Cognitive biases in moral judgments that affect political behavior

Jonathan Baron

Abstract   Cognitive biases that affect decision making may affect the decisions of citizens that influence public policy. To the extent that decisions follow principles other than maximizing utility for all, it is less likely that utility will be maximized, and the citizens will ultimately suffer the results. Here I outline some basic arguments concerning decisions by citizens, using voting as an example. I describe two types of values that may lead to sub-optimal consequences when these values influence political behavior: moralistic values (which people are willing to impose on others regardless of the consequences) and protected values (PVs, values protected from trade-offs). I present evidence against the idea that voting is expressive, i.e., that voters aim to express their moral views rather than to have an effect on outcomes. I show experimentally that PVs are often moralistic. Finally, I present some data that citizens' think of their duty in a parochial way, neglecting out-groups. I conclude that moral judgments are important determinants of citizen behavior, that these judgments are subject to biases and based on moralistic values, and that, therefore, outcomes are probably less good than they could be.

Keywords   Cognitive biases · Moral judgment · Moralistic values · Protected values · Parochialism · Citizenship

J. Baron (✉)
Department of Psychology, University of Pennsylvania, Philadelphia, USA
e-mail: baron@psych.upenn.edu

⁂ Springer

1 Introduction

Political behavior such as voting has special properties. Although the collective input of citizens is a major determinant of their standard of living as a group, the behavior of a single citizen has almost no effect. Some citizens seem to think of politics as a kind of market, in which voting is analogous to buying, so that more popular "products" are produced more. Yet, in a market, you get what you buy, but in an election, you have a minuscule chance of affecting the outcome.

The idea of the market is roughly incorporated into the rational choice theory of political behavior. (Green and Shapiro 1994, critique this tradition.) The essential idea is that voters vote their interests. If this happened, then politics would lead to outcomes that roughly maximize total utility. The main exceptions would occur when the losses for minorities exceed the gains for majorities. Hence, some sort of protection of minority rights is needed.

Since the work of Downs (1957), it has been generally accepted that this approach cannot work quite as advertised. Consumers in the market have good reason to inform themselves about their purchases, because they will suffer the consequences. Citizens do not have sufficient incentive to be well informed about political choices, exactly because their behavior has so little effect on their self-interest. Downs used the term "rational ignorance." Recently, Cohen (2007) has extended this argument to include "rational irrationality," which goes beyond mere ignorance toward systematic bias. Cohen argues that people have preferences for beliefs and that they engage in wishful thinking, which tends to reinforce the beliefs that they prefer.

Here I want to go one step further. Wishful thinking is not the only systematic distortion that affects political behavior. There are other biases as well. To define bias, we need first to have an ideal standard, a normative model (in the specific sense of "normative" used here, which refers to a standard of evaluation). I shall adopt the standard of utilitarianism. Thus, I shall compare political judgments to those that would follow a utilitarian theory. Because utilitarianism has the standard of maximizing good, any departure from that standard should yield outcomes that are less good, i.e., worse. Biases are thus defined as systematic patterns of judgment that make things worse on the whole, for everyone.

The use of utilitarianism as a standard is useful even if it is not the correct theory of what is morally right. In particular, it may be approximately true, in politics, that, when we aim for something, we are more likely to get it than when we aim for something else. This can be stated as what I call the *simple-effect hypothesis*: decisions made by trying to follow principle P usually yield outcomes more consistent with P than are the outcomes of decisions made in other ways. If this is approximately true, then trying to maximize utility leads to outcomes that are better on the whole than trying to do something else. This hypothesis is of course empirical; I have argued the empirical case elsewhere (Baron 1998). To the extent to which it is incorrect, we get better results by, for example, trying to follow rules even when we do not think they

will lead to the best outcomes.[1] To the extent to which the simple-effect hypothesis is correct, however, then knowledge of how people are biased away from utilitarian judgments can help to explain why political outcomes are less good than they might be. If these "biases" are the result of following correct but non-utilitarian moral rules, then at least we know the cost of those rules in terms of outcomes. I shall assume that the simple-effect hypothesis is approximately true, at least for the kinds of cases that I discuss here.

1.1 The utilitarian citizen

Let us assume that citizenship consists of opportunities to vote, which are all relatively low-cost so that sacrifice of self-interest for mere participation is not an issue. I this put aside the cases of serious activism, although these can be analyzed by extension. Likewise, assume that self-interest is not relevant to information gathering (despite the argument for rational ignorance). The latter assumption is plausible if utilitarian citizens pool resources to support an agent that collects information for them, essentially telling them how to vote. Under these assumptions, the utilitarian voter as I define her would vote for whichever candidate, and whichever ballot propositions, maximize total good in the long run for everyone.

In this article, I shall focus on three aspects of utility maximization. The first is that utility maximization requires considering the goals of those affected, not just the moral judgments of decision makers about what these goals ought to be. The second is that it involves trade-offs. Sometimes we must give up some good in order to do more good, especially when the latter involves more people or a higher probability. The third is that total good includes everyone, including those who do not vote. In particular, it includes children, the not-yet-born, and foreigners. Utilitarian citizens are willing to use the little power they have to help humanity, even at some cost to their own political unit.

I shall examine some possible deviations from utilitarian behavior. First, I consider the possibility that political behavior, being of little consequence, is "expressive" of emotional responses and moral values other than the utilitarian value of better consequences (Brennan and Hamlin 2000; Brennan and Lomasky 1993). The idea here is that voters, knowing that their vote doesn't really count, use it to express their reactions. An obvious problem with this account is the low vote for third-party candidates who have no chance of winning. (Janowski 2007, reviews the literature on this point.) If voters were expressive, they would vote for these candidates. Instead, they seem to behave as if their votes mattered. Likewise, they turn out more in close elections (Edlin et al. 2007). I shall present additional evidence against expressive voting. A crucial question is whether choices are any different for voters than for decision makers.

Next, I note that some values are sometimes thought of as absolute, so that they cannot be traded off. Such a conception makes citizens unwilling to compromise.

---

[1] If we are reasonably sophisticated, however, we might know that we follow the rules exactly because they lead to better outcomes, in which case it may be said that we are indeed trying to bring about the best outcomes, but not trusting our initial judgment of how to do that.

Many of these values are moralistic, that is, values that do not necessarily correspond to the values of those affected. Thus, moralistic citizens try to impose their values on others.

Finally, I consider the concept of citizens' duty. Many people believe that their duty is to their group. In the realm of world affairs, such an attitude of "parochialism" can lead to overall harm, because the citizens of each nation try to bring about what is best for that nation even if it harms outsiders more than it helps insiders, relative to an alternative policy.

My main task is thus to contribute to the understanding of how voters' thinking can lead to consequences that are far from the best possible outcomes. Such understanding provides, in a way, its own prescriptive model for improving the situation. When people learn from scholarship, they may come to think differently. People can actually learn to think differently in lasting ways (Larrick et al. 1990; Nisbett et al. 1987). Changes in law and other institutions may also help people to work around their biases, so as to produce better consequences (e.g., Sunstein and Thaler 2003; Baron et al. 2005).

1.2 Special properties of political action

Political action has some special properties as a decision problem. Most of it 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. 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.

Voting (like cooperation in a social dilemma) is typically not justifiable by the narrow self-interest of the voter (Downs 1957). Popkin (1991) uses the term "low information rationality," suggesting that political behavior is influenced by cognitive short cuts that might not be applied (for example) to consumer decisions with real consequences for the decision maker. 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 (Brennan and Hamlin 2000; Brennan and Lomasky 1993).

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, intuitive judgments, and, I shall suggest, moral (or moralistic) principles. Political behavior might therefore be more subject to decision biases, fallacies and errors than is market-oriented behavior.[2]

---

[2] Camerer (1987) gives an example of how market discipline can reduce biases in repeated situations. But see Beattie and Loomes (1997) and Camerer and Hogarth (1999) for evidence that incentives sometimes have little effect.

