

Representational Altruism: The Wary Cooperator as Authoritative Decision Maker

Kevin B. Smith University of Nebraska

What drives a policymaker to put the interests of others above their own? If human nature is inherently selfish, it makes sense to institutionalize incentives that counter decision makers' temptations to use their positions to benefit themselves over others. A growing literature rooted in evolutionary theories of human behavior, however, suggests that humans, under certain circumstances, have inherent predispositions towards "representational altruism," i.e., to make an authoritative decision to benefit another at one's own expense. Drawing on Hibbing and Alford's conception of the wary cooperator, a theoretical case is made for such behavioral expectations, which are confirmed in a series of original laboratory experiments.

A central problem of governance is establishing some means to insure that policymakers will use their decision-making power for public interests and not just self-interests. Men are not angels, as James Madison famously noted. A selfish nature combined with a position of discretionary authority is expected to produce self-interested choices, not individual sacrifice for the collective. Recognizing the individualistic streak of human nature, democracies have long sought to institutionalize a set of restraints on the temptations for selfish behavior. Regular elections promote accountability to the electorate, division of powers restrains the power of faction, and numerous rules within and between the institutions of government—veto threats, due process requirements etc.—provide incentives to limit the assumed selfish nature of man.

A growing literature, however, suggests that most people already have internal, "hardwired," restraints on selfish behavior, and that these operate independently of external, institutionalized rules. The procedural justice literature, for example, has convincingly demonstrated that people are deeply concerned about more than the outcomes of decision making. It is not just what they get from decisions, but whether they perceive process of decision making as fair that leads people to view the decisions as legitimate (e.g., Tyler 1990, 1994, 2001; Tyler, Rasinski, and Spodick 1985). If human nature is inherently selfish, why are people as concerned with how decisions are

made as much as they are with what they gain or lose from those decisions? One answer to this question is that viewing human nature as inherently selfish is an overly narrow assumption on which to base theories of political behavior. Hibbing and Alford (2004; Alford and Hibbing 2004) argue that people are wary cooperators, meaning they are inherently predisposed to be group oriented, highly sensitive to being taken advantage of, and willing to incur costs to punish others who are perceived as putting themselves above the group (see also Fehr and Gächter 2000; Guth and Tietz 1990). These characteristics are used to explain why people are so sensitive to process considerations and not just to outcomes.

The theoretical implications of the wary cooperator stretch beyond providing an answer to why people are so sensitive to nonoutcome factors of authoritative decision making. The framework rests on the argument that humans have a set of "hardwired" behavioral predispositions that are legacies of selective forces from our evolutionary past, a claim also made by a number of other recent high-profile studies of political behavior (e.g., Alford, Funk, and Hibbing 2005; Orbell et al. 2004). If, as Hibbing and Alford (2004) suggest, the behavioral predispositions of the wary cooperator are a parsimonious, universal basis for hypothesizing what preferences drive behavior in a given environmental context, this theory should be capable of explaining and predicting the choices of authoritative decision makers. Yet

Kevin B. Smith is professor of political science, University of Nebraska, Lincoln, NE 68588 (Ksmith1@unl.edu). (402) 472-0779

The author would like to thank the anonymous referees and acknowledge the invaluable input and assistance of John Hibbing and Christopher Larimer.

American Journal of Political Science, Vol. 50, No. 4, October 2006, Pp. 1013–1022

©2006, Midwest Political Science Association

ISSN 0092-5853

while Hibbing and Alford provide a convincing theoretical and empirical case for how and why the wary cooperator responds to decisions, they do not pursue their theory's implications for decision making itself. In short, while they explicitly use the wary cooperator framework to explain why people focus on the justice of procedures, the framework also implies an explanation for outcome distributions.

What happens when a wary cooperator becomes a decision maker? In this paper it is argued that when in an authoritative decision-making position, wary cooperators will have strong incentives to engage in "representational altruism," i.e., to make choices that incur individual costs in order to benefit others. This is an argument that also finds support in other evolutionary-based theories of political behavior (see esp. Orbell et al. 2004). A series of original laboratory experiments that confirm the behavioral expectations of the wary cooperator as authoritative decision makers, successfully predicting when decision makers will act selfishly, and when they will engage in representational altruism. The results suggest the theory of wary cooperation may provide a robust platform for understanding what circumstances prompt policymakers to put the benefit of others above their own self-interest, and independent of external, institutionalized incentives such as the threat of a veto.

The Theory of Wary Cooperation

The theory of wary cooperation is distinct from mainstream social science behavioral theories in that it is built from neither axiomatic first principles (such as classical rational choice) nor from the cumulative inductive understanding that comes from observing how individuals respond to environmental stimuli (such as classic behaviorism). It takes as its starting point evolutionary psychology.

