COALITIONAL BARGAINING OVER FAIRNESS

By

John Kubinski

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Psychology—Masters of Arts

2016

ABSTRACT

COALITIONAL BARGAINING OVER FAIRNESS

By

John Kubinski

Moral attitudes are influenced by a self-interest bias, yet theory suggests that people should value coordination with others in the moral domain. Thus, in deciding what moral propositions to endorse, people have to balance the pursuit of self-interest with the preservation of coordination. In bargaining over contested moral issues, the ability to recruit support from a coalition with a shared moral agenda influences bargaining strength. Accordingly, moral self-interest bias should be moderated by the strength of the coalition supporting one's interests. The current study found qualified support for this account in a group economic game where players assigned to one of two roles with competing interests vote on how to divide a reward. The current study also tested whether the self-interest bias arises due to universal self-deceptive processes and/or individually differentiated strategic tendencies and found evidence that is consistent with both of these mechanistic routes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Moral debate is a ubiquitous phenomenon and at some point in their lives nearly all people will take an interest in the following question: what rules ought to govern how we live? The answers people give to this question, and the subsequent moral disagreements that occur, may be influenced by many factors, but among them self-interest is particularly important. Moral debates routinely center on zero-sum social conflicts, with different interest groups preferring different moral rules to resolve particular conflicts. For example, political arguments about taxation and welfare policies pit the interests of wealthy people against those of poor people, and as expected, factors related to the present and future ability to obtain wealth—income and education—are positively associated with economic conservatism (Johnson & Tamney, 2001). However, given that people's morality is not identical to their self-interest, the question arises as to what psychological processes adjust morality in a self-interested direction and what factors moderate such processes. The current research focuses on the social context of moral advocacy and tests the theory that the influence of self-interest on an individual's moral agenda is constrained by the ability to obtain consensus among the people to whom that agenda would apply (Tooby & Cosmides, 2010). Specifically, drawing on a coalitional bargaining perspective, the current study examines whether individuals calibrate their moral attitudes based on the strength of their interest-based coalition.

The Effect of Self-Interest

The effect of self-interest can be seen in the associations between real world moral attitudes and relevant demographic characteristics. Egalitarian political attitudes are related to race, gender, social class and income in ways predicted by the self-interest thesis, such that advantaged individuals are less likely to endorse egalitarianism (Ritzman & Tomaskovic-Devey,

1992). Among men, factors that predict the ability to benefit from social inequality such as muscularity and attractiveness are negatively related to egalitarian attitudes (Price, Brown, Dukes, & Kang, 2015; Price, Kang, Dunn, & Hopkins, 2011). Experimentally manipulating subjective perceptions of one's socioeconomic status by inducing social comparison processes causes self-interested changes in one's attitudes towards economic redistribution, as well as changes in one's support for the ideologies that justify such policies (Brown-Iannuzzi, Lundberg, Kay, & Payne, 2015). While income tends to predict economic policy preferences, attitudes about policy issues that restrict or enhance opportunity on the basis of group identity (e.g. immigration and government assistance for minorities) are predicted by membership in the relevant dominant or subordinate groups as well as the ability to be socially competitive under meritocracy, such that less educated members of dominant groups are most opposed to promoting opportunity for subordinate groups (Weeden & Kurzban, 2014).

Self-interest effects are also found in laboratory studies of moral judgment. For example, one study found that fairness judgments in resource division games became more self-interested when individuals played games for real rather than hypothetically, and that the amount of change in what range of behavior one considered fair was correlated with the amount of change between one's hypothetical and real behavior (Rustichini & Villeval, 2014). Clearly, self-interested morality is aimed at obtaining and justifying one's preferred outcomes in the social world. But when is it adaptive to propose self-serving moral rules and what are the ecological constraints on this process? When should individuals use morality to get their way, and how much should individuals demand?