The social science literature generally supports the conclusion that voting is predicted better from moral beliefs than from narrow self-interest (Brodsky and Thompson 1993; Sears and Funk 1991; Shabman and Stephenson 1994). This conclusion is not necessarily true, but it will tend to be true if people understand that voting has only a probabilistic effect on themselves individually. If people vote out of a sense of moral obligation, they will vote according to their moral beliefs rather than their self-interest. Of course, this is only a tendency. People do vote their self-interest, for interesting reasons, and often their interests and their moral beliefs, conveniently, coincide.

2 Types of goals

The analysis of voting (and other political behavior) requires some further distinctions between types of goals that voting might achieve, or not achieve (Baron 2003). 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 depend on the goals of others. This is the distinction between altruism and self-interest.

The other dimension is whether they concern others' voluntary behavior or not. When goals concern others' behavior, you want other people to behave in certain ways. These are the goals that lead to "endorsement" of principles in the most general sense, which includes teaching principles to children or students, setting an example of following the principles, gossiping about people who violate them, doling out punishment and reward in a way designed to encourage people to follow the principles, and so on. Such endorsement is, I argued, a relevant piece of morality.[3]

	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.[4] 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 X are thus a replica in you of X'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.[5] For present purposes, the critical feature of altruism is that the utility of some option, for the individual who chooses it, would be

---

[3] The present distinction is similar to Sen's (1977) distinction between sympathy and commitment (p. 326). Dworkin (1977) also distinguishes internal and external preferences, the latter being those who hold for others.

[4] I do not assume that goals cease with death. Indeed, I have argued that they may continue (Baron 1996).

[5] 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.

lower than it is if it did not help to achieve the goals of others. A clear test of altruism, in rational people, is that they will not do something at all if they do not think it does any good for others.

The most interesting type of goals in politics are those you have for the voluntary behavior of other people, particularly the moralistic goals, which do not depend on their goals. "Behavior" includes their goals as well as their overt behavior.[6] People could want others (and themselves) not to engage in homosexual behavior, or not to desire it. Other examples abound in public discourse: antipathy to drug use; enforcement of particular religions against other religions; promotion of certain tastes in fashion, personal appearance, or artistic style against other tastes; and so on.[7] Moralistic goals have the character of being "objective moral obligation" in the sense of Turiel (1983) and Miller et al. (1990): people must follow these rules whether they think they should or not. They are also roughly equivalent to other ideas in the literature, such as Sen's (1970) concept of "meddlesome preferences" and Congleton's (2007) "moral voter."

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" [Jervis 1976; Baron 1998]). Talk of consequences can appeal to those who do not share the moralistic goals. For example, opponents of homosexuality claim that the behavior increases mental disorders, and this argument, if true, is relevant to those who do not find homosexuality inherently repugnant.

The people affected by moralistic goals may or may not be limited to some group. Your goals for others' behavior could be limited to others who share your religion or nationality.

Unlike altruism, moralism, as I define it, is independent of the goals of others. Moralistic goals are interesting exactly because they can go against the goals of others. (Such conflict is more likely in heterogeneous societies.) When moralistic goals play out in politics, they interfere with people's achievement of their own goals. That is, if we define "utility" as a measure of goal achievement, they decrease utility.

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 be unrelated to your true underlying goals (Baron 1996). Undoubtedly moralists often believe that they are altruists in just this way. The experiments I present later, 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.

---

[6] And it excludes involuntary or coerced behavior. When we endorse behavior for others, we want them to want to choose it.

[7] 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.

What I call a "moral" goal is the goal that people behave so as to achieve each others' goals. We could call this the utilitarian goal (Baron 1996). Moral goals are also similar to moralistic goals in that both may 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. Moral goals are achieved to the extent to which each person behaves so as to achieve people's goals (including his own).

Each of us benefits from the altruistic goals of others (assuming that others pursue these goals rationally). And we also benefit from others' moral goals, if these goals are effectively achieved. In that case, people encourage other people to be altruistic and to endorse moral goals, a kind of virtuous circle.

The benefit from moralistic goals is less sure. If the moralistic goals of others happen to lead people to promote my own goals (of all sorts), I am in luck, but it is also possible that moralistic goals could impede my goals at every turn.

These goals—all four types—are the criteria by which we evaluate the individual rationality of behavior. Voting for something is rational if it achieves the sum total of these goals better than the alternatives.

Note, however, that political behavior often involves expression of opinions about what should be done. Expressions of opinion are not the same as goals. They are a form of behavior. Failure to distinguish expressions of opinion from goals (sources of utility) is a frequent source of misunderstanding of utilitarianism. For example, students often say, "Utilitarianism says that slavery is acceptable if enough people favor it." Although there is an issue here, "favoring" is not it. Rather, the question is whether the harm to the slaves can be outweighed by benefits of slavery to others. (See Hare 1963, 1981, for discussion.)

Because voting cannot be sustained by narrow self-interest alone, moralistic, moral, and altruistic goals will surely affect it, when they exist in sufficient strength. Although moralistic goals should count as part of people's utility, they may reduce the total utility from political behavior, even compared to the hypothetical outcome that would result if everyone voted according to self-interest alone.

3 Rational political action

The foregoing analysis of goals allows us to discuss the rationality of voting or other political action. Voting (or other political action) is unlikely to be rational on the basis of narrowly self-interested goals, as I have defined them. We might think that voting is rational if the voter gets a good feeling from doing it, but such good feelings seem to me to be dependent on the existence of altruistic, moral, or moralistic goals.

Voting can be rational for you, however, if you are sufficiently altruistic (Baron 1997a). Roughly, the probability that your vote will have an effect is proportional to $1/N$, where N is the number of voters (Edlin et al. 2007; Margolis 1983, p. 325; Mulligan and Hunter 2001), but the benefit of your vote is roughly proportional to N, assuming that you vote for better candidates and that the number of voters is a fixed proportion of the number affected by good policies. Thus, simplifying further, if the utility of a better policy for each person is U, the expected effect of a vote on adoption

of a policy is $1/N$ (rather than just being proportional to $1/N$, as I assumed), and you are completely altruistic, then your expected utility of voting is $U \cdot N/N$, or U. If the benefit of the better policy for the average person is greater than your cost of voting, you should vote. The argument becomes stronger if the population affected is much larger than the number of voters. If Americans, to take an example, assumed that their votes had effects on the rest of the world and on people not yet born or not old enough to vote, the utilitarian benefits of voting could increase by orders of magnitude.

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 N people into altruists just like you, who would then vote altruistically, your influence would be multiplied by N. Of course, this is unlikely.

Voting may 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 each other person 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 their votes are expected only to reduce total utility.

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

Voting on the basis of moral or altruistic goals can lead to very different sorts of voting than those actually observed. If we assume that altruism and morality is not limited to your own political unit, then you should vote to achieve the best outcome for everyone. You should vote as a utilitarian, concerned with the whole world now and in the future. In the real world, altruism may be more limited, and that limitation would yield outcomes that are worse than the utilitarian optimum just described, although still possibly better than the current situation.

4 Empirical evidence on political action

Lots of evidence indicates that people generally do *not* simply vote their economic self-interest.[8] Political behavior in general is determined more by ideological commitments—by moral intuitions, as I have called them here. People vote on behalf of property rights, aid for the poor, tax reduction, economic nationalism, or whatever,

---

[8] Brennan and Lomasky (1993), Brodsky and Thompson (1993), Sears and Funk (1991), Shabman and Stephenson (1994).

when it is consistent with their moral view of the world, almost without regard to how it affects their pocketbook.[9]

Several experiments support two related conclusions, both of which are important in what follows. (Because the same experiments often support both conclusions, I shall review them together.) One conclusion is that voting is more heavily influenced by moral considerations than is decision making that has a direct and large influence on self-interest. The second conclusion is that voting is at least somewhat independent of self-interest.

Eichenberger and Oberholzer-Gee (1998) asked subjects to play two games, in a variety of conditions. In the "dictator game," one player (the dictator) in each pair of players was given some money (7 Swiss Francs) to divide between himself and another player. The other player had no choice. In some conditions, the dictator was selected as a result of better performance on a homework assignment, which gave him some right to the money. In this basic condition, the average dictator gave 34.6% of the money to the other player.