Briefly, evolutionary psychology argues that the human brain evolved to solve adaptive problems, i.e., repeatedly encountered problems whose solutions had implications for survival and reproduction (for a primer see Workman and Reader 2004). The result is a flexible set of behavioral predispositions that provide context-sensitive solutions to adaptive problems. Evolutionarily speaking, these are far superior to deterministic, fixed-decision rules dictating selfish or altruistic behavior (Fehr and Gächter 2002; Rauch 2002). Political behavior in primates generally, and the evolution of human cognitive architecture in particular, are widely attributed to the evolutionary pressures created by the adaptive problems associated with group living (e.g., De Waal 1996; Massey 2002). Group liv-

ing offers numerous fitness maximizing advantages and humans, unsurprisingly, have strong, innate tendencies to be group oriented (Tooby and Cosmides 1992; Rubin 2002). These advantages, however, depend on cooperative predispositions vulnerable to the free riding of selfish individuals. Group living also creates conflict and competition among individuals (for status, access to mates, and division of scarce resources), which in turn creates complex political environments. Such environments create opportunities to maximize fitness by acting selfishly rather than putting the group first. Temptation for selfish behavior may be particularly strong for those at the upper end of a social dominance hierarchy (i.e., those more likely to be authoritative decision makers), because power and status increase opportunities for individual gain.

Group living and its opportunities for fitness maximization through contradictory behavioral temptations (to be selfish and other regarding) favors the evolution of "Machiavellian intelligence." This postulates that, "that group living selects for whatever cognitive capacities facilitate an individual's successful negotiation of the competitive and complex social environment of the group" (Orbell et al. 2004, 2). Orbell et al. (2004) suggest a likely adaptive response to group living is the evolution of these characteristics: a preference for cooperation, a modest level of mistrust, an ability to persuade others of one's own good faith, and an ability to detect lack of good faith in others. These act as a social lubricant, making cooperation easier and reducing the potential for individuals to selfishly maximize at the expense of the collective. The end result is beings that are demonstrably not rational in action, i.e., their behavior does not comport with the behavioral expectations of rational choice. Instead, humans are rational in design: their behavioral predispositions constitute an adaptive fit to problems of navigating a complex social environment (see also Cosmides and Tooby 1994).

The wary cooperator proposed by Alford and Hibbing (2004; Hibbing and Alford 2004) extends the basic argument that human behavioral predispositions are an evolved response to the pressures of group living. Hibbing and Alford reason that human behavioral predispositions are likely to be heavily oriented toward a specific feature of their environments: other human beings. Wary cooperators follow two basic rules in social interaction: cooperate with ingroup members who also cooperate and punish noncooperators.¹ Wary cooperators have an intense desire to appear as good, other-regarding members of the group, and an equally intense desire to avoid

¹A third basic rule is to not cooperate with outgroup members. As outgroups are not a particular concern to this study, this element of the wary cooperator is not pursued.

being taken advantage of by selfish group members (Alford and Hibbing 2004, 65; see also Elster 1996; Massey 2002). These reflect an adaptive response to the problems of group living, and frequently result in behavior based on emotional reaction rather than introspective calculation, and lead people to incur individual costs to benefit the collective.

This is an intrinsically social conception of behavior that claims to provide a basic theory of preferences, something that more mainstream theories of human behavior pointedly do not claim to do (Lichbach 2003, 43; Rubin 2002, 16). Wary cooperation seeks a clear separation from the rational choice paradigm by arguing individual maximization is an incomplete conception of what motivates purposive behavior, and by rejecting methodological individualism outright. For a wary cooperator the interests of others are not merely strategic concerns that have to be dealt with in order to maximize individual gain. Intentionality in decision making, the very core of goal-seeking behavior, is social rather than individualistic. Thus the underlying preferences that drive behavior are not necessarily predicated on maximizing individual gains and are often based on minimizing potential losses, both in terms of reputation within the group and in terms of being played for a sucker (the latter postulate has considerable cross-disciplinary empirical support, e.g., see Hoffman, McCabe, and Smith 1996; Kahneman and Tversky 2001).

The end result is a theory that predicts cooperation, but not necessarily altruism; and also competition, but not necessarily individual self-interest. Rather than utility maximization, what drives the behavior of wary cooperators is “sucker aversion.” The default behavioral pattern in a given situation is to take the action that appears fair to others in one’s ingroup, and be ready to punish any action on the part of others—especially those in positions of authority—that might be self-serving.

Building from these premises Hibbing and Alford (2004) hypothesize that individuals confer acceptance and legitimacy on outcomes based on their perceptions of the how and why of decision making, not just on what they gain or lose from the outcomes. They find strong empirical support for their claims. The process of decision making and the characteristics of the decision maker clearly exercise a role independent of outcomes in shaping how people perceive and respond to those outcomes. As Hibbing and Alford conclude, people react to decisions not on the basis of “what will I get?” but, “have I been screwed?” (2004, 7). When people believe they’ve been played for a sucker they cease cooperating and are willing to incur costs to punish; when people believe others are using positions of authority for self-serving ends they view the decisions as illegitimate, independent of the ultimate outcomes and who benefits from them (see also

Fehr and Gächter 2002). What Hibbing and Alford do not address is the implications of the wary cooperator framework for explaining outcomes themselves, i.e., why decision makers make the choices that they do.

