VALUE ANALYSIS OF POLITICAL BEHAVIOR—SELF-INTERESTED : MORALISTIC :: ALTRUISTIC : MORAL

JONATHAN BARON[†]

I distinguish four types of goals: self-interested, altruistic, moralistic, and moral. Moralistic goals are those that people attempt to impose on others, regardless of the others' true interests. These may become prominent in political behavior such as voting because such behavior has relatively little effect on self-interested goals. I argue (sometimes with experimental evidence) that common decisional biases concerned with allocation, protected values, and parochialism often take the form of moralistic values. Because moralistic values are often bundled together with other values that are based on false beliefs, they can be reduced through various kinds of reflection or "de-biasing."

> If your morals make you dreary, depend upon it they are wrong. I do not say "give them up," for they may be all you have; but conceal them like a vice . . . .[1]

INTRODUCTION

The quality of government is a major determinant of the quality of people's lives. Underdeveloped countries suffer from misguided economic policies, such as price controls and subsidies, overinvestment in arms and underinvestment in health and education, excessive regulation, and corruption. Developed countries also suffer from unbalanced national budgets, inefficient subsidies, and other ills, but the developed countries generally have better government. World problems, such as depletion of fisheries, suffer from a lack of regulation and international institutions.[2]

[†] Useful comments were provided by Matthew Adler, Steven Shavell, students in a class at Harvard Law School, participants in the University of Pennsylvania Decision Processes Brown Bag, and attendees of the University of Pennsylvania Law School Symposium on Preferences and Rational Choice. The experiments were supported by grants from the National Science Foundation and the Russell Sage Foundation.

[1] ROBERT LOUIS STEVENSON, A CHRISTMAS SERMON 19 (1906).

[2] *See* JONATHAN BARON, JUDGMENT MISGUIDED: INTUITION AND ERROR IN PUBLIC DECISION MAKING 1-20 (1998) (arguing that government decision making is flawed and advocating for improvement in the way decisions are made), *available at* http://www.sas.upenn.edu/~baron/vbook.htm; MAX H. BAZERMAN ET AL., YOU CAN'T ENLARGE THE PIE: SIX BARRIERS TO EFFECTIVE GOVERNMENT, at xvi-xx (2001) (presenting six ways in which government decision making causes societal harm).

In the long run, and sometimes in the short run, government policy depends on the political actions and omissions of citizens. But the individual citizen typically has little influence. Thus, one of the most important determinants of our well-being is controlled almost completely by our collective behavior but is affected very little by our individual behavior. This situation contrasts with that in a free market for goods and services, where our well-being is also affected by our decisions, but individual decisions have a direct effect on individual outcomes. In political behavior—such as voting, responding to polls, trying to influence others, writing letters, or making contributions of time and money—our little actions are pooled together to make one huge decision that affects us all.

This situation opens up the possibility that political behavior is less sensitive to its consequences for the decision maker, and more sensitive to other factors such as emotions and, I shall argue, moral (or moralistic) principles.[3] Political behavior might therefore be more subject to decisional biases, fallacies, and errors than is market-oriented behavior.[4] This insensitivity to consequences, if it happens, is interesting for the study of rational decision making. It raises special questions, and I want to address some of these issues. First, I describe a distinction among types of values, which is the main point of this Article. Next, I summarize some experimental evidence about how people think about their values.

I assume utilitarianism as a normative theory, as a standard against which I compare people's judgments and decisions. This theory is defended elsewhere.[5] Utilitarianism enjoins us to make decisions that

---

[3] For discussions of the idea that political behavior is "expressive," see GEOFFREY BRENNAN & ALAN HAMLIN, DEMOCRATIC DEVICES AND DESIRES 30-33 (2000); GEOFFREY BRENNAN & LOREN LOMASKY, DEMOCRACY AND DECISION: THE PURE THEORY OF ELECTORAL PREFERENCE 35-38 (1993).

[4] *See* Colin F. Camerer, *Do Biases in Probability Judgment Matter in Markets? Experimental Evidence*, 77 AM. ECON. REV. 981, 981-97 (1987) (giving an example of how market discipline can reduce biases in repeated situations). *But cf.* Jane Beattie & Graham Loomes, *The Impact of Incentives upon Risky Choice Experiments*, 14 J. RISK & UNCERTAINTY 155, 165-66 (1997) (testing the hypothesis that financial incentives are "not a crucial factor in encouraging" particular choices in pairwise choice problems, and concluding that there may be situations in which incentives affect the decision-making process); Colin F. Camerer & Robin M. Hogarth, *The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework*, 19 J. RISK & UNCERTAINTY 7, 7-42 (1999) (analyzing experimental data and concluding that incentives often do not improve performance).

[5] *See* JONATHAN BARON, MORALITY AND RATIONAL CHOICE (1993) [hereinafter BARON, MORALITY] (countering arguments that utilitarianism cannot explain people's decisions); JOHN BROOME, WEIGHING GOODS: EQUALITY, UNCERTAINTY AND TIME 1-21

produce the best consequences on the whole, for everyone, balancing the gains and losses. If we fall short of this standard, then we make decisions that make someone worse off—relative to the standard—without making anyone better off to an extent sufficient to balance the harm. Utilitarianism, then, has at least this advantage: if we want to know why decisions sometimes yield consequences that are less good than they could be on the whole, one possible answer is that people are making decisions according to principles that are systematically nonutilitarian, and they are getting results that follow their principles rather than results that yield the best outcome.[6]

A. *Types of Goals*

Decisions are designed to achieve goals, or (in other words) objectives or values. Goals may be understood as criteria for the evaluation of decisions or their outcomes.[7] I use the term "goals" very broadly to include (in some senses) everything a person values. Behavior is rational to the extent to which it achieves goals. Rational political behavior will not happen without goals.

Goals, in the sense I describe, are criteria for evaluating states of affairs. They are reflectively endorsed. They are the result of thought and are, in this sense, "constructed" in much the way that concepts are the result of reflection. Goals are not simply desires, and very young children might properly be said to have no goals at all, in the sense at issue. Your goals fall into four categories: self-interested, altruistic, moralistic, and moral. These correspond to a two-by-two classification. One dimension of this classification is whether or not your goals de-

---

(1991) (defending multiple assertions about the structure of "good," including utilitarianism); RICHARD M. HARE, MORAL THINKING: ITS LEVELS, METHOD AND POINT 1-24 (1981) (arguing for a form of utilitarianism); Jonathan Baron, *Norm-Endorsement Utilitarianism and the Nature of Utility*, 12 ECON. & PHIL. 165, 165-82 (1996) [hereinafter Baron, *Norm-Endorsement*] (rejecting moral intuition and arguing that utilitarianism justifies moral claims); *cf.* LOUIS KAPLOW & STEVEN SHAVELL, FAIRNESS VERSUS WELFARE 11 (2002) (arguing that an assessment of individuals' well-being is essential to policy analysis).

[6] Aiming directly at utilitarian outcomes is not always the best way to reach the most beneficial outcomes. *See* HARE, *supra* note 5, at 48-49 (asserting that even though the best outcomes are not always achieved, the principles underlying utilitarianism still hold true). Sometimes it may be better to follow simple rules, such as, "Don't kill noncombatants," even when violating the rules appears to be for the best. For the cases I discuss, however, it may be sufficient to be aware of such situations so that we knowingly follow rules *because* we believe they lead to the best outcomes.

[7] Baron, *Norm-Endorsement*, *supra* note 5, at 166.

pend on the goals of others. The other dimension is whether they concern others' voluntary behavior.

	For Your Behavior	For Others' Behavior
Dependent on Others' Goals	Altruistic	Moral
Independent of Others' Goals	Self-Interested	Moralistic

The idea of dependence on others' goals assumes that goals are associated with the individuals who have them. Your goals are contingent on your existence. If you were never born, no goals would be yours.[8] Your self-interested goals are those that are yours, in this sense. Altruistic (and moral) goals are goals for the achievement of others' goals. Your altruistic goals concerning another person are thus a replica in you of the other person's goals. Altruism may be limited to certain people or certain types of goals. But it rises or falls as the goals of others rise and fall.[9]

We have goals for what other people do voluntarily. The behavior of others is a kind of consequence, which has utility for us. But this type of consequence has a special place in discussion of social issues. It is these goals for others' behavior that justify laws, social norms, and morality itself. When you have goals for the behavior of others, you apply criteria to their behavior rather than to your own. But of course these are still your own goals, so you try to do things that affect the behavior of others.[10] When we endorse behavior for others, we want them to want to choose it.