Evolutionary psychology attempts to place self-interest in social context and understand the logic underlying the strategies people use to realize their agendas. While it would obviously

be fitness-enhancing to compel others to act in ways that are perfectly consistent with our wishes, attempts at such absolute social influence are unlikely to be successful. Instead, people pursue partial fulfillment of their social goals through negotiations and concessions. Rejection behavior in the ultimatum game, which involves turning down money that is unfairly divided, shows that potential social partners will refuse interactions when the proposed relationship terms are insufficiently considerate of their interests (Pillutla & Murnighan, 1996). In the absence of massive power disparities, individuals typically can only access the benefits of social exchange through mutual, consensual coordination. However, within the space of possible negotiation solutions, there is room for social conflict about the exact balance of interests expressed in the terms of the relationship. Interaction partners will frequently have asymmetric bargaining positions on account of differences in the ability to inflict costs on each other or withhold benefits from each other (Sell, 2011). Theories grounded in the notion of bargaining, such as the recalibrational theory of anger, have led to successful predictions about positive correlations between determinants of social formidability (e.g. physical strength for men and physical attractiveness for women) and the tendency to demand more favorable treatment from others (Sell, Tooby, & Cosmides, 2009). In support of the idea that bargaining may influence contested moral issues, male upper-body strength predicts opposition to economic egalitarianism among wealthy men but support for economic egalitarianism among poor men, suggesting that moral self-interest bias may be moderated by formidability-related factors that influence self-perceived bargaining power (Petersen, Sznycer, Sell, Cosmides, & Tooby, 2013). In resource division games, self-interested discrepancies between actual and hypothetical behavior are more likely when the game puts participants in a position of power (Rustichini & Villeval, 2014). Aside from personal characteristics, the ability to marshal social support is an important component of

human bargaining. For example, in an economic game where two individuals can only succeed by coordinating on an unequal division of benefits, the physical presence of one player's in-group increases the likelihood of that player obtaining his or her preferred solution (Charness, Rigotti, & Rustichini, 2007). To the extent that morality involves bargaining, cues of social support for one's agenda may influence moral perception, not unlike how the presence of allies influences threat perception (Fessler & Holbrook, 2013).

Moral Coalitions

Morality, particularly in its political manifestations, can be usefully viewed through the lens of coalitional bargaining. Because coalitional behavior requires coordination, common knowledge of a shared interest is likely to be a critical first step in the formation of moral coalitions (Thomas, DeScioli, Haque, & Pinker, 2014). When individuals become mutually aware of a shared interest, they may moralize that interest so as to influence social institutions in ways consistent with their shared agenda (Tooby & Cosmides, 2010). The problem of dominant alpha males is a probable ancient context where selection may have favored the ability to form moral coalitions, and the solution to the dominance problem – the moral condemnation of power-seeking – is commonly found in traditional hunter-gatherer cultures (Boehm, 1999). While the dominant alpha male case involves a conflict of interest between one and many individuals, moral controversies typically involve conflicts of many versus many. Opposing moral camps, drawn around opposing sets of interests, emerge and compete for influence over the rules, norms and institutions that govern social life. And ultimately, whichever moral coalition can obtain a larger and more powerful constituency is most likely to establish its preferred rules in society. Indeed, among non-human primates, coalition size is an important factor in deciding conflict outcomes within and between groups (Silk, 2007; Wrangham, 1999). However, human conflict,

4

and moral conflict in particular, is distinguished from animal conflict by an enhanced ability and motivation to resolve disputes in a manner that preserves social coordination.

Morality as an Intuitive Coordination Device

There are general coordination benefits to having moral rules. Consensus about right and wrong facilitates smooth social transactions and prevents disputes, which may involve entangling alliances, from spiraling out of control (DeScioli & Kurzban, 2013). Under the theory proposed by DeScioli & Kurzban (2013), morality is a coordination device that communities use to solve the coordination problem of choosing whether to condemn certain actions. Although moral norms may arise through a social process, they are intuitively perceived as objective (Haidt, Koller, & Dias, 1993). The intuitive ontology underlying morality involves the concept of permissibility. An act that violates a moral norm is perceived to be morally impermissible and judged as wrong. Wrongness judgments are important because they determine whose actions can be condemned. When an actor has done something morally wrong, that actor becomes, in the eyes of the community, a legitimate target of condemnation and potential punishment. The community's consensus about wrongness enables coordinated condemnation of the actor, which is critical to stabilizing the enforcement of the moral norm. Unlike conventions, which are recognized to be arbitrary and culture-bound, morality is intuitively perceived as universal, such that the wrongness of actions that violate established moral norms is non-negotiable (Goodwin & Darley, 2008; Haidt et al., 1993; Smetana & Braeges, 1990). Somewhat paradoxically, however, moral norms can of course change. On the coalitional account, moral norms are updated as support for new moral agendas coalesces within communities. Specifically, bargaining processes are posited to underlie this updating, and when moral norms change, individuals should adjust their personal moral beliefs so as to maintain coordination with their community. While a