The Wary Cooperator as Decision Maker

If the wary cooperator is, as claimed, a universal notion of human nature, it should be capable of shedding explanatory light on the underlying motivations of decision making and not just reactions to decisions made by others. What does a wary cooperator do when making decisions that are binding on others? If a universal conception of human nature, then we should expect the basic cognitive architecture of the wary cooperator to also apply to authoritative decision makers. Authoritative decision makers should thus be intensely sensitive to how their decisions might be perceived by others; they will want their decisions to appear fair and will be willing to incur costs to keep up those appearances. In short, the other-regarding, social aspect of decision making that drives the preference to appear fair should take precedence over any preference for individual gain. This base behavioral predisposition is internal and “hardwired” and will be resistant to any external temptations to reverse the order of these preferences.

The wary cooperator thus provides a concept of human nature distinct from other frameworks that explain the actions of authoritative decision makers on the basis of selfish individualism and incentives provided by the external environment. Most concepts of human nature deriving from the enlightenment notion of rationality—i.e., as individualistic, goal seeking, introspectively calculating, and with the task environment providing motivation for virtually all behavior—see the need for formal constraints on authoritative decision makers because human nature is inherently selfish. James Madison, for example, certainly did, arguing in *Federalist #10* that containing man’s selfish nature was the central problem of constructing a government. The solution he prescribed was a formal system of checks and balances that balanced interest against interest. The same basic idea is found in the prescriptive implications of a wide range of studies suggesting that external incentives (e.g., credible vetoes and enforcement of threats) are the key to coupling the self-interest of authoritative decision makers to the general welfare.

The wary cooperator is perfectly capable of acting selfishly, but suggests informal—actually, *inherent*—constraints are just as likely to prevent decision makers from taking advantage of their privileged positions for

selfish gain. Anthropologists have long noted that humans have a strong aversion to what they term “big man behavior”; i.e., people who act as though they believe they deserve special treatment, elevated status, and the freedom to put their individual interests above the group (Boehm 1999). This claim is generalizable from hunter gatherer groups to the most sophisticated political systems; people in positions of authority who are seen as desiring power in order to serve their self-interested ends are mistrusted, engender resentment and suspicion, and often find themselves toppled from their positions by coalitions of others or just the plain old rank and file (examples from contemporary U.S. politics might include Newt Gingrich, Richard Nixon, and Tom Delay). This is exactly what we would expect from wary cooperators when they perceive big man behavior: cease cooperating and cut the big man down to size, even if doing so incurs individual costs.

The Machiavellian mind, however, suggests that no one should be more aware of the costs of big man behavior than the big man himself (if Boehm is correct, those blind to such costs will create the very forces that remove them from power). Acting as a representational altruist, i.e., to be seen (and being seen is critically important) as incurring individual costs in order to benefit others, is obviously a powerful way to inoculate an authoritative decision maker against the big man problem. Again, this is just what we would expect from a wary cooperator: when in doubt, appear as other-regarding, even if it hurts to do so. It is important to note that Hibbing and Alford’s theory sees this is an *internal* preference; it is a predisposition “hardwired” into our biology rather than generated by environmental incentives such as institutional constraints or degree of discretionary authority. We should not expect Lord Acton’s famous aphorism that power corrupts and absolute power corrupts absolutely to fully apply to a wary cooperator. Even with no formal restraints, i.e., given complete freedom of discretionary authority with no veto or other check, a wary cooperator will be loath to act in a way that will appear selfish to members of their ingroup, even if those members are wholly subordinate.

Authoritative decision makers, in other words, will be sensitive to how others perceive their actions, and it is this internal, hardwired sensitivity that provides powerful behavioral incentives. Those in positions of power will go out of their way to convince even the weak that they are playing fair and will incur costs to satisfy this preference even if there is no formal constraint or incentive to do so. As an authoritative decision maker the wary cooperator will be less concerned with “what can I get away with?” than “what will people in my ingroup think?” There already exists empirical evidence from the laboratory of

just such behavior (e.g., van Dijk and Vermunt 2000). It is also recognized that fair versus selfish behavior is highly context sensitive, with proximity to other human beings playing a significant role in influencing decision making. Anonymity and social distance tends to lead people to make more selfish decisions, while the reverse situations make people more egalitarian (Hoffman, McCabe, and Smith 1996; Hoffman et al. 2000).

While this is all broadly accepted, incorporating the empirical findings—succinctly, that people act selfishly and people act altruistically—into mainstream theoretical frameworks has been challenging. The rational choice framework, for example, has no problem accepting that people have a preference for appearing to be fair and incorporating that preference into theoretical models of behavior (e.g., Kagel, Kim, and Moser 1996). Rational choice, however, does not explain why people have this preference or why it is so context dependent. Rational choice simply takes preferences as given, and it is accepted that, for whatever reason, people seem to have different preferences in different contexts. Wary cooperation, on the other hand, squarely addresses where these preferences come from and in doing so offers a robust platform for generating a priori behavioral hypotheses.