What I call "moral goals" are goals concerning the behavior of others so as to achieve their goals as well as your own. These are "moral" in the utilitarian sense only. In fact, they are the fundamental

---

[8] I do not assume that goals cease with death. Indeed, I have argued that they may continue. *See id.* at 177-79 (arguing that we should honor people's wishes even after they are dead).

[9] We could imagine negative altruism, in which *X* wants *Y*'s goals to be frustrated in proportion to their strength. Clearly such goals exist and influence political behavior, but I do not discuss them here.

[10] "Behavior" includes their goals, and all their thinking, as well as their overt behavior (just as self-interested goals may concern your own mental states), and excludes involuntary or coerced behavior.

goals that justify the advocacy of utilitarianism.[11] I shall return to these goals.

By contrast, moralistic goals are goals for the behavior of others that are independent of the others' goals. People could want others (and themselves) not to engage in homosexual behavior, or not to desire it.[12] Other examples abound in public discourse, including: antipathy to drug use; enforcement of particular religions against other religions; and promotion of certain tastes in fashion, personal appearance, or artistic style against other tastes.[13]

Often the public discourse about such things is expressed in the language of consequences. Moralistic goals usually come bundled with beliefs that they correspond to better consequences (a phenomenon that has been called "belief overkill").[14] For example, opponents of homosexuality claim that the behavior associated with it increases mental disorders, and this argument, if true, is relevant to those who do not find homosexuality inherently repugnant.

The people whose behavior is the concern of moralistic goals could be limited to some group. Your goals for others' behavior could be limited to others who share your religion or nationality, for example. In a sense, altruism can be limited, but truly moral goals cannot. You can have altruistic goals toward only certain other people, not caring about the rest. But if you have goals for others' behavior based only on these altruistic goals, these goals are actually moralistic, not altruistic, because they are independent of the goals of those outside of your sphere of altruism. This is just a consequence of the way I have defined "moralistic." We might want to distinguish moralistic

---

[11] *See* Baron, *Norm-Endorsement*, *supra* note 5, at 173-82 (providing examples of the kinds of fundamental goals that serve people's self-interests).

[12] It is a moralistic goal if we want people who have homosexual desires to override them through conscious effort. But it is not a moralistic goal if we simply want to prevent homosexual behavior through coercion, e.g., through round-the-clock supervision of boys in a boarding school. The point is that the overriding is voluntary, even though the desire is still present.

[13] Clearly the goals of fanatics who carry out terrorist attacks, and their supporters, are at least partly moralistic. Terrorists do not engage in ordinary political behavior, to be sure, but many of their financial supporters do.

[14] *See* BARON, *supra* note 2, at 13-14 (defining this effect as the distortion of the rational arguments for and against controversial issues through "wishful thinking" that ignores strong arguments contrary to the agent's beliefs); ROBERT JERVIS, PERCEPTION AND MISPERCEPTION IN INTERNATIONAL POLITICS 129-36 (1976) (providing some historical examples in which policymakers adjusted their beliefs to agree with policies already chosen).

goals that come from limited altruism from those that come from self-interest alone (which I shall discuss shortly).

In sum, unlike moral goals, moralistic goals can go against the goals of others. When moralistic goals play out in politics, they can interfere with people's achievement of their own goals. That is, if we define "utility" as a measure of goal achievement, moralistic goals decrease the utility of other goals. They need not do this if enough people have the same goals for themselves as others have for them. But the danger is there, especially on the world scale or in local societies with greater diversity of goals.

Moral goals may also involve going against the goals of some in order to achieve the goals of others. But moral goals are those that make this trade-off without bringing in any additional goals of the decision maker about the behavior of others.

Altruism and moralism are difficult to distinguish because of the possibility of paternalistic altruism. A true altruist may still act against your stated preferences, because these preferences may depend on false beliefs and thus may be unrelated to your true underlying goals.[15] Undoubtedly, moralists often believe that they are altruists in just this way. The experiments I review later,[16] however, will suggest that some moralism does not take this form. People think that their values should sometimes be imposed on others even when they (ostensibly) agree that the consequences are worse and that the others involved do not agree with the values being imposed. Moreover, even when people think that they are being paternalistically altruistic, they may be wrong. Although underlying goals are difficult to discover and somewhat labile, their existence is often a matter of fact.

Because other people's moralistic goals—in contrast to their altruistic and moral goals—are independent of our own individual goals, each of us has reason to want to replace moralistic goals with altruistic goals. Specifically, you will benefit from the altruistic and moral goals of others, and you will also benefit from their moralistic goals, if these agree with your own goals. But you will suffer from moralistic goals that conflict with your goals.

It is conceivable that most moralistic goals have greater benefits than costs, relative to their absence, and that they could not be replaced with other goals—such as moral goals—that have even greater

---

[15] *See* Baron, *Norm-Endorsement*, *supra* note 5, at 174-75 (discussing the relationship between individuals' erroneous goals and paternalistic policies).

[16] *Infra* Part III.

benefits. But it is also conceivable that moralistic goals arise from the same underlying motivation that yields moral goals and that the difference is a matter of belief, which depends on culture, including child rearing and education. Some of the beliefs in question may be false and subject to correction. To the extent that moralistic goals can be replaced with more beneficial moral goals, we have reason to try to make this happen. More generally, if we had only altruistic and self-interested goals, and only true beliefs, we might have no reason to adopt any moralistic goals at all. Moralistic goals serve neither self-interest nor altruism.

B. *A Norm-Endorsement Argument*

"Norm endorsement" is what I call the activity we undertake to try to influence the behavior of others. I have argued that norm endorsement is a fundamental moral activity in terms of which morality can be defined.[17] By this view, what should count as moral is what we each have reason to endorse for others and ourselves to follow, assuming that we have put aside our current moral views. By this account, moralistic goals are nonmoral. But we all have reason to endorse moral and altruistic goals. (Specifically, our reasons come from both self-interest and altruism.)

The basis of this argument is an analytic scheme in which we ask about the purpose of moral norms, or, more generally, norms for decision making. The idea is that the act of endorsing norms is fundamental for a certain view of what the relevant question is. There are surely other questions worth answering, but the question of what norms we should endorse for decision making is of sufficient interest, whether or not it exhausts the meaning of "moral" or "rational."

The motivation for norm endorsement can be altruistic, moralistic, or moral. If we want to ask about what norms we should endorse for others, we are concerned about their behavior. Hence, we want to put aside the norms that arise from our current moral and moralistic goals to see whether we can derive these goals from scratch. Again, this is not the only question to ask, even within this scheme, but it is a useful question. The answer provides a justification of our goals concerning others' behavior, without the circularity of justifying those goals in terms of goals of the same sort.

---

[17] *See* Baron, *Norm-Endorsement*, *supra* note 5, at 167-70 (explaining the role of norm endorsement in defining morality).

Our remaining altruistic goals are sufficient to motivate the endorsement of norms for others' decision making. If we successfully encourage others to rationally pursue their self-interest and to be altruistic toward others, they will better achieve their own goals, satisfying our own altruism toward them. Note that our altruistic goals concern the achievement of their goals, but we have now used this to justify moral goals, in which we endorse altruistic goals for others.

The upshot of this argument is that we should have moral goals. That is, we should endorse voluntary altruism and moral behavior. We should want others to be altruistic, to endorse altruism, and to endorse the endorsement of both of these behaviors. This contrasts with moralism because it depends completely on the goals that exist in the absence of any goals for others' voluntary behavior.

I. RATIONAL POLITICAL ACTION

The analysis of goals is relevant to decisions about political action, such as voting. Most political action is low cost, with little effect on the decision maker, but it is part of a system in which the little actions of a great many people taken together have large effects on everyone. In this regard, it is much like many large-scale public goods problems (or social dilemmas) in which each person is faced with a choice of cooperation or defection in the support of a public good, e.g., constraining water use when water is short.[18]

Voting, like cooperation in a social dilemma, is typically not justifiable by the narrow self-interest of the voter.[19] The low cost of voting and the lack of significant effect on the voter's interests suggest that voting will typically be motivated by moral beliefs or by expressive concerns, rather than by rational calculation of self-interest.[20] Social science research generally supports the conclusion that voting is predicted better from moral beliefs than from narrow self-interest.[21]

---

[18] This analysis may apply to one side of a political divide, but it cannot apply to both sides. If one side is the side of defection, its adherents will think that it is actually cooperation.

[19] *See* ANTHONY DOWNS, AN ECONOMIC THEORY OF DEMOCRACY 160-63 (1957) (evaluating the conflict between an individual's self-interest and the interest of a larger group, namely her party).

