaplow and Steven Shavell*

Abstract

In this article, we ask what system of moral rules would be best from a consequentialist perspective, given certain aspects of human nature. This question is of inherent conceptual interest and is important to explore in order better to understand the moral systems that we observe and to illuminate longstanding debates in moral theory. We make what seem to be plausible assumptions about aspects of human nature and the moral sentiments and then derive conclusions about the optimal consequentialist moral system — concerning which acts should be deemed right and wrong, and to what degree. We suggest that our results have some correspondence with observed moral systems and also help to clarify certain points of disagreement among moral theorists.

*Harvard Law School and National Bureau of Economic Research. We thank Christopher Kutz and participants in a workshops at the University of California at Berkeley and Harvard University for comments, Anurag Agarwal, Anthony Fo, Felix Gilman, Allon Lifshitz, Christopher Lin, Damien Matthews, Jeff Rowes, Laura Sigman, and Leo Wise for research assistance, and the John M. Olin Center for Law, Economics, and Business at Harvard Law School for financial support.

Human Nature and the Best Consequentialist Moral System

Louis Kaplow and Steven Shavell

I. INTRODUCTION

What system of moral rules is best from a consequentialist perspective? That is, if one seeks to induce individuals to behave in a manner that maximizes a conception of social welfare, or best promotes some other objective, how does one determine which acts should be deemed right and wrong — which acts should lead individuals to feel guilty and be subject to social disapprobation and which acts should lead individuals to feel virtuous and to experience praise — and to what extent?[1] In addressing these questions, we will take into account how the answers are influenced by certain aspects of human nature, notably the difficulty of inculcating morality, limits on individuals' capacity to experience moral sentiments, such as feelings of guilt and virtue, and individuals' tendency to organize their understanding of the world and their guiding principles into categories that group together related but not identical situations.

We examine these questions for a number of reasons. One is the inherent interest in explaining what we observe. In human societies, we find that there exist moral rules — categories of acts viewed as right or as wrong, various prohibitions, permissions, duties, obligations, and so forth — and that these rules vary in strength. As individuals, we have related moral instincts and intuitions that are revealed by contemplation of actual or hypothetical situations that we may confront. Correspondingly, we have moral sentiments — emotions of sorts — and we are also subject to external expressions of moral approval and disapproval. For example, we might feel guilty if we tell a lie, even if no one finds out about it; we might be ashamed if others do find out; and we might be subject to painful disapprobation if others find out and respond. Taken together, these interdependent phenomena constitute an elaborate social and psychological system. It is reasonable to consider whether this system might serve instrumental purposes. Indeed, consequentialists such as Hume (1739, 1751), Mill (1861), and Sidgwick (1907) advanced the view that society's moral rules and individuals' moral sentiments tend to promote welfare.[2]

---

[1]As should be clear from the context, our use of the concept of "virtue" refers to feelings or moral sentiments — the positive analogue to feelings of guilt — and thus is distinct from the use of the term in virtue ethics.

[2]Consider, for example, the following statements. "In all determinations of morality, this circumstance of public utility is ever principally in view. . . ." (Hume, 1751: 81.) "Whatever be the origin of our notion of moral goodness or excellence, there is no doubt that 'Utility' is a general characteristic of the dispositions to which we apply it: and that, so far, the Morality of Common Sense may be truly represented as at least unconsciously

We also observe that the moral system exhibits both important similarities and differences across cultures and over time. It is particularly challenging to attempt to explain the variation. In doing so, instrumental accounts should be considered because, if the moral system is functional, its features will depend upon the particular circumstances of a society. Such an inquiry, it should be noted, is descriptive, social scientific in character: Particular circumstances are observed, perhaps by anthropologists, sociologists, psychologists, or philosophers, and analysts seek to determine what sort of functionalist explanation might be offered for it.

Our inquiry also has normative relevance. It is intellectually interesting to ask — from a consequentialist perspective, just as from various other perspectives — what would be the best moral system if one were free to design a moral system for society. This question is of conceptual interest regardless of whether the existing moral system in one or another society is consequentialist or is ideal under any other standard. Moreover, the answer to this question can be used critically, as a guide for reform, if one accepts the consequentialist normative view and believes that there is a plausible way to go about implementing it; but these matters do not concern us here.

We also are motivated to investigate an ideal consequentialist moral system because the analysis of such a system may help to illuminate strands of the longstanding debate between consequentialists and deontologists. As mentioned, prominent consequentialists have advanced the claim that our moral system promotes welfare. But their view has proved to be controversial, in part because of the existence of situations in which the implications of consequentialism (often, utilitarianism) seem to conflict with our moral intuitions and with many societies' moral codes. For example, if the benefit from breaking a promise precisely equals any harm that would be caused, a consequentialist would be indifferent to whether the promise is broken, but a deontologist would hold that there is at least some moral weight in favor of keeping the promise. Another type of example involves the apparent requirement of utilitarianism that individuals who have the means to do so should expend a substantial portion of their resources to help others — such as destitute individuals in far-off lands — yet in fact the failure to engage in such saintly behavior is not generally taken to be morally wrong, subjecting the transgressor to disapprobation. As a further illustration, consequentialist approaches typically do not distinguish between acts and omissions whereas our moral rules and intuitions often do. Examples such as these are offered by Ross (1930), Williams (1973, 1981), and many others as critiques of utilitarianism. To be sure, there have been responses, such as by Smart (1973) and Hare (1981), but the matter remains in dispute. Furthermore, it seems that an important aspect of the disagreement concerns what system of morality is in fact entailed by consequentialism, but the answers given by participants on both sides of the debate tend to have an ad hoc character.

The criticisms of consequentialism associated with the foregoing examples are, we believe, symptomatic of a more fundamental shortcoming in leading consequentialist accounts of moral rules and

---

Utilitarian." (Sidgwick, 1907: 424.)

related moral instincts and intuitions.[3] Namely, consequentialists generally have not systematically elaborated how an ideal moral system should be specified; instead, they have tended to be reactive, offering rationalizations of existing moral rules or responses to particular conundrums put forward by critics.[4] For example, consequentialists sometimes invoke various assumptions about human nature to explain certain imperfections in the moral system or to make sense of particular, problematic examples. Yet, no matter how plausible such arguments are in a given context, one is left wondering whether the consequentialist's assumptions are employed consistently across contexts, and, more fundamentally, what would be the conclusions if one thoroughly investigated the assumptions' implications. In addition, consequentialists often merely identify a tendency for a given moral rule to advance welfare without deriving what would be the best way to design the moral system in order to address the range of behavior in question. Furthermore, consequentialist explanations often take as given identified catalogs of virtues and vices — corresponding to permitted behavior that is commendable and to prohibited behavior. As a result, they do not ask why, say, a given act (such as helping others) deemed to be a virtue should be viewed as such, rather than instead seeing the opposite act (perhaps an omission, failing to help others) as a vice, subject to moral prohibition. This limitation may, for instance, help to explain the difficulty consequentialists have in making sense of the act/omission distinction.

We believe that both the general questions that naturally arise about moral systems and the particular aforementioned disagreements among contending philosophical camps can be illuminated by a direct inquiry concerning what the best consequentialist moral system would look like. We advance this view both on a priori grounds and because we believe that our very preliminary exploration yields some useful insights. Their importance is for the reader to judge, but we hope that in any event our discussion will provoke further study of these questions.

We proceed as follows. In section II, we present the assumptions that we will employ in our analysis. There are four main elements: the influence of a moral system on individuals' behavior, the costs involved in inculcating morality, certain limits on individuals' capacities to be motivated by moral concerns, and the grouping of related (but not identical) situations for purposes of applying moral rules. Each of these elements, as will be seen, involves assumptions about human nature. We draw upon various literatures to motivate our particular assumptions, but their approximate truth is fundamentally an empirical matter.[5]

---

[3]The deficiency that we identify is not, of course, equally present in all work by consequentialists. For example, Brandt (1996) inquires directly into the features of an optimal moral system.

[4]Regarding descriptive shortcomings of consequentialism, even if analysts derive the best consequentialist moral system, it will not necessarily be reflected in a society's (ours or another's) moral code, as we discuss intermittently below.

[5]Scholarship in many fields, including psychology, sociology, anthropology, economics, neurobiology, sociobiology, and philosophy, has addressed the role of moral concerns (sometimes described as social norms) in regulating behavior. Early writers include Hume (1739, 1751), Smith (1790), Darwin (1872, 1874), and Sidgwick (1907). Among modern writers are Alexander (1987), Barkow, Cosmides, and Tooby (1992), Baron (1994), Becker (1996), Ben-

In section III, we use these assumptions in determining how a moral system might best be designed to advance welfare.[6] Not surprisingly, our analysis is consistent with the general thrust of prior consequentialists, who indicate that it tends to be advantageous to employ a moral system to encourage (reward) socially desirable behavior and to discourage (punish) socially undesirable behavior. However, because moral rules tend to be categorical (and, as we will discuss, must be so, on account of human nature), the correspondence will often be rough. For example, it is possible that particular acts that would in fact be socially desirable will be deemed wrong — with the result that individuals committing such acts will feel guilty and be subject to disapprobation or that some individuals will, because of the anticipation of such consequences, abstain from committing acts that may both advance their own (narrow) self-interest and be socially desirable. Relatedly, some undesirable acts may be deemed right, or at least permissible, and accordingly may be committed. In addition, various acts deemed right and that are in fact socially desirable may be abstained from, and some wrong, undesirable acts may be committed, in spite of the dictates of the moral code. These conclusions from our analysis will be reminiscent of rule-utilitarian claims and, more generally, those associated with two-level moral theories.

Another set of conclusions concerns the domains of virtues and of vices in a system of common morality. As previously noted, most consequentialist analysis is silent on the question whether a socially desirable act should be deemed right or failing to commit it should be deemed wrong. We analyze the issue and find that it tends to be advantageous to deem socially undesirable acts (or omissions) to be wrong when most individuals can be induced to comply with such a moral command. Thus, cutting in line or punching someone who is rude should be deemed wrong, rather than abstention being viewed as virtuous, because the prospect of guilt feelings and disapprobation will be sufficient to deter most individuals from committing these acts. Likewise, we find that it tends to be beneficial to deem socially desirable acts (or omissions) to be virtuous, worthy of praise, when few individuals can be induced to commit them (for example, rescuing another at great risk to oneself or devoting one's life to helping others), rather than treating abstention as wrongful behavior. The reason for these conclusions, as we will explain, derives from certain limitations of human nature.

In section IV, we relate the conclusions in section III to the observed use of guilt and virtue in our moral system. In section V, we elaborate on our argument, with particular attention to the

---

Ner and Putterman (1998), Campbell (1975), Daly and Wilson (1988), Damasio (1994), Elster (1999), Frank (1988), Gibbard (1990), Hechter and Opp (2001), Izard (1991), Kagan (1984), LeDoux (1996), Mackie (1985), Pinker (1997), Trivers (1971), E.O. Wilson (1975), and J.Q. Wilson (1993).

[6]Obviously, consequentialism is broader than welfarism; we focus on welfare for concreteness, because most consequentialists are, at least in part, welfarists, and because a plausible descriptive account of social evolution would in some manner be related to individuals' well-being. In this preliminary inquiry, further refinement seems inappropriate, but we will on occasion note how different maximands may be pertinent in interpreting our results.

In addition, when speaking of maximization, we often will use the language of addition (as under utilitarianism); here, the motivation is concreteness.

assumptions that we introduce in section II. In particular, we reconsider the grouping of acts into moral categories, examining the possibility of exceptions to moral rules and also overlapping, and thus potentially conflicting, moral rules. We examine how our analysis is affected by the extent to which moral systems are products of biological evolution rather than socialization. We distinguish internal and external aspects of sanctioning for norm violation — for instance, whether and how it matters that one may experience guilt rather than or in addition to being subject to disapprobation for wrongful behavior. We address the importance of heterogeneity among actors (for example, in the extent to which they have internalized a society's moral code) in understanding moral systems that we observe. And finally, we comment on how rules of prudence — which purport to address self-regarding behavior — relate to our discussion of morality — which largely focuses on behavior that affects other individuals.

In section VI, we draw on our analysis to explore certain disputes in moral philosophy. Specifically, we discuss the relationship between self-interest and moral motivation, the independent importance of acting morally (that is, independent of any consequences of doing so), the use in debates about consequentialism of counter-examples that draw on our moral intuitions, the problem of unlimited individual obligation under consequentialism, and the act/omission distinction. Although we do not attempt to resolve these disputes, we do believe that some commonly-offered arguments may be illuminated by our analysis.

Before we begin, it may be helpful for us to remark on what positions we are and are not advancing in this article. We are arguing that it is important to ask the questions that we raise and to attempt to answer them with the general approach that we employ. More tentatively, but not without force, we believe that our most general conclusions probably have an important element of truth, both as a purely conceptual matter (identifying some of the features of an ideal consequentialist moral system) and as a descriptive matter (indicating some of the rough contours of observed moral systems). Nevertheless, our inquiry is preliminary. Our results depend in varying degrees on our assumptions about human nature, and these are subject to reasonable disagreement and, more importantly, to further understanding and revision in light of subsequent research in other disciplines. Moreover, in using our analysis of an ideal moral scheme to examine any actual moral system, we are both extrapolating from a simple model and making the implicit assumption — which we in fact believe to be only partially valid — that social forces will lead to optimality. A more complete analysis would require attending to questions about evolutionary processes, social and biological, and developing more precise ways of identifying what moral system actually exists in a given society.

Finally, because this article is almost entirely conceptual and descriptive, it should not be interpreted as advancing any view about what society should seek to accomplish in formulating or inculcating moral principles.[7] With respect to normative matters, our only claim is that debates between consequentialists and deontologists can be illuminated — not resolved — by having a better understanding of what a consequentialist system of morality would entail.

[7]In Kaplow and Shavell (2002), we advance a welfarist view.

II. ASSUMPTIONS

In this section, we present and provide some justification for the four main assumptions that we will use in our analysis. They are as follows: First, individuals can be motivated in some manner (at least some of the time) to follow moral rules, even when doing so is against their narrow self-interest. Second, there exists some process of instilling moral rules in individuals and establishing them in a society, although this process is socially costly. Third, there are limits on the extent to which individuals can be induced to behave morally; for example, most individuals could not be induced constantly to make great sacrifices for others, even if that is what a moral code instructed them to do. Fourth, there are limits on the manner in which moral rules can be formulated; in particular, they must, to an extent, be general (categorical) in character.