An example of how these hypotheses might play out is suggested in Table 1. This shows a basic ultimatum/dictator game experiment similar to those employed in behavioral economics, but with some important modifications. In this framework decision makers are asked to divide something of value between themselves and someone else, but they have varying levels of discretionary authority and “insider information” about the resource itself. In the first row the decision maker (or allocator) has unilateral power—they divide the resource and the recipient must accept that division. In the second row, the allocator’s decision is subject to a veto—the recipient can either accept or reject the proposed division, with a rejection meaning both players get nothing. Thus the rows reflect classic dictator (first row) and ultimatum (second row) games widely used in laboratory experiments on decision

TABLE 1 Framework of a Choice Experiment

	Resource More Valuable to Allocator than Recipient	Resource More Valuable to Recipient than Allocator
No Veto (Dictator Game)	A	B
Veto (Ultimatum Game)	C	D

making. The columns introduce information asymmetry into these games. In the first column, the allocator knows something about the value of the “pot” that the recipient does not; the individual units of the resource are worth more to the allocator than to the recipient. In the second column, the allocator again has inside information, but the value condition is reversed—the individual units of the resource to be divided are worth more to the recipient than the allocator (for similar asymmetry approaches in dictator and ultimatum games see Guth, Huck, and Okenfels 1996; Kagel, Kim, and Moser 1996; Roth 1995).

The research design thus creates four distinct decision-making situations (represented by the cells A, B, C, and D) with varying conditions of decision-making authority and information asymmetry. Assuming it is rational to maximize individual gain of the valued resource, one would expect a difference in outcomes between the rows. If an individual is motivated primarily by maximizing their share of the pot, what will force that decision maker be attendant to the interests of others is the external environmental incentive, i.e., the threat of a veto. Lacking this external constraint the base preference of a rational actor as traditionally conceived is to choose “more” over “less” (Buchanan and Tullock 1962, 18). To the wary cooperator, however, the degree of discretionary authority is secondary—indeed, here it would be all but irrelevant. Wary cooperators are inherently other regarding (though not necessarily altruistic). What drives a wary cooperator is how their decisions appear to others, and in this decision-making situation that perception is not affected by the degree of discretionary authority. For the wary cooperator it is the columns, not the rows, providing the all-important context for making decisions. This is because the information asymmetry plays a direct role in shaping how others are likely to view the decision of the allocator.

In the first column, where the resource is worth more to the allocator, a wary cooperator can go for an even split and end up with more than the a recipient (an even split in experimental games is universally considered a fair division; see Fehr and Fischbacher 2003). Here a wary cooperator can satisfy the contradictory predispositions for being selfish and being other regarding. They get to have their cake and eat it too—they get to appear fair, while in reality getting more than the recipient. In the second column, however, appearing fair brings a steep cost. To appear fair, the allocator has to act as a representational altruist; in effect, screwing themselves for the benefit of the recipient. An even split here means the recipient gets more than the allocator. Making such a decision would obviously demonstrate a strong preference for appearing to be fair and an intense sensitivity to how others perceive

their actions; so intense that the preference for appearing fair not only takes precedence over individual gain, but even over an actual fair split of the resource.

If such allocations are made independent of degree of discretionary authority, it would support the notion that decision making is being motivated by an internal preference rooted in a sensitivity to the perceptions of others, rather than as a response to external incentives to maximize individual gain. If an individual is strategically pursuing maximum personal gain, the rows clearly provide critically important incentives. On the other hand, if socially generated incentives drive decision making, the columns provide the key behavioral incentives. Even splits in the first column would show that people in positions of power are perfectly willing to take advantage of their positions for individual gain, as long as they can maintain a reasonable appearance of fairness. Even splits in the second column would show that people in positions of power will also benefit others at a considerable cost to themselves. Exactly such outcomes are expected if people act as wary cooperators when they become authoritative decision makers. In short, the wary cooperator would put the appearance of being fair above individual gain, and because this order is “hardwired” internally, it will be resistant to external incentives to put individual gain in front of appearing fair.

An Experimental Test

To operationalize the research design outlined in Table 1, a group of $N = 132$ undergraduates at a large Midwestern university were recruited for the experiment. The payoff promised for participation was extra credit points for the course they were taking (participation was voluntary, not participating had no impact on an individual’s grade, and at the end of the experiment all students, regardless of the divisions made in the games, received the same, maximal extra credit payoff). Subjects were asked to divide a “pot” of extra credit points between themselves and another student (the latter were laboratory fictions) in any way they saw fit. The extra credit points were physically represented in the form of 15 tokens. Allocators were given the tokens, two small envelopes, and one large envelope. The division was made by dividing the tickets into the two small envelopes, the allocator designating which envelope contained their share of the pot by printing an identification number on the relevant envelope. Both envelopes were then placed in the large envelope and returned to a designated drop site. Allocators were told the large envelopes would be collected, given randomly to a group of students who would put their identification numbers on

the unmarked envelop and return everything to a drop site. Allocators in the ultimatum game were told recipients had the option of rejecting the division. Identification numbers were used to maintain anonymity—from the allocator’s perspective, neither the other “student” nor the laboratory assistant had any visible means to ascertain whether the contents of a particular (small) envelope were from an allocator or a recipient and therefore to attach an individual to a particular decision.

The research design deliberately seeks to encourage individually maximizing behavior. This is a one-shot game, with no option for “learning” cooperation during an iterative process. Decisions are made anonymously and to the greatest extent possible, all obvious costs to maximizing individual gains were eliminated. The stakes here are “real” and highly valued by individuals (virtually all research subjects indicated they wanted extra credit points to improve their grade). The idea was to make it as easy as possible for research subjects to *not* behave as wary cooperators, especially in the first row where there is no external incentive to take into account the interests of others.