[20] *See, e.g.*, BRENNAN & HAMLIN, *supra* note 3, at 129-55 (discussing voting and elections, especially voters' expressive concerns); BRENNAN & LOMASKY, *supra* note 3, at 90-96 (comparing expressive and public choice accounts of voting behavior).

[21] *See, e.g.*, David M. Brodsky & Edward Thompson III, *Ethos, Public Choice, and Referendum Voting*, 74 SOC. SCI. Q. 286, 297-98 (1993) (finding that voters in a local transit-funding referendum in Chattanooga, Tennessee, behaved in a selfless, public-

Most behavior of interest to legal scholars and economists is primarily self-interested, but most political behavior is not primarily self-interested in the narrow sense. Voting is unlikely to be rational on the basis of purely self-interested goals, as I have defined them. Reasons for individual departures from rationality might include the good feeling of participation, insofar as that feeling is not itself dependent on altruistic, moral, or moralistic goals. It is hard to imagine such a feeling.

Voting can be rational for you, however, if you are sufficiently altruistic.[22] For similar reasons, voting and other political behavior can be rational if they achieve moral goals by affecting other people's behavior. In principle, such effects could be larger than the effects on purely altruistic goals. If you are the only altruist in the world, and your vote made a certain number of people into altruists just like you, who would then vote altruistically, your influence would be multiplied by that number of people. Of course, this is unlikely.

Voting may also be rational for someone with sufficiently strong moralistic goals. Moralistic goals may not be so dependent on the number of people who behave consistently with them, but they can also be more powerful than altruistic goals. Most altruistic goals for other people are weaker than self-interested goals, usually quite a bit weaker. Moralistic goals are not limited in this way. People who understand, however vaguely, that voting is not justified by self-interest may still feel that their moralistic goals require them to vote, and they are not irrational to feel this way, in terms of their own goals, even when they are immoral from a utilitarian perspective.

In sum, voting is rational when it is supported by moral, altruistic, and moralistic goals. The trouble is that the moralistic goals are often the strongest. This leads people, through their political behavior, to impose on others policies that sometimes subvert the other group's individual goals.

---

regarding way rather than in a self-interested, private-regarding way); David O. Sears & Carolyn L. Funk, *The Role of Self-Interest in Social and Political Attitudes*, *in* 24 ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY 1, 76-78 (Mark P. Zanda ed., 1991) (concluding that empirical evidence shows that voters base their decisions on more than self-interest); Leonard Shabman & Kurt Stephenson, *A Critique of the Self-Interested Voter Model: The Case of a Local Single Issue Referendum*, 28 J. ECON. ISSUES 1173, 1184 (1994) (reporting that voters in a Roanoke, Virginia, bond referendum were motivated by expected benefits to the community, as well as by self-interest).

[22] *See* Jonathan Baron, *Political Action Versus Voluntarism in Social Dilemmas and Aid for the Needy*, 9 RATIONALITY & SOC'Y 307, 307 (1997) (concluding that for a "somewhat selfish and rational utilitarian, under specified assumptions . . . political action is sometimes worthwhile and superior to voluntarism").

II. WHAT CAN WE DO WHEN MORALISTIC GOALS CAUSE TROUBLE?

I argued for putting aside moralistic goals when thinking about what norms we would endorse if we did not already have norms. But real people do not put these goals aside. Although you may not like other people's moralistic goals, especially when they conflict with your own goals, they are goals nonetheless. If you are altruistic toward others, then you want their goals to be achieved, whatever they are. In the extreme, moralistic goals are sadistic. Altruism, hence utilitarianism, must count sadistic goals as well.[23]

Sadistic goals, by definition, go against the goals of others, and moralistic goals often do this as well. Thus, when we are altruistic toward people who have moralistic and sadistic goals, we should also be altruistic toward those with conflicting goals. Utilitarianism advocates balancing the two considerations so as to maximize total utility in the long run. We cannot simply discount sadistic or moralistic goals because of their nature, but, in any given case, these goals run into resistance from other goals that we must consider.

We do not want people to interfere with our liberties because of their misguided moral opinions. But we also do not want to interfere with what might be a well-targeted but unpopular moral opinion. The problem is exacerbated by the fact that the goals that support political behavior are weak, and, without moralistic goals, fewer people would participate at all.

Consider, as an example, whether the Food and Drug Administration should approve a rotavirus vaccine.[24] The benefits are the prevention of a serious disease (especially in poor countries, whose approval of the vaccine may depend on what the FDA does), while the costs include various side effects, some as serious as the disease itself.[25] But another cost is that citizens are bothered by the idea of putting some at risk in order to help others. Should this source of opposition to the vaccine count as part of the cost-benefit analysis?

---

[23] *But see* HARE, *supra* note 5, at 140-42 (rejecting the utilitarian idea that "we should have to give as much weight to the pleasures or preferences of the Marquis de Sade as to those of Mother Teresa").

[24] *See* Jon Cohen, *Rethinking a Vaccine's Risk*, 293 SCIENCE 1576, 1576-77 (2001) (discussing a pharmaceutical company's decision to voluntarily pull a vaccine off the market in response to serious side effects, but suggesting that the risk-benefit analysis would actually weigh in favor of using the vaccine in developing countries where the virus is more common and more deadly).

[25] *Id.*

The opposition to the vaccine may be broken down into two parts. One is an opinion about what should be done. This is not properly considered as a component of utility, because it does not involve the goals of those who hold this opinion. We can, in principle, have opinions about what others should do without actually caring about what they should do, i.e., without having that as a goal of our own decisions, although of course we usually do care.

The other part is in fact a matter of caring, a goal. It is a moralistic goal because it concerns the behavior of others—such as the FDA, parents, and pediatricians—and it is held without regard to whether these others share the goal. A utilitarian cost-benefit analyst would take this goal into account, because it is part of the utility of those who hold it.

A. *Qualifications: Why We Might Ignore or Suppress Moralistic Goals*

This conclusion is qualified by two considerations. One is that neglecting certain goals might have future benefits in the form of discouraging people from having those goals. When a person's goals interfere with others' pursuit of their goals, we have reason to suppress the former, if we can.[26]

Some goals may even hurt the person who has them. Many goals are generally desired even when they will not be fully achieved. Most people, I suspect, would rather have a sex drive than not, even if they know it will not be satisfied. The same can arguably be said for goals concerning love, friendship, food, drink, the arts, music, and sports. These are goals to be cultivated. (I put aside the interesting question of what makes them this way.) Many moralistic goals go in a class that we might call primarily negative. Once we have these goals, the failure to satisfy them can make us unhappy, but achieving them is not usually positive. Due to the lack of positive benefit, in the absence of such goals, most people would probably not want them.[27] In this regard, the discouragement of moralistic goals might even benefit those who would otherwise have them. This is also a possible distinction between self-interested and moralistic goals. Self-interested goals often interfere with the achievement of others' goals, in the same way that moralistic goals do. But most self-interested goals are of the positive

---

[26] *Cf.* BARON, MORALITY, *supra* note 5, at 35 (arguing that sadistic goals should be discouraged); HARE, *supra* note 5, at 142 (arguing for the adoption of prima facie principles that discourage sadism and encourage "less harmful pleasure").

[27] I am assuming this. It is an empirical question, and I'm not sure of the answer.

type, so there is a greater cost to discouraging their formation and maintenance.

The other qualifying consideration is that utilitarian analysis should (arguably) count only fundamental goals, and it should ignore goals based on false beliefs.[28] Some moralistic goals might be based on false beliefs (such as the belief that the Golden Rule as stated in The Bible concerns actions but not omissions, if that is indeed false).

To take another example, consider opposition to homosexuality. Whatever the origin of this in feelings of disgust, the belief that homosexuality is wrong is supported by a host of related beliefs: that it is unnatural, that it is controllable, and that it has harmful effects on other aspects of people's lives. Thus, opposition to homosexuality is a means to the end of avoiding the unnatural through things we can control. To the extent to which the evidence says homosexuality is not unnatural, controllable, or harmful, moral opposition is based on false beliefs. Beliefs, once formed, often resist evidence of their falsity, but if the evidence had the effect it should, then this belief would be eliminated or reduced in strength.