In the "gangster game," the player without the money, the gangster, could take as much as she wanted. The average gangster gave 24.1% to the other player. In a "democracy game," subjects of each type voted on different proposals for how much money to leave the players of the other type (with or without the money). The amount left by gangsters (those without) was 48.2%, much higher than the amount they left individually. (The dictators did not change significantly, possibly because their original allocations were close to what subjects judged to be fair when they did not know which group they were in.) When gangsters were asked in a survey what they would do, the resulting allocations to the other player (if made) would amount to 48.9% in the individual condition (the original gangster game) and 60.2% in the democracy game.

The groups in the democracy game were small enough (a small classroom) so that each vote could affect the outcome, so self-interest was involved somewhat more than in real politics. Yet, the reduced role of self-interest made the gangster subjects in the democracy game behave more like those asked hypothetical questions than those who played "for real." And these responses were close to what the subjects judged to be fair, from a moral point of view.

Kroll et al. (2001) asked subjects to play a prisoner's dilemma game under two conditions. In one condition, the usual prisoner's dilemma (PD), players played other players as individuals for 25 games. In each game, the payoffs were 25 points each if both players cooperated, 15 each if both defected, and 35 and 5, respectively, if one player defected and the other cooperated. In a voting condition (DD, for "direct democracy"), subjects played the same game in two groups of five, each with partners in the other group. If a majority voted to cooperate, then everyone in the group was counted as making a cooperative response, and likewise for defection. Subjects were more likely to vote cooperatively in the DD condition: cooperation declined

---

[9] Actual correlations between voting and monetary self-interest may even be explained in terms of people choosing their course of life so that their financial interests tend to coincide with their ideologies. Left-wingers rarely become investment bankers, and right-wingers rarely go into social work. Also, people who become investment bankers tend to take on the attitudes of their associates.

essentially to zero in the PD condition, over the course of 25 games, while it remained at about 20% in the DD condition even at the end of the series of games.

One explanation of this result is that the DD condition is low-cost. In the, PD condition, you gain 10 points by defecting, regardless of what the other player does. In the DD condition, the benefit of defecting (voting for defection) depends on the probability that your vote is critical, which is probably less than .5 (and is probably perceived as low). The benefit of cooperating *for others* is also diluted in the same way. A "weighted utilitarian," who followed utilitarianism except for giving himself a constant extra weight, would not behave differently in PD and DD. But the expressive value of doing the right thing may be less dependent on its expected consequences (Margolis 1982).

Tyran (2002) did a similar study in which the "right thing" was contributing real money (from the experimental payment) to a charity. In all conditions, contributions to the charity were decided by vote. The experimental conditions varied in the number of votes required for the contributions to be made, and in whether "no" voters had to contribute or not. If no-votes did have to contribute, then the effect of your vote on whether you had to contribute was small; your vote determined your contribution only if it was pivotal. If no-votes did not have to contribute, your vote affects you in a much broader range of conditions, specifically, if enough others vote "yes."

Voting in favor did not depend on whether no-votes had to contribute, but they increased as a function of the subject's expectation that the proposal would be approved. Evidently, subjects wanted to do their part if others voted for contributing. Most importantly, subjects ignored the effect of their vote on the outcome for themselves.

In a related study, Tyran and Sausgruber (2002) gave subjects the chance to vote, in groups of five, in favor of redistributing payments from the two rich members of each group to the single poor member. One result of interest was that even the two rich members voted for the redistribution about 34% of the time, and the two middle-class voters (who did not stand to gain or lose) voted for it 70% of the time. A second result is that voters were no more likely to favor redistribution when they thought that their vote was pivotal than when they did not; in this case, self-interest effects are undetectable.

Although self-interest itself does not seem to play a large role in voting decisions, people may think that it does. They may deceive themselves into thinking that their vote really does affect outcomes for themselves, so that, therefore, they should vote according to what is best in terms of their self-interest. Baron (1997b) found that people thought of cooperation in social dilemmas—such as fishers cutting back their catch in order to preserve the fishery—was consistent with self-interest, even when numbers were available showing that cooperation, while benefiting everyone in total, hurt the individual cooperator. It is as if people do not like the idea that morality and self-interest conflict, so, when they side with morality, they convince themselves that the conflict does not exist. (And they probably do this too when they side with self-interest.) Such an illusion of morality as self-interest seems to be at work as well in what I call parochialism: making choices that favor one's group even when the total effect is negative, taking outsiders into account (Baron 2001). People think, "What helps my group helps me, because I am a member of my group. Therefore, my own self-sacrifice helps me." Again, they fail to do the arithmetic to see that cooperation with their own group is a net loss for themselves as well as for everyone (counting outsiders).

These results together suggest that self-interest plays a small role in political behavior, although people sometimes think that it does. Political behavior seems to be more influenced by moral judgments.

Note that the claim that voting is influenced by moral judgments, rather than self-interest, is not the same as the claim that voting is expressive. Moral voting could still be concerned with consequences of the vote itself. For example, a moral voter could vote strategically, for the less immoral of the candidates who have a chance of winning. An expressive voter would vote for the morally best, regardless of the chance of winning.

5 Web studies

This article reports several experiments, all done using questionnaires on the World Wide Web. The subjects in all studies were members of a panel recruited over a 10-year period, mostly through their own efforts at searching for ways to earn money by completing questionnaires. Approximately 90% of respondents were U.S. residents (with the rest mostly from Canada). The panel is roughly representative of the adult U.S. population in terms of income, age, and education but not in terms of sex, because (for unknown reasons) women predominate in our respondent pool (typically about 75%). Although the sample is not a random sample of Americans, the population of interest is people in general, so a random sample is impossible, and at least the sample here is quite varied. Other studies of political attitudes with this panel suggest that they range from extreme anti-government conservatives to near-communists who want the tax system to equate income for everyone. The panel includes evangelical Christians and atheists.

For each study, I sent email to about 500 members of the panel, saying how much the study paid and where to find it on the World Wide Web. Each study was a series of separate web pages, programmed in JavaScript. The first page provided brief instructions. Each of the others presented a case and some questions, with a space for optional comments. Otherwise the subject had to answer all questions in order to proceed. The order of cases was randomized separately for each subject. The study was removed when about 75 responses had been submitted, with a target of 80. The studies are all available at http://www.psych.upenn.edu/~baron/ex/.

Subjects were paid, typically $3 or $4 for a study that takes 15–20 min. They were also timed on every response, and those who responded very quickly (defined as outliers in the distribution of times for the fastest half of the pages) were omitted and not paid.

Like much of political behavior itself, the responses in these studies are "cheap talk," with no consequence for the subjects.

6 Trust, expressiveness, intuition and ethics

The first experiment addressed the question of expressiveness. People may know that their political behavior is expressive, less concerned with consequences than is ideal for their society. They may be inclined to trust government agencies that are less subject to intuition and bias than they themselves are in their own political behavior.

I report an experiment here that address this issue by asking people about decisions for themselves, voting, other sorts of expression, and trust. Each decision contrasts a biased option and an optimal option. The general conclusion is that the hypothesis just described received little support. However, people gave wider latitude to government agencies than would be inferred from their own decisions or voting. And, faced with a contrast between what is "ethical" and what is based on "reason," decisions of all sorts are typically somewhere in between.

Baron (2008, Chap. 17) discuss several non-consequentialist biases that affect judgments about allocation. These biases often appeal to intuition, and they sometimes seem more ethical. The term "ethics," invokes a view of morality that tends to be more deontological than utilitarian. People may thus see a conflict between, on the one hand, emotion, intuition, and ethics, and, on the other hand, reason, optimality, and effectiveness.

The question asked in the present experiment is whether people can trust government agencies that optimize, even when ethics and intuition favor a competing policy. I used several different biases (nine, in fact, listed below).

On each trial of the experiment, subjects saw a choice between two policies for vaccinations. One policy had better consequences in terms of illness or death. The other policy appealed to some well-known bias (except in two cases which were a control condition). The subjects indicated which policy was more ethical, which appealed more to reason, intuition, and emotion, which led to better outcomes, and which inspired more trust in the agency that chose it.