Allocators were randomly assigned to one of the four decision-making conditions. To create the information asymmetry allocators were told the tokens were worth twice as much to them as the recipient, or vice versa. Specifically they were told the tokens were worth 1 point to an allocator (recipient) and a half-point to the recipient (allocator). Random assignment allows us to attribute any differences between the groups to the treatment conditions.²

This research design is aimed at empirically testing two hypotheses:

H1: The mean division of *tokens* in of the decision making situations A, B, C, or D outlined in Table 1 will be an even split, i.e., allocators should keep half of the tokens, and offer half to the recipient (i.e., $A = B = C = D = 7.5$).

Such a finding would support the argument that individuals have a strong preference for appearing fair, even when they have to incur costs in order to do so.

H2: The mean division of *points* should vary by column rather than by row, i.e., points ultimately received by the allocator should follow this pattern: $A = C$ and $B = D$.

²Allocators completed a survey that included questions on a broad range of potentially confounding factors (everything from religiosity to partisanship to age). In a regression analysis of allocation outcomes the only significant variables were the treatment groups.

TABLE 2 Point Predictions and Observations of Mean Token Divisions

	Cell A	Cell B	Cell C	Cell D
Mean Tickets Kept	7.48	9.12*	7.58	7.91
Even Split	7.5	7.5	7.5	7.5

*= $p < .05$ (two-tailed t test).

Such a pattern would indicate that it is not the external constraint (i.e., the veto threat) that is driving the decision, but an overriding concern about the appearance of fairness to others. From a recipient’s perspective an even split of tokens in cells B and D appears fair, even though the allocator’s inside information makes it clear such a division gives the recipient the lion’s share of the extra credit points. In short, this finding would support the notion that authoritative decisions are driven by an internal behavioral predisposition that is centered on others, rather than on individualistic gain. In other words, this tests the notion of representational altruism, i.e., the idea that authoritative decision makers will in predictable circumstances incur costs to benefit others.

Initial results of this experiment are reported in Table 2. In three out of the four treatment conditions there is no significant difference between the predicted and observed allocations. Statistically, the observed pattern of divisions by cell is $A = C = D \neq B$, where A, C and $D = 7.5$.³ The exception to the fair division rule, however, is important.

This exception comes in the no-veto condition (i.e., the dictator game) where the tokens are worth twice as much to the recipient. This suggests power, i.e., the external constraint, may be playing a key role in the deviation, which suggests a concern about individual gain is overriding concerns about how the outcome appears to others. It is only when the allocator has unilateral authority to assign tokens that they deviate from appearing fair. This deviation, on average, still nets the allocator fewer points than the recipient, which is consistent with the behavior of a wary cooperator (allocators kept an average of 9.12 tokens, which results in 4.56 points for the allocator and 5.88 for the recipient). Yet the mean allocation is significantly different from the other treatments, and it raises the possibility that it is the degree of decision-making authority driving this difference.

An ANOVA examining the differences between the treatment groups, however, strongly suggests this is not

³Both individual hypothesis tests (two-tailed t, $p < .05$) examining the difference between cell means and an even split, and a joint hypothesis test examining differences between cell means ($F = 2.8$, $p = .40$) confirm this basic pattern.

TABLE 3 Two-Factor ANOVA of Ticket Divisions

	Resource More Valuable to Allocator than Recipient	Resource More Valuable to Recipient than Allocator
No Veto	M = 7.48 SD = 3.2	M = 9.12 SD = 1.6
Veto	M = 7.58 SD = 2.5	M = 7.91 SD = 2.5

Cells report mean number of tickets kept by allocator and standard deviations.

Main effect for power: $F = 1.57$ ($p = .21$).

Main effect for information: $F = 4.85$ ($p = .03$).

Interaction effect: $F = 2.12$ ($p = .15$).

$N = 132$.

the case. The results presented in Table 3 show that the degree of discretionary authority (i.e., dictator or ultimatum game) has no statistically significant impact on the mean allocation of tokens. Nor is there any evidence of an interaction effect. The variance is accounted for by information asymmetry. Thus the evidence is consistent with the conclusion that it is not the ability to act selfishly but concern about appearing to be fair that is driving allocation decisions. This inference is supported by a separate ANOVA analyzing the actual point values represented by the mean allocation of tokens (this is the formal test of Hypothesis 2). This provides clear statistical confirmation of the expected patterns on point divisions, i.e., $A = C$ and $B = D$. The results are presented in Table 4. When the tokens are worth twice as much to the allocator

TABLE 4 Two-Factor ANOVA of Point Allocations

	Resource More Valuable to Allocator than Recipient	Resource More Valuable to Recipient than Allocator
No Veto	M = 7.48 SD = 3.2	M = 4.5 SD = .80
Veto	M = 7.47 SD = 2.65	M = 3.9 SD = 1.29

Cells report mean number of points received by allocator as a result of ticket divisions and standard deviations.

Main effect for power: $F = .64$ ($p = .42$).

Main effect for information: $F = 68.76$ ($p = .00$).