One issue here is whether these two qualifications apply when moralistic goals and self-interested goals coincide. It could turn out that people's moralistic goals coincide with the self-interested goals of almost everyone. Surely this happens sometimes. For example, many people think that human cloning is repulsive and that nobody should do it. These people provide a number of consequentialist reasons, but it is clear that these reasons have little to do with the moral view in question, as they all may disappear in time without reducing their proponents' opposition to cloning. Why not ban cloning if, for example, only one percent of all people would ever think of having themselves cloned? It might even be that ninety percent of people would favor a ban as a self-control device. They may some day want to clone themselves, e.g., if they need a bone marrow transplant that a clone could provide, and they may want the temptation to do so re-

---

[28] *See* Baron, *Norm-Endorsement*, *supra* note 5, at 174 (endorsing "a rule that favors taking account of fundamental goals, not erroneous subgoals"). More generally, we should ignore what Keeney calls "means values," that is, goals that are subgoals, or means to the achievement of other goals. *See* RALPH L. KEENEY, VALUE-FOCUSED THINKING: A PATH TO CREATIVE DECISIONMAKING 34-35 (1992) (defining a "means objective" as one having interest "because of its implications for the degree to which another (more fundamental) objective can be achieved"). These goals exist because they are connected to more fundamental goals by beliefs, which may or may not be correct.

moved by law. Similarly, in the case of a vaccine, it may be that almost everyone weighs the side effects more than the disease.

In some cases, one or both of these two considerations may make neglect of moralistic goals desirable. For example, it might be reasonable to put them aside when performing some cost-benefit analyses. When utilities are elicited in surveys, we must take care to separate utilities for outcomes from both moralistic goals and opinions about what others should do. When we ask about vaccines, we should ask about the consequences without saying which are caused by the natural disease and which are caused by the vaccine. Similarly, when we ask about the value of damage to the environment, we should not say whether the damage was caused by people or by nature itself.[29] It might seem possible to argue that all moralistic goals are based on false beliefs. The argument would take the form of the argument that I sketched for utilitarianism itself. Namely, if we did not already have such moralistic goals, the creation of such goals (itself an act that is somewhat under our control) could not be motivated by any nonmoralistic goals that we already have. Therefore, they must arise from confusion.

In particular, people may initially form moralistic goals on the basis of their own reaction to some new situation, such as the possibility of cloning. On the basis of this reaction, they decide that some behavior, such as getting cloned, is not for them. They then go further to conclude that anyone who thinks otherwise must be misguided and in need of protection from her own delusions. Thus, out of parentalism (the gender-neutral version of "paternalism"), people endorse the general prohibition of the behavior, even in those who benefit from it. The impulse for moralistic goals is thus the same as the impulse for altruistic and moral goals, but is based on a false belief about the true goals of others. This is an especially difficult judgment to make, however, because there are surely cases in which such parentalism is justified, such as when parents try to inculcate in their children a taste for education or classical music. Difficulty of a judgment, however, does not imply that there is no such thing as a correct judgment in a given case. An apparent problem with this argument is that moralistic goals can be motivated by self-interested goals. One self-interested goal is to want other people to be like us, to share our religion, for example.

---

[29] *See* Jonathan Baron, *Biases in the Quantitative Measurement of Values for Public Decisions*, 122 PSYCHOL. BULL. 72, 80 (1997) (finding that measures of values are influenced by irrelevant factors, such as whether the damage was caused by humans or by nature).

When we live in a community of like-minded others, it is easier for us to pursue our personal commitments. Thus, moralistic goals may rationally arise from self-interested goals. Moralistic goals thus become a kind of rent seeking in the domain of social norms. At least when we look at moralistic goals in this way, the claim of some people that moralistic goals should trump other people's goals looks more like a tough negotiating stance than a high-minded moral assertion. We must give moralistic goals their due, but simply because they are related to self-interested goals, not because they have any special status, despite the claims made for them. Some portion of these goals may well depend on false beliefs about others.

More to the point, people may come to understand and accept the kind of argument I have made about the nature of moralistic goals. To the extent to which they accept such an argument, they would realize that much of the support for their moralistic goals is based on a false belief, specifically the belief that moralistic goals are truly moral, because they are good for those on whom they are imposed. People may accept this proposition unreflectively, without fully putting themselves in the position of those who are affected. The kind of argument made here may help people identify those cases in which they are rationalizing, rather than being truly parental toward others.

III. SOME EXAMPLES OF BIASES

In the rest of this Article, I argue that several biases in judgments about public policy take the form of moralistic goals to some extent. These are biases away from utilitarianism, and hence toward worse overall consequences if the decisions have their intended effects. I will illustrate these principles from my research on people's judgments about hypothetical policy decisions. When these principles are moralistic goals, they serve to provide a rational motivation for voting, because voting serves these goals. One question of my research is to find out the form of these goals—in particular whether they are moralistic and knowingly nonutilitarian—and another is to explore the possibility of modifying judgments through de-biasing. I address three types of principles: allocation principles, protected values, and parochialism.

Of primary interest is whether these represent moralistic values, at least in part. That is, are people willing to impose these principles on others, even when the others disagree with the principles and prefer some other option out of self-interest?

A. *Allocation*

One type of principle concerns allocation of goods or bads. These principles may thus involve intuitions about fairness. These intuitions generally support policies that lead to worse consequences for some people, and potentially everyone.[30] Interestingly, people's goals for distributional properties like fairness are moralistic goals. They go against the goals of others, with no compensating advantage in achieving other goals, except for the moralistic goals of others who support them.

1. Omission Bias

Ilana Ritov and I have examined a set of hypothetical vaccination decisions modeled after real cases like the rotavirus vaccine,[31] (although the actual inspiration was the diphtheria, tetanus, and pertussis (DTP) vaccine).[32] In one experiment, subjects were told to imagine that their child had a 10 out of 10,000 chance of death from a flu epidemic, a vaccine could prevent the flu, but the vaccine itself could kill some number of children.[33] Subjects were asked to indicate the maximum overall death rate for vaccinated children at which they would still be willing to vaccinate their child,[34] and most subjects answered well below 9 per 10,000.[35] Of the subjects who showed this kind of reluctance, the mean tolerable risk was about 5 out of 10,000, half the risk of the illness itself.[36] These results were also found when subjects were asked to take the position of a policymaker deciding for

---

[30] *See* KAPLOW & SHAVELL, *supra* note 5, at 52 (arguing that fairness does not depend exclusively on the well-being of individuals, and therefore "satisfying notions of fairness can make individuals worse off").

[31] Ilana Ritov & Jonathan Baron, *Reluctance to Vaccinate: Omission Bias and Ambiguity*, 3 J. BEHAV. DECISION MAKING 263, 266-77 (1990) (presenting four experiments where subjects were asked to decide whether to vaccinate a child or require vaccination by law, given different probabilities that either the vaccine or the disease would kill the children).

[32] For information regarding DTP vaccination, its risks, and modern alternatives to the traditional vaccine, see U.S. DEP'T OF HEALTH & HUMAN SERVS., CTRS. FOR DISEASE CONTROL & PREVENTION, DIPHTHERIA TETANUS & PERTUSSIS VACCINES: WHAT YOU NEED TO KNOW (2001), *available at* http://www.cdc.gov/nip/publications/VIS/vis-dtp.pdf.

[33] Ritov & Baron, *supra* note 31, at 267-68.

[34] *Id.* at 268.

[35] *Id.* at 270.

[36] *Id.*

large numbers of children.[37] When subjects were asked for justifications, some said that they would be responsible for any deaths caused by the vaccine, but they would not be as responsible for deaths caused by failure to vaccinate.[38]

Biases such as this are apparently the basis of philosophical intuitions that are often used as counterexamples to utilitarianism, such as when analyzing the problem of whether one should kill a single person to save five others. Many people feel that they should not kill even one person and that, therefore, utilitarianism is incorrect.[39] They assume that their intuitions arise from moral truth, usually in some obscure way. An alternative view, one I have defended elsewhere, is that intuitions arise from learning simple rules that usually coincide with producing good consequences, but, in the critical cases, do not.[40] In terms of consequences, one dead is better than five.[41]

Omission bias is related to issues of public controversy, such as whether active euthanasia should be allowed,[42] whether vaccines should be approved or recommended when they have side effects as bad as (but less frequent than) the diseases they prevent,[43] or whether we are morally obligated to help alleviate world poverty.[44]

Omission bias is somewhat labile. It can be reduced by the instructions to take the point of view of those who are affected, e.g., "If you were the child, and if you could understand the situation, you would certainly prefer the lower probability of death. It would not matter to you how the probability came about."[45]

---

[37] *Id.* at 266-67.

[38] *Id.* at 275.

[39] For an overview of the premise that people prefer harmful omissions to less harmful acts, and this premise's relation to consequentialism and utilitarianism, see, for example, JONATHAN BARON, THINKING AND DECIDING (3d ed. 2000); Jonathan Baron, *Nonconsequentialist Decisions*, 17 BEHAV. & BRAIN SCI. 1 (1994).