5

psychology of convention could in principle solve the problem of coordinating condemnation, DeScioli & Kurzban (2013) discuss why a moral psychology is a more efficient solution. In particular, the perceived objectivity of moral norms increases coordination efficiency by aiding in the creation of common knowledge about who is a legitimate target of condemnation. The non-negotiable nature of moral judgments also facilitates coordination by ensuring that moral considerations override interpersonal favoritism in the condemnation process. If individuals sometimes chose not to condemn friends and family who violated moral norms, then coordination would be undermined because the attribution of condemnation status would be unreliable. To summarize, morality is a psychological system that prepares coordinated condemnation of norm violators and outputs wrongness judgments that are perceived as objective. For an act to fall under the moral domain, both the potential for that act to be met with social condemnation and the perceived objectivity of that act's wrongness are necessary conditions. Under the coalitional theory, moral activism is defined as an attempt to establish a new norm—i.e. propose that an act be met with condemnation and perceived as objectively wrong. The fate of an introduced norm should be decided by the balance of power between coalitions who support and oppose the norm.

Bargaining in the Moral Domain

Because morality provides important coordination benefits, coalitional bargaining efforts in the moral domain are predicted to be cautious, but calibrated to the strength of the coalition. A coalition that is excessively demanding in its moral agenda may provoke withdrawal, protest or direct conflict from its opponents—that is, uncompromisingly selfish morality threatens to disrupt coordination. It follows that the key challenge for a moral advocate is identifying the range of candidate moral norms (i.e. coordination solutions) that are within the envelope of

viable equilibria and then rallying a coalition around the sub-region of this envelope that is most consistent with one's self-interest. The problem of adopting a morality within the constraints of what other people are likely to accept is analogous to the problem of adopting a mode of self-presentation that other people are likely to believe. The tendency to bias self-presentation in a positive direction must be tempered by plausibility; successful self-enhancement exploits ambiguities in the public evidence about oneself, but does not generate completely unbelievable claims (Schlenker, Britt, & Pennington, 1996). Similarly, the self-interest bias in morality should be constrained by which moral attitudes are socially plausible—that is, acceptable to a critical fraction of one's moral community. This may explain why, within a certain range, people's fairness judgments change to justify their selfish behavior, but beyond that range—when the behavior is too clearly deviant from the fairness norm—people do not judge their selfish behavior to be fair (Rustichini & Villeval, 2014). In essence, adaptive moral advocacy should balance the need to maintain the general benefits of coordination against the goal of securing the specific, zero-sum benefits that the particular content of moral rules can create for oneself or one's interest group.

A coalitional perspective suggests that moral attitudes have an important action preparation function. When alternative institutions are socially plausible, internal moral attitudes prepare people to advocate for institutions that are in their self-interest, not only through lone activism but by creating or lending support to a like-minded moral coalition. Thus, the content of moral attitudes should be regulated by factors that make attitudes more or less useful for participating in a moral coalition. The self-interest bias becomes non-adaptive when it leads to moral beliefs that are so partial and extreme that they substantially reduce the probability of finding an adequately supportive coalition. To choose an adaptively self-interested moral

attitude, people need to cognitively monitor the coalitional potential of different moral positions, which shifts across time and place. If people make coalitional considerations and adjust their attitudes accordingly, then the relationship between the private self-interest of individuals and the publicly stated morality of individuals should be moderated by the degree to which individuals believe that moral rules consistent with their private agenda are capable of attracting popular support in the relevant social milieu.