6.1 Method

One hundred and seventy subjects completed the questionnaire. Nineteen were omitted either because of excessive errors in the control condition (described below) or extremely fast response times. Of the remaining 151 subjects, ages ranged from 18 to 72 (median 39), 23% were male, and 15% were students. The introductory page read as follows:

> Vaccinations
> This questionnaire concerns decisions made by a government agency about vaccinations against diseases, both those that might arise naturally and those that might arise from terrorism.
> Each question involves a choice of two programs. *The government agency cannot afford both.*
> You answer 7 questions about each of 20 choices. It usually does not make sense to give the same answer to all 7 questions.
> Some questions assume that there is a conflict between "reason" and something else, such as "intuition" or "emotion." Do your best with these.

Here are the items. The names were not presented to subjects. Notice that B is always the optimal item. The phrases in brackets after each item were used in more specific questions about bias and optimality. Items 19 and 20 were control conditions without any bias.

1. Omission
   A. One vaccine will reduce the number of people getting serious flu from 50 out of every 1,000 people to 25. [has less harm caused by the vaccine]
   B. Another vaccine will reduce the number of people getting serious flu from 50 out of every 1,000 to 10, but it will cause a similar type of flu in 10 (out of 1,000). [lead to fewer cases of flu]
2. Omission
   A terrorist attack of smallpox is expected to infect 1,000,000 people.
   A. One vaccine will prevent 500,000 of the cases. [has less harm caused by the vaccine]
   B. Another vaccine will prevent 750,000 of the cases but cause a similar (and equally serious) disease in 200,000 people. [lead to fewer cases of serious disease]
3. Human
   A. A serious bacterial disease is resistant to antibiotics because of antibiotics given to farm animals as a preventative measure. A vaccine will reduce the annual number of cases (in people) by 750, from 1,000 to 250. [does better at undoing harms that people cause]
   B. An equally serious (natural) bacterial disease is resistant to antibiotics because no antibiotic has been discovered that will cure it. A vaccine will reduce the annual number of cases by 800, from 1,000 to 200. [lead to fewer cases of bacterial disease]
4. Human
   A. A serious flu was created artificially in the laboratory as part of an experiment. The virus escaped by accident. A vaccine will reduce the annual number of cases by 750, from 1,000 to 250. [does better at undoing harms that people cause]
   B. An equally serious natural form of flu has been around for decades. A vaccine will reduce the annual number of cases by 800, from 1,000 to 200. [lead to fewer cases of flu]
5. Ex-ante
   A. One vaccine is given to everyone. It reduces the number of people getting serious flu, in an epidemic, from 5.0% of the population to 3.0%. [is fairer in giving people an equal chance]
   B. Another vaccine is in short supply, and is given to half of the population picked at random. It is completely effective for those who get it, so it reduces the number of people getting serious flu from 5.0 to 2.5% of the whole population. (The 2.5% are people who did not get the vaccine.) [lead to fewer cases of flu]
6. Ex-ante
   A. One vaccine is given to everyone. It reduces the number of people getting smallpox, in a terrorist attack, from 5.0% of the population to 3.0%. [is fairer in giving people an equal chance]
   B. Another vaccine is in short supply, and is given to half of the population picked at random. It is completely effective for those who get it, so it reduces the number of people getting smallpox from 5.0 to 2.5% of the whole

population. (The 2.5% are people who did not get the vaccine.) [lead to fewer cases of smallpox]

7. Zero
   A. One vaccine, given to everyone, cuts the prevalence of one type of serious flu from 2.5% of the population to zero. [does better at eliminating risk completely]
   B. Another vaccine, given to everyone, cuts the prevalence of another type of (equally) serious flu from 10.0% of the population to 7.0% of the population. [prevents more cases of flu]

8. Zero
   Two anthrax epidemics have been started by a terrorist.
   A. One epidemic will cause 10,000 cases of anthrax. A vaccine, given to everyone, will eliminate all of these cases. [does better at eliminating risk completely]
   B. The other epidemic will cause 100,000 cases of anthrax. Another vaccine, given to everyone, will eliminate 11,000 of these 100,000 cases. [prevents more cases of anthrax]

9. Catastrophic
   A. One vaccine removes the risk of a life-threatening viral disease that will infect *1,000,000* people in a country of 100,000,000. The disease spreads in *outbreaks of 200,000 people at once*. [prevents the more catastrophic disease]
   B. Another vaccine removes the risk of a life-threatening viral disease that will infect *1,100,000* people in a country of 100,000,000. The disease affects unrelated people, *one at a time*. [prevents more cases of disease]

10. Catastrophic
    A. An outbreak of viral meningitis will affect 100,000 people all starting on the same day. A vaccine will prevent all of these cases. [prevents the more catastrophic disease]
    B. Another epidemic of viral meningitis will affect 110,000 people, one at a time, over the course of year. Another vaccine will prevent all of these cases. [prevents more cases of meningitis]

11. Ambiguity
    A. One vaccine removes the risk of a life-threatening viral disease. The number of people infected could be as high as 1,500,000 or as low as 500,000. All numbers in this range are equally likely. The average of these possibilities is 1,000,000. [reduces the more ambiguous risk]
    B. Another vaccine removes the risk of a life-threatening viral disease that will infect 1,100,000 people. The disease is well known and has affected about the same number each year. [omitted unintentionally]

12. Ambiguity
    A. One vaccine will prevent whooping cough (which can be life threatening for infants and others). But it will cause a disease like whooping cough in 11% of those who get the vaccine. [has the less ambiguous risk of side effects]
    B. Another vaccine will prevent whooping cough. It will cause a disease like whooping cough in 20% of those in a risk group. Half of the population is in

the risk group. (The test for membership in the risk group is not available.) On the whole, then, this vaccine will cause the disease in 10% of those who get the vaccine. [leads to fewer cases of whooping cough]

13. Distribution

There are two types of serious flu this year, A and B. Each will affect 50 out of every 1,000 people.

A. One vaccine reduces the risk of type A from 50 to 30. It reduces the risk of type B from 50 to 40. [is more fairly distributed between groups]

B. Another vaccine reduces the risk of type A from 50 to 10. It has no effect on type B (50 people). [leads to fewer cases of flu]

14. Distribution

Two groups, each with 1,000 people, have been exposed to smallpox, and everyone without a vaccination in each group will get smallpox. A vaccine is 80% effective in group A, 50% effective in group B. There is enough vaccine for 1,000 people.

A. Divide the vaccine equally between the two groups. (Pick who gets it with a lottery.) The vaccine will prevent 400 cases in group A, 250 in group B. [is more fairly distributed between groups]

B. Vaccinate group A only. The vaccine will prevent 800 cases of smallpox. [prevents more cases of smallpox]

15. Familiarity

A. One vaccine removes the risk of a *new, previously unknown*, life-threatening viral disease that is expected to infect 100,000 people. [prevents the less familiar disease]

B. Another vaccine removes the risk of a *well-known*, life-threatening viral disease that is expected to infect 110,000 people. [prevents more cases of disease]

16. Familiarity

A. A vaccine used for 10 years is 50% effective against a type of influenza that infects 1,000,000 people each year. [uses a more familiar vaccine]

B. A new vaccine, tested in research with 10,000 people but never used on a large scale, is certain to be 55% effective against the same type of influenza. [prevents more cases of influenza]

17. Immediacy

A. One vaccine removes the risk of a life-threatening epidemic that will begin *immediately* and kill 1,000 people before it dies out. No other vaccine can be given for this disease, if this one is used. [reduces the more immediate risk]

B. Another vaccine removes the risk of a life-threatening epidemic that will begin *in 6 months* and kill 1,100 people before it dies out. No other vaccine can be given for this disease, if this one is used. [prevents more cases of disease]

18. Immediacy

A type of serious influenza affects 1,000,000 people every year.