[40] For further discussion of this point, see sources cited *supra* note 39.

[41] Of course, once the intuition starts running, it is bolstered by all sorts of imaginary arguments, some of them about consequences.

[42] *See, e.g.*, PETER SINGER, PRACTICAL ETHICS 202-13 (2d ed. 1993) (articulating the arguments for and against active euthanasia, ultimately concluding that controlled euthanasia should be permitted).

[43] *See* Cohen, *supra* note 24, at 1576-77 (describing the current discussion about reintroducing a vaccine for rotavirus infection—a disease causing diarrhea that kills nearly 800,000 children per year—that was pulled off the market because of fatal side-effects in a small number of children).

[44] *See* SINGER, *supra* note 42, at 229-46 (arguing that individuals have an obligation to help reduce absolute poverty).

[45] Jonathan Baron, *The Effect of Normative Beliefs on Anticipated Emotions*, 63 J. PERSONALITY & SOC. PSYCHOL. 320, 323 (1992).

2. Proportionality and Zero Risk

People worry more about the proportion of risk reduced than about the number of people helped. This may be part of a more general confusion about quantities. Small children confuse length and number. As a result, when we ask them to compare rows of objects for length or number, they will answer in terms of length (or number), regardless of the question, even if the shorter row has more objects.[46] The literature on risk effects of pollutants and pharmaceuticals commonly reports relative risk, the ratio of the risk with the agent to the risk without it, rather than the difference. Yet the difference between the two risks, not their ratio, is most relevant for decision making: if a baseline risk of injury is one in a million, then twice that risk is still insignificant; but if the risk is one in three, a doubling of the risk matters much more.

Several studies have found that people want to give priority to large proportional reductions of small risks, even though the absolute risk reduction is small relative to other options.[47] I have suggested that these results were the result of quantitative confusion between relative and absolute risk.[48]

The extreme form of this bias is the preference for zero risk. If we can reduce risks to zero, then we do not have to worry about causing harm. This intuition is embodied in the infamous Delaney clause, a provision of the Food, Drug, and Cosmetic Act outlawing any food additive that increases the risk of cancer by any amount.[49] Other laws

---

[46] Jonathan Baron et al., *Effects of Training and Set Size on Children's Judgments of Number and Length*, 11 DEVELOPMENTAL PSYCHOL. 583, 583-88 (1975) (concluding that young children did not differentiate between dimensions of length and number).

[47] *See, e.g.*, Jonathan Baron, *Confusion of Relative and Absolute Risk in Valuation*, 14 J. RISK & UNCERTAINTY 301, 307-08 (1997) (finding that subjects were willing to pay for saving others' lives when the proportion of lives saved was great but not when it was low, even if the same number of persons were saved); David Fetherstonhaugh et al., *Insensitivity to the Value of Human Life: A Study of Psychophysical Numbing*, 14 J. RISK & UNCERTAINTY 283, 297-99 (1997) (discussing a series of studies in which subjects were prone to "psychological numbing," whereby they valued saving a fixed number of persons from a small population more than the same number of persons from a large population); T.L. McDaniels, *Comparing Expressed and Revealed Preferences for Risk Reduction: Different Hazards and Question Frames*, 8 RISK ANALYSIS 593, 593-604 (1988) (analyzing studies that reveal a tendency to spend disproportionate amounts on relative, rather than absolute, risk reduction); Eric R. Stone et al., *Risk Communication: Absolute Versus Relative Expressions of Low-Probability Risks*, 60 ORG'L BEHAV. & HUM. DECISION PROCESSES 387, 402-06 (1994) (showing that relative expressions of risk avoidance are misleading).

[48] Baron, *supra* note 47, at 308.

[49] Food Additives Amendment of 1958, Pub. L. No. 85-929, § 4, 72 Stat. 1784, 1786

favor complete risk reduction, such as the 1980 Superfund law, which concerns the cleanup of hazardous waste that has been left in the ground.[50] Justice Stephen Breyer has argued that most agencies are inclined to spend exorbitant amounts to clean up "the last 10%," implying that, for most purposes, the 90% cleanup is adequate.[51] In many cases, cleanup costs are so high that the process is proceeding very slowly.[52] It is very likely that more waste could be cleaned up more quickly if laws and regulations did not encourage perfection.

Subjects in questionnaire studies show the same bias toward reducing risk to zero.[53] They are willing to pay more for a smaller risk reduction if the reduction causes the risk to become zero.[54] They also prefer a smaller reduction over a larger one if the former reduces some risk to zero.[55]

3. Ex Ante Equity

The ex ante bias is the finding that people want to equate ex ante risk within a population even when the ex post risk is greater. For example, many people would give a screening test to everyone in a group of patients covered by the same HMO if the test would prevent 1000 cancer cases rather than give a test to half of the patients (picked

---

(codified as amended at 21 U.S.C. § 348(c)(3)(A) (2000)) ("[N]o additive shall be deemed to be safe if it is found to induce cancer when ingested by man or animal . . . .").

[50] Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (CERCLA), Pub. L. No. 96-510, 94 Stat. 2767 (codified as amended at 42 U.S.C. §§ 9601-9607 (2000)).

[51] *See* STEPHEN BREYER, BREAKING THE VICIOUS CIRCLE: TOWARD EFFECTIVE RISK REGULATION 10-19 (1993) (arguing that, due to administrative and regulatory "tunnel vision," the majority of costs for regulating health risks goes to the final 10% of the project's targeted problem).

[52] *See id.* at 18-19 (citing various estimates for the cost of toxic waste cleanup, including one possibility reaching as high as one trillion dollars); *cf. id.* at 11-12 (describing a particular instance of litigation and cleanup efforts continuing despite no evidence to suggest that the site posed additional risk).

[53] *See, e.g.*, Jonathan Baron et al., *Attitudes Toward Managing Hazardous Waste: What Should Be Cleaned up and Who Should Pay for It?*, 13 RISK ANALYSIS 183, 190-91 (1993) (discussing subjects' bias toward zero risk); W. Kip Viscusi et al., *An Investigation of the Rationality of Consumer Valuations of Multiple Health Risks*, 18 RAND J. ECON. 465, 468-69 (1987) (describing the concept of a "certainty premium").

[54] *See* Viscusi et al., *supra* note 53, at 478 (finding that subjects were willing to pay extremely large premiums to eliminate risk).

[55] *See* Baron et al., *supra* note 53, at 190 (indicating that subjects preferred a total cleanup of one waste site that saved fewer lives than a partial cleanup of two waste sites).

at random) that would prevent 1100 cancer cases.[56] Peter Ubel, David Asch, and I found that this bias was reduced when the group was expanded.[57] When subjects were told that the HMO actually covered people living in two states and that the test could not be given in one of the states, some subjects switched their preference to the test that prevented more cancers.[58] It was as though they reasoned that, since the "group" was now larger, they could not give the test to "everyone" anyway and might as well aim for better consequences, given that "fairness" could not be achieved.[59] This result illustrates a potential problem with some nonutilitarian concepts of distributive justice, namely that the distributions they entail can change as the group definition is changed. If groups are arbitrary, then the idea of fair distribution is also arbitrary.

4. Are Allocation Biases Moralistic?

Recently I have been exploring the role of these allocation biases in low-cost political behavior, such as answering questions in polls. The experiments present subjects with hypothetical decisions, and the subjects say how they think the decisions should be made. One series of studies asks whether people are willing to impose their allocation judgments on others, even when the others disagree. That is, do these biases form the basis of moralistic goals? I shall summarize one such study.[60]

In this study, ninety-one subjects completed an online questionnaire designed to test three types of biases: omission bias, zero bias, and ex ante bias. Subjects found the questionnaire page through links in other web pages and through search engines.[61] Their ages ranged from 18 to 74 (with a median of 38), 29% were male, 14% were students. The questionnaire provided a general description of the scenario, which concerned a health insurance company that had

---

[56] Peter A. Ubel et al., *Cost Effectiveness Analysis in a Setting of Budget Constraints: Is It Equitable?*, 334 NEW ENG. J. MED. 1174, 1175-76 (1996).

[57] *See* Peter A. Ubel et al., *Preference for Equity as a Framing Effect*, 21 MED. DECISION MAKING 180, 180-86 (2001) (finding that preference for equitable testing was subject to framing effects).

[58] *Id.* at 184-85.

[59] *Id.* at 187.