Interaction effect: $F = .57$ ($p = .448$).

$N = 132$.

as the recipient (first column) there is almost no variation between the ultimatum and the dictator games. When the reverse is true, allocators essentially reverse the underlying point (as opposed to token) division of the pot—giving themselves a lower payoff than the recipient. This payoff is slightly higher in the dictator condition, but only slightly (a post hoc test shows no significant difference between 4.56 and 3.95 reported in the second column of Table 4). In short, when presented with a stark choice between self-interest or assigning to themselves what amounts to a sucker's payoff in order to be seen as "doing the right thing," subjects chose the latter. The results, in short, are consistent with the notion of representational altruism.

Overall, the findings are highly consistent with the behavioral expectations of a wary cooperator in an authoritative decision making position. In essence, decision makers are highly sensitive to the perceptions of others, and it is this sensitivity that supplies the underlying motivations for behavior—the internal constraint appears to be driving behavior, not the external constraint. This holds true even in a laboratory situation specifically designed to provide a strong incentives to behave individualistically. The results fit with the idea that authoritative decision makers, in effect, try to view outcomes from the perspective of others, and it is that assessment that drives decisions.

What the preceding experiment does not shed light on is the role of decision maker characteristics. Hibbing and Alford (2004) show reactions to decisions are determined not just by how decisions are made, but in part by assessments of *who* makes the decisions. Decisions made by those who crave power are viewed with less legitimacy than those who do not overtly seek positions of authority. Are decision makers themselves sensitive to such appearances?

Existing empirical studies would suggest yes. For example, it has been demonstrated that those who believe they come to a decision-making position on the basis of merit—regardless of how flimsy the measure of merit—are more likely to act self-interestedly than those randomly selected to such positions (Hoffman and Spitzer 1985; Hoffman et al. 2000). This does not necessarily contradict the wary cooperator framework, but it is an incomplete test of the core theoretical expectation. A wary cooperator should be sensitive to how they come to a decision-making process, but this sensitivity should cut both ways. If an individual gains position through personal effort and diligence, it may be reasonable to expect others to view that individual as deserving of the rewards. Under such conditions a wary cooperator can go for an uneven split and still expect others to see this as fair. If selected by the luck of the draw, others may view it as unseemly to overly benefit individually from a position of authority, and, recognizing

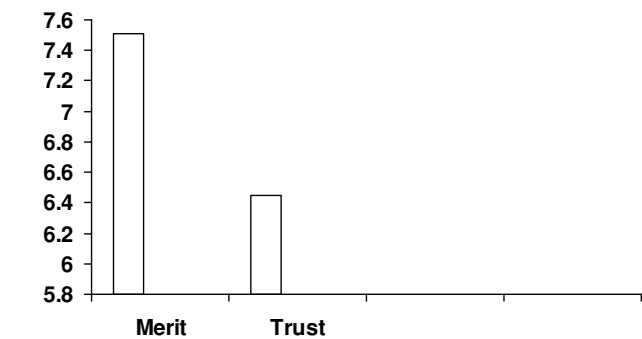
this, a wary cooperator will opt for an even split. What has not been tested is whether an individual will choose *less* if they believe they have been specifically entrusted with interests of the group. For a wary cooperator this form of representative altruism is a reasonable expectation. If someone comes to a position of power with an explicit expectation of acting in the best interests of others then doing anything less creates the appearance of not being other regarding—something a wary cooperator is inherently predisposed to avoid.

To test this possibility a second experiment using the same basic procedure as described above was conducted. However, in this experiment there were only two groups, there was no information asymmetry (the tokens were worth one point in all cases), and there was no veto. Research subjects ($n = 94$) were randomly assigned to two groups and instructed to divide the tokens however they wished.⁴ The only difference between the groups came in a single sentence in their instructions. The first group was told: “You have been selected as an allocator on the basis of merit.” The second group was told: “You have been selected as an allocator and given the trust and responsibility for dividing the tokens.” These sentences constituted the manipulation of the independent variable, i.e., one sentence served to signal that research subjects came to their position as authoritative decision makers by dint of deserving effort, the other that research subjects were explicitly entrusted with interests of others.

The result of this experiment is presented in Figure 1. As can be seen there is a clear difference between the two groups. Those who gained their positions on the basis of “merit” went for slightly above an even split. Those who came to their positions on the basis of “trust” went for an uneven split; an uneven split that gave them *fewer* points than the recipient. In short, though there was complete anonymity, no chance of reciprocity in another game, and no external constraint on discretionary authority, those in the trust category *chose less over more*. This fits with the wary cooperator argument that individual preferences are tied to a strong sensitivity to the perceptions of others and that this sensitivity leads to representational altruism. In this experiment there was no explanation given for the basis of “merit” or “trust and responsibility.” Simply creating the expectation about how subjects would be viewed as decision makers seemed to be enough to get at least one group of decision makers to use their decision-making power to put the interests of others above their own.

⁴The sample in the second experiment is completely separate from the first experiment, i.e., no research subject participated in both experiments.