Mechanisms: The Self-Deception and Strategizing Hypotheses

Separate from the question of what factors moderate self-interested moral behavior is the question of what mechanisms produce it. In general, evolutionary theory suggests that moral cognition is adaptive insofar as it causes strategically correct behavior in social situations (Trivers, 1971). In the context of coalitional bargaining, moral attitudes are adaptive if they cause individuals to take self-interested moral positions that one's coalition can successfully bargain for. An evolved computational system that fulfilled this bargaining function would have to compute and integrate several regulatory variables: what is in one's self-interest; what is in the interests of others; how are different sets of interests affected by different moral rules; to what extent are different moral rules attracting coordinated advocacy; what is the relative formidability of competing moral coalitions, and so on (John Tooby, Cosmides, Sell, Lieberman, & Sznycer, 2008). The final output of such a computational system would be a moral attitude that causes optimal coalitional bargaining behavior. On the computational account, adjustment of the self-interest bias could occur automatically and unconsciously, with various social cues supplying the values of the relevant regulatory variables. In fact, there is reason to hypothesize that the self-interest bias is designed to function outside of conscious awareness. Individuals expend effort to cultivate the impression of impartiality, which can be compromised if others

8

perceive an individual's moral decisions to be biased by factors unrelated to justice (Shaw, 2013). To the extent that one's moral credibility is increased by the appearance of being unbiased, bias should operate in a self-deceptive manner—that is, people who are influenced by coalition strength should consciously believe they are not influenced. Self-deception is thought to increase credibility because speakers will have more cognitive fluency when they hold only one version of events in working memory rather than two contradictory versions (von Hippel & Trivers, 2011). Hereafter, the model of self-interest bias as arising from an unconscious computational system will be referred to as the self-deception hypothesis.

Alternatively, strategic adjustments to one's moral attitudes or behavior may occur through other mechanisms that operate with varied levels of automaticity and awareness. Moral adjustment could be driven by specific motives, such as the desire to maximize one's personal or relative gains. Individuals high on such motives (who are called 'proselfs') are more willing to maximize their payoffs at the expense of others and are more likely to navigate cooperation contexts in ways that track the situation's explicit incentives (Boone, Declerck, & Kiyonari, 2010). In the coalitional context, awareness and comprehension of the social and political environment may be cognitively important for producing moral adjustment (Ferris, 2005). For example, the personality trait of Machiavellianism, which is related to social astuteness, might predict the successful use of coalitional information (Dahling, Whitaker, & Levy, 2008). High Machs also report more willingness to engage in opportunistic economic behavior and are capable of strategically adjusting their moral self-presentation on the basis of situational factors, traits that might bolster the self-interest bias in general (Bereczkei, Birkas, & Kerekes, 2010; Sakalaki, Richardson, & Thépaut, 2007). Finally, explicit, controlled reasoning processes could play a general role in producing strategically optimal behavior. Experimentally promoting

reflection and discouraging intuition leads to less cooperation in one-shot group interactions as well as increased utilitarian judgments of moral dilemmas (Paxton, Ungar, & Greene; Rand, Greene, & Nowak, 2012b). Although there is some heterogeneity among the mechanistic accounts discussed thus far as alternatives to the self-deception hypothesis, for the purpose of this paper they will collectively be referred to as the strategizing hypothesis.

The self-deception and strategizing hypotheses are not mutually exclusive, as the mechanisms proposed by each may operate in parallel. However, testing the predictions of each hypothesis may help shed light on how moral adjustment operates. The self-deception hypothesis predicts that coalition strength will affect the self-interest bias even among individuals who are low on measures of strategizing. The strategizing hypothesis predicts that individual differences in strategizing will moderate the impact of coalition strength. Although these predictions have been articulated in the context of explaining coalitional effects on the self-interest bias, these hypotheses are applicable to the self-interest bias *per se* regardless of its moderators. Thus, tests of the coalitional theory and the mechanistic hypotheses, while potentially mutually informative, are also somewhat independent. In general, the self-deception hypothesis predicts that the self-interest bias will manifest even among individuals low on measures of strategizing, while the strategizing hypothesis makes the non-competing prediction that measures of strategizing will positively predict the degree of self-interest bias. In the results below, supplementary analyses of whether the self-interest bias is moderated by strategizing are provided.