Ultimately, all of the assumptions that we present in this section are based on empirical conjectures about human nature, largely within the provinces of sociology, psychology, and neurology. In each instance, the assumptions that we make will be fairly general in character and thus not dependent on the resolution of many existing areas of uncertainty in these fields. Nevertheless, what we say is hardly beyond dispute. In any event, we will attempt to make clear in our analysis how various conclusions depend on our assumptions so that those who disagree with one or more of our starting points will have some idea of what destination would be reached had we taken a different path.

A. Influence of Morality on Behavior

Our central assumption is that moral considerations affect individuals' behavior. Thus, for example, an individual who would otherwise tell a lie because it would advance his self-interest, narrowly construed, might abstain if lying is morally prohibited. We wish to elaborate this point along a number of dimensions.

First, why might individuals deviate from their narrow self-interest in order to be moral? This question, emphasized by Hume (1739, 1751) and others, has occupied a wide range of thinkers, and we do not purport to offer any novel insights into the matter. Our assumption will be that individuals are motivated to follow moral rules because of what might be referred to as moral rewards and punishments, associated with moral emotions.[8] See, for example, Izard (1991). These may be purely internal: An individual may be deterred from wrongful behavior because he would otherwise feel guilty (regardless of whether others would learn of his misdeed); likewise, an individual may be motivated to do the right thing because doing so would make him feel virtuous. Incentives may be external:

[8]Our use of the term "moral emotions" may seem inapt to some. Little of our argument depends on whether the phenomena under discussion are emotions per se. Nevertheless, we note that there is a substantial psychological literature, some of which we cite throughout, that identifies guilt and related notions as emotions. See generally Haidt 2001). Moreover, there is direct evidence from brain scans showing that many classic moral dilemmas primarily activate emotional areas of the brain. See Greene et al. (2001).

Individuals may be deterred from wrong and induced to behave righteously by the prospect of disapprobation and approbation from others.[9] There are also mixed cases, such as when an individual would feel ashamed if others knew of his misdeed, without the others having to take any action to express their disapproval. We find it convenient to focus on — and write using the language of — the internal sanctions involving feelings of guilt and virtue; much of what we say, however, is more broadly applicable, and we will return to the distinctions among types of moral rewards and sanctions below.

More precisely, we assume that individuals' behavior is determined by a weighing of narrow self-interest against moral rewards and punishments. Thus, in a given situation, if an act greatly advances narrow self-interest and is associated only with slight feelings of guilt (whether because the violation is small or because the rule being violated is not a very serious one), individuals will tend to commit the act. Contrariwise, if the degree to which the act serves individuals' narrow self-interest is small, but guilt feelings would be great, individuals will tend to refrain from the action. (We sometimes speak explicitly of tendencies in order to emphasize that not all individuals are alike and even a single individual may not always be consistent, matters we discuss further below; to simplify, however, we will sometimes state that individuals will, or will not, commit a given act.)

Our claim that individuals decide how to behave based upon a balancing of narrow self-interest and moral rewards and punishments may strike some readers as constituting an impoverished or inaccurate account of why individuals often act in accordance with moral rules. A contrary view of the nature of individuals' moral decisionmaking and behavior is advanced by Kant (1785) and has been pursued by many others.[10] But our characterization can be given a number of interpretations that are consistent with seemingly different (and, to some, more plausible) accounts without fundamentally affecting the analysis to follow.

We first state the most straightforward interpretation (that some may reject), which is that guilt and virtue are elements of individuals' utility, just like other sources of pain and pleasure, and individuals simply act to as to maximize the excess of pleasure over pain. Note that this formulation does not require that guilt and virtue be qualitatively similar to other sources of disutility and utility or that they are produced by the same mental process as other sources. The analysis only requires that guilt and virtue are *a* source of something that can broadly be referred to as disutility or utility, that they are produced by *some* mental process, and that individuals ultimately make decisions reflecting the relative weights of

---

[9]Relatedly, individuals may understand, for example, that it is against their narrow self-interest to behave badly if this would lead others — also motivated purely by narrow self-interest — not to deal with them in the future, perhaps because certain misdeeds reveal their perpetrators to be untrustworthy. In this article, however, we focus solely on considerations related to the morality of behavior per se.

[10]See, for example, Smith's (1790) discussion of Mandeville and Hobbes, and also Hutcheson's (1725-1755) attempts to distinguish acts based on self-interest from those based on obligation or benevolence. For modern examples, see Anderson (2000), Foot (1972), Scheffler (1992), Sen (1977), and Woods (1972).

different sources of utility, as suggested by the earlier examples.[11]

Many would suggest, however, that individuals' conscious experience of moral decisionmaking is different: Individuals do not weigh and balance different sources of utility, but rather reflect on whether contemplated acts would be right or wrong and act accordingly. (For example, individuals' thinking may be of a Kantian sort or reflect an acceptance of divine authority.) This view, when amplified in ways that seem plausible, turns out to have fairly similar implications (at least for present purposes) to those of the utility-based view just mentioned.

The main reason for this similarity is that most individuals do not treat morality as absolute. There may be conflicts among moral rules, there may be exceptions (or individuals may, on the spot, create exceptions), and individuals are not moral saints. Moreover, regardless of these considerations, it is acknowledged that moral rules vary in their strength or importance and thus in the weight that individuals would give them. For example, few individuals who view themselves as moral would sacrifice their careers or the future prospects of their children to avoid slight violations of most moral rules. At the same time, many such individuals would abstain from acts that produce only modest personal benefits while involving serious moral violations. Thus, it is apparent that individuals who understand themselves as making decisions based on their sense of right and wrong will tend to act in a manner that reflects a comparison of the magnitude of their personal gains and the importance they associate with any pertinent moral rules.

Although the foregoing observations are sufficient to justify our assumption, it is also interesting to explore the matter further by asking why it is that individuals, if they are not assumed to engage in an explicit balancing of narrow self-interest and moral rewards and sanctions, are motivated to follow moral rules at all. If the answer is not that considered previously — that there is some direct force, a species of utility or something akin to utility — there must be some sort of indirect force (and its indirectness would help to explain why individuals may not experience their moral decisionmaking as involving a pure utility calculus).

One possibility, which seems plausible as a matter of human psychology, is that moral behavior can become a habit, much as do a wide range of other sorts of behaviors (ranging from brushing one's teeth to saying "thank you").[12] Having internalized the rule that one should not steal or tell lies, one may simply no longer think about it on most occasions. Perhaps as a child, the rule was inculcated using rewards and punishments of various sorts, but they are no longer so necessary. Of course, habits are not entirely rigid and, given sufficient contrary motivation, habits will not be followed. As temptation

---

[11]This point implies that some seemingly different views, such as Scheffler's (1992) endorsement of a Freudian account under which guilt is a punishment deriving from the superego, need not be distinguished from a standard utility-based interpretation for present purposes.

[12]This view of why individuals behave morally is advanced, among others, by Hume (1751), Mill (1861), and Darwin (1874).

increases, it may be that the feelings of guilt and virtue associated with immoral and moral behavior will be important in inducing individuals to follow their moral habits, if that is what they are. And if the potential gain from deviation is quite large and an individual does not in fact feel guilty about breaking the habit, compliance with the moral rule would be less likely.

Another (related) interpretation is that feelings of guilt and virtue stand behind moral rules generally even though these moral rewards and sanctions are not explicitly contemplated. One way to express this idea is that individuals *explicitly* associate various acts with degrees of compliance with or violations of their moral code. Then, if the moral consideration is sufficiently serious, it will control behavior. What remains *implicit* is that there is a further association between compliance with or violations of individuals' moral codes and experiencing feelings of virtue and guilt. This latter, tacit relationship is necessary to understand why individuals tend to behave morally, but it is not required that individuals routinely think in such terms in deciding how to behave.[13]

In sum, our assumptions with regard to individuals' behavior can be restated follows: First, individuals are led, at least to some extent, to follow moral rules. Second, that to be so led, there must be some sort of motivation that is (sometimes) sufficient to counter other motivations, notably, narrow self-interest. Third, whatever is the precise internal, conscious experience of moral decisionmaking (and it may well vary among individuals and for a single individual in different situations), we can simply use the terminology of guilt and virtue to describe whatever it is that leads individuals to follow moral rules. In this respect, our claim is more of a tautology than an assumption. (Perhaps because we are both economists, we find most felicific a choice of language that is suggestive of the initial interpretation, under which individuals are imagined to engage in an express calculation based on competing sources of utility, those reflecting narrow self-interest and those based on the moral emotions or sentiments, but hopefully it is clear at this point that there is substantial overlap among competing descriptions of the decisionmaking process at issue.[14] In any event, we further examine whether there is a distinction between acting from self-interest and acting from moral obligation below.)

---

[13]This idea can be expressed formally. Suppose that the prospect of feeling guilt, denoted g, underlies moral violations, the moral weight of which is indicated by m. The implicit relationship between g and m could be indicated by the functional expression $m(g)$. Now decisions d could be related to personal benefit or cost (utility) u and moral weight m, using the functional expression $d(u,m)$. This expression can be more fully stated as $d(u,m(g))$, reflecting that m is implicitly a function of g. Now, to complete the argument, one can simply define an alternative decision function $d\mathbb{N}$, such that $d\mathbb{N}(u,g) = d(u,m(g))$. That is, given any u and g, one can find the corresponding u and m using the function $m(g)$, use those values to determine the decision d, and assign that same decision to the constructed function $d\mathbb{N}$. In this manner, stating that decisions depend on u and m is formally equivalent to stating that decisions depend on u and g. (One might therefore say that individuals who decide based on u and m behave as if they decide based on u and g, and vice versa.)

[14]For example, in the following subsection, when we speak of inculcating a certain level of guilt, one could interpret our statement as referring to the inculcation of a moral rule of a given degree of importance.

B. Inculcation of Morality

We assume that it is possible to inculcate moral rules and, relatedly, to do so in a manner that leads individuals to feel a specified level of guilt for violating various moral rules or virtue for following them. This assumption is both strong and innocuous. It is strong (and unrealistic) because there is obviously no person or entity (except possibly a divine one) with such a capacity. To the extent that moral rules are inculcated in a society and similarly transmitted across generations, the inculcation and transmission is done by parents, teachers, religious figures, peers, governments, and so forth; these processes are imperfect and possibly conflicting. And to the extent that moral rules arise through biological evolution, there is at best only an imperfect, invisible hand guiding the process of natural selection.

Our assumption is benign given that our main purpose here is conceptual. Our central question concerns what a system of moral rules would be, and what use would be made of guilt and virtue (and approbation and disapprobation, and so forth), *if* the moral system were designed entirely on consequentialist grounds (maximizing welfare, for concreteness). Thus, the present assumption merely corresponds to stating the necessary hypothetical premise. Of course, our society — and, it would seem, all others that have been studied — does have a moral system, so there is some factual basis for the assumption that such a system may somehow be created (whether through social evolution, divine intervention, or otherwise). We will return later to the question of how the source of a moral scheme matters for descriptive purposes.

We further assume that the inculcation process is costly. Specifically, to inculcate, say, the rule that lying is wrong and the corresponding tendency to feel guilty for telling lies — or virtuous for telling the truth — involves a cost. Moreover, we suppose that this cost is greater the stronger is the moral rule, that is, the more guilt or virtue is to be inculcated in association with the rule. The motivation for this assumption is that, whatever inculcation process one imagines, there will be time, effort, and other costs associated with inculcating morality. For example, parents cannot simply snap their fingers and have their children absorb a moral code; rather, there must be instruction, the offering of examples, uses of reward and punishment, and so forth. Furthermore, because time with one's children is finite, there is an opportunity cost: The more effort parents spend inculcating one moral rule, the less time they will have for other endeavors — whether teaching their children to read, going on picnics, or inculcating other moral rules.[15]

[15]Hooker (2000) emphasizes the importance of considering inculcation costs in determining what moral rules consequentialism requires. We further note that even to the extent that some moral rules arise from evolution in the biological sense, there is a sort of scarcity in natural selection, so that if significant selection is occurring on some dimensions, selection on other dimensions will tend to be less sharp; hence, just as when inculcation costs are taken into account, there would be a preference for moral rules that control more important behaviors (those having greater effects and those that arise more frequently).

C. Limits on the Force of Morality

It is not just that inculcation is costly, but it also seems true that there are limits on the extent to which most individuals can be influenced by morality. That is, there are limits to individuals' capacities to experience (directly or indirectly) feelings of guilt and virtue, to be moved by concerns about disapprobation and approbation, and so forth. Thus, we suppose that a person cannot feel boundless guilt or virtue all of the time. If someone already feels terribly guilty from one misdeed, there may be less room to feel guilty from another. With regard to internal feelings, emotions of sorts, it is a regularity of human psychology that reaction to a common type of stimulus diminishes with repetition, a numbing effect.[16] Likewise, it is familiar that one type of reaction can crowd out others: If one's foot is smashed, one no longer notices the pain from a previously bruised elbow. With regard to external moral rewards and sanctions, it seems that there are limits on how much effort third parties can expend meting out praise and blame and on how effective such praise and blame can be. Concerning the latter, if everyone, or even a single individual, is constantly praised, the impact of such praise will presumably diminish.

In sum, we assume that the effectiveness of those forces that potentially motivate individuals to behave morally is subject to diminishing returns. Moreover, we suppose that such limits affect not only single types of moral behavior but also, to an extent, moral behavior as a whole. Thus, if one already feels tremendously virtuous and has already received unending praise from doing a certain type of good deed, further feelings of virtue and approbation have less ability to motivate similarly large sacrifices to do another type of good deed.[17]

D. Generality of Moral Rules

Our final assumption is that moral rules must have some degree of generality. Thus, a moral rule might prohibit all lies, or perhaps all lies of a given type or in a given context. But it is assumed to be impossible to inculcate moral rules of unlimited specificity; thus, a different moral rule — with a different associated level of guilt or virtue — could not be instilled for every conceivable situation in which an individual might have the opportunity to tell a lie.

Although there is room for disagreement about the extent to which this assumption accurately characterizes what can be inculcated in a particular individual or as a part of a society's moral code, it seems clear that nontrivial limits on specificity exist. Since the point is familiar, we merely sketch some

---

[16]See, for example, Frederick and Loewenstein (1999). We note that, although there is a substantial regularity to the tendency of reactions to stimuli to fall as the stimuli are repeated, there are some exceptions.