FIGURE 1 Mean Token Allocations by Merit and Trust



Mean merit = 7.52
 Mean trust = 6.45
 $t = 2.146, p < .05$ (two-tailed)
 $N = 94$

Discussion

The above results suggest the wary cooperator framework can be usefully expanded from explaining why the legitimacy of outcomes is tied to perceptions of process to explaining what motivates authoritative decision makers to choose particular outcomes. Such an expansion has a number of important implications for systematically understanding why those in power make the decisions they do.

Most importantly, those in power will not be wholly driven by a preference for individual gain. In a position of authority, wary cooperators retain the key behavioral predispositions suggested by the underlying theory, predispositions that can lead towards altruistic or selfish behavior depending on context. The theory itself explains the motivating forces generated by environmental context (specifically in terms of sensitivity to the perceptions of others) and thus creates a robust platform for generating hypotheses about how people make authoritative—i.e., political—decisions.

Second, finding evidence of representational altruism suggests the conditions under which authoritative decision makers are most likely to act in the best interests of others rather than acting selfishly. Specifically, these conditions have less to do with external constraints on behavior such as veto threats or institutional limits on power and more to do with an internal constraint on any action the decision maker believes will be viewed as unfair. What stops authoritative decision makers from being greedy or selfish, at least in part, is that they do not want to be seen engaging in “big man” behavior. Wary

cooperators are perfectly willing to act individualistically, but only if they do not appear so to others. What ultimately seems to be driving the choices between maximizing individual gain, acting fairly, and acting altruistically (i.e., incurring a loss to benefit others), is the extraordinary sensitivity humans have towards other humans. The theory of wary cooperation explains why that sensitivity exists and provides a clear set of behavioral expectations in given environments.

Third, while it may be hard to institutionalize the conditions that elicit representational altruism, transparency in decision-making processes is almost certainly an effective means to limit selfish behavior by policymakers. An interesting extension of the first experiment reported above would be to repeat the experiment for conditions B and D (tokens worth half as much to the allocator as the recipient), but tell the allocator that the recipient knew of the differential values of the tokens. This would likely make representational altruism disappear—the expectation would be that allocators would make the necessary adjustment in tokens to produce a fair division in points, a division that would also be seen as fair by the recipient. What all this suggests is that what drives the wary cooperator to act selfishly is not the power to act unilaterally. It is the belief that people will not know they are being played for suckers. It is not what is actually gained that ultimately drives choices, it is how others perceive those choices. All else equal, policymakers are more likely to put the interests of others at least on an equal footing with their own if they hold an office that carries an explicit expectation of public trust and responsibility, and they also have a clear view of what others see as fair.

The theory of wary cooperation and other evolutionary-based models of political behavior such as Orbell et al.'s notion of the Machiavellian mind clearly can contribute to our understanding of the choices made by authoritative decision makers. Yet it is also important to point out some potentially important limitations to the implications sketched here. The key objective of this study was to assess whether wary cooperation could provide a robust platform for explaining the mean behavior of authoritative decision makers. Wary cooperation did this very well; yet a key problem in generalizing these findings beyond the laboratory is that actual policymakers may not reflect mean behavior. Authoritative decision makers, after all, are not randomly assigned from a representative population. Typically they differ on a number of dimensions from the groups they represent (the average member of Congress, for example, is socioeconomically differentiated from their constituents, and, moreover, has by definition a greater ambition for authoritative decision-making

power). The theory of the wary cooperator recognizes and accepts a lot of variation in cooperative tendencies, but argues there is enough universality to explain and predict mean behavior. Yet what if policymakers differ significantly from these central tendencies?

For example, take the basic characteristics of the Machiavellian mind and assume that people who actually end up in office tend to be in the tail ends of the distributions of these characteristics. Someone who is better at convincing others of their good faith (even if they are actually playing them for suckers) and better at detecting lack of faith in others, might have a good career as a political leader. It is not clear, though, that their behavior would follow the mean expectations of the wary cooperator. Representational altruism would only be expected by such individuals in extreme circumstances, if at all.

A second important issue is the role of the groups whose perception of fairness authoritative decision makers will be sensitive to. Implicit in the experiments presented here is the notion that all the subjects belong to the same group (the same class, or at least the same school). This served the purposes of the experiment because of what was being tested; i.e., the notion that preferences are generated by how we think that others in our ingroup perceive our actions. Not tested is the important question of how authoritative decision makers identify the ingroup they are sensitive to. There are innumerable potential ingroups for a policymaker: constituency, party, ideological base, even the nation as a whole. What seems fair could be very much in the eye of the beholder to different groups, which obviously complicates how the internal behavioral motivation argued for here is translated into behavior. It may even be easier, at least under some circumstances, to respond to outgroup perceptions than ingroup perceptions, a possibility (deliberately) not contemplated in this study.

Such issues, however, should be read as possibilities rather than caveats. Ultimately evolutionary-based frameworks such as the theory of wary cooperation offer the basis for a broad-ranging research agenda that can—and should—be pursued on a number of questions of fundamental interest to the discipline.

References

- Alford, John R., and John R. Hibbing. 2004. "The Origin of Politics: An Evolutionary Theory of Political Behavior." *Perspectives on Politics* 2(4):707–23.
- Alford, John R., Carolyn L. Funk, and John R. Hibbing. 2005. "Are Political Orientations Genetically Transmitted?" *American Political Science Review* 99(2):153–68.