A. One vaccine can be given right now and provides immunity for 3 years. (Those immune will not get the disease during the three years.) No other

vaccine can be given for this disease, if this one is used, and this vaccine can be given only once. [reduces the more immediate risk]
  B.   Another vaccine will be ready one year from now. It provides immunity for 4 years. No other vaccine can be given for this disease, if this one is used, and this vaccine can be given only once. [prevents more cases of disease]
19.   Control
  A.   One vaccine for flu is 80% effective in preventing the flu. [provides the greater risk reduction]
  B.   Another vaccine is identical except that it is 90% effective in preventing the same flu. [leads to fewer cases of flu]
20.   Control
  A.   One vaccine for viral meningitis is 80% effective in preventing the disease. [provides the greater risk reduction]
  B.   Another vaccine is identical except that it is 90% effective in preventing the same disease. [leads to fewer cases of meningitis]

   The questions were listed after each item, as follows. The choices were, again, A, B, or "No difference."

Bias.   Which program [text in brackets for A, above]?
Optimal. Which program [text in brackets for B]?
Self.   If you could chose to be in one program or the other *for yourself*, which would you choose?
Ethical.   Which program seems more ethical to you?
Vote.   If you could vote, which program would you vote for?
Emotion. In the conflict between emotion and reason, which program seems more emotionally appealing?
Reason.   In the conflict between emotion and reason, which program appeals more to reason?
*Intuition*. In the conflict between intuition and reason, which program seems more intuitively right?
Trust.   Which choice would make you trust the government agency more, so that you would be more willing to see it have authority over decisions like this in the future?

For even items (or odd items for half the subjects), the order of Bias and Optimal was reversed, and the order of the remaining items was also reversed. The names are for use in describing the results and were not presented to subjects.

6.2 Results

Figure 1 shows the distribution of responses to the questions, for all items except the control items. The black part of each bar is the proportion of responses favoring the Optimal option (B), the gray part is "no difference," and the white part favors the Biased option (A).

Responses as a function of question and bias
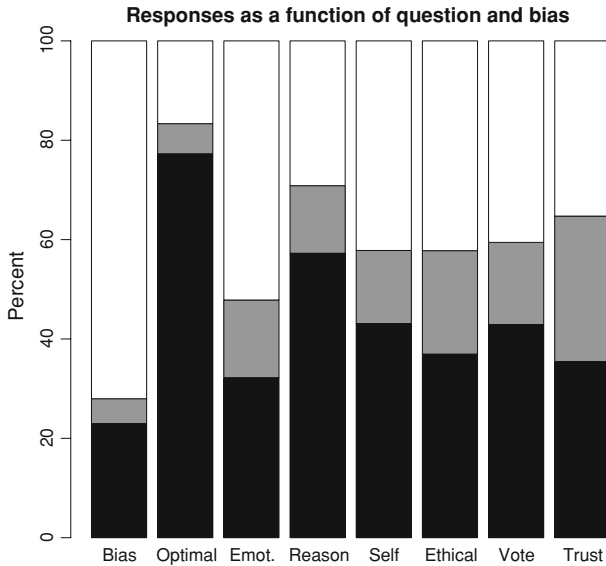
Fig. 1 Responses to questions in expressiveness study

The Trust question showed less bias than the Ethical question, ($t_{150} = 2.72$, $p = 0.0074$, across subjects; $t_{17} = 2.18$, $p = 0.0440$, across the 18 items).[10] The effect was in the number of No-difference responses ($t_{150} = 4.93$, $p = 0.0000$) and in the number of Bias responses ($t_{150} = 5.11$, $p = 0.0000$). The difference in Optimal responses was not significant.

Most importantly, Vote responses were somewhat more optimal than Self responses. This was not significant overall, but it was significant when the analysis was restricted to those cases (38 vs. 10% in the opposite direction) in which the Reason response was more optimal than the Emotion response ($t_{150} = 2.47$, $p = 0.0146$). Thus, voting seemed to be based more on the subjects' impression of the dictates of reason as opposed to the lure of emotion. This is the opposite of the conclusion we would expect from the hypothesis that voting is more expressive than decisions for the self.

The Ethical question was more closely related than the Trust question to Emotion and Bias (mean within-subject multiple $r = .59$ and $.56$ respectively), and the Trust question was more closely related than the Ethical question to Reason and Optimal ($.57$ vs. $.55$; $t_{145} = 2.36$ for the interaction, $p = 0.0197$). Trust is therefore based to some extent on perceived consequences, not just ethics and intuition.

6.3 Follow-up

The last experiment suggests that voting is based more on consequences than would be suggested by the idea that voting is expressive, more like a poll. In a follow-up study, I presented the same items but changed the questions.

---

[10] There was an order effect, but the difference between questions did not reverse significantly in either order.

Choice. If the decision were up to you, which program would you choose?
Emotion. [unchanged]
Reason. [unchanged]
Vote. If you could vote, which program would you vote for?
Bias. [unchanged]
Optimal. [unchanged]
Poll. If you were asked in an opinion poll which program you supported, which program would you favor?

Half the cases (odd or even, depending on the subject) switched the order of Choice and Poll, and of Emotion/Reason and Bias/Optimal. The latter two pairs served to separate the three critical questions, Choice, Vote, and Poll, where were thus presented in a counterbalanced order.

The 95 subjects were 76% female, and their ages ranged from 21 to 73 (median 40).

On a scale where 1 favors the optimal option, 0 is neutrality, and -1 favors the biased option, the means for the critical conditions were .209 for Vote, .201 for Choose, and .172 for Poll. Across subjects Vote was higher than Poll ($t_{94} = 2.50$, $p = 0.0142$), and Choose was higher than Poll ($t_{94} = 2.12$, $p = 0.0368$), but Vote and Choose did not differ. Thus, voting is based just as much on consequences as choice is, and both are less "expressive" than answers in an opinion poll. Although these effects are small, the questions were nearby, and subjects might have felt that it was reasonable to answer them all the same, as subjects did in 84% of all cases.

6.4 Discussion

The last two experiments do not support the hypotheses that voting is more expressive and that people are willing to trust consequence-maximizing authorities even when the authorities go against the people's own intuitive judgments.

However, people consistently give authorities a wider latitude than they give themselves in their own responses and votes. They are much more likely to choose the "no difference" option. I have observed this effect in several other studies using different methods and response options. This effect is optimistic for those who want government to be more optimal, and who think that it can be more optimal, yet still retain the people's trust (such as Breyer 1993 and Sunstein 2002). If we want to find out the limits of people's trust, it seems that we have to ask them about trust specifically, rather than about what they would do or what they would vote for.

It is also of interest that people are willing to go against what is "ethical" for the sake of better consequences, in decisions made for themselves, in voting, and clearly in what makes them trust an agency. Ethics itself is not, in people's minds, the only relevant criterion, and in some cases it competes directly with consequences and loses.

7 Protected values

People think that some of their values are protected from trade-offs with other values (Baron and Spranca 1997; Tetlock et al. 1996). 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 substituted for another is infinite. For example, no amount of money can replace the loss of a pristine ecosystem.

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 consequences, so it has value of its own.

Omission bias is greater when PVs are involved (Ritov and Baron 1999; Baron and Ritov, in press). For example, when people have a PV for species, they are even less willing to cause the destruction of one species in order to save even more species from extinction. Thus, PVs apply to acts primarily, as opposed to omissions.

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 his own choices but not (as much) to his obligations with respect to others' choices. So it is better for him not to buy the stock, even if his not buying it means that someone else will buy it.

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 (Ritov and Baron 1999). They say this more often than they say the same thing for violations of non-protected values (NPVs).

Several researchers have noted that PVs cause problems for quantitative elicitation of values, as is done in cost-benefit analysis or decision analysis (Baron 1997c; Bazerman et al. 1999). PVs imply infinite values.[11]

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, whatever they do in their own behavior. When people learn that they have violated some value they hold, they may regret their action rather than revising the value.

I am assuming here that values (or utilities) are criteria for evaluating states of affairs. Values 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. Values are not simply desires, and very young children might properly be said to have no values at all, in the sense at issue.

If values are constructed, like concepts, we can ask whether they are constructed well, just as we can ask whether concepts are formed well. It is possible that PVs

---

[11] As argued by Baron (1997c), we should not necessarily take PVs at face value. For one thing, we need to distinguish strong opinions about what should be done from strong personal utilities for such action. For another, utilitarian calculation based on value measurement can (arguably) ignore values that are based on false beliefs, as some PVs surely are.

result from the same kind of unreflectiveness that leads to overgeneralized concepts of the sort I have been discussing. People may agree with the claim that "all apples are red" without pausing to consider counterexamples. Likewise, they may endorse the statement that "no benefit is worth the sacrifice of a pristine rain-forest" without thinking much about possible benefits (a cure for cancer or malaria?). 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. Baron and Leshner (2000) indeed found that PVs are reduced by asking subjects to think of counterexamples.