Equality vs. Equity

The current study aims to test the coalitional theory of morality as well as the self-deception and strategizing hypotheses using the issue of fairness in a group economic game. Because there are competing conceptions of fairness, it is an ideal issue for testing coalitional

influences on moral attitudes. On the equality conception of fairness, resources should be distributed equally regardless of potential differences in deservingness, such as different levels of productive work. On the equity conception of fairness, resources should be distributed such that people get what they deserve based on their productive inputs. Self-interest may lead individuals to prefer the equality or the equity conception. Experimental research on attitudes about fairness in the workplace indicates that people rate performance-based pay schemes as more fair than equal pay schemes when they are asked to imagine that they perform better than their colleagues, but not when asked to imagine the reverse situation (van Yperen, van den Bos, & de Graaff, 2005). Similarly, a study using false feedback about performance in an investment economic game found that, compared to individuals told they performed exceptionally worse than others, individuals told they performed exceptionally better than others reported less support for a rule to make player payoffs more equal and even rated the American economic system itself as more just (Brown-Iannuzzi et al., 2015).

Most pertinent to the current study, DeScioli, Massenkoff, Shaw, Petersen, & Kurzban (2014) demonstrated that morality and fairness ratings of resource divisions in a dyadic economic game, hereafter referred to as the Checker Typist Game (CTG), can be biased by short-term assignment to different roles (Checker or Typist) that differentially stand to benefit from the principles of equality and equity. Specifically, in a joint task where one role performs more work than another, individuals assigned to do more work (Typists) judged an equitable division of a task-related monetary reward as more fair than did individuals assigned to do less work (Checkers), who judged an equal division to be more fair. Importantly, assigning the roles to do equal work, thereby removing the Typist's plausible justification for a self-serving division (i.e. eliminating the self-interested option's coordination potential), dramatically reduced the

11

self-interest bias in fairness judgments as well as actual decisions on how to divide the reward, with only 22% of Typists choosing the self-favoring division as opposed to the 70% or more Typists who chose the self-favoring division when the unequal work justification was available. If coordination viability is a boundary condition for the self-interest bias, then self-interest effects should be positively moderated by the introduction of a coalition that shares one's self-interested agenda.

Study Overview

The current study extends the CTG paradigm of DeScioli et al. (2014) to test whether the degree to which an individual's self-interest influences moral attitudes and decisions in a resource division context depends on the presence of others who are willing and able to support that individual's agenda. In DeScioli et al. (2014), the CTG consisted of two players working together to complete a text transcription task. Players assigned to the Checker role would perform less transcription work than players assigned to the Typist role, who had the power to decide how to divide a monetary bonus for completing the task. In the current study, the CTG is played in a group, with some players assigned to the Typist role and others to the Checker role. Rather than having a single Typist determine the bonus division, the division is determined through a voting process that involves all players. Because players have the option to reject unfavorable divisions, the current study resembles an ultimatum game whereas the DeScioli et al. (2014) study resembles a dictator game (Forsythe, Horowitz, Savin, & Sefton, 1994).

The self-interest of players is defined by their roles. Typists transcribe three paragraphs of text to complete the task, while Checkers transcribe only one of these paragraphs for the purpose of allowing Typists to gauge their accuracy by checking their work against an independent transcription. The CTG is played for multiple rounds, with groups offered a

collective bonus for every round they complete. Since Typists perform more work, it is in their interest to divide the bonus according to the principle of equity, while it is in the interest of Checkers to favor the principle of equality. Players are able to influence the bonus division through a voting process that occurs prior to each round. One player proposes a division rule (with options ranging from Typists get 100% to Typists get 0%; see methods below for more details), and then all players in the group vote on whether to accept or reject the rule. If less than half the players accept the proposed division rule, then the round is skipped and the bonus for that round is permanently lost.