- Boehm, Christopher. 1999. *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*. Cambridge: Harvard University Press.
- Buchanan, James, and Gordon Tullock. 1962. *The Calculus of Consent*. Ann Arbor: University of Michigan Press.
- Cosmides, Leda, and John Tooby. 1994. "Better Than Rational: Evolutionary Psychology and the Invisible Hand." *American Economic Review* 84(2):327–32.
- De Waal, Frans. 1996. *Good Natured*. Cambridge: Harvard University Press.
- Elster, Jon. 1996. "Rationality and the Emotions." *The Economic Journal* 106(438):1386–97.
- Fehr, Ernst, and Simon Gächter. 2000. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review* 90(4):980–94.
- Fehr, Ernst, and Simon Gächter. 2002. "Altruistic Punishment in Humans." *Nature* 415(10 January):137–40.
- Fehr, Ernst, and Urs Fischbacher. 2003. "The Nature of Human Altruism." *Nature* (23 October):785–91.
- Guth, Werner, and Reinhard Tietz. 1990. "Ultimatum Bargaining Behavior: A Survey and Comparison of Experimental Results." *Journal of Economic Psychology* 11(4):417–49.
- Guth, Werner, Steffen Huck, and Peter Ockenfels. 1996. "Two-Level Ultimatum Bargaining With Incomplete Information: An Experimental Study." *The Economic Journal* 106(436):593–604.
- Hibbing, John R., and John R. Alford. 2004. "Accepting Authoritative Decisions: Humans as Wary Cooperators." *American Journal of Political Science* 48(1):62–76.
- Hoffman, Elisabeth, Kevin McCabe, and Vernon Smith. 1996. "Social Distance and Other-Regarding Behavior in Dictator Games." *American Economic Review* 86(3):653–60.
- Hoffman, Elisabeth, and Matthew Spitzer. 1985. "Entitlements, Rights and Fairness: An Experimental Examination of Subjects' Concepts of Distributive Justice." *Journal of Legal Studies* 14(2):259–97.
- Hoffman, Elisabeth, Kevin McCabe, Keith Shachat, and Vernon L. Smith. 2000. "Preferences, Property Rights, and Anonymity in Bargaining Games." In *Bargaining and Market Behavior*, ed. Vernon L. Smith. Cambridge: Cambridge University Press.
- Kagel, John H., Chung Kim, and Donald Moser. 1996. "Fairness in Ultimatum Games with Asymmetric Information and Asymmetric Payoffs." *Games and Economic Behavior* 13(1):100–10.
- Kahneman, Daniel, and Amos Tversky, eds. 2001. *Choice, Values and Frames*. New York: Cambridge University Press.
- Lichbach, Mark. 2003. *Is Rational Choice Theory All of Social Science?* Ann Arbor: University of Michigan Press.
- Massey, Douglas S. 2002. "A Brief History of Human Society: The Origin and Role of Emotion in Social Life." *American Sociological Review* 67(1):1–29.
- Orbell, John, Tomonori Morikawa, Jason Hartwig, James Hanley, and Nicholas Allen. 2004. "Machiavellian Intelligence as a Basis for the Evolution of Cooperative Dispositions." *American Political Science Review* 98(1):1–16.
- Rauch, Johnathan. 2002. "Seeing Around Corners." *The Atlantic Monthly* 289(April):35–48.
- Roth, Alvin E. 1995. "Bargaining Experiments." In *The Handbook of Experimental Economics*, ed. John H. Kagel and Alvin E. Roth. Princeton: Princeton University Press.
- Rubin, Paul H. 2002. *Darwinian Politics: The Evolutionary Origins of Freedom*. New Brunswick: Rutgers University Press.
- Tooby, John, and Leda Cosmides. 1992. "The Psychological Foundations of Culture." In *The Adapted Mind*, ed. Jerome H. Barkow, Leda Cosmides, and John Tooby. New York: Oxford University Press.
- Tyler, Thomas R. 1990. *Why People Obey the Law*. New Haven: Yale University Press.
- Tyler, Thomas R. 1994. "Psychological Models of the Justice Motive: Antecedents of Distributive and Procedural Justice." *Journal of Personality and Social Psychology* 67(5):850–63.
- Tyler, Thomas R. 2001. "The Psychology of Public Dissatisfaction with Government." In *What Is it About Government that Americans Dislike?* ed. John R. Hibbing and Elizabeth Theiss-Morse. Cambridge: Cambridge University Press, pp. 227–42.
- Tyler, Tom R., K. Rasinski, and N. Spodick. 1985. "The Influence of Voice on Satisfaction with Leaders: Exploring the Meaning of Process Control." *Journal of Personality and Social Psychology* 48(1):72–91.
- van Dijk, Erik, and Riel Vermunt. 2000. "Strategy and Fairness in Social Decision Making: Sometimes it Pays to be Powerless." *Journal of Experimental Social Psychology* 36(1):1–25.
- Workman, Lance, and Will Reader. 2004. *Evolutionary Psychology: An Introduction*. New York: Cambridge University Press.