It should not be surprising that people think of PVs as rules that are independent of consequences for people. An experiment reported by Baron (2003) demonstrates that PVs are sometimes moralistic. Examples of items that were most likely to evoke PVs were:

1. testing a fetus for IQ genes and aborting it if its expected IQ is below average,
2. testing a fetus for sex and aborting it if it is the same sex as all its brothers or sisters,
3. cloning someone with desired traits so that these may be passed on, such as an athletic champion or brilliant scientist,
4. modifying the genes of an embryo so that, when it is born, it will have a higher IQ,
5. giving a drug (with no side effects) to enhance school performance of normal children,
6. cloning an adult human, who had difficulty having a child, to produce a child who would be raised by the adult,
7. cloning an adult human to produce a fetus, which would be aborted and used for cells to save the adult's life,
8. cloning an adult human to produce a fetus, which would provide cells to save the adult's life and which would then be born and raised as the adult's child,
9. using artificial insemination to have a child with a desired trait, by choosing a donor with this trait, such as intelligence.

On each screen, an item was presented, followed by a series of questions, with possible answers printed on buttons (shown here as rectangles), with the proportions of answers.[12]

Should the government allow this to be done?

This should be allowed, so long as other laws are followed.	40%
This should sometimes be allowed, with safeguards against abuse.	32%
This should never be allowed, no matter how great the need.	28%

---

[12] Questions not discussed here concerned whether it was morally acceptable to allow the action, whether the consequences of allowing it would be better on the whole than the consequences of banning it, and whether it should be banned if most people opposed it.

These responses of course reflect the sample of items we used. The nine items listed above had 61% "never allowed" responses. For the rest of the items, the proportions of responses are limited to the actions that "should never be allowed."

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?

| Yes. | 8% | No. | 68% | I cannot imagine this. | 24% |

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?

| Yes, allowed. | 10% | No, banned. | 76% | I cannot imagine this. | 14% |

Of interest, subjects favored bans in 76% of the cases that they wanted to ban, 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 and produce consequences that they themselves acknowledged were worse.

7.1 Parochialism, protected values, moralistic goals, and moral realism

As another example, I present a study of nationalistic attitudes, which represent a form of "parochialism." As a technical term, parochialism is a bias toward choices that benefit an in-group even when the total effect on everyone (including the out-group) is to make things worse on the whole. More specifically, it is an under-weighing of the consequences for the out-group, compared to the utilitarian standard of weighing all those affected equally. Assuming the simple-effect hypothesis, parochial national policies will do, on the whole, less good than utilitarian policies.

Parochial values can be protected from tradeoffs, moralistic (imposed on others regardless of consequences and others preferences), and objective. By "objective," I mean that they are thought of as true, as an example of moral realism in the sense of Goodwin and Darley (2008; see Greene 2002, for a critique of moral realism in this sense). I assess moral realism using questions similar to those used by Goodwin and Darley.

A social norm (as defined by Bicchieri 2006) is a moral commitment that affects behavior contingently on others making the same commitment. Although a social norm can be understood as morally real, we might expect that social norms would be more often understood as subjective, since their implementation is contingent on whether people endorse them.

7.1.1 Method

Moral realism was defined as follows in the introduction to the experiment:

One question is whether the morality of the action is objective. This means that its truth or falsity does not depend on anybody's judgment. The opposite

of "objective" is "subjective." Subjective judgments can differ from person to person.

Examples of objective statements are: "2+2=4", "2+2=5" (which is objectively false), "the earth orbits the sun", and "a Hummer gets 28 mpg."

Examples of subjective statements are: "hot pepper tastes good", "Brigitte Bardot was the most attractive movie star of all time", "rap music is annoying", and "Woody Allen is funny."

The experiment consisted of 15 pages, each concerning a different behavior. A typical page read as follows (with percent endorsements in brackets):

Private universities in the U.S. accept foreign students while rejecting some U.S. students who are almost as well qualified.
What do you think of this action?
1. It is not a moral issue. [25%]
2. It is morally acceptable. [10%]
3. It is a moral issue, but I cannot say in general whether it is wrong or not. [12%]
4. It is morally wrong, but it should be allowed. [9%]
5. It is morally wrong, and it should be banned in most cases. [20%]
6. It is morally wrong, and it should be banned in all cases, regardless of the benefits to the outsiders. [13%]
7. …regardless of the benefits to the outsiders and citizens. [11%]
Is the moral rightness or wrongness of this objective or subjective?
    subjective [32%]
    objective [33%]
    not sure [34%].
Suppose a poll found that 80% of the voters thought that good citizens should vote to ban the action. Would this information increase your obligation to vote for a ban?
    Yes. This figure means that other citizens care, and that matters. [18%]
    No. It would have no effect on my obligation to vote for a ban. [78%]
    No. It would reduce my obligation to vote for a ban. [4%]
Suppose a poll found that 80% of the voters thought that good citizens should vote to *allow* the action. Would this information *reduce* your obligation to vote *for* a ban?"
    Yes. Other citizens do not care about their nation, and that matters. [6%]
    No. It would have no effect on my obligation to vote for a ban. [78%]
    No. It would *increase* my obligation to vote for a ban. [16%]

The 86 subjects ranged in age from 19 to 78 (median 40.5); 78% were female.

7.1.2 Results

Table 1 shows some selected actions and the percent of subjects who endorsed "ban in all cases, regardless …" indicating a PV. The selection is not based on the high-

Table 1 Actions and % "ban in all cases regardless of the benefits to outsiders" and "…and citizens" (selected)

Companies hire foreigners, helping them immigrate, while some citizens who are almost as qualified do not have jobs	23	9
Companies open new facilities in foreign countries rather than their own country, even though the foreign cost is only a little less	14	9
Non-governmental disaster-relief organizations send more help in response to a foreign disaster than to a domestic one, even though the domestic need is almost as great	13	7
Private universities in the U.S. accept foreign students while rejecting some U.S. students who are almost as well qualified	15	8
The national government gives research grants to foreign scientists, while rejecting applications from domestic scientists that are almost as worthy	16	5

est responses but on the basis of items that showed the clearest examples of PVs for parochialism. Although PVs are rare, they are present in some respondents.

Of the cases in which the subject wanted to ban the action at all, I examined the questions relevant to social norms. In 20% of these cases, greater support of others increased the subject's obligation, indicating a social norm at work. In only 3%, lower support reduced the subject's obligation.

Subjects thought that their "ban" judgments were objective in 45% of the cases. Subjects differed substantially in their responses to these items (Cronbach's α .92 for the 15 items, including all responses). Contrary to my expectation, these judgments were unrelated to judgments of social norms.

Subjects thought that their "ban" judgments were what a good citizen should do in 78% of the cases. I shall return to this issue of citizens' duty.

In sum, some people endorse parochialism even when they know that the consequences are worse for everyone. They think it is their moral duty. They are sometimes unwilling to trade off this principle whatever the consequences, and they often see it as morally real.

8 Parochialism and duty

Parochial choices may be seen as a duty. People could think that they are morally obliged to advance the interests of their nation, or some other group, with little regard for the effects of policies on outsiders. I describe here a preliminary test of this possibility.[13]

I designed hypothetical policy changes that had purely financial effects on two groups, one being the subject's group. The changes had different effects on different groups, and the groups varied in size. It was thus possible to manipulate the benefits to the in-group and out-group independently. To save the subjects the trouble of doing the arithmetic, I provided a summary of each choice.

---

[13] This study is actually a follow-up of another study, described in Baron, in press. Further studies are reported in Baron et al. (2008).

Half of the cases were designed to remove the effects of self-interest. Thus, proposals that helped the subject's group, the in-group, did not help the subject herself. Self-interest is almost surely an insufficient reason for voting for one's group, because of the low probability of having an effect, but the subjects might have disagreed with this conclusion.