Applying a coalitional theory of morality suggests that the degree of self-interest bias observed in proposed division rules will be proportional to the size of one's role-based coalition. To test this account, the current study varied the number of Typists and Checkers in the group, thereby varying the relative bargaining strength of each coalition. As one's coalition grows in size, it becomes easier to secure the votes necessary to pass a self-interested division rule or reject a self-harming division rule. The ability of a coalition to protest the proposed rule by skipping the round may be thought of as analogous to the ability of workers to go on strike when offered a pay grade they feel to be inadequate. Essentially, voting to reject is strategically similar to withholding social exchange benefits in a bargaining context. Thus, as a coalition gains more votes, its ability to withhold benefits grows and it is in a position to propose more self-interested offers. In addition to testing this behavioral prediction, the current study tested whether individuals calibrate their moral attitudes to the strength of their coalition. If moral attitudes prepare individuals to participate in coalitions, then ratings of the morality and fairness of equal and equitable divisions should track what outcome an individual's coalition can feasibly obtain by bargaining, which varies depending on coalitional strength.

13

METHODS

Subjects ($N = 674$) were recruited from the psychology subject pool at Michigan State University to participate in a group session with up to six people (69% female; mean age = 19.34, $SD = 1.56$). Players were seated at individual computers, and the CTG was implemented using zTree software (Fischbacher, 2007). At the beginning of the study, the experimenter gave a brief overview of the game and then players completed a tutorial at their computers. Computers were separated by dividers to prevent participant interaction. The tutorial truthfully informed participants that they would be completing a transcription task each round, that they would be permanently assigned to the role of Typist or Checker, that Typists transcribe three paragraphs while Checkers only transcribe one paragraph, that submitted work would be shared within the Typist-Checker dyad, that the entire group could earn a collective bonus of 60 points each round, that the bonus points would be redeemed for raffle tickets for a $50 Visa gift card at the end of the study, that a single randomly selected player would propose how to divide the bonus between the roles before each round, that the group would vote on the proposed bonus division, that divisions would pass if the votes tied but fail if a majority voted to reject the proposal, and that the round would be skipped and the bonus points for the round would be permanently lost if the proposal was rejected.

The tutorial also primed participants to think about fairness by describing the bonus division process as giving players the chance to divide the bonus in a way they think is fair. The tutorial mentioned that players may prefer to divide the bonus equally or based on the amount of work performed by each role. Following a practice round, players completed up to five rounds of the CTG, answered a survey, and, based on the points they earned in the game, received raffle tickets for the $50 Visa gift card prize. The number of rounds was always limited by the number

of participants in the session because each player was only allowed to propose a maximum of one time. The number of rounds was also limited by session time during some phases of data collection; in time-limited sessions, the number of rounds was set at a maximum of three.

Game Structure

Before the game starts, the computer randomly and permanently assigns participants to the role of Typist or Checker. At the beginning of each round, players vote on a bonus division rule, and if the rule passes, then players perform their assigned work for that round. Specifically, an image of text appears on each player's screen, and players are instructed to type out the text in the image as accurately as possible. Three paragraphs are shown to a Typist, while only one paragraph is shown to a Checker. When participants have finished typing the transcription, their work is recorded in a text box, which is then made visible to their partnered Checker or Typist. Players may compare their work to that of their partners, but there is no material opportunity for revisions. The fact that Typists and Checkers both exchange their work creates social responsibility for fulfilling the duties of one's role. In some instances, a Typist may be paired with multiple Checkers. Aside from exchanging work, proposing a division, and voting, players never communicate with each other.

Stimulus Materials

Text images were obtained from archive.org. Passages presented were authored by Martin Heidegger, Seymour Martin Lipset, Hugh Didascalicon, Mario Pei and Samuel Briggs.

Voting Procedure

At the beginning of each round, the computer randomly selects one participant to propose a division rule for the bonus in order to determine how the 60 point bonus will be allocated between roles. Each player acts as the proposer no more than one time. While making their

decision, proposers are shown the number of Typists and the number of Checkers in the group. The proposer enters a percentage between 0% and 100% that corresponds to the share an individual Typist will receive as compared to an individual Checker. Specifically, proposers are asked, "Imagine there was one Typist and one Checker dividing the bonus. Enter the proposed percent of the bonus that will go to the Typist." Essentially, proposers are deciding on the wage per worker for each role. To help players understand the division, proposers are shown a table that indicates how much each Typist and each Checker would get at divisions ranging from 10 to 100 percent, separated by 10 percent intervals. For example, in a group of six with three Typists and three Checkers, the table would show that a proposal of 70 percent means that each Typist gets seven points and each Checker gets three points. After a division rule has been proposed, it is made visible to all players, who then vote on whether to accept or reject the proposal. The role of the proposer is not revealed, and voting behavior is similarly anonymous. Players are notified of whether the vote passed, but they are not shown how members of each role voted. If the vote passes, then players proceed to complete their work for the round. At the end of the session, participants receive one raffle ticket for every 10 points earned in the game.