[17]We acknowledge that there could be exceptions. Individuals vary and some, perhaps never having done a good deed before, may not have appreciated how good it makes one feel. Our assumption is merely that, for most individuals in most circumstances, there comes a point of diminishing marginal effectiveness of guilt and virtue and of disapprobation and approbation.

of the standard reasons.[18] First, there are limits on what any individual can be taught, reflecting limits on precision in our language, in time available, in mental processing ability, and so forth.[19] Moreover, these limits assume particular importance given that a significant part of moral education (and probably the most effective part) takes place in early childhood. Second, there are limits on individuals' ability to apply complex rules successfully. Complexity adds to decision costs, requires information that may not be available, and itself produces error. Third, although error can be unmotivated, motivated error is especially important for moral rules: If an individual would like to be able to lie, because it promotes his narrow self-interest, he would like to convince himself that it would be moral to do so, for then he would avoid feeling guilty (and so forth); now, if moral rules are complex, admitting myriad context-specific exceptions, the capacity to rationalize and misperceive pertinent information makes it likely that individuals would frequently err in favor of their own self-interest. This would substantially undermine the purpose of moral rules. Fourth, because moral rewards and sanctions are in part administered by others — those expressing approbation and disapprobation — it is important that rules be sufficiently simple (in particular, not requiring information available only to the actor) that other members of society can do their part in enforcing the moral system.

In the analysis to follow, we will simply assume that there is a given division of acts into various categories and that, although acts within a given category will tend to be similar, they will not be identical. Thus, if the category is lies (or some plausible subset thereof), it will be the case that some lies in the category cause more harm to third parties than do other lies (indeed, some lies may be beneficial to others) and that there will be variation in the extent to which an individual's narrow self-interest will be advanced by telling a lie. Later, we will return to the subject of the generality of moral rules, to consider the possibility of inculcating exceptions to moral rules as well as the possibility that the various categories, and thus the associated moral rules, may overlap and conflict.

III. ANALYSIS[20]

A. Framework

There are various situations in which individuals may find themselves. In each situation, we

---

[18]For further elaboration and some references, see subsection V(A). The analysis could be extended in a number of ways. One important consideration is that our ability to perceive events and act in light of them is limited by the structure and operation of our brains. In addition, because many of the relevant systems of the brain (visual perception, for example) have a multiplicity of purposes and may have evolved primary to serve other purposes (such as identifying prey or alerting individuals to physical danger), our mental categories may well not be ideally suited to making moral decisions. See, for example, Kosslyn and Koenig (1992) and Pinker (1997).

[19]In addition to strict limits, there are also considerations of cost. Even if it were feasible to inculcate moral rules situation by situation, the costs of doing so would be tremendous. Put another way, there are substantial economies of scale to be realized through the wholesale inculcation of moral rules, for categories of situations.

[20]A formal version of the analysis appears in Kaplow and Shavell (2001).

assume for simplicity that individuals have only two choices: They may either act or refrain from acting. For example, in one situation, the choice may be whether or not to tell a lie; in another, whether or not to rescue someone. Each act in each situation has two characteristics: The individual may gain or lose from committing the act, relative to not committing the act. We refer to this gain or loss as narrow self-interest. And the act may (but need not) result in a harm or benefit to third parties.

Furthermore, we suppose that these acts fall into natural groupings, as previously described. Thus, all lies, or perhaps all lies of a given general type or in a certain set of circumstances, fall into a single group.

If a moral rule is inculcated with regard to a particular category of behavior (say, a moral prohibition on lying), a social cost will be incurred, one that, as noted, is assumed to rise with the strength of the rule. For example, the more guilt is to be associated with lying — or the greater the feeling of virtue that is to be associated with telling the truth — the higher will be the corresponding cost of inculcation. Furthermore, we also noted that there are limits or diminishing returns regarding the extent to which individuals can actually experience feelings of guilt or virtue (and to which social disapprobation and approval can be expressed by others and be felt by an actor). It will sometimes be convenient to treat this limitation as a literal constraint, an upper limit on the extent to which guilt or virtue can be realized, recognizing that a more precise statement would reflect that, the more guilt or virtue is experienced for one type of act, the less effective will be the guilt or the virtue associated with other types of acts.

We now return to our first assumption: Individuals' behavior is taken to be determined by a weighing of their narrow self-interest along with any guilt or virtue (and disapprobation or approval) associated with an act. If, for example, guilt were associated with telling a lie and virtue with telling the truth, the individual would tell a lie if, but only if, his narrow self-interested gain (if any), minus the effect of the guilt he would experience for telling the lie, exceeded the virtue he would feel if he told the truth. (This is equivalent to stating that an act will be committed if and only if the narrow-self-interested gain from doing so exceeds the sum of guilt associated with committing the act and virtue associated with abstention.)[21]

Finally, we need to state the consequentialist objective. For concreteness, we will speak in terms of the maximization of social welfare.[22] From a consequentialist perspective, therefore, the goal is

---

[21]In section II(A), we stated our assumption that individuals weigh narrow self-interest along with guilt and virtue in making decisions. The present statement, in which we treat guilt and virtue as if they can be subtracted from or added to the gains or losses associated with acts, is not materially different, for we have made no statement about how any of the units are measured. In any event, any such arithmetic statements can readily be translated into statements regarding whether guilt and virtue are or are not sufficiently large to outweigh narrow considerations of self-interest.

[22]If the maximand were otherwise, see note 6, we could simply redefine the harm or benefit to others in terms of the alternative maximand and make corresponding adjustments, and the analysis would be similar.

to choose moral rules to associate with different categories of acts — which is to say, to choose levels of guilt and of virtue to associate with different categories of acts — so as to maximize welfare.[23] The constituents of welfare are as follows: the gains or losses, in terms of narrow self-interest, to individual actors; the harms or benefits caused to others by the acts; the costs of inculcating the moral system; and the effect on actors due to the actual experiencing of guilt and virtue.

Two points deserve emphasis. The first pertains to the inclusion of guilt and virtue in welfare. Because it is assumed that individuals who commit acts or abstain from them may experience guilt or virtue (indeed, it is this assumption that underlies their behavior) and because it is assumed that the consequentialist objective is to maximize welfare from all sources, it follows that this particular source of individuals' welfare is included in the social calculus.[24]

The second more broadly concerns the difference between the aforementioned statement of the consequentialist objective and that which is often encountered (both in the writing of consequentialists and of their critics). It is often assumed — usually implicitly — that whether an act is desirable under a consequentialist framework is determined by considering the effects of the act (its consequences) on the actor and on other parties, wherein attention is confined to what we have referred to as narrow self-interest and harm and benefit to others. Our formulation differs by including both inculcation costs and the effects of experiencing guilt and virtue.[25] As a result, it is possible that a moral rule that seems best from a conventional consequentialist perspective will be inferior to a different moral rule when one takes a more inclusive view of the consequentialist objective.

B. General Moral Rules

We now consider what moral rules — and associated levels of guilt and virtue — would best be associated with situations in a given category.[26] We begin by considering a moral system enforced

---

[23]As should be apparent, we are seeking to determine what use of guilt and virtue would be optimal under the implicit assumption that the moral system is the only check on self-interested behavior, thereby omitting other regulators, notably, the legal system. For explorations of law and morality, see, for example, Ellickson (1991), Posner and Rasmusen (1999), Shavell (2002), Sidgwick (1897), and Sunstein (1996).

[24]One could make different assumptions about the maximand or about individuals' behavior under which this conclusion would not follow; it will generally be clear which of our results would be affected (and how) if this element of welfare were thus excluded.

[25]Others, such as Mill (1861) and Sidgwick (1907), have stated that the experiencing of moral sentiments should be included in welfare, but they did not pursue the implications of that assumption.

[26]More precisely, we will consider what is best with respect to a given category, taking as given how the moral system will treat other categories of behavior. Because it is assumed that different moral rules, and levels of guilt and virtue, may be employed for different categories, this procedure is acceptable. (Our assumption that morality must operate at a categorical level proscribes fine-tuning morality situation by situation within a given category, but not across categories.) The only point of importance here is that, if more guilt or virtue is to be employed with respect to the category under consideration, it must be kept in mind that this tends to involve an

by guilt alone — a system consisting only of prohibitions of varying strengths — and then we introduce virtue as well. We proceed in this fashion because many of our overall conclusions can be derived by confining attention to guilt; the introduction of virtue makes the analysis more complicated, even though it ultimately does not affect many of the results.

Guilt. — Suppose that the consequentialist must decide how much (if any) guilt to inculcate with respect to a category of acts. Consider first the behavioral effect of inculcating guilt, that is, the deterrence of acts in the pertinent category.[27] For this behavioral effect to constitute a social benefit, it is necessary that some of the acts in the category are overall socially undesirable. Indeed, it must be that the preponderant effects of deterred acts would be undesirable if the acts were committed; otherwise, with regard to behavior, it would be better to forgo imposing guilt. Note that for acts to be socially undesirable, it must be that their net effect is negative. Thus, although ordinary driving behavior causes noise and air pollution, congestion, and the possibility of an accident, all harms to third parties, typically these costs are believed to be a good deal less than the benefits to drivers, so driving would not be socially undesirable from this consequentialist perspective.

Second, apart from behavior, there are other possible consequences of inculcating guilt, and each of these involves a disadvantage. There are the direct costs of inculcation. In addition, when not all individuals are deterred from committing acts in the category, guilt will be experienced, which reduces the welfare of those individuals. (For example, some people may litter when in a hurry and no trash receptacle is available, and they may feel guilty as a consequence.) Finally, when not all individuals are deterred and guilt is experienced, the limited capacity to experience guilt will be depleted to an extent, tending to reduce the effectiveness of guilt in controlling other types of behavior. (Individuals who already feel guilty about their littering and other moral infractions may not feel as much additional guilt from telling lies and thus may be more likely to do so.)

Overall, it makes sense to inculcate guilt only when there is an advantage, involving the deterrence of acts that are overall undesirable, and this advantage outweighs the costs of using guilt. In general, it is desirable to instill more guilt the greater is the extent to which acts are undesirable, the lower are the costs of inculcation, and the greater is the portion of individuals who will thereby be deterred from committing acts. Relatedly, when determining the extent to which guilt should be inculcated, an important consideration will be that raising the level of guilt affects the total amount of

---

implicit cost, given that the total capacity to realize guilt and virtue is assumed to be constrained and hence using more guilt and virtue in one category will reduce the effectiveness of guilt and virtue used elsewhere.

[27]For convenience, we will discuss the possibility of associating guilt with committing acts in a given category and virtue with abstention from acts in the category. We could as well discuss the assignment of guilt to abstention and virtue to action — which would make sense if the acts in the category were generally desirable and omissions undesirable. Allowing for this possibility would make our exposition more cumbersome without changing any of our analysis. The possibility that there may be an intrinsic distinction between acts and omissions which bears, from nonconsequentialist perspectives, on whether acts and omissions can be viewed symmetrically, is discussed below.

guilt that is experienced, and in a somewhat complicated manner. (Raising the level of guilt results in less guilt being experienced on account of more individuals being deterred from committing acts, but it results in more guilt being experienced by those who remain undeterred. A priori, either factor could be more significant.[28])

These factors cannot be viewed in isolation. Thus, even though greater harm, taken by itself, favors the use of guilt, it will often be sensible to inculcate guilt for acts that do not cause significant harm to third parties. Consider fairly minor misbehavior, like cutting in line. When even a small amount of guilt is instilled with respect to such behavior, it seems that nearly everyone is deterred from cutting in line most of the time. Though the benefits from such deterrence are modest, the costs apparently are negligible. Inculcation costs are probably rather low, and since few actually cut in line and thus experience guilt for having done so, the other costs seem to be very small as well.

The foregoing discussion has a number of implications regarding the use of guilt. Clearly, acts in a given category can be socially undesirable, yet it would not be advantageous to employ guilt. That is, it may be appropriate to deem behavior morally permissible even when, ideally, it would be better if it were not engaged in. And, even when it is advantageous to inculcate guilt, it may be too costly or otherwise undesirable to inculcate it at so high a level so as to deter all undesirable acts.[29] In that case, individuals may commit undesirable acts, despite the moral prohibition, and experience guilt as a result.

It is also possible that desirable acts will be subject to guilt. The basic reason is that acts in a given category are heterogeneous. Accordingly, when most acts in the category are sufficiently undesirable (perhaps lies, or lies of a given type), it is advantageous to instill guilt — to deem acts in the category morally impermissible — but this will mean that guilt will be associated as well with the atypical, desirable acts in the category (those occasional lies whose social benefits exceed the harm they cause). Two results are possible in this case: Desirable acts may be deterred (if the level of guilt is high relative to the actor's personal benefit from committing the act), and desirable acts may be committed (in the contrary case), in which event individuals will experience guilt as a consequence of committing desirable acts.[30]

---

[28]If the deterrence effect is relatively small, it may be best to use little or no guilt, despite the acts being socially undesirable, on average. If the deterrence effect is relatively large, it may be best to set guilt at a high level, even at a level where, at the margin, most of the acts deterred are socially desirable ones (see the discussion in the following note).

[29]The reason it may otherwise be undesirable is that it may be that the undeterred acts would include many that are in fact desirable. (For example, if the harm to third parties from a type of act is rather uniform for all acts in a category, but individual actors' benefits vary, the acts of those with the highest personal benefits — and those will tend to be the socially desirable acts — would constitute most of those that remain undeterred when guilt is set at a moderate level.)

[30]The results described in the text hold if the term "desirable acts" is understood conventionally, to refer to acts for which the personal gain to the actor exceeds any harm caused to others. As we have explained, however, a complete account of welfare in this setting is more inclusive. Hence, in asking whether it would be ideal for a given

Guilt and virtue. — We now suppose that the consequentialist must decide how much virtue as well as guilt to inculcate with respect to a category of acts. One possibility is simply not to use virtue, in which case the foregoing analysis would indicate how much (if any) guilt would be best to employ. Another possibility is to use just virtue, and, of course, the remaining possibility is to use some combination of the two.

As we have previously suggested, when guilt and virtue are employed together, individuals tend to act when their personal gain minus the guilt they would experience exceeds the virtue associated with not committing the act, which is equivalent to the statement that individuals act when their personal gain exceeds the combined effect of guilt and virtue. Therefore, in determining the extent to which an individual will comply with a moral rule, it does not matter what combination of guilt and virtue is employed, only the total of the two. Guilt and virtue, viewed as incentive devices used to induce individuals to comply with moral rules and thus to behave in a more socially desirable manner, can thus be seen as substitutes.

If this were the complete story, the remaining analysis would be straightforward because the foregoing discussion of guilt could be applied to virtue. Notably, it would tend to be advantageous to employ virtue when it leads individuals to behave in a more socially desirable manner and when the costs of inculcating virtue are low. Also, the general nature of moral rules is such that, when virtue is employed, it will sometimes be associated with undesirable behavior. To illustrate the latter, just as guilt may be associated with telling lies even though some lies may be socially desirable, virtue may be associated with being truthful even though sometimes telling the truth may be socially undesirable.