8.1 Method

In the introduction, subjects were told:

> In each case, imagine that you are voting in a referendum on proposals for changes in tax law. These changes will reduced taxes by some percent of current taxes. The reductions are available because of a budget surplus, and they will help the economy. Assume that more tax reduction is always better.
> Different groups of taxpayers are taxed according to somewhat different rules. In the U.S., for example, the tax rates depend on whether you are married or single, whether you have dependents, whether you are a dependent yourself, and whether you are retired.
> Your own taxes may also be affected differently from the effect on your group. The concern of this study is your obligation as a voter. Thus, it will not specify WHAT your group is, because we do not want you to think about the needs of particular groups.
> Assume that your group and other groups are treated fairly now.

Four of the items were as follows, with the statement that requires self-sacrifice to vote for the group's interest, and with A having the most benefits. I report data mainly from these 4 items. The other items either had equal benefits (as in the last experiment) or no restriction on self-interest, or both.

Your group is 1/3 of the population.		
	Reduction for your group	Reduction for all others
Proposal A:	2%	4%
Proposal B:	4%	2%
Group size	1/3	2/3

Notice that:

* Proposal A has more total benefits.

* Proposal B is better for your group.

[The following was included in half of the pages:]

Because of a quirk in the new law, your own taxes will not change with those of your group. They will be reduced by 1% with Proposal A and will not change at all with Proposal B. Thus, Proposal A is better for you, even though it is worse for your group.

How would you vote?

This was followed by four additional questions. Note that the answers were separate; it was possible to agree with all of them.

Indicate whether you agree or disagree with each statement *for a case like this*.

- My duty is to vote for what is in my personal self-interest.
- My duty is to vote for what is in my group's interest.
- My duty is to vote for what is best for everyone.
- Each group's duty is to vote for what is in its interest.
- Each group's duty is to vote for what is best for everyone.
- If everyone voted on the basis of self-interest, the best outcome will be achieved.
- If each group protected its own interests, the best outcome will be achieved.
- If everyone voted for what is best for everyone, the best outcome will be achieved.

The 73 subjects ranged in age from 23 to 84 (median 40); 81% were female. Six other subjects were omitted because they answered yes to every question.

8.2 Results

When self-interest was aligned with group-interest, 64% of the votes favored the proposal that was best for the group. When self-interest was opposed to group-interest, 40% favored this proposal, still a substantial number. The difference between the two indicates that some voters do think that voting is justified by self-interest, and some even think that it is their duty to vote for their own interest.

Figure 2 shows the proportions of endorsement of each item, broken down according to the answer to the first question. The important result is that group voters' (black) duty is to their group (top), but they acknowledge that voting for "all" would be best (bottom). This result was highly significant statistically. Importantly, when subjects voted for the proposal that was best for their group, they still tended to think that the best outcome would result if they voted for the proposal that was best for all ($t_{49} = 2.72$, $p = 0.0091$, tested across subjects, based on the mean for each subject of available responses).

The question about what would be best indicates that many subjects were not making any sort of simple justification of parochial voting in terms of serving the greater good. Such a justification might be possible, e.g., if everyone believed that the best results overall are achieved, as an empirical fact, when every group looks after its own interests only. But subjects who thought this would, I assume, endorse the item corresponding to this justification. In sum, some people have a concept of group duty that they themselves understand to be in conflict with the best interests of all.

9 Conclusion

Self-interest is insufficient to justify political behavior such as voting, but voting can be rational for a voter who is sufficiently altruistic and who considers the expected effects of voting on a large enough population.

Voters may fail to understand the rationality of voting in terms of altruism toward a large group. They may think that voting is justified by self-interest, that voting for their in-group serves their self-interest, or that their duty is to vote for their group's interest,
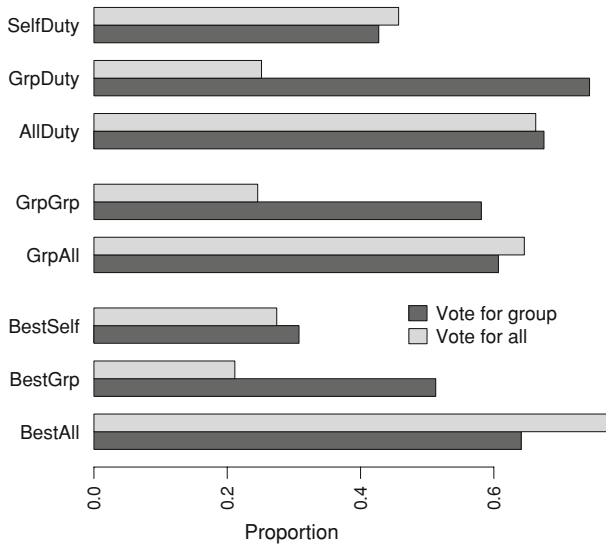
Fig. 2 Responses to questions about duty as a function of responses to the first question about choice

even when doing so reduces total good. The argument for the rationality of voting based on altruism does not justify voting in terms of self-interest or group-interest.

The theory that voting expresses moral views has not fared well. In addition to its problem explaining strategic voting (not voting for sure losers) and increased turnout in close elections, it is also contradicted by the experiment I have reported, which indicates that people tend to think of voting as more like a real decision than like an opinion poll.

If some people do not understand the argument from altruism, and do not vote expressively, why do they vote? Much of the answer seems to hinge on moral concerns other than simple altruism. Many of these concerns are moralistic. Many voters vote to advance their moral values. In many cases these values coincide with the values of altruistic voters. In other cases, it is difficult to distinguish paternalistic altruism from moralism. However, many people are willing to impose their moral values on others even when they go against the good of others even as judged by the voters themselves. These are true moralistic values.

Some of these moralistic values arise from more general non-consequentialist biases. I have argued that such non-consequentialist biases do not typically work out for the best (Baron 1998). Rather, when people make decisions in terms of values other than producing the best consequences, the decisions typically lead to consequences that are not the best. Thus, if we are concerned about consequences, we ought to worry about these biases in others.

Moralistic values may be involved in many current controversies. The examples from my experiments suggest many of them. Almost all advocates of cloning try to make a distinction between reproductive cloning and "therapeutic cloning" (or some substitute word designed to remove the stigma). Few ask what is so bad about reproductive cloning—once the bugs are out of it—so bad that it must be outlawed immediately

in ways that will be difficult to undo. Such values also infuse debates over abortion, birth control, and other issues of the sort that account for large amounts of variance in how people vote.

Many of these values are "protected" from trade-offs in people's minds. In part, such thinking may result from the difficulty of comparing vaguely defined objectives. For example, if cloning is seen as wrong because it violates human "dignity," and if we are willing to accept some reduction in dignity for the sake of some other value, such as saving a life (as when a tissue donor is cloned) or respecting the right of infertile people to reproduce, how do we compare the loss of dignity with the other gains? It is easier simply to say that dignity trumps everything else, at least until a clear counterexample comes along. Experiments on protected values suggest that people are, indeed, sensitive to such counterexamples (Baron and Leshner 2000). It may also be possible to reduce the moralistic aspect of protected values independently of their absoluteness, by asking people to put themselves in the position of others who do not accept the values in question. This is for future research.

In the long run, we might be able to de-bias moralistic values. Even now, these values are locked in a political struggle with various forms of altruism, including utilitarian altruism that considers outsiders as well as co-nationals. Many moralistic values might arise partly as errors based on false beliefs about the good of others. People may think that others are more like them than they are. People's good may really differ. If people could come to see moralistic values as possible errors, they would be more open to discussion about them, and more open to evidence about the true nature of other people's good, and about how they could use their power as citizens to advance the good of others.

In the meantime, we have one protection against the imposition of moralistic values: liberty as a general political commitment, limited necessarily by consideration of externalities (including those that result from the imposition of moralistic values themselves). When people make decisions for themselves, they are unlikely to be influenced much by the moralistic goals of others. When governments remove the power to make decisions, limiting choice, they open up the possibility that groups of people with moralistic goals can influence the law, usually in the guise of paternalism, by which people are protected from the consequences of their own decisions. To the extent that liberty is salient as a general political issue, however, those who are afraid of such intrusions will join together to defend liberty as such, even when they might agree with some particular moralistic effort. Note that the defense of liberty need not restrict government's efforts to promote the general good (Thaler and Sunstein 2008).