[60] The general approach applies to all the studies described in this Article and is reported in an unpublished manuscript currently on file with the author.

[61] *See* Jonathan Baron, Questionnaire Studies, *at* http://www.psych.upenn.edu/~baron/qs.html (last modified Mar. 21, 2003) (providing updates about, and links to, my ongoing research).

to decide which expensive treatments to cover. Each screen presented a choice of two treatments, *A* and *B*. An example of a screen testing the omission bias was:

> Treatment *A* cures 50% of the 100 patients per week who have this condition, and it causes no other conditions.
>
> Treatment *B* cures 80% of the 100 patients per week who have this condition, but it causes a different (equally serious) condition in 20% of them.

The test questions were:

> Which treatment leads to fewer sick people? Which treatment should the company choose (if it didn't know what its members favored)?
>
> Now suppose that 75% of the members favor Treatment *B*. They think *B* leads to fewer sick people and that is what they care about. Which treatment should the company choose?
>
> Now suppose that 100% of the members favor Treatment *B* . . . .
>
> Now suppose that 75% of the members favor Treatment *A* . . . .
>
> Now suppose that 100% of the members favor Treatment *A* . . . .

Each question was followed by a four-point response scale: "Certainly *A*," "Probably *A*," "Probably *B*," and "Certainly *B*." The subject had to answer the test question correctly in order for the answers to be recorded. The text of the conditions testing the zero and ex ante biases read:

> *Zero Bias*:
>
> Treatment *A* is for a type of the condition that affects 50 patients per week. It cures 100% of them.
>
> Treatment *B* is for a type of the condition that affects 100 patients per week. It cures 60% of them.
>
> *Ex ante Bias*:
>
> Treatment *A* can be given to all 100 patients per week who have this condition, and it cures 30% of them.
>
> Treatment *B* is in short supply. It can be given only to 50 of the 100 patients per week with this condition, picked at random. It cures 80% of these 50.

The experiment included a control condition with the same wording as the ex ante condition, but with no difference in the number of patients who could get the treatment.

The main result was that bias was present even when 100% of the members wanted the optimal treatment covered. For example, in the omission bias condition, the mean bias was 0.19 (on a scale where 0 represents neutrality between the two treatments and 1.5 is the maximum omission bias possible), in contrast to a mean bias of 0.16 for the

control condition when subjects think the more effective treatment leads to fewer sick people and that is what they care about. In sum, some people are willing to impose their allocation judgments on others, even when it is clear that the consequences for others are worse and that the others do not favor the allocations in question.

5. De-Biasing Allocation Biases

Another series of studies has explored de-biasing. In one study, Andrea Gurmankin and I explored two general methods of de-biasing. The "minimal" method involved stripping away information and focusing on consequences alone, a minimal description. We tried this method on four different biases: omission bias, zero-harm bias, preference for ex ante equality, and preference for group equality (even when these made consequences worse). Here is an example of the testing condition for the omission bias:

> Treatment *A* cures 50 people out of 100 who come in with condition *X* each week, and it leads to no other conditions.
>
> Treatment *B* cures 80 of the people with condition *X*, but it leads to condition *Y*, occurring randomly in 20 of the 100 patients. *X* and *Y* are equally serious.

The following was added for the experimental condition: minimal de-biasing. In other words, treatment *A* leads to 50 people with condition *X* and nobody with any other condition, and treatment *B* leads to 20 people with condition *X* and 20 people with condition *Y*, which is equally serious. In general, this minimal de-biasing manipulation reduced bias on several different measures. The effects were small but statistically significant. Most subjects thought that the summary ("In other words . . . .") was fair.

A second method, like that just described, involved expansion by providing additional information that might change the bias. A typical item, for the zero-risk bias, was:

> *Zero-Risk Bias*:
>
> *X* and *Y* are two kinds of cancer, equally serious. Each year, 100 people get cancer *X* and 50 get cancer *Y*.
>
> Treatment *A* is given to the 100 people with cancer *X*. It cures 60 of them.
>
> Treatment *B* can be given to the 50 people with cancer *Y*. It cures all 50 of them.

The following was added in the de-biasing condition:

> The total number of cancer cases of all types, including *X* and *Y*, is 1000 each year. Treatment *A* thus cures 60 out of 1000 cases, and treatment *B* cures 50 out of 1000.

Subjects were asked which option should be covered by an HMO if it had to choose one. In one study, the expansion statement led subjects to favor the optimal treatment, *A*, and most subjects did not think that it was inconsistent with the initial statement.

These results have two implications. First, as a practical matter, biases can be reduced by restating the issue. Second, as a theoretical matter, biases are subject to framing effects—subject to change with redescription of the same situation—and should thus not be taken as people's last, most-considered view on the issues.

B. *Protected Values*

People think that some of their values are protected from trade-offs with other values.[62] Many of these values concern natural resources such as species and pristine ecosystems. People with protected values (PVs) for these things do not think they should be sacrificed for any compensating benefit, no matter how small the sacrifice or how large the benefit. In an economic sense, when values are protected, the marginal rate at which one good can be substituted for another is infinite. For example, no amount of money can substitute for certain types of environmental decline.

PVs concern rules about action, irrespective of their consequences, rather than consequences themselves. What counts as a type of action, e.g., lying, may be defined *partly* in terms of its consequence (false belief) or intended consequence, but the badness of the action is not just that of its consequence, so the action has value of its own.

Omission bias is greater when PVs are involved.[63] For example, when people have a PV for biological species generally, they are even

---

[62] *See* Jonathan Baron & Mark Spranca, *Protected Values*, 70 ORG'L BEHAV. & HUM. DECISION PROCESSES 1, 1-16 (1997) ("Protected values are those that resist trade-offs with other values, particularly economic values."); Philip E. Tetlock et al., *Revising the Value Pluralism Model: Incorporating Social Content and Context Postulates*, *in* THE PSYCHOLOGY OF VALUES: THE ONTARIO SYMPOSIUM 25, 36-39 (Clive Seligman et al. eds., 1996) (explaining that "sacred" values resist trade-offs).

[63] *See generally* Ilana Ritov & Jonathan Baron, *Protected Values and Omission Bias*, 79 ORG'L BEHAV. & HUM. DECISION PROCESSES 79, 93-94 (1999) (concluding that experimental subjects with protected values had a large bias against harmful acts that go against the value as opposed to harmful omissions).

less willing to cause the destruction of one species in order to save even more species from extinction. Thus, PVs primarily apply to acts, as opposed to omissions.[64]

People think that their PVs should be honored even when their violation has no consequence at all. People who have PVs for forests, for example, say that they should not buy stock in a company that destroys forests, even if their purchase would not affect the share price and would not affect anyone else's behavior with respect to the forests. This is an "agent relative" obligation, a rule for the person holding the value that applies to her own choices but not (as much) to her obligations with respect to others' choices. So it is better for her not to buy the stock, even if her not buying it means that someone else will.

PVs are at least somewhat insensitive to quantity. People who hold a PV for forests tend to say that it is just as bad to destroy a large forest as a small one.[65] They say this more often than they say the same thing for violations of nonprotected values (NPVs).[66]

Several researchers have noted that PVs cause problems for quantitative elicitation of values, as is done in cost-benefit analysis or decision analysis.[67] PVs imply infinite values.

Notice that the issue here is not behavior. Surely, people who endorse PVs violate them in their behavior, but these violations do not imply that the values are irrelevant for social policy. People may want public decisions to be based on the values they hold on reflection, regardless of their actual behavior. When people learn that they have violated some value they hold, they may regret their action rather than revising the value.

1. PVs Are Moralistic Too

It should not be surprising that people think of PVs as rules that are independent of consequences for people. I did an experiment to

---

[64] *See id.* at 80 ("PVs . . . show stronger omission bias than other values. Omission bias is the tendency to be less concerned with harms caused by omission than with identical harms caused by action.").

[65] *Id.* at 86-88.

[66] *Cf. id.* at 86 ("Although omission bias was greater for PVs, it remained for [NPVs].").

[67] *See* Baron, *supra* note 29, at 72-88 (noting that some PVs cause distinct cost-benefit problems because there is no way for the market to price some preferences); Max H. Bazerman et al., *The Human Mind as a Barrier to Wiser Environmental Agreements*, 42 AM. BEHAV. SCIENTIST 1277, 1286-88 (1999) (observing that when dealing with certain sacred or protected values, it becomes difficult to determine the actual value of a belief or at what price it is tradeable).

demonstrate this, using examples such as: testing a fetus for "IQ genes," and aborting it if its expected IQ is below average; cloning someone with desired traits, such as an athletic champion or a brilliant scientist, so that these may be passed on; modifying the genes of an embryo so that, when born, it will have a higher IQ; and giving a drug (with no side effects) to enhance school performance of normal children. On each screen, the item was presented, followed by a series of questions with possible answers printed on buttons (shown here as rectangles). The significant questions, absent the mnemonic names in brackets, were:

Should the government allow this to be done? ["Allow"]

This should be allowed, so long as other laws are followed.
This should never be allowed, no matter how great the need.
This should sometimes be allowed, with safeguards against abuse.