That both guilt and virtue are potentially available, and that they are substitutes with regard to controlling behavior, raises the question whether there is any basis for a consequentialist to make greater use of one or the other with regard to a particular category of acts — that is, whether socially undesirable acts should be morally prohibited, the opposite acts should be morally encouraged, or some combination of the two. The most obvious reason for differentiation is that it is possible that, in terms of inculcation costs or relative scarcity (given individuals' limited capacities to experience guilt and virtue), it may be more advantageous to use more of one or the other, as the case may be. Regarding individuals' limited capacities, whichever of guilt and virtue was already used to a greater extent to control compliance with other moral rules would tend to be subject to more substantially reduced effectiveness and would tend to involve greater costs from further use in terms of interfering with the control of other types of behavior.

Until this point, we have discussed guilt and virtue as substitutes. There are, however, two qualitative differences between them, each of which may bear on which should be employed. First,

---

act to be committed, one would also include the reduction in welfare due to the actor's experiencing guilt (if the act is committed) as well as the negative effect of this phenomenon in using the scarce capacity to experience guilt. If "desirable act" were defined by reference to this inclusive conception of welfare, the two results in the text would still hold.

when guilt is used to control behavior and is successful in doing so, guilt is *not* experienced, whereas when virtue is used and is successful, it *is* experienced. That is, when individuals do not commit an act because they would feel guilty if they did, they by assumption will not feel guilty. By contrast, when individuals do not commit an act because they would feel virtuous as a consequence of abstention, they by assumption will feel virtuous. Because individuals' capacities actually to experience both guilt and virtue are limited, this difference is important. In particular, it favors using guilt to deter acts in a given category when most individuals will successfully be deterred, because then the stock of guilt is not much depleted. Likewise, it favors using virtue to induce individuals not to commit acts in a given category when only a few individuals will be successfully induced to do so, because then the reservoir of virtue is not heavily drawn down.

For example, if it is supposed that most individuals could readily be deterred from cutting in line, the present consideration suggests that it makes sense to accomplish this result by inculcating guilt. Since the prospect of experiencing guilt leads most individuals to behave appropriately, few ever experience guilt, so there is little depletion of the limited capacity to experience guilt. If, however, society relied instead on virtue, then when individuals routinely abstained from cutting in line, they would feel virtuous as a consequence. Using virtue to control such everyday activity would quickly deplete individuals' capacity to experience virtue, making it unavailable to control other behavior that may be more important.

Now consider socially desirable behavior that few individuals may be induced to undertake, such as rescuing others or assisting starving people in far-away lands, either at great personal sacrifice. If guilt were used in an attempt to induce individuals to thus help others, by assumption few would do so, with the consequence that the capacity to experience guilt would be rapidly depleted. But if virtue were used, since only a few can be induced to behave virtuously, the depletion of the capacity to experience virtue would be less.[31]

A second factor — which introduces some further complication into our analysis but we believe does not ultimately change it in a significant manner — is that guilt *negatively* affects the welfare of the individual who experiences it whereas virtue *positively* affects welfare. On this account, it would seem that it is preferable, ceteris paribus, to employ virtue rather than guilt to induce individuals to behave morally. Indeed, because experiencing virtue is beneficial in and of it self, it might even seem that it

---

[31]In both of these examples, it is useful to recall the analogue to our assumption with respect to external moral rewards and sanctions, approbation and disapprobation. If approbation were used whenever a person failed to cut in line, most everyone would quickly be exhausted, whereas it takes relatively little effort to bestow praise on those few who engage in unusual self-sacrifice to help others. Likewise, if disapprobation were expressed whenever an individual failed to abandon his self-interest to save the world, everyone would quickly become exhausted, whereas we are taxed far less if we only bestow disapprobation on those few who violate moral rules that are followed by almost everyone. One might further reinforce this argument with the psychological point that individuals would experience dissonance if they constantly expressed disapproval of the very behavior that they routinely engage in themselves; similarly, it would seem strained to bestow sincere praise on behavior that was routine and that one regularly engaged in oneself.

would be beneficial to instill virtue willy-nilly — even for acts that have no effects (positive or negative) on third parties and perhaps even for acts that are (only slightly) undesirable, for the benefit of experiencing virtue may exceed any loss due to inducing otherwise undesirable behavior. We do not emphasize these conceivable implications regarding virtue because we believe that they are self-limiting. The more that virtue is instilled, the greater the extent of diminishing returns regarding its effectiveness in controlling behavior. As long as humans' capacity to experience virtue is not especially large (relative to the myriad types of behavior it would be ideal to control), it will make sense to use virtue only where it improves behavior. If, for example, one viewed individuals' capacity to experience virtue as a specific, fixed amount, the main question regarding virtue would be where to use it — not how much of it to use — and, for a consequentialist, it would best be used where the behavioral benefits are greatest.[32] If all important sorts of behavior were effectively controlled and if the capacity for virtue was only lightly taxed, it may make sense to use virtue to control less important behavior, or even to use it purely so that individuals could benefit by the experience of virtue. But our empirical conjecture and observation is that it does not seem so easy to induce most individuals to behave in a socially ideal manner in all respects; thus, we are unlikely, in a consequentialist moral system, to have any virtue to spare, much less a significant amount. (And, relaxing the assumption in the preceding discussion that the capacity to experience virtue is literally fixed, assuming instead that virtue becomes increasingly less effective as more of it is used, the basic conclusion that virtue should be focused where it is most useful in controlling behavior would still follow.) Furthermore, for it to make sense to use virtue just for the sake of the benefit of individuals' experiencing it, the costs of inculcating virtue would have to be sufficiently low, and it hardly seems clear that this condition would be met in practice.

In all, it seems plausible that, in a consequentialist moral system, virtue, like guilt, would best be employed only where it would result in a significant behavioral benefit. Nevertheless, because feelings of virtue contribute to welfare whereas feelings of guilt detract from welfare, it will on this account be optimal in a consequentialist moral system to use more virtue and less guilt than otherwise would be best, taking into account the costs of inculcating virtue and guilt and their decreasing marginal effectiveness.

C. Remark on Specific Moral Rules

In this subsection, we briefly shift our focus to consider the hypothetical, admittedly unrealistic, but conceptually interesting question of how our analysis changes if we drop our assumption that moral rules must be general to some extent, and assume instead that it is possible to tailor the use of guilt and virtue perfectly to each possible situation. Briefly exploring this case will reinforce our understanding of the problem of designing a consequentialist moral system and highlight the significance of our assumption about the categorical nature of moral rules. To simplify our discussion, we will confine our

[32]There in fact remains the question of how much virtue to use because inculcating virtue is costly. If, however, inculcation costs led to using less than the available supply of virtue, the argument that one should use virtue where its benefits are greatest — and not pervasively — is even stronger.

attention to the case in which only guilt is used to enforce moral rules.[33]

First, it would be best to employ guilt only with respect to undesirable acts. (Above, the only reason that guilt might be employed with respect to desirable acts is that they were grouped with undesirable acts, but now we are assuming that this is no longer the case.) After all, inculcating guilt is costly; if it deterred a desirable act, that would be a further cost to instilling guilt; and if it failed to deter the desirable act, then the guilt would be experienced, which itself would involve a further cost.

Second, when guilt is properly employed, it would be inculcated at a level that is sufficient — and just barely sufficient — to deter an undesirable act. If the level of guilt were insufficient to deter the act, inculcating it would be pointless: Inculcation costs would be incurred and guilt would be experienced, without any benefit being produced. And it would not be advantageous to instill more guilt than necessary to deter the act, for any excess guilt that was inculcated would entail a further cost with no corresponding benefit.

Third, a direct implication of the foregoing point is that guilt would never actually be experienced. As stated, guilt is only properly employed when it is set at a level sufficient to deter the undesirable act.

Finally, guilt would not be employed to deter all undesirable acts, but only those for which the benefits of deterrence exceed the inculcation costs.

It is apparent that some of the foregoing results differ from those we obtained for the case of general moral rules — namely, that guilt is employed only with respect to undesirable acts; that whenever guilt is used, it is successful in deterring undesirable acts; and that guilt is never experienced. These differences in results follow, intuitively and directly, from the difference between the two cases. This reinforces the importance of our assumption that moral rules are categorical. Even when the types of situations included in a category are largely similar, there will usually exist cases in which the results regarding general moral rules differ qualitatively and significantly from those in the hypothesized ideal situation in which moral rules can be tailored specifically to every possible situation that may arise.

IV. THE OBSERVED USE OF GUILT AND VIRTUE

Our analysis indicates how guilt and virtue would be used to enforce moral rules if, in fact, moral rules were designed to maximize social welfare. Some conclusions are straightforward, and these seem to be in accordance with what we observe, such as that guilt and virtue are employed to prevent acts that typically reduce welfare (lying, breaking promises, harming others) and to encourage acts that benefit others (rescuing someone in distress). Of course, it is not always true that a given type of act is

[33]Considering guilt and virtue is largely analogous, subject to the qualifications noted in subsection III(B). We provide a complete and formal analysis of both cases in Kaplow and Shavell (2001).

invariably desirable or undesirable; for example, some lies may be desirable. Nevertheless, as is well-known, we still may feel guilty when we tell such a lie. This too is implied by our analysis because guilt and virtue must be inculcated for categories of acts (here, the category of lying), fine-tuning being costly or impossible.

Our analysis also produces some novel conclusions. Notably, these concern the fact that guilt and virtue will actually be experienced by individuals who commit acts with which guilt and virtue are associated. As we explain, if either guilt or virtue would succeed in inducing most individuals to act in a socially desirable manner, it will be better to use guilt (because guilt will not usually be suffered, whereas if virtue were employed, the limited pool of virtue would be squandered). Likewise, when few individuals can be induced to act properly, it will be better to use virtue (because little virtue will be consumed, whereas if guilt were employed, much would of the scarce stock would be consumed). These conclusions do seem to be in accord with the observed use of guilt and virtue. On one hand, individuals who do a range of undesirable acts — from cutting in line to physically assaulting those with whom they have disagreements — generally feel guilty, and indeed these are acts that most individuals are successfully deterred from committing most of the time. But individuals do not, it seems, feel especially virtuous when they abstain from such acts, since it is expected that everyone will do so.[34] On the other hand, individuals who rescue others at great personal sacrifice and those who devote their lives, say, to helping the poor in less developed countries, feel virtuous and are objects of praise, whereas the substantial majority of us who do not routinely give most of our time or resources to helping strangers (and could not readily be induced to do so) do not generally feel terribly guilty and are not subject to social disapproval. An implication related to the foregoing is that we do not ordinarily see significant use of both guilt and virtue with regard to the same decision, which also is suggested by our analysis.

In all, it appears that our analysis helps us to understand why guilt may be primarily associated with some acts and virtue mainly with others. Let us compare this situation with another seemingly plausible possibility: a moral system under which significant virtue is associated with all (or most) good choices among actions and significant guilt is associated with all (or most) bad choices. If this were generally the case, then abstaining from bad behavior in everyday situations in which almost everyone would abstain would be associated with individuals' feeling highly virtuous, and failing to behave in a manner that raises total welfare at significant personal sacrifice when most others would similarly fail to do so would be associated with substantial guilt. Neither seems generally to be true. In sum, our analysis does seem to offer an explanation for the differential use of guilt and virtue, a distinction that

[34]There are, of course exceptions. For example, one who abstains despite unusual provocation may feel virtuous or be subject to approbation from others. But, interestingly, this is precisely a type of situation in which most individuals would not act in a desirable manner, so this apparent exception is itself consistent with our analysis. As an illustration, there is often an exception to the moral injunction against aggression for cases of self defense. This rule — in addition to the benefit of the prospect of retaliation in deterring aggression — has the advantage that, since most individuals will not be able to restrain themselves in certain settings, a needless use of guilt is avoided.

does not readily seem capable of explanation on other consequentialist grounds.

Additionally, we note that guilt seems to be much more on people's minds than is virtue. (This, admittedly, is a casual empirical conjecture. It is prompted by what seems to us to be greater attention in the literature to guilt and disapprobation than to virtue and approbation and the relatively higher frequency with which individuals seem to discuss the prospect of feeling guilty than the prospect of feeling virtuous.) To explain this, we observe initially that, because both guilt and virtue are actually experienced only when individuals behave atypically, it is the contemplation of acts that are not ultimately committed that is most relevant here. And, with regard to contemplated acts, guilt discourages acts that advance individuals' narrow self-interest and thus that they would otherwise commit; because such temptation may be frequent (for example, one may often be tempted to lie or to cut in line), the prospect of guilt would often come to mind. In contrast, virtue encourages acts that individuals would not otherwise commit — acts that are against individuals' narrow self-interest (such as acts involving self-sacrifice to aid others); because most individuals may not frequently be inclined to contemplate committing such acts, virtuous feelings would seldom be pondered.[35]

Another implication of our analysis is that, the greater is the deterrent effect, the greater is the benefit of raising the level of guilt or virtue in inducing more individuals to behave in a socially desirable manner. For example, behaviors that are more automatic, less conscious, will be less subject to the use of moral sanctions and rewards. Consider the cases of young children and adults of limited mental capacity.[36] (An exception would be where behavior is automatic due to habit formation, when following one's moral emotions is part of what produced the habit in the first instance.)

Furthermore, our analysis may help to explain variations in moral rules across cultures, over time, and within societies (comparing different groups).[37] Such variation has long been recognized and is said to pose difficulties for some normative theories. As a positive matter, however, our analysis suggests that, even if moral rules in different settings were optimal from a consequentialist perspective, their content — whether and the extent to which there would be guilt and virtue associated with assorted types of behavior — may well differ. Methods of inculcation as well as the identity and thus the interests of inculcators will vary (role of organized religion, form of government, existence of formalized education, mobility across communities), as will the relative frequency of different types of

---

[35]Another explanation is that the cost of inculcating virtue may be much higher than that of inculcating guilt or that the limitations on the use of virtue may be much greater than those on using guilt, so there is simply less virtue than guilt employed in enforcing moral rules.

[36]The treatment of children is somewhat complicated. On one hand, due to their less developed ability to conform their behavior to moral rules, we tend to excuse them. On the other hand, one cannot wholly refrain from applying moral rules if one is hoping to inculcate them. An implication is that one should defer the use of moral sanctions, especially guilt, until children reach an age where there is a reasonable prospect of achieving success over a period of time that is not unduly prolonged.

[37]See, for example, Miller (2001).

situations and the magnitude of personal benefits or costs and harms or benefits to third parties associated with acts in those situations. And, even if two types of behavior were identical in frequency, harm, and personal benefits in two settings, optimal moral rules may nonetheless differ, for example, if limitations on the use of guilt or virtue are more constraining in one society due to the greater need to use guilt or virtue to regulate some other type of behavior. That a consequentialist moral system is consistent with variations in moral rules across societies is, of course, not a novel suggestion, for it is recognized that consequentialist morality is contingent upon the circumstances under consideration.