Manipulation

The numerical composition of roles within groups was manipulated between sessions. Within a session, the number of Typists and Checkers was permanently fixed at the start of the game. Because of variable attendance, overall group size varied across sessions and thus the exact numerical conditions depended on group size. In a group of 6, the number of Typists ranged between 2 and 5. In a group of 5, the number of Typists ranged between 2 and 4. In a group of 4, there were either 2 or 3 Typists. In a group of 3, there was either 1 or 2 Typists. In a group of 2, there was one Typist. Changing the relative size of the Typist and Checker interest

groups was intended as a manipulation of the bargaining power each role has in the voting process. Because the number of Typists and Checkers is shown during the proposal process, proposers have the chance to incorporate coalitional information into their decision.

Survey

At the end of the session, participants completed a survey which began with comprehension check questions and four items about the morality and fairness of equal as opposed to equitable bonus divisions. For example, participants were asked, "In your opinion, how fair is it to divide the bonus equally for each role, 50% for a Typist and 50% for a Checker?" Participants also judged the morality and fairness of dividing the bonus according to roles based on the amount of work (i.e. equitably), 75% for a Typist and 25% for a Checker. Fairness responses were recorded on a 7-point scale ranging from "Very unfair" to "Very fair" with a midpoint labeled "Neutral." Morality items asked participants to judge how morally justified equal and equitable divisions were, with endpoints labeled "Very Morally Wrong" and "Very Morally Justified." The survey also contained items about basic demographics as well as potential moderators and relevant individual differences. An item measuring the explicit influence of coalition strength asked, "When you proposed the bonus division in the task, was your decision influenced by the number of votes your role had as compared to the other role?" (1 = Not at all influenced; 5 = Very much influenced). As shown in subsequent analyses, the explicit influence measure seemed to index a general tendency to behave strategically, potentially similar to the proself construct that tracks one's sensitivity to extrinsic incentives (Boone et al., 2010). Thus, explicit influence proved useful in tests of the strategizing hypothesis. Participants also completed a 16 item Machiavellian Personality Scale (Dahling et al., 2008), with item agreement measured on a 5-point sale ($\alpha = .85$). Of interest to the current study, the

Machiavellianism Personality Scale contains items measuring amorality (e.g. "I am willing to be unethical if I believe it will help me succeed") and distrust of others (e.g. "Team members backstab each other all the time to get ahead"). Lastly, another item measured political beliefs (1 = Strongly liberal; 7 = Strongly conservative).

Analysis

Participants were excluded from both proposer and moral judgment analyses if they answered a comprehension question about the voting system incorrectly ($n = 81$) and were excluded from the moral judgment analysis if they indicated that their survey responses were bogus ($n = 7$). For the survey data, this left a final sample of $N = 549$. Due to time constraints, some participants did not receive a turn to propose in the game ($n = 94$), leaving a final sample of $N = 493$ for the proposer behavioral data. In exploratory analyses reported below, the group of participants who did not propose was exploited to see whether the moral self-interest bias differed as a function of proposing.

Coalition strength was the key predictor in confirmatory analyses. Coalition strength was computed to be the percent of group participants belonging to one's own role, such that Typists and Checkers receive a coalition strength score on a single common scale. A coalition strength score of 50 indicates that one's role represented half the group. The sample sizes per coalition strength cell, along with other sample size information, are displayed in Table 1.