If the circumstances were (or are) present so that the consequences of allowing this were better than the consequences of banning it, should it be allowed? ["If-Better"]

Yes.	No.	I cannot imagine this.

If these circumstances were present, and if almost everyone in a nation thought that the behavior should be allowed, should it be allowed in that nation under these circumstances? ["If-Pro"]

Yes, allowed.	No, banned.	I cannot imagine this.

The mean proportions of endorsement are shown in the following tables:

	Allow	Sometimes	Never
"Allow"	0.40	0.32	0.28

	Allow	Ban	Can't Imagine
"If-Better"	0.71	0.20	0.09

	Allow	Ban	Can't Imagine
"If-Pro"	0.72	0.22	0.06

Of interest, subjects favored bans 22% of the time, even though, in these cases, they could imagine that the action was better and that the vast majority favored it. In other words, subjects were willing to impose their values on others.

2. Parentalism

A question that arises here is whether protected values, in their moralistic form, are parentalistic. That is, whether people who impose their values on others think that they are really going against the values of others. Surely this must happen because of belief overkill. People do not like their beliefs to conflict, so they deceive themselves into agreement. They are thus likely to think that their values are really good for everyone, whether others know it or not. However, this distortion of beliefs may be incomplete; we might expect some cases in which people know that their moralistic values go against the interests of others.

To test this possibility, I did another experiment, in which I attempted to find actions that were done by a specific person and that benefited that person. For example:

> A mother gives a drug (with no side effects) to her child that will improve the child's school performance (which is otherwise average). ["IQ Drug"]
>
> A single, childless woman has a child by cloning herself because she has no prospect for marriage. ["Clone"]
>
> A young widow has a child by cloning her husband, using cells taken from him as he was dying. ["Widow"]
>
> A man in his fifties has himself cloned to produce an embryo that will yield cells that will prevent him from getting Alzheimer's disease. ["Alzheimer's"]

PVs were defined roughly as in the last study I described. The main question of interest was:

> (8) If the person were stopped from doing this, how would it affect [him/her], on the whole (considering all effects together)?

Stopping it would be good for [him/her] on the whole, and [he/she] would see it as good, immediately.
Stopping it would be good for [him/her] on the whole, and [he/she] would see it as good, eventually.
Stopping it would be good for [him/her] on the whole, even if [he/she] thought it was bad.

Stopping it would be good for [him/her] on the whole, because it would go against what [he/she] wants.

Stopping it would be good for [him/her] on the whole, because it would infringe on [his/her] rights.

The following table shows the main results as proportions for the four cases shown above. (Other results were similar.)

	PV	Stop-Effect	Stop/PV
"IQ drug"	0.404	0.509	0.196
"Clone"	0.588	0.421	0.104
"Widow"	0.500	0.509	0.158
"Alzheimer's"	0.404	0.605	0.152

In the table, stop-effect is the proportion of answers in which stopping the act was acknowledged to be bad for the actor—the last two options of question (8)—and Stop/PV is the proportion of PV responses in which stopping the act was acknowledged to be bad. Although Stop/PV is much lower than Stop-effect, as predicted by the belief-overkill hypothesis,[68] it is not zero. In a substantial number of cases the subjects did acknowledge that their imposed values were harmful to the actors.

3. Challenging PVs

PVs may be unreflective overgeneralizations. People may endorse the statement that, "No benefit is worth the sacrifice of a pristine rainforest," without thinking much about possible benefits, e.g., a cure for cancer or malaria. Or, when people say that they would never trade off life for money, they may fail to think of extreme cases, such as crossing the street (hence risking loss of life from being hit by a car) to pick up a large check, or failing to increase the healthcare budget

[68] For an evaluation of this hypothesis, in which "unreflective overgeneralizations provide one possible avenue for challenging PVs to make compromises and trade-offs possible," see Jonathan Baron & Sarah Leshner, *How Serious Are Expressions of Protected Values*, 6 J. EXP'L PSYCHOL.: APPLIED 183, 184 (2000).

enough to vaccinate every child or screen everyone for colon cancer. Such unreflective overgeneralizations provide one possible avenue for challenging PVs in order to make compromise and trade-offs possible. If PVs are unreflective in this way, then PVs should yield to simple challenges. In an experiment conducted in 2000, Sarah Leshner and I gave subjects questions about whether they would regard certain outcomes, such as, "Electing a politician who has made racist comments," as against their values to such an extent that no benefit would be sufficient to justify actions that caused such outcomes.[69] Then, when values were protected in this way, we challenged the subjects by asking them to think of counterexamples. We found that PVs do sometimes respond to such challenges.[70]

We also found that the effects of counterexamples can transfer to measures of omission bias, the bias toward harm caused by omissions when that is pitted against harm caused by acts.[71] The last two experiments found that PVs are not honored when the probability and magnitude of harm is low enough.

In a more recent experiment, I have found that PVs have less of an effect on decisions when subjects are asked to put themselves in the position of a government decision maker with sole responsibility for the decision, as opposed to putting themselves in the position of a citizen responding to an opinion survey.[72] The latter position is close to what the subjects are actually doing, so it is not hard to imagine. The former may be more difficult, but subjects were significantly more willing to make trade-offs when they were asked to imagine themselves making the actual decisions.[73] This result provides further evidence for the lability of PVs.

The results of the experiments described above suggest that PVs are strong, moralistic opinions that are weakly held. They are strong in the sense that they express infinite trade-offs—holders of these values assert that they are so important that they should not be traded off for anything. This assertion yields to a variety of challenges. After yielding, of course, the value may still be strong in the sense that a large amount of benefit is required to sacrifice the value.

---

[69] *Id.*
[70] *Id.* at 185.
[71] *Id.* at 185, 187-88.
[72] Baron, *supra* note 60.
[73] *Id.*

C. *Parochialism*

The tendency of people to favor a group that includes themselves, at the expense of outsiders and even at the expense of their own self-interest, has been called parochialism.[74] We may think of parochialism as an expression of both altruistic and moralistic goals. It is altruistic toward comembers. It may be moralistic in its effects on outsiders. The outsiders are being asked to help achieve the goals of insiders, in effect, whether this is consistent with their own goals or not. (What is not clear is whether they are being asked to do this voluntarily, or whether coerced behavior would suffice, in which case the values are not truly moralistic.) More likely, though, parochialism is moralistic in its application to insiders, who are expected to be loyal to the group.

A prime example is nationalism, a value that goes almost unquestioned in many circles, just as racism and sexism went unquestioned in the past. Nationalists are concerned with their fellow citizens, regardless of the effect on outsiders. Nationalists are willing to harm outsiders, such as in war, for the benefit of conationals. This sort of nationalism is moralistic to the extent to which nationalists want outsiders to behave willingly in ways that benefit their conationals, e.g., cede territory, stop trying to immigrate, or allow foreign investment. Nationalists typically want others in the group to be nationalists as well. The idea that one should vote for the good of humanity as a whole, regardless of the effect on one's own nation, would make total sense to a utilitarian (and it would require little self-sacrifice because voting has such a tiny effect on self-interest); however, such a vote would be considered immoral by the nationalist.

An experiment by Bornstein and Ben-Yossef illustrates the parochialism effect. Subjects came in groups of 6 and were assigned at random to a red group and a green group, with 3 in each group.[75] Each subject started with 5 Israeli Shekels (IS), about $2.[76] If the subject contributed this endowment, each member of the subject's group, including the subject, would get 3 IS. This amounts to a net loss of 2 IS for the subject, but a total gain of 4 IS for the group. However, the

---

[74] *See* Peregrine Schwartz-Shea & Randy T. Simmons, *Egoism, Parochialism, and Universalism*, 3 RATIONALITY & SOC'Y 106, 107 (1991) (defining parochialism as "the norm that one owes cooperation to one's 'solidarity group'").

[75] Gary Bornstein & Meyrav Ben-Yossef, *Cooperation in Intergroup and Single-Group Social Dilemmas*, 30 J. EXP'L SOC. PSYCHOL. 52, 58-59 (1994).