It would be difficult to pursue the foregoing descriptive claims for two reasons. First, the ability to measure the relevant phenomena is limited and further speculation at this stage in the investigation of the subject seems premature. Second, a number of complications discussed below also bear on the observed use of guilt and virtue. Most important are that the optimality of actual moral rules is hardly assured and that the rules that tend to emerge may promote an objective that is different from social welfare.

V. DISCUSSION OF ASSUMPTIONS AND EXTENSION OF THE ANALYSIS

A. The Categorical Nature of Moral Rules

Right and wrong versus social desirability. — Under the consequentialist moral system that we have outlined, whether any particular act is "right" — and accordingly is associated with the actor feeling virtuous and being subject to approbation — or "wrong" — and associated with the actor feeling guilty and being subject to disapprobation — depends on the *category* of acts in which the particular act falls. Now, for a given category of acts, the behavior that is determined to be right or wrong in the sense just described depends, in turn, on the typical, average effects of acts in the category. For example, if most lies are socially undesirable, lying will be treated as a wrong under an optimal consequentialist moral system. As we have emphasized, this will be true even though some lies are socially desirable. Hence, as long as moral rules have any tendency to be general — and we have suggested that, due to human nature, they must be so to an extent — there will sometimes be a conflict between what is right and wrong under a consequentialist moral system and what is understood, in principle, to be socially desirable.

This contrast between which acts are deemed to be right and wrong and which acts are ultimately socially desirable or undesirable is, as we have noted, a familiar one. Writers such as Hume (1739, 1751), Austin (1832), Mill (1861), and Sidgwick (1907) made this sort of distinction in advancing what is now described as a two-level view of morality.[38] At the first (higher) level is the

[38]Two-level views are often associated with rule utilitarianism, in contrast to act utilitarianism, but discussions of rule versus act utilitarianism often fail to illuminate because there is a lack of agreement about the meaning of each version of utilitarianism and whether, at a deep level, they can be distinguished at all. Twentieth-century two-level accounts that seek to address these issues include Brandt (1979, 1996), Hare (1981), Harrod (1936), Rawls (1955), and Sartorius (1972).

ultimate criterion of judgment, which for these writers was social utility. At the second (lower) level are the moral rules. Because the choices at the second level are limited — due to given facts of human nature — one does not expect any moral rule to generate ideal behavior in all cases.[39] Relatedly, under two-level theories, blame and praise (externally administered analogues to guilt and virtue, which receive less attention in many contemporary discussions) are viewed instrumentally: Whether an act should be deemed blameworthy, it is argued, should depend not on intrinsic features of the act or even on whether the act produces undesirable consequences, but rather on whether the practice of blaming those who commit the act will itself promote welfare. (Thus, as noted, blame might be associated with some welfare-promoting acts if the practice of blaming acts of that general type is, as a whole, socially desirable.) As we note in our introduction, in some respects our attempt to present a systematic analysis of a consequentialist moral scheme builds upon the insights of these prior thinkers.

Rules and exceptions. — As discussed in subsection II(D), there are strong reasons that we are led to group acts in various ways, and, to the extent that guilt and virtue are inculcated, one would expect there to be significant savings involved in inculcating these moral sentiments with respect to groups of similar acts. Inevitably, the natural groupings will not be ideally suited for the particular purposes of regulating behavior. For example, there may be a category such as lies, or even a particular type of lies, such that the situations in the category tend to be similar; but it is unlikely that there will be complete uniformity, and, in particular, even if most lies are undesirable, not all of them may be.

Of course, our minds have a good deal of flexibility and are susceptible to some forms of instruction. Hence, if some natural category for which we would like to inculcate guilt is overinclusive — perhaps an important subset of acts in the category is desirable or is difficult to deter — we might expend additional resources to inculcate an exception.[40] For example, we might be taught that white lies are not wrongs, sparing us from the prospect of feeling guilty about or being deterred from committing lies that would be innocuous or even beneficial, such as one told to lure an individual to a surprise party. To take another example, self-defense in certain types of circumstances might be excepted from the prohibition on aggression.

One might suppose instead that society could simply inculcate guilt over a smaller set — aggression that is not in self-defense — rather than inculcate guilt over the broader set and then expend additional effort to inculcate an exception. Whether this is feasible, we submit, is largely exogenously determined; sometimes there will be a narrower natural set that is rather homogenous regarding the best level of guilt or virtue to instill, and sometimes the natural set will be broader and quite heterogenous in

---

[39]Similarly, psychologists have suggested that moral rules function as decisionmaking heuristics that are subject to error in application due to overgeneralization. See, for example, Baron (1994), Spranca, Minsk, and Baron (1991).

[40]More generally, the boundaries between categories of situations could be made endogenous in the analysis, so that the inculcation process itself was concerned in part with instilling certain categorizations.

this regard. Other times, there may be a choice whether to inculcate guilt or virtue act by act (more realistically, small cluster by small cluster) or to inculcate over a larger set. The larger set may not allow as precise a match of guilt and virtue to particular situations, but the scale economies realized through more wholesale inculcation may warrant the use of grosser classifications. Indeed, the savings in inculcation costs stand as an independent reason for moral rules to operate on a categorical level rather than to be specified for every possible type of act. This is especially true in light of the fact that many types of situations may, in isolation, be unlikely to arise: It would not make sense to expend the effort to teach separate moral lessons for each of the large number of such cases, whereas it may be highly desirable to teach general moral lessons that would each cover many situations, however unlikely each might be.

Overlaps and conflicts among moral rules. — Another important feature of categories of acts is that they may overlap, and indeed it is routinely observed that more than one moral rule may apply in a given case, raising the possibility of conflicts among moral rules. For example, there could be one set of acts — pushing another individual out of one's way — that is subject to guilt and another set of acts — aiding others in distress — that is associated with virtue. But this raises the question of what happens when one pushes someone out of one's way in order to help someone else who is in distress. One possibility is that individuals simply combine all sources of guilt and virtue in making their decisions. Hence, a prospective rescuer may help a person in distress and thereby feel virtuous, but still feel guilty for having pushed someone out of the way.[41] Or, the prospect of that guilt when combined with the rescuer's own direct costs of aiding another may exceed the virtue he would feel, thus deterring the act of assistance. Another possibility is that certain emotions may trump or at least dull others, so perhaps the rescuer would not feel guilty after all if he pushed someone out of his way in the course of rescuing someone else. How such overlaps and conflicts are resolved is an empirical question about the nature of the categories in our minds and the manner in which our minds actually function.[42] To an extent, the outcome may also be socially determined, for society could choose to inculcate an exception to one or another moral rule in cases of conflict, and sometimes this seems to be done.

We now consider one particular manner in which categories of situations may overlap: In addition to moral rules for particular types of acts (such as lying or stealing) there exist moral rules that apply very broadly, notably, the Golden Rule, which directs individuals always to take into account the

---

[41]Brandt (1996) and Ross (1930) suggest that, when individuals follow the stronger moral obligation, they nevertheless feel compunction about having neglected the weaker obligation.

[42]There is considerable debate about the nature of moral conflicts. On one hand, it is widely acknowledged that apparent conflicts exist and that conflict is perceived to exist at the psychological level — and this is all that is relevant for our purposes. On the other hand, it is contested whether, as a matter of moral truth (or, under a two-level moral theory, at the higher, critical level), there can be moral conflicts. For differing views, see the essays in Gowans (1987). Statman (1995) argues that there can be moral conflicts (two moral principles favor different outcomes) but no true dilemmas (conflicts that yield no correct answer); nevertheless, he suggests that when individuals have internalized moral principles they may, as a consequence, experience feelings of guilt or regret when they act correctly in cases of moral conflict.

effects of their behavior on others.[43] One can understand such a rule as associating guilt with all undesirable acts and/or virtue with all desirable acts, perhaps with the level of guilt or virtue rising with the extent of harm or benefit caused to third parties. It seems clear that such broad rules do exist, although it is equally clear that they exist along side the categorical rules that we have been discussing thus far and not in lieu thereof.

This raises the question of why society does not simply inculcate the Golden Rule or some variant, eschewing all other rules, and thereby require all individuals always to act in a socially ideal manner. Reflecting on the factors we discuss in subsection II(D), there are good reasons why this is not how moral systems operate: It would be difficult to inculcate the command to engage in complex calculations concerning all behavior to young children, even as adults the application of such a rule would be costly and difficult, and there would arise the problem of rationalization (that individuals would miscalculate in their own self-interest to avoid the restraining force of guilt).[44] Moreover, our analysis suggests that, even if successful, such a broad rule would be problematic if the associated levels of guilt or virtue were high, because of the constraints on the ability to experience guilt and virtue. Thus, even with the Golden Rule in force, many individuals would still commit undesirable acts, which would consume the scarce pool of guilt, making it more difficult to control other acts that it may be more important to deter; likewise, if virtue were instilled for all good acts, virtue would quickly be consumed on routine good behavior, leaving little to encourage certain types of behavior that may be particularly valuable. In sum, broad rules like the Golden Rule, as a supplement to more specific (but still fairly broad) rules are likely to be valuable precisely because of their breadth (they may cover acts that fall in the gaps between other moral rules) and their flexibility (they are directly sensitive to the unique features of particular situations). Nevertheless, due to their limitations, they ideally would be associated with only modest levels of guilt and virtue, and they would be supplemented by the kind of moral rules that we have emphasized throughout our discussion.

We now consider how the foregoing conceptual analysis of overlaps and conflicts among categories of situations regulated by moral rules helps to explain observed aspects of feelings of guilt and virtue and certain features of philosophical argument. Regarding the former, it is obvious that individuals sometimes do feel conflicted about what behavior is morally correct. Moreover, conflicting feelings about morality often seem to arise in instances in which two or more moral rules plausibly apply in the same situation.

Likewise, much discussion in moral philosophy is concerned with cases in which our moral

---

[43]Similar analysis would apply to intermediate cases, such as a rule enjoining all acts that harm others or all acts that intentionally harm others.

[44]These and related factors have been emphasized by many writers, including Smith (1790), Austin (1832), Brandt (1979, 1996), Sartorius (1972), Hare (1981), Mackie (1985), and Hooker (2000). In addition, Cosmides and Tooby (1994) suggest that the human mind is better at specialized than general problem solving, implying that we are more capable of properly applying rules targeted to particular contexts than a broad command like the Golden Rule.

instincts and intuitions seem to be in conflict, such as in cases in which one must inflict harm on one individual in order to help others (who are greater in number or who are affected to a greater extent).[45] Indeed, if a command akin to the Golden Rule is inculcated, then such conflicts among our moral intuitions will arise whenever a categorical moral rule requires welfare-reducing behavior, a phenomenon that does seem to fit many discussions of consequentialism. That is, many arguments concern cases in which an act promotes welfare and nevertheless violates some moral rule.

Furthermore, many analyses of such cases fail to acknowledge that, even without regard to overlap, the categorization of acts implies that some behavior will be subject to feelings of guilt or virtue even when, if the act were viewed in isolation, that would be inappropriate. It also should be noted that, when there is overlap, it does not follow that whichever category seems to exert a stronger pull on our intuition — perhaps the category for which the absolute magnitude of the feeling of guilt or virtue is greater — is the one whose rule would lead to ideal behavior. For example, helping others may have little virtue attached to it because virtue is costly to instill, because few would in fact help others, or because typical instances of helping others do not involve nearly as great a benefit as would the particular act in question; but none of these reasons suggest that commission of the particular act, which may also be in another category to which guilt is assigned, would be socially undesirable.

B. Evolution and Inculcation

Both evolution and inculcation (nature and nurture) seem to play important roles in determining the use of guilt and virtue in enforcing moral rules.[46] Initially, we observe that the general capacity to feel guilt and virtue — as distinct from how that capacity may be employed in a given society — obviously has an evolutionary origin, just as does any other capacity we might have. See Darwin (1874), E.O. Wilson (1975), and Izard (1991).[47] Likewise, the manner by which guilt and virtue may be inculcated and associated limitations or costs must have a biological foundation in the way that our brains process information and in the mechanisms by which various emotions are triggered. It is further study of the neurological foundations of aspects of human psychology that have the potential to illuminate the mechanisms currently under investigation; indeed, further examination of the existing

---

[45]See, for example, the discussion in subsection VI(E) on the act/omission distinction and related doctrines.

[46]Many of the ideas discussed in this subsection are developed in the literature cited in note 5. That both evolution and inculcation determine behavior in general is a theme of Barash (1982). See also Tooby and Cosmides (1990), who explore the evolutionary interdependence between genes and the influence of the environment.

[47]See also Darwin (1874) and de Waal (1996), who suggest that certain other species exhibit aspects of morality and conscience, and see Darwin (1872), who argues at length that the facial expressions that correspond to different emotions are universal in humans and evident in some other species and hence must have an evolutionary origin (which seems necessarily to imply that the emotions being expressed must too have an evolutionary origin). Thus, although critics of sociobiology, such as Lewontin, Rose, and Kamin (1984), would give a heavier role to cultural determinants in most domains, it seems difficult to deny that biology has an important role at least in explaining the existence of features of the human brain that enable us to experience moral emotions.

scientific literature should be illuminating.[48]

Our capacity to feel guilt and virtue is a flexible one. Such flexibility would seem to confer an evolutionary advantage because it would allow subsequent generations to adapt to changed circumstances. Moreover, regardless of the particular explanation, we certainly do observe substantial efforts to inculcate guilt and virtue to enforce various moral rules — in the rearing of children, in organized religion, in educational institutions, and in some acts of government. This is particularly apparent in extreme cases, in which feelings of patriotism or fidelity to a religious belief are able to motivate individuals or groups to engage even in suicidal behavior. The possibility of inculcation, moreover, is important in attempting to explain cross-cultural variation in moral rules as well as their rate of change over time (which seems greatly to exceed the rate of biological evolution).[49]

It also seems plausible that some of our particular feelings of guilt and virtue are not entirely the product of inculcation. Consider, for example, the guilt that we associate with stealing. No doubt society instills guilt in this case, but it also seems possible that some of the guilt we feel is a product of evolution in the biological sense. An instinctive reluctance to steal may well help to overcome acquisitive urges that, if acted upon, would be met with retaliation, which can prove very costly to the initial aggressor.

The extent to which moral rules and associated feelings of guilt and virtue are inculcated rather than purely evolved has normative and positive implications. On the normative side, to the extent that rules can be manipulated, society can attempt consciously to design policy to improve our moral system, whereas if everything were hard-wired and fixed, there would be little that could be done, short of eugenics.