References

Baron, J. (1996). Norm-endorsement utilitarianism and the nature of utility. *Economics and Philosophy, 12*, 165–182.

Baron, J. (1997a). Political action vs. voluntarism in social dilemmas and aid for the needy. *Rationality and Society, 9*, 307–326.

Baron, J. (1997b). The illusion of morality as self-interest: A reason to cooperate in social dilemmas. *Psychological Science, 8*, 330–335.

Baron, J. (1997c). Biases in the quantitative measurement of values for public decisions. *Psychological Bulletin, 122*, 72–88.

Baron, J. (1998). *Judgment misguided: Intuition and error in public decision making*. New York: Oxford University Press. http://www.sas.upenn.edu/~baron/vbook.htm.

Baron, J. (2001). Confusion of group-interest and self-interest in parochial cooperation on behalf of a group. *Journal of Conflict Resolution, 45*, 283–296.

Baron, J. (2003). Value analysis of political behavior—Self-interested : Moralistic :: altruistic : Moral. *University of Pennsylvania Law Review, 151*, 1135–1167.

Baron, J. (2008). *Thinking and deciding* (4th ed.). New York: Cambridge University Press.

Baron, J. (in press). Parochialism as a result of cognitive biases. In A. K. Woods, R. Goodman, & D. Jinks (Eds.), *Understanding social action, promoting human rights*. Oxford: Oxford University Press (To appear).

Baron, J., Altman, N. Y., & Kroll, S. (2005). Parochialism and approval voting. *Journal of Conflict Resolution 49*, 895–907.

Baron, J., & Leshner, S. (2000). How serious are expressions of protected values. *Journal of Experimental Psychology: Applied, 6*, 183–194.

Baron, J., & Ritov, I. (in press). Protected values and omission bias as deontological judgments. In B. H. Ross (Series Ed.), D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.), *Psychology of learning and motivation, Vol. 50: Moral judgment and decision making.* San Diego, CA: Academic Press.

Baron, J., Ritov, I., & Greene, J. (2008, November). *The duty to support nationalistic policies*. Poster for presentation at the meeting of the Society for Judgment and Decision Making, Chicago (Also a draft of a paper).

Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes, 70*, 1–16.

Bazerman, M. H., Moore, D. A., & Gillespie, J. J. (1999). The human mind as a barrier to wiser environmental agreements. *American Behavioral Scientist, 42*, 1277–1300.

Beattie, J., & Loomes, G. (1997). The impact of incentives upon risky choice experiments. *Journal of Risk and Uncertainty, 14*, 155–168.

Brennan, G., & Hamlin, A. (2000). *Democratic devices and desires.* Cambridge: Cambridge University Press.

Brennan, G., & Lomasky, L. (1993). *Democracy and decision: The pure theory of electoral politics.* Cambridge: Cambridge University Press.

Breyer, S. (1993). *Breaking the vicious circle: Toward effective risk regulation*. Cambridge, MA: Harvard University Press.

Brodsky, D. M., & Thompson, E. (1993). Ethos, public choice, and referendum voting. *Social Science Quarterly, 74*, 286–299.

Camerer, C. F. (1987). Do biases in probability judgment matter in markets? Experimental evidence. *American Economic Review, 77*, 981–997.

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty, 19*, 7–42.

Congleton, R. D. (2007). *The moral voter hypothesis: Economic and normative aspects of public policy formation within democracies.* Manuscript, Center for Study of Public Choice, George Mason University. http://ssrn.com/abstract=1017512.

Downs, A. (1957). *An economic theory of democracy*. New York: Harper and Row.

Dworkin, R. (1977). *Taking rights seriously*. Cambridge: Harvard University Press.

Edlin, A., Gelman, A., & Kaplan, N. (2007). Voting as a rational choice: Why and how people vote to improve the well-being of others. *Rationality and Society, 19*, 293–314.

Eichenberger, R., & Oberholzer-Gee, F. (1998). Rational moralists: The role of fairness in democratic economic politics. *Public Choice, 94*, 191–210.

Goodwin, G. P., & Darley, J. M. (2008). The psychology of meta-ethics: Exploring objectivism. *Cognition, 106*, 1339–1366.

Green, D. P., & Shapiro, I. (1994). *Pathologies of rational choice theory: A critique of applications in political science.* New Haven, CT: Yale University Press.

Greene, J. D. (2002). *The terrible, horrible, no good, very bad, truth about morality and what to do about it*. Doctoral dissertation, Department of Philosophy, Princeton University. http://www.csbmb.princeton.edu/~jdgreene/NewGreene-WebPage_files/Greene-Dissertation.pdf.

Hare, R. M. (1963). *Freedom and reason*. Oxford: Oxford University Press (Clarendon Press).

Hare, R. M. (1981). *Moral thinking: Its levels, method and point*. Oxford: Oxford University Press (Clarendon Press).

Janowski, R. (2007). Altruism and the decision to vote: Explaining and testing high voter turnout. *Rationality and Society, 19*, 5–34.

Jervis, R. (1976). *Perception and misperception in international politics*. Princeton, NJ: Princeton University Press.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–291.

Kroll, S., List, J. A., & Mason, C. F. (2001). The prisoner's dilemma as a two-level game: An experimental investigation. Working paper, California State University, Sacramento.

Larrick, R. P., Morgan, J. N., & Nisbett, R. E. (1990). Teaching the use of cost-benefit reasoning in everyday life. *Psychological Science, 1*, 362–370.

Margolis, H. (1982). *Selfishness, altruism, and rationality: A theory of social choice.* New York: Cambridge University Press.

Margolis, H. (1983). The Banzhaf fallacy. *American Journal of Political Science, 27*, 321–326.

Miller, J. G., Bersoff, D. M., & Harwood, R. L. (1990). Perceptions of social responsibilities in India and in the United States: Moral imperatives or personal decisions? *Journal of Personality and Social Psychology, 58*, 33–47.

Mulligan, C. B., & Hunter, C. G. (2001). The empirical frequency of a pivotal vote. National Bureau of Economic Research, Working paper 8590. http://www.nber.org/papers/w8590.

Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P. W. (1987). Teaching reasoning. *Science, 238*, 625–631.

Popkin, S. L. (1991). *The reasoning voter: Communication and persuasion in presidential elections*. Chicago: University of Chicago Press.

Ritov, I., & Baron, J. (1999). Protected values and omission bias. *Organizational Behavior and Human Decision Processes, 79*, 79–94.

Sears, D. O., & Funk, C. L. (1991). The role of self-interest in social and political attitudes. *Advances in Experimental Social Psychology, 24*, 1–91.

Sen, A. K. (1970). The impossibility of a Paretian liberal. *Journal of Political Economy, 78*, 152–157.

Shabman, L., & Stephenson, K. (1994). A critique of the self-interested voter model: The case of a local single issue referendum. *Journal of Economic Issues, 28*, 1173–1186.

Sunstein, C. R. (2002). *Risk and reason: Safety, law, and the environment.* New York: Cambridge University Press.

Sunstein, C. R., & Thaler, R. H. (2003). Libertarian paternalism. *American Economic Review, 93*(2), 175–179.

Tetlock, P. E., Lerner, J., & Peterson, R. (1996). Revising the value pluralism model: Incorporating social content and context postulates. In C. Seligman, J. Olson, & M. Zanna (Eds.), *The psychology of values: The Ontario symposium* (Vol. 8). Hillsdale, NJ: Erlbaum.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* New Haven, CT: Yale University Press.

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. New York: Cambridge University Press.

Tyran, J.-R. (2002). Voting when money and morals conflict: An experimental test of expressive voting. Working paper 2002–07, University of St. Gallen. http://papers.ssrn.com.

Tyran, J.-R., & Sausgruber, R. (2002). A little fairness may induce a lot of redistribution in democracy. University of St. Gallen Department of Economics Working Paper Series 2002 2002-30, Department of Economics, University of St. Gallen.