[76] *Id.* at 56.

contribution would also cause each member of the *other* group to *lose* 3 IS. Thus, taking both groups into account, the gains for one group matched the losses to the other, except that the contributor lost the original 5 IS. The effect of this 5 IS loss was simply to move goods from the other group to the subject's group. Still the average rate of contribution was 55%, and this was substantially higher than the rate of contribution in control conditions in which the contribution did not affect the other group (27%). Of course, the control condition was a real social dilemma in which the net benefit of the contribution was truly positive.

Similar results have been found by Peregrine Schwartz-Shea and Randy Simmons.[77] Notice that the parochialism effect is found despite the fact that an overall analysis of costs and benefits would point strongly toward the opposite result. Specifically, cooperation is truly beneficial, overall, in the one-group condition, and truly harmful in the two-group condition, because the contribution is lost and there is no net gain for others.

This kind of experiment might be a model for cases of real-world conflict in which people sacrifice their own self-interest to help their group at the expense of some other group. We see this in strikes and in international, ethnic, and religious conflict, when people even put their lives on the line for the sake of their group and at the expense of another group. We also see it in attempts to influence government policy in favor of one's own group and at the expense of other groups, through voting and contributions of time and money. We can look at such behavior from three points of view: the individual, the group, and everyone (the world). Political action in favor of one's group is beneficial for the group but, in these cases, costly to both the individual and the world.

In part, parochialism seems to result from two illusions. In one, people confuse correlation and cause, thinking that they influence others becausethey think, "I am just like others, so if I contribute, they will too."[78] In the second illusion, people think that self-sacrifice for

---

[77] *See* Schwartz-Shea & Simmons, *supra* note 74, at 124-25 (finding that discussion increased parochial behavior and that outgroup identity affected the extent of such behavior); Peregrine Schwartz-Shea & Randy T. Simmons, *The Layered Prisoners' Dilemma: Ingroup Versus Macro-Efficiency*, 65 PUB. CHOICE 61, 69-80 (1990) (testing the effects of discussion, outgroup identity, and decision-making structure on ingroup cooperation).

[78] *See* George A. Quattrone & Amos Tversky, *Causal Versus Diagnostic Contingencies: On Self-Deception and on the Voter's Illusion*, 46 J. PERSONALITY & SOC. PSYCHOL. 237, 244 (1984) (hypothesizing that people will make choices to "induce" other like-minded

their own group is in their self-interest because their contribution "comes back." They do not work through the arithmetic, which would show that what comes back is less than what goes in.[79] People who sacrifice on behalf of others like themselves may be more prone to the self-interest illusion because they see the benefits as going to people who are like themselves in some salient way. They think, roughly, "My cooperation helps people who are *X*. I am *X*. Therefore it helps me." This kind of reasoning is easier to engage in when *X* represents a particular group than when it represents people in general. Thus, this illusion encourages parochialism.[80]

1. Parochialism Is Moralistic

If parochialism is a moralistic value for co-citizens, then people are willing to impose it on others, even when they disagree and even when it goes against the greater good. This tendency might be greater when nationalism can be justified by perceived unfairness. In an experiment to test whether parochialism is moralistic, I manipulated unfairness by telling subjects that "your nation," i.e., the subject's nation, had already made a substantial contribution to some public good.[81] Even though it would be best for all if the nation continued to contribute, people might feel that it was another nation's turn. The experiment involved two kinds of situations (with four examples of each): one like a prisoners' dilemma in which each of two nations decided independently on its action, and one in which an external authority could impose a solution. A typical screen began:

> This case involves a dispute about contributions to a peacekeeping force, which needs reinforcements. It is best for each nation to maintain its current contribution, whatever the other nation does. But casualties will rise from 1% to 5% per year without reinforcements.

---

people to do the same).

[79] *See* Jonathan Baron, *The Illusion of Morality as Self-Interest: A Reason to Cooperate in Social Dilemmas*, 8 PSYCHOL. SCI. 330, 334 (1997) ("People may fail to attempt a quantitative comparison between the costs of cooperation and the benefits to the self that result from the effect of cooperation . . . . The latter benefits exist, but they are typically . . . small compared with the costs.").

[80] *See* BAZERMAN ET AL., *supra* note 2, at 104-06 (explaining that this type of reasoning is the cause behind much special-interest group action).

[81] These experimental results are reported in the unpublished manuscript cited *supra* note 60.

In the unfairness condition, the following was added:

> Your nation has already contributed an additional 50% and has committed somewhat more troops and equipment than the other nation. (Further contributions are based on your nation's current level.)

The experiment continued:

> Consider the following three proposals for a choice to be made by your government. [In the unfairness condition, the decision would be made by an international agency.]
>
> *A*: Your nation contributes 40% more. The casualty rate will remain at 1%, even if the other nation does nothing.
>
> *B*: Your nation contributes 20% more. The casualty rate will rise to 3% if the other does nothing, and it will stay at 1% if the other nation also contributes 20%.
>
> *C*: Your nation does nothing. The casualty rate will rise to 5% if the other nation does nothing, 3% if it contributes 20% more, and 1% if it contributes 50% more.

When the decision was made by the international agency, the three options all maintained the best outcome but varied the contributions. For example:

> Your nation contributes 40% more. The casualty rate will stay at 1%.
>
> Both nations contribute 20% more. The casualty rate will stay at 1%.
>
> The other nation contributes 50% more. The casualty rate will stay at 1%.

Five questions followed. The two of greatest interest, (4) and (5), are shown here, with alternatives in brackets. (The term "approved" will be explained shortly.)

> (4) What should *your government* [alternatively, "*the international agency*"] choose if almost everyone voted for (approved) *B* in an advisory referendum and almost nobody voted for (approved) *A* or *C*? [Options were *A*-*C*.]
>
> (5) What should *your government* [alternatively, "*the international agency*"] choose if almost everyone voted for (approved) *A* in an advisory referendum and almost nobody voted for (approved) *B* or *C*? [Options were *A*-*C*.]

Option *A* was always worst for the subject's nation, but it required less sacrifice than option *C* would require for the other nation.

Even when a majority favored one of the proposals other than the one that favored the group, i.e., Self, subjects still favored the Self proposal: 4% in the fair condition when the decision was national, 6% when made by an external authority, 9% when unfair and national, and 12% when unfair and external. The effects of the type of

decision and fairness were both significant. These results support the view that parochial values are sometimes moralistic.

2. De-Biasing Parochialism with Approval Voting

The experiment also contrasted two types of voting. In approval voting, voters say yes or no to each of several candidates or proposals. The option with the most approvals wins. By contrast, in standard plurality voting, voters vote for one option, and the option with the most votes wins. Approval voting has many well-known advantages over plurality voting.[82]

Approval voting could reduce parochialism if people could see themselves as members not only of their own group, but also of the larger group that includes affected outsiders; they would then approve proposals consistent with both views. Such voters may be torn between the greater good for all and the demands of the self-interest illusion for their narrow group.[83] In the present experiment, like others I have done, approval voting generally yielded more votes for the most efficient proposal (counting approval votes).

An important question is whether the use of approval voting itself reduces moralistic, i.e., Self, responses about how government should respond to the votes of others. The experiment provided some evidence for this. At least in subjects who showed an effect of approval voting in other questions, the number of Self responses to the two questions of interest was lower with approval voting than without it. In sum, approval voting reduced the expression of parochial moralistic values themselves. It may be that nationalists are more tolerant of others who vote for the greater good when the voting is by approval voting.

CONCLUSION

The main argument of this Article is that biases away from utilitarianism may take the form of moralistic values, in which people seek to impose their values on the behavior of others, sometimes explicitly ignoring the nature of others' good (utility). I have presented experiments showing this for allocation biases and protected values and

---

[82] *See* STEVEN J. BRAMS & PETER C. FISHBURN, APPROVAL VOTING 3-11 (1983) (describing the advantages that approval voting has over plurality voting and considering possible objections).

[83] *See* Baron, *supra* note 79, at 330-34 (reporting experimental results that examine the self-interest illusion).

for parochialism. Moralistic values are often supported by beliefs that make them appear to be altruistic or moral. The beliefs are necessarily incorrect and are thus amenable to correction in varying degrees. I have discussed some experiments showing that nonconsequentialist biases can be reduced.

In the long run, we might be able to de-bias moralistic values as well. Many of these values might arise partly as errors based on false beliefs about the good of others. A group may think that others are more like them than they are; their good may actually differ. If people could come to see moralistic values as possible errors, they would be more open to discussion about those values, and more open to evidence about the true nature of other people's good.