Regarding positive implications, we offer two comments.[50] The first concerns what is being maximized. Evolution tends to maximize survival (more precisely, replication of the pertinent genes) whereas inculcation, particularly in a society not on the brink of subsistence, may more plausibly be

---

[48]For example, one recent study measures the extent to which contemplation of various moral dilemmas activates different areas of the brain. See Greene et al. (2001). See also Haidt's (2001) discussion of the application of various literatures in psychology and other fields to the study of moral emotions.

[49]See, for example, Nisbett and Cohen (1996), who identify cultural differences regarding what they refer to as a "culture of honor" between inhabitants of northern and southern states. Izard (1991) elaborates the view that the capacity to experience guilt and shame has an evolutionary origin but that the association between these emotions and particular acts is produced by internalization as a consequence of parental activity and other forms of social learning and thus varies across societies. See also Tangney and Fischer (1995).

[50]One of the best ways to assess the validity of positive claims about moral systems is to examine differences in morality among different societies and over time and to see whether they can be explained in a manner consistent with such claims. For an interesting examination of another culture, see Brandt (1954).

concerned with maximizing welfare.[51] In examining various questions — such as how bad it is to tell a lie or how good it is to help others in particular circumstances — it seems clear that the answers may be different if the goal is different. If controlling aggression was (in the relevant evolutionary period) far more important to survival than helping others pursue their ambitions, and if the pattern of moral emotions is determined primarily by evolution, one would predict a heavy use of guilt to control aggression but little use of guilt or virtue to induce individuals to assist others' attempts to realize their objectives. But some acts of helping may have been important to survival, such as sharing food among members of one's tribal group (as long as they did not shirk), as a form of insurance. If so, guilt or virtue might be used heavily to encourage cooperative, sharing behavior. On the other hand, if inculcation can affect the situations in which cooperation can be induced, this human capacity can usefully be employed to serve a wider range of purposes and thus be more adaptive to modern circumstances.

Second, the tendency for moral systems to be optimal — by reference to whatever is being maximized — may differ depending upon the relative importance of evolution and inculcation. Neither process assures optimal results. With evolution, there is the familiar point that selection is fundamentally at the level of individual genes, so, for example, traits that would benefit a group as a whole do not tend to emerge (although they may arise to some extent through kin selection, reciprocity, and so forth). Also, with evolution, there must be a feasible path for a desirable trait to emerge. With inculcation, there is the problem that inculcators do not bear all the costs and benefits of their actions. For example, parents may fail to inculcate guilt concerning a type of behavior that does not harm other family members or contribute to the ability to establish a reputation. Likewise, when there are multiple inculcators, each may be excessively inclined to draw upon individuals' limited capacity to experience guilt and virtue, a sort of common pool problem.

In most of our discussion, we take the view that guilt and virtue are to a substantial extent the product of inculcation and thus may be regarded as more plausibly concerned with welfare than with survival (although there remains the question of which groups' welfare is likely to be maximized). To an extent, the moral system that we observe seems to be consistent with this view. There do, however, seem to be important aspects of existing morality that may well have evolutionary explanations. Notably, there is a strong tendency to limit altruism, and related feelings of virtue, to one's kin. Nevertheless, the explanation may be mixed, for even if moral emotions are largely a product of inculcation, one's parents and other relatives may have a disproportionate influence on the inculcation process; to the extent that they act in large part out of self-interest, limits on caring about the welfare of others would tend to be self-perpetuating.

Another matter concerns the capacities for guilt and virtue themselves. From a welfare-maximizing view, a large capacity for feelings of virtue would be highly desirable, for not only could one

[51]There are, of course, limits to the latter, in that societies less successful in ensuring survival, especially in competition with other societies, will tend to die out.

use the large reservoir of virtue better to control behavior, but there is the direct utility benefit from experiencing virtue. Indeed, would it not be a wonderful world if we could all feel incredibly virtuous every time we did not cut in line or each instance in which we refrained from punching someone who was rude? As a matter of survival, however, virtue and guilt may be equally useful, for all that matters is controlling behavior, not whether it is controlled by the prospect of rewards or of punishments.[52] Moreover, if there are limits on the extent to which capacities for moral emotions can be developed, then having a modest pool each for guilt and for virtue rather than, say, only guilt or only virtue (with a total pool of equal size), has the benefit previously described: When individuals can usually be induced to refrain from an undesirable type of act, then little guilt needs to be used to accomplish this (the threat suffices for most); but when individuals cannot usually be induced to commit a good type of act, virtue is superior because much less of it needs to be used.[53]

C. Internal and External Views of Right and Wrong

Most of our discussion refers to feelings of guilt and virtue, which we interpret as internal punishments and rewards for following moral rules. There are, as noted in our introduction and intermittently throughout, corresponding external sanctions and rewards as well, disapprobation and approbation, blame and praise.[54] Until now, we have loosely suggested that these external sanctions (hereinafter taken to include rewards) are encompassed by our analysis; hence, if guilt is properly associated with a particular set of acts, so would disapprobation or blame.

Despite their similarities, a more complete analysis would also take into account the differences between internal and external sanctions. External sanctions require the actions of third parties, sometimes one's victim (or, in the case of helpful acts, beneficiary) and often individuals with little or no direct relationship to the victim (beneficiary). There are three prerequisites for external sanctions to be effective: The individuals imposing the sanctions need information about the actor's behavior; they must be motivated to mete out the sanctions; and the actor must care about others' expressions of blame and praise.[55] The third element seems quite closely related to the internal sanctions and rewards of guilt and

---

[52]If guilt were so extensive and so often experienced that the level of emotional pain induced individuals to commit suicide, the situation would be different, but, short of that, there seems to be little difference between guilt and virtue in this respect.

[53]Other factors would seem to have an evolutionary explanation. Notably, guilt and virtue are part of a larger system of emotions that serves many functions; moral emotions are plausibly an application of this more general system and thus would have its attributes even if they might not be ideal for the task of enforcing moral rules. For example, there may be limits on the extent of our emotions, and extreme emotions might be reserved for acts more directly related to our survival (reproduction, caring for offspring, self-protection).

[54] Smith (1790) devoted significant attention to the similarities and differences between internal and external moral sanctions and rewards.

[55]There are also mixed or intermediate cases. For example, one might feel ashamed, and thus suffer a decline in utility, if others find out about one's act, without others having to engage in any particular behavior (such as

virtue: It would appear that those who would feel guilty committing an act would usually feel badly if others express disapproval, and vice versa. The second element, individuals' motivation to impose sanctions on actors, cannot be taken for granted.[56] One explanation for individuals' motivation in this regard is that the very process by which, for example, guilt may be inculcated for committing a particular type of act would lead an individual to express disapproval of others' commission of the same type of act.[57] The first element, third parties' information about the actor's behavior, is an independent factor; in some contexts, certain third parties will automatically learn about behavior; in others, they may learn about it indirectly, such as through gossip (which itself requires information and motivation).

Although there are important differences between internal and external sanctions, one can still analyze them in a similar manner. To begin with, the expression of disapprobation or approbation is itself an act of sorts, and this type of act could be analyzed much as we analyzed primary acts, such as breaking promises or coming to the aid of others. Rules about when individuals should express disapprobation or approbation may be inculcated, with accompanying moral emotions. For example, failing to express disapproval of someone who behaved badly — perhaps by continuing to greet him cheerfully rather than by shunning him — might itself be a bad act for which guilt (or, perhaps more apt, a proclivity to feel disgust) could be inculcated.[58] And, as previously suggested, there may well be synergies: If guilt is to be inculcated for committing a particular type of act, it may not add much cost, if any, simultaneously to inculcate a sense of disgust at others' commission of that type of act, which in turn would induce one to express disapprobation.[59] Considering our other assumptions, we would suppose that there would be costs associated with inculcating these lessons (which may be small for the reason just given) and, importantly, limits on the ability to employ disapprobation and approbation (for

---

expressing disapprobation) in response to their learning about the act.

[56]Some external sanctions are motivated by ordinary self-interest, such as when one chooses not to deal with a third party known to be unreliable. We view this as distinct from the expression of disapprobation for its own sake, which may include refusal to deal with an unreliable party even when it would be in one's interest to do so in spite of their unreliability. Of course, reputational sanctions motivated by self-interest, narrowly and conventionally understood, sometimes reinforce moral sanctions. (Interestingly, even when reputational sanctions operate, morality may be at work, for the third party's misbehavior is, one supposes, taken as a signal of the degree to which the actor is a moral person. See our discussion of heterogeneity, below.)

[57]These phenomena need not, of course, be the same, and there are independent moral rules that govern expressing approval or disapproval of others' behavior, such as rules about when one should mind one's own business.

[58]Moreover, society might employ external sanctions to enforce third parties' enforcement against primary behavior, and so forth. See, for example, Axelrod (1986), McAdams (1997), and Pettit (1990).

[59]To complete the analogy with regard to acts of expressing disapprobation or approbation, the benefit or harm to third parties associated with the acts would be the acts' effects on welfare through enforcing or undermining, as the case may be, the moral rules that directly govern primary behavior — under the assumption that those subject to blame or praise care about this and accordingly will be induced to comply with moral rules by the prospect of external sanctions. Finally, there may be feelings of guilt and virtue (or other moral sentiments) associated with conveying information about others' behavior.

example, there are undoubtedly limits on the extent to which individuals can be perpetually upset at third parties' behavior and on the ability of individuals to express their disapproval in a manner that influences others).

In sum, although it would be an oversimplification simply to treat internal and external moral rewards and punishments as if they were the same, there are important similarities in how they should be analyzed. We believe, therefore, that there is some basis for our preliminary conjecture that implications of our analysis for the proper use of guilt and virtue will often be suggestive with regard to disapprobation and approbation.

D. Heterogeneity of Individuals

Our discussion for the most part refers to how individuals as a whole would tend to behave. However, individuals may differ, for example, in the extent of their personal gains or losses from committing various acts, how much harm or benefit their acts will cause to others, or their likelihood of being in one or another type of situation. Such differences help to explain why it might be that, when a given moral system is in place, some individuals act morally more often than others.

Another important source of heterogeneity is that different individuals may be differentially susceptible to feelings of guilt and virtue. This could be due to differences in their constitution or differences in their upbringing. Izard (1991) indicates genetic differences in individuals' susceptibility to emotions. With regard to inculcation, since much of it is done by parents or in local institutions, the potential for variation is substantial. Thus, to the extent that one can speak of a social decision — or an evolved tendency — for guilt or virtue of a specific magnitude to be associated with a class of acts, one will be speaking about averages, not about the moral constitution of each and every individual.

The primary effect of heterogeneity on our analysis would be to magnify the impact of the grouping of acts that themselves are heterogeneous. For example, when we described the possibility that a given level of guilt might deter most but not all acts in a given category, one could think of an additional reason being that some individuals, when committing acts in that category, would fail to be deterred, not because their particular situation involves an unusually high level of personal gain from committing the act, but rather because they experience atypically low levels of guilt. Individual heterogeneity combined with the grouping of acts helps to explain why guilt is sometimes experienced and why even modest levels of inculcated virtue will induce some individuals to do desirable acts that most individuals could not be induced to commit even by the prospect of great rewards.

Heterogeneity in the extent to which guilt and virtue are experienced helps to explain other features of observed behavior.[60] Clearly, there are many undesirable acts that very few individuals

[60]Additionally, as suggested in note 56, heterogeneity helps to explain certain responses to others' past behavior, such as refusing to deal with someone who is of an untrustworthy type (which might be translated as the person having little capacity to experience guilt or as the person not having been well inculcated with respect to

would commit, and we sometimes classify individuals who would commit such acts as psychopaths. One possibility is that these individuals have little capacity for feeling guilt. At the other extreme, there are a handful of individuals — such as Mother Theresa — who seem unusually willing to make significant personal sacrifices to help others. One might suppose that such individuals either experience less direct disutility from self-sacrifice or experience stronger feelings of virtue from committing such acts. Finally, we observe that heterogeneity, particularly with regard to different experiences of inculcation, helps to explain the moral disagreement that we observe among individuals in a given society.

E. Prudence

Many acts involve no (or only trivial) effects on third parties. Accordingly, there would seem to be no role for the use of guilt and virtue to regulate them because, in the absence of moral sanctions, individuals would commit such acts if and only if their own benefit from doing so was positive, and this behavior would be socially best from a consequentialist (welfarist) perspective. Nevertheless, discussions of virtue and vice over the ages have often included categories of acts that seem to involve no obvious effect on others, and psychologists indicate that individuals experience guilt when they act in ways that harm themselves. See Izard (1991). For example, individuals may be urged to save for a rainy day, not to overeat, and otherwise to protect themselves from their own folly, and individuals who fail to do so may feel guilty.

Can one offer a consequentialist explanation for the use of a system of morality — or at least of mechanisms that seem very similar to such a system — for the regulation of self-regarding acts? One possibility is that third parties are affected after all. Others may feel badly when individuals act in ways that harm themselves; moreover, such others might be motivated to expend resources to aid those who have fallen victim to their own imprudence. Indeed, some level of general altruism may be supported by the moral sentiments themselves. Moreover, certain inculcators, notably parents, will feel altruistically toward their children and thus be motivated to use available means, including the inculcation of moral rules, to encourage more prudent behavior.

Another explanation is that individuals may lack self-control. (This explanation is particularly important because it constitutes a reason that imprudent behavior might arise in the first place.) In particular, many instances in which guilt and virtue seem to be associated with self-regarding behavior involve problems of myopia. As Schelling (1984) and others have suggested, these problems can be

---

certain moral rules). This, in turn, can explain certain signalling behavior and, relatedly, our tendency to make associational decisions based on what may otherwise seem to be irrelevant characteristics, such as whether a prospective business associate is philanthropic or sexually abuses subordinates. Compare, for example, Posner (2000).

Yet another possibility is that an analysis that incorporates heterogeneity could address how others' behavior influences an individual's susceptibility to the moral emotions. For example, it may be more difficult to inculcate or maintain the effectiveness of guilt for committing an act — as well as a social practice of expressing disapprobation — if too many other individuals commit the act.

thought of as involving two selves — in the case of myopia, a present self whose decisions affect a future self. Under such a formulation, the behavior of the present self does affect another — the future self — and hence our analysis suggesting the potential benefits of employing guilt and virtue can be applied. For example, the prospect of guilt deters the present self from harming the future self. As a consequence, we do not regard subjecting such personal choices to the same type of moral mechanisms used for activity affecting others as inconsistent with our analysis of moral rules. It remains, however, to consider the extent to which the actual association of morality (or a moral-like system) to matters of prudence is consistent with the implications of our analysis.

VI. COMMENTS ON SOME PARTICULAR DEBATES

In this section, we describe how the foregoing analysis may help to illuminate certain familiar debates, mainly between deontologists and consequentialists (often, utilitarians). As we have stated previously, nothing we will say here purports to resolve any of these disputes, our comments being limited to how particular components of the pertinent discussions might be recast.