Confirmatory analyses focus on two constructs, proposal self-interest and moral self-interest. Since all participants made proposals regarding the Typist's percent, proposal self-interest was computed by re-coding the proposal variable to obtain the percent allocated to one's own role. Thus, the proposal self-interest variable allowed proposal data from each role to be put on a common scale, where a score of 50 indicates an equal division between roles and higher

numbers represent a more self-interested proposal. Moral self-interest was computed from the moral judgment survey data in a conceptually similar way.

A principal axis factor analysis of the four moral judgment items yielded only a single factor with an eigenvalue greater than 1 (first three eigenvalues = 2.61, .86, .30), confirming the unidimensionality of the moral items. This single factor accounted for 65.22% of the total variance, and factor loadings were uniformly high (> .7). Reverse-coding the equality items allowed for the computation of the internal reliability of the four measures (α = .82). However, the analysis employed the difference score between equity endorsement and equality endorsement as the moral judgment dependent variable, rather than a simple composite of the four items.

For all individuals, endorsement of equity and endorsement of equality were computed by averaging the two items measuring the morality and fairness of each division type. Next, the degree to which equity was endorsed over equality (relative equity endorsement, hereafter) was computed by subtracting the equality endorsement score from the equity endorsement score. Finally, moral self-interest was computed by re-coding relative equity endorsement to obtain the relative endorsement of the fairness conception that is consistent with the interests of one's role (i.e. equity for Typists and equality for Checkers). A score of 0 on moral self-interest indicates indifference towards equality versus equity, and higher numbers represent a more self-interested judgment. Moral self-interest had a potential range of -6 to 6. While the predictions of the coalitional theory were tested using proposal and moral self-interest as outcome variables, some trends reported below are clarified by examining associations with raw proposal or relative equity endorsement variables as a function of role.

All analyses were conducted using a linear model with a random intercept to account for the fact that data is nested within experimental sessions. All continuous predictors were mean-centered prior to analysis.

RESULTS

Descriptive statistics and correlations for study variables are presented in Table 2. The

average proposal (i.e. the percent offered to the Typist) was 56.51 (SD = 14.69). The three most

common proposals were 50 (38.7% of proposals), 60 (29.8% of proposals) and 70 (13.8% of

proposals). Proposals less than 50 only occurred 7.7% of the time. Turning to voting behavior,

93.5% of proposals were accepted. In general, higher proposals were more likely to be accepted

($r = .30$, $p < .001$), while the degree to which a proposal was self-interested had no relation to its

likelihood of acceptance ($r = -.07$, $p = .14$). At the individual level, 81.3% of all votes were in

favor of the proposal.

Proposal self-interest increased in later periods of play ($r = .14$, $p < .01$), a trend which

will be discussed more below. Additionally, as an unintended byproduct of the game structure,

period of play was confounded with coalition strength ($r = .22$, $p < .01$). For models of proposal

behavior, both period and coalition strength were always included as covariates. Variation in

overall group size ($M = 4.94$, $SD = .86$) did not affect any of the results reported below and is

therefore not included in any of the models presented. Finally, unless noted otherwise, the

reported effects were not moderated by role. In confirmatory models, only significant

theoretically predicted interactions are shown. Non-significant predicted interactions are

mentioned in the text, but not shown in tables.

Confirmatory Analyses – Testing the Coalitional Theory of Morality

Were proposals biased by self-interest? Yes. There was an effect of role such that Typists

made higher proposals ($b = 5.78$, $\beta = .39$, $p < .001$; Table 7).

Was the self-interest bias in proposal behavior predicted by coalition strength? No. Contrary to prediction, coalition strength had no main effect on proposal self-interest ($b = .04$, $\beta = .05$, $p = .35$; Table 3).

Were morality and fairness judgments biased by self-interest? Yes. There was an effect of role such that Typists had higher levels of relative equity endorsement ($b = 1.46$, $\beta = .54$, $p < .001$; Table 8, Model 1).

Was the self-interest bias in moral judgment predicted by coalition strength? No. Contrary to prediction, coalition strength had no effect on moral self-interest ($b = -.01$, $\beta = -.04$, $p = .45$; Table 4).

Confirmatory Analyses – Testing the Self-Deception and Strategizing Hypotheses

Did Machiavellianism, by itself or interacting with coalition strength