A. Acting from Self-interest versus Acting from Moral Obligation

Many writers, following Kant (1785), are emphatic that acting out of a sense of obligation or duty is distinct from acting out of self-interest.[61] This, of course, raises the question addressed by Hume (1751), Mill (1861), and Sidgwick (1907), among others: If an act is against self-interest and is nevertheless committed because it is morally the right thing to do, what is it that, as a positive matter, can explain such behavior? Of course, an important possibility is that the moral sentiments — feelings of guilt and of virtue and, relatedly, concern for the disapprobation or approval of others — provide the explanation. When the utility effects of the moral emotions outweigh the utility associated with the act per se, individuals will behave differently. As we discuss at some length in subsection II(A), this view can be reinterpreted in a number of ways that render it substantially consistent with seemingly different understandings of individuals' conscious experience of moral decisionmaking, such as the view that individuals who "do their duty" do not obtain pleasure thereby, but rather they feel compelled to do the morally correct act.[62]

---

[61]See, for example, the sources cited in note 10.

[62]We believe that our discussion of an ideal consequentialist moral system is also pertinent to one particular strand of the debate on whether moral motivation can be subsumed under the concept of self-interest (with the latter construed sufficiently broadly). It is often believed that proponents of the self-interest view rest their arguments on positive utility, such as from altruistic feelings toward others or simply from the feeling of virtue associated with doing the right thing. Against this argument, it is suggested (plausibly, in our view) that individuals would in fact have preferred that the situation, in which they have to sacrifice "ordinary" personal gain in order to comply with the dictates of morality, had never arisen. Thus, it is argued that it is not any personal benefit from doing one's duty that motivates moral behavior. This response, however, ignores an alternative, plausible interpretation: Perhaps it is not pleasure that motivates moral behavior in such situations, but rather the desire to avoid pain; that is, the moral rules in question may be enforced by guilt rather than by virtue. (And, indeed, most of the moral rules that are addressed in this literature seem to be those that are enforced by guilt.) Clearly, if one would,

B. The Independent Importance of Acting Morally

A common objection to consequentialist accounts of morality is that they cannot make sense of our intuition that compliance with the dictates of morality has weight independent of the consequences of our actions. For example, Ross (1930) suggests that an individual following a consequentialist morality would be indifferent to whether a promise should be kept when the balance of benefits and harm was precisely equal, whereas our moral intuition is that at least some weight should be accorded to keeping the promise. Arguments about whether consequentialism can account for our instinct that promise-keeping is independently important often involve consequentialists identifying indirect consequences of breaking promises (such as by setting a bad example that will affect others' behavior) and deontologists posing hypothetical examples in which such effects are absent (a promise to a dying person that will never become known to anyone else) or suggesting that such additional effects be subsumed in the balance of benefits and harm and asking whether our intuition about promise-keeping still seems to carry weight.

Without entering into the particulars of prior debates, such as that about promise-keeping, we nevertheless observe that our framework of analysis has a straightforward implication regarding whether consequentialism can account for the moral intuition that morality has independent weight. Specifically, in the consequentialist moral system that we have described, anything morally prohibited is associated with feelings of guilt (and possibly shame and the prospect of being subject to disapprobation) and anything morally encouraged is associated with feelings of virtue (and an expectation of approbation). If this scheme has even modest descriptive accuracy, then an explanation has been offered: The tendency for moral behavior to be associated with moral sentiments is not contingent on an independent consequentialist assessment of such behavior, but rather is associated with all behavior subject to the pertinent moral rule. Moreover, when moral rules operate at a categorical level, all acts in the relevant category — such as that consisting of situations in which we decide to break a promise — will lead, say, to the experiencing of guilt, even if the act happens to be, on balance, socially neutral or desirable in its consequences. Thus, in the sort of situation posed by Ross, in which the direct effects of breaking a promise are neutral, it would be true under a consequentialist system of morality that promise-keeping would be favored — by the weight of the guilt associated with breaking a promise. Indeed, many who have written about the moral sentiments in times past, such as Sidgwick (1907), suggest a link between moral sentiments and moral instincts and intuitions. And to the extent that moral emotions tend to have an autonomous, not entirely conscious character, being triggered by particular actions and being anticipated by the mere contemplation thereof, their nature does seem similar to that of an instinct or intuition of the sort that Ross invokes.

---

in a given situation, be induced to sacrifice personal benefit to avoid guilt feelings of a greater magnitude, one would wish that the situation had never arisen. Hence, understanding individuals' moral behavior to be motivated by considerations of guilt and virtue and by related concerns about disapprobation and approbation is not at all inconsistent with the common suggestion that individuals would prefer that situations requiring painful moral choices never arise.

C. Moral Intuitions, Counter-Examples, and the Generality of Moral Rules

A common form of argument against consequentialism involves offering specific (and often unusual) examples in which its implications seem inconsistent with our moral instincts or intuitions. For example, Williams (1973) presents a case in which an individual would have to take one person's life to save others (Jim must kill one captive, or else Pedro will kill twenty). Or, as many have posed, a sheriff may frame an innocent person to satisfy an angry mob that otherwise will riot and kill many. Again, many arguments pro and con have been offered concerning which act in each situation really is implied by consequentialism, with consequentialists giving reasons why what seems intuitively right (not taking a life, not framing an innocent person) would in fact be best in terms of its effects and with deontologists disagreeing or offering modified hypothetical examples in which certain effects would be absent, thereby restoring the tension.

As some writers who advance two-level views have noted, however, there is another sort of response available under a consequentialist view.[63] Namely, since moral rules must, as we have suggested here, be general in nature, it is inevitable that there will be instances of conflict between what such rules command and what act would actually be best in a given situation. Thus, as we have suggested, there may be a prohibition on all lies (or all lies in a given class) even though, given inevitable heterogeneity, there will be some lies in the class that would have on-balance desirable consequences. Likewise, we may suppose that killing an innocent person and framing an innocent person are each generally prohibited categories of acts. This suggests that committing either act— even in atypical circumstances in which the act might (as deontologists offering certain examples assume would be the case) have overall positive effects — would be associated with feelings of guilt and shame and expectations of disapprobation and thus would grate against our moral intuition.

In this regard, it is worth recalling some of our previous discussion about our assumptions and their rationale. As we explain in subsection II(D), a moral system's operation at the level of categories of acts — in which case-specific exceptions may not freely be made when consequences would be atypical — is an unavoidable aspect of human nature and, given this fact, a consequentialist system would not attempt to function otherwise. Relatedly, we have argued that feelings of guilt tend to be automatic, not contingent on the particulars of the act in question. And, in discussing external moral sanctions, we observed that onlookers need to be induced to express disapprobation, which may well be accomplished by making them such that they would feel guilty if they fail to do so. Indeed, the very tendency to express disapprobation upon hearing of another's violation of a general moral rule would lead one to predict that any well-socialized individual would have a negative reaction to the very sorts of examples commonly presented in debates about consequentialism. And, as we note in the preceding subsection, the nature of our moral sentiments seems similar in important respects to that of moral instincts and intuitions.

[63]See, for example, Harrod (1936), Rawls (1955), Sartorius (1972), and Hare (1981).

In sum, it is a hallmark of two-level consequentialist accounts that it is simultaneously possible for an act in a specific situation to be overall desirable as an ideal matter and to be prohibited by a sound moral rule. Moreover, in the sort of consequentialist system that we have outlined here, the latter carries the implication that the actor will feel guilty and that onlookers will feel negatively about the act (even if all realize that the balance of effects of positive). Finally, given the general nature of moral rules, which apply to groups of acts rather than act-by-act, it is inevitable that some of the time these sorts of conflicts will arise. In fact, since the moral system must operate at a categorical level given the inherent constraints of human nature, it will inevitably be possible to identify such conflicts, and any consequentialist who purports to be clever enough to refute all such counter-examples should not be believed.

D. The Problem of Unlimited Individual Obligations under Consequentialism

In this subsection and the next, we consider two specific conundrums that are often posed in debates between consequentialists (utilitarians, in particular) and deontological moral philosophers. One critique raised by the latter group against the former is that a consequentialist or utilitarian moral criterion is too demanding: Everyone would have to be a Mother Theresa; individuals in richer countries would have to donate most of their income to help poor people in less developed nations; and so forth. Such implications, it is said, are inconsistent with our moral intuitions, which in turn is said to demonstrate that consequentialist moral philosophy is fundamentally defective.[64] Implicit or explicit in such arguments is the claim that our moral intuitions constitute (or at least indicate some of the contours of) an ideal moral system.

Our analysis, as we have already suggested, offers an answer to this question, that is, an explanation for how a consequentialist view can be reconciled with our seemingly inconsistent moral intuitions. According to the posited consequentialist criterion, it would be a good thing if individuals behaved as stated. But, given human nature, it is unrealistic to expect this. Thus, even if one attempted to inculcate a high degree of guilt for failing to make substantial sacrifices to help others, it would probably be insufficient to induce most individuals to so behave. The consequence would be that more individuals would frequently suffer guilt, which is itself undesirable; moreover, the crowd-out of guilt in other realms would impose serious social costs. Accordingly, it is not advantageous to use guilt to encourage such behavior. In other words, a proper consequentialist analysis would oppose, not favor, deeming the failure to engage in such highly altruistic behavior as a moral wrong.

Our consequentialist analysis further suggests that it may be appropriate to use virtue: Such use may encourage some highly desirable acts and, because most will not act accordingly, there will be little depletion of the limited capacity for virtue and thus little social cost with regard to regulating other behavior. Hence, on consequentialist grounds, one would indeed not use guilt but instead use virtue.

---

[64]See, for example, Williams (1973). See Hooker (2000), Kagan (1989), and Murphy (2000) for further discussions (and Brandt 1996 for brief remarks) from a consequentialist and utilitarian perspectives and Heyd (1982) for a broad exploration of supererogation under a range of religious and moral theories (including utilitarianism).

This conclusion as well is consistent with the categorization of such acts as ones that individuals are not morally obligated to perform, but that it would nevertheless be morally worthy to do.

Finally, it is useful to draw a contrast by considering the particular case of the act of helping others in distress when there would be little cost to the actor — for example, calling for help or rescuing someone when only slight inconvenience would be involved. Here, many non- (or at least non-strict) consequentialists would seem to agree that, when we consult our moral intuitions, it seems that there is (or at least may be) a moral duty to act. This view is also suggested by our framework: One would probably use guilt to induce such behavior because little guilt would required and because, since most individuals would be induced to behave properly, there would be little use of the scarce reservoir of guilt.

Combining these two cases, what has appeared to be a puzzle — reconciling our moral intuition that tells us that there is no duty in the former case with our intuition that there is a duty in the latter — can be solved by a more explicit consequentialist analysis of how guilt and virtue should be employed, rather than focusing exclusively on the characteristics of the acts themselves, whether on their consequences or on the intrinsic properties the acts may be believed to have.[65]

E. The Act/Omission Distinction

Another set of debates concerns the act/omission distinction (and related doctrines).[66] The

---

[65]Relatedly, many philosophers suppose that there must be a qualitative distinction between the acts — since there is moral obligation in one case and not the other — whereas there seem to be only differences in degree, namely, in the cost of self-sacrifice. (Sometimes these differences are nevertheless described in qualitative terms, such as mere inconvenience versus disturbances to the integrity of one's self-defined mission in life.) In our analysis, however, differences in degree can translate into differences in kind. In particular, as the fraction of individuals who will not behave in an ideal manner increases, at some point it is no longer desirable to employ guilt. Such a change can be discontinuous in the formal sense, which is consonant with viewing the difference as one in kind.

[66]Debate over the act/omission distinction is closely related to that concerned with the doctrine of double effect, the doctrine of doing and allowing, and other principles that draw similar distinctions. For differing views, see, for example, Bennett (1995), Foot (1967), Kagan (1989), and Williams (1973). For differing views on whether the intuition underlying the distinction can be explained by Kahneman and Tversky's prospect theory, see Horowitz (1998) and Kamm (1998). For the suggestion that the distinction involves overgeneralization of an otherwise useful decisionmaking heuristic, see, for example, Spranca, Minsk, and Baron (1991). Interestingly, recent evidence from brain scans indicates that it is emotional areas of the brain that are activated when individuals presented with certain classic moral dilemmas decide to refrain from acts that cause harm but nevertheless produce greater good. See Greene et al. (2001).

We note that the act/omission distinction relates to the foregoing subject of unlimited individual obligations under consequentialism because many moral theorists attribute that alleged problem to consequentialism's failure (inability) to distinguish between acts and omissions, whereas if duties are largely limited to affirmative acts, individuals' obligations can more readily be limited. See Bennett (1995), who discusses the failure of attempts by the classical utilitarians to deal with the problem.. Subsequently, as discussed in Heyd (1982), a number of theorists have considered whether "negative utilitarianism" (under which some variant of the

problem, as often put, is that there is another important type of conflict between our moral intuitions and the implications of a consequentialist approach. Namely, one can consider two situations in which acting in the first situation has the same consequences as not acting (an omission) in the second. Moreover, in some such cases, our moral intuitions would distinguish the two situations, forbidding the act in one situation but permitting (or requiring) the corresponding omission in the other, such as when the act and omission both raise welfare — perhaps five lives would be saved at the expense of one. It is suggested that our moral intuition is such that it is impermissible to act to take one life in order to save five, whereas it would be permissible (or even mandatory) to refrain from an act if abstention would result in one death but would save five (that is, one should not act to save one if doing so would kill five).

Our framework offers a number of possible ways to reconcile this seemingly inconsistent character of common morality with a consequentialist moral system. One explanation involves the grouping of acts: As observed above, the groupings that naturally arise are based upon characteristics of different types of behavior that relate to how we perceive the world and organize it in our minds, and these need not correspond to the groupings that would be ideal from the perspective of formulating moral rules. Thus, a particular act may be condemned not because it is itself socially undesirable, but because it shares characteristics with other acts that together form a cluster for which it makes sense to inculcate guilt.[67] For example, the rare situation in which killing one individual will save five may be grouped in the general category of killing, which does not typically raise social welfare.

Another possibility, which we have not raised previously, concerns the stimuli necessary to trigger feelings of guilt and virtue.[68] Because these emotions need to be reasonably automatic to function, it must be that they are experienced not as a result of careful contemplation and reflection, but rather due to particular patterns being identified. Perhaps it is the case that acts are more naturally capable of triggering feelings of guilt and virtue than are omissions, for committing an act may result in a more identifiable stimulus than an omission. A related problem is that many omissions are ongoing (failure to devote more resources to helping the poor) — every instant in which we could have acted but did not is an omission — whereas we do not have the capacity to experience constant flows of guilt or virtue that will register in a meaningful way.[69] Hence, the underlying psychology of the operation of

---

act/omission distinction is embraced) can be a plausible or appealing moral theory.

[67]Unger (1996) advances a similar argument.

[68]A third, related point is that, to the extent that the labeling of behaviors as involving acts or omissions (is failing to hold open a door for the next person an omission, or the commission of the act of releasing the door so as to impose the risk of injury on another?) may be inculcated at the same time guilt and virtue are being inculcated, it may be that we encourage people to perceive as distinct acts only those behaviors for which we are going to inculcate guilt or virtue. Thus, whenever our analysis suggests that it is not advantageous to employ guilt, there is little point in teaching individuals to recognize the corresponding undesirable behavior as a distinctive act.

[69]Of course, some omissions are distinctive, so the failure to call for help when one sees a person drowning is different from the ongoing failure to donate half of one's income to charity. Bennett (1995) offers further examples

moral systems in the human mind may impose important constraints on their use that help to explain the system of moral rules and the use of moral sentiments that we observe, but in a manner that does not imply that behaviors are benign (from an ideal perspective) simply because they are not associated with guilt or virtue.

VII. CONCLUSION

The purpose of this article is to illuminate our understanding of morality by attempting to answer the question of what moral system would be best from a consequentialist perspective. We posited a number of assumptions about human nature — about the motivation to follow moral rules, the process of inculcating principles of right and wrong, limits on the extent to which individuals can be induced to behave morally, and the need for moral rules to be formulated in a categorical manner. Given those assumptions, we derived a number of properties of a consequentialist moral system. These included the inculcation of feelings of guilt and virtue, and use of disapprobation and approbation, so as to induce individuals to behave better than they would if guided solely by their narrow self-interest; the tendency of moral rules, right and wrong, not to correspond perfectly to ideal behavior (so that, for example, guilt will be associated with some desirable acts); and the determination of which undesirable behavior should be morally prohibited, rather than permitting it but morally rewarding contrary, desirable behavior. We also described how our conclusions need to be amended in light of a number of further considerations.

Our results, we have suggested, are potentially helpful in explaining the moral systems that we observe in various societies, may be normatively relevant (to the extent one endorses a consequentialist approach, a subject we do not address here), and can illuminate (though not resolve) certain arguments in longstanding debates in moral theory. Each of these claims is tentative and qualified, given the preliminary nature of our analysis, uncertainty surrounding our assumptions about human nature, various reasons to doubt that observed moral systems would be optimal, and so forth.[70] Nevertheless, we believe that it is valuable to ask the central question that we pose, that there is reason to believe that the general features of our main conclusions are robust to some extent, and that these conclusions do seem to offer some insight into the subjects that we address. In all, there seems to be basis for pursuing this

---

in which what would seem to be omissions, as ordinarily defined, are viewed in moral discussions as if they were acts. And, as noted in the preceding discussion, it seems that some such omissions — notably, failing to aid another when the sacrifice to oneself is trivial and the benefit to the third party is great — are indeed associated with feelings of guilt.

[70]One particular further qualification is that we have considered what would be the optimal consequentialist system of morality under the implicit assumption that there is no other means of regulating behavior, whereas in fact there is the legal system (which we take to encompass civil and criminal law, regulation, taxes and subsidies, and any other government apparatus). In a more complete analysis, the optimal moral system may be different from what we derive here. Moreover, in such an expanded setting, one would have a distinction between legality and illegality, to go along with the distinctions we have discussed between acts that are moral and immoral (right and wrong) and acts that are socially desirable and undesirable; just as the latter two distinctions are not, in general, the same under our analysis, each of the three distinctions may sometimes differ in the broader setting.

type of investigation, wherever it may ultimately lead.

References

Alexander, Richard D. 1987. *The Biology of Moral Systems.* New York: Aldine de Gruyter.

Anderson, Elizabeth. 2000. Beyond Homo Economicus: New Developments in Theories of Social Norms. *Philosophy and Public Affairs* 29: 170-200.

Austin, John. 1832. *The Province of Jurisprudence Determined*. Wilfrid E. Rumble, ed. Cambridge: Cambridge University Press (1995).

Axelrod, Robert. 1986. An Evolutionary Approach to Norms. *American Political Science Review* 80: 1095-1111.

Barash, David P. 1982. *Sociobiology and Behavior*. Elsevier: New York, 2nd ed.

Barkow, Jerome H., Leda Cosmides, and John Tooby, eds. 1992. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford: Oxford University Press.

Baron, Jonathan. 1994. *Thinking and Deciding*. Second edition. Cambridge: Cambridge University Press.

Becker, Gary S. 1996. *Accounting for Tastes*. Cambridge: Harvard University Press.

Ben-Ner, Avner, and Louis Putterman, eds. 1998. *Economics, Values, and Organization*. Cambridge: Cambridge University Press.

Bennett, Jonathan. 1995. *The Act Itself*. Oxford: Oxford University Press.

Brandt, Richard B. 1954. *Hopi Ethics: A Theoretical Analysis*. Chicago: University of Chicago Press.

Brandt, Richard B. 1979. *A Theory of the Good and the Right*. Oxford: Oxford University Press.

Brandt, Richard B. 1996. *Facts, Values, and Morality.* Cambridge: Cambridge University Press.

Campbell, Donald T. 1975. On the Conflicts Between Biological and Social Evolution and Between Psychology and Moral Tradition. *American Psychologist* 30: 1103-26.

Cosmides, Leda, and John Tooby. 1994. Better than Rational: Evolutionary Psychology and the Invisible Hand. *American Economic Association Papers and Proceedings* 84: 327-32.

Daly, Martin, and Margo Wilson. 1988. *Homicide*. New York: A. de Gruyter.

Damasio, Antonio. 1994. *Descartes's Error: Emotion, Reason and the Human Brain*. New York: Putnam.

Darwin, Charles. 1872. *The Expression of the Emotions in Man and Animals*. Paul Ekman, ed., third edition. Oxford: Oxford University Press (1998).

Darwin, Charles. 1874. *The Descent of Man; and Selection in Relation to Sex*. Second edition. Amherst, NY: Prometheus Books (1998).

de Waal, Frans. 1996. *The Origins of Right and Wrong in Humans and Other Animals*. Cambridge: Harvard University Press.

Ellickson, Robert. 1991. *Order without Law: How Neighbors Settle Disputes*. Cambridge: Harvard University Press.

Elster, Jon. 1999. *Alchemies of the Mind: Rationality and the Emotions*. Cambridge: Cambridge University Press.

Foot, Philippa. 1967. The Problem of Abortion and the Doctrine of Double Effect. *Oxford Review* 5: 5-15.

Foot, Philippa. 1972. Reasons for Action and Desires II. *Proceedings of the Aristotelian Society* Supp. Vol. 46: 203-210.

Frank, Robert H. 1988. *Passions within Reason*. New York: W.W. Norton & Co.

Frederick, Shane, and George Loewenstein. 1999. Hedonic Adaptation. In Daniel Kahneman, Ed Diener, and Norbert Schwarz, eds., *Well-Being: The Foundations of Hedonic Psychology*. New York: Russell Sage Foundation.

Gibbard, Allan. 1990. *Wise Choices, Apt Feelings; A Theory of Normative Judgment*. Cambridge: Harvard University Press.

Gowans, Christopher, ed. 1987. *Moral Dilemmas*. New York: Oxford University Press.

Greene, Joshua D., R. Brian Sommerville, Leigh E. Nystrom, John M. Darley, and Jonathan D. Cohen. 2001. An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science* 293: 2105-08.

Haidt, Jonathan. 2001. The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment. *Psychological Review* 108: 814-34.

Hare, R.M. 1981. *Moral Thinking: Its Level, Method, and Point*. Oxford: Oxford University Press.

Harrod, R.F. 1936. Utilitarianism Revised. *Mind* 45: 137-56.

Hechter, Michael, and Karl-Dieter Opp, eds. 2001. *Social Norms.* New York: Russell Sage Foundation.

Heyd, David. 1982. *Supererogation: Its Status in Ethical Theory*. Cambridge: Cambridge University Press.

Hooker, Brad. 2000. *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality.* Oxford: Oxford University Press.

Horowitz, Tamara. 1998. Philosophical Intuitions and Psychological Theory. *Ethics* 108: 367-385.

Hume, David. 1739. *Treatise of Human Nature*. Buffalo: Prometheus Books (1992).

Hume, David. 1751. *An Enquiry Concerning the Principles of Morals*. Tom L. Beauchamp, ed. Oxford: Oxford University Press (1998).

Hutcheson, Francis. 1725-1755. *Philosophical Writings*. R.S. Downie, ed. London: J.M. Dent (1994).

Izard, Carroll E. 1991. *The Psychology of Emotions.* New York: Plenum Press.

Kagan, Jerome. 1984. *The Nature of the Child*. New York: Basic Books.

Kagan, Shelly. 1989. *The Limits of Morality.* Oxford: Oxford University Press.

Kamm, F.M. 1998. Moral Intuitions, Cognitive Psychology, and the Harming-versus-Not-Aiding Distinction. *Ethics* 108: 463-488.

Kant, Immanuel. 1785. *Groundwork of the Metaphysics of Morals*. Translated and edited by Mary Gregor, Cambridge: Cambridge University Press (1997).

Kaplow, Louis, and Steven Shavell. 2001. Moral Rules and the Moral Sentiments: Toward a Theory of an Optimal Moral System. John M. Olin Center for Law, Economics, and Business, Harvard Law School, Discussion Paper No. 342.

Kaplow, Louis, and Steven Shavell. 2002. *Fairness versus Welfare*. Cambridge: Harvard University Press.

Kosslyn, Stephen M., and Olivier Koenig. 1992. *Wet Mind: The New Cognitive Neuroscience*. New York: Free Press.

LeDoux, Joseph E. 1996. *The Emotional Brain: The Mysterious Underpinnings of Emotional Life.* New York: Simon & Schuster.

Lewontin, R.C., Steven Rose, and Leon J. Kamin. 1984. *Not in Our Genes: Biology, Ideology, and Human Nature*. New York: Pantheon Books.

Mackie, J.L. 1985. *Persons and Values: Selected Papers, Volume II.* Joan Mackie and Penelope Mackie, eds. Oxford: Oxford University Press.

McAdams, Richard H. 1997. The Origin, Development, and Regulation of Norms. *Michigan Law Review* 96: 388-433.

Mill, John Stuart. 1861. *Utilitarianism*. Edited by Roger Crisp, New York: Oxford University Press (1998).

Miller, Joan G. 2001. Culture and Moral Development. In David Matsumoto, ed., *The Handbook of Culture and Psychology*. New York: Oxford University Press.

Murphy, Liam B. 2000. *Moral Demands in Nonideal Theory*. Oxford: Oxford University Press.

Nisbett, Richard E., and Dov Cohen. 1996. *Culture of Honor: The Psychology of Violence in the South.* Boulder: Westview Press.

Pettit, Philip. 1990. *Virtus Normativa*: Rational Choice Perspectives. *Ethics* 100: 725-55.

Pinker, Steven. 1997. *How the Mind Works.* New York: W.W. Norton & Co.

Posner, Eric A. 2000. *Law and Social Norms.* Cambridge: Harvard University Press.

Posner, Richard A., and Eric B. Rasmusen. 1999. Creating and Enforcing Norms, with Special Reference to Sanctions. *International Review of Law and Economics* 19: 369-82.

Rawls, John. 1955. Two Concepts of Rules. *Philosophical Review* 64: 3-32.

Ross, W.D. 1930. *The Right and the Good*. Oxford: Oxford University Press.

Sartorius, Rolf. 1972. Individual Conduct and Social Norms: A Utilitarian Account. *Ethics* 82: 200-18.

Scheffler, Samuel. 1992. *Human Morality.* New York: Oxford University Press.

Schelling, Thomas C. 1984. *Choice and Consequence*. Cambridge: Harvard University Press.

Sen, Amartya K. 1977. Rational Fools: A Critique of the Behavioral Foundations of Economic Theory. *Philosophy and Public Affairs* 6: 317-44.

Shavell, Steven. 2002. Morality versus Law as Regulators of Conduct. *American Law and Economic Review* 4 (forthcoming).

Sidgwick, Henry. 1897. Law and Morality. In *The Elements of Politics*. Second Edition. Reprinted in Henry Sidgwick, *Essays on Ethics and Method*, Marcus G. Singer, ed. Oxford: Oxford University Press (2000).

Sidgwick, Henry. 1907. *The Methods of Ethics*. Seventh edition. Indianapolis: Hackett Publishing Company (1981).

Smart, J.J.C. 1973. An Outline of a System of Utilitarian Ethics. In J.J.C. Smart and Bernard Williams, eds., *Utilitarianism: For and Against*. Cambridge: Cambridge University Press.

Smith, Adam. 1790. *The Theory of the Moral Sentiments.* Sixth edition. Oxford: Oxford University Press (1976).

Spranca, Mark, Elisa Minsk, and Jonathan Baron. 1991. Omission and Commission in Judgment and Choice. *Journal of Experimental Social Psychology* 27: 76-105.

Statman, Daniel. 1995. *Moral Dilemmas*. Atlanta: Rodopi.

Sunstein, Cass. 1996. Social Norms and Social Roles. *Columbia Law Review* 96: 903-68.

Tangney, June Price, and Kurt W. Fischer, eds. 1995. *Self-Conscious Emotions: The Psychology of Shame, Guilt, Embarrassment, and Pride*. New York: Guilford Press.

Tooby, John, and Leda Cosmides. 1990. On the Universality of Human Nature and the Uniqueness of the Individual: The Role of Genetics and Adaptation. *Journal of Personality* 58: 17-67.

Trivers, Robert L. 1971. The Evolution of Reciprocal Altruism. *Quarterly Review of Biology* 46: 35-57.

Unger, Peter. 1996. *Living High and Letting Die: Our Illusion of Innocence*. New York: Oxford University Press.

Williams, Bernard. 1973. A Critique of Utilitarianism. In J.J.C. Smart and Bernard Williams, eds., *Utilitarianism: For and Against*. Cambridge: Cambridge University Press.

Williams, Bernard. 1981. *Moral Luck: Philosophical Papers, 1973-1980*. Cambridge: Cambridge University Press.

Wilson, Edward O. 1975. *Sociobiology*. Cambridge: Harvard University Press.

Wilson, James Q. 1993. *The Moral Sense*. New York: Simon & Schuster.

Woods, Michael. 1972. Reasons for Action and Desires I. *Proceedings of the Aristotelian Society* Supp. Vol. 46: 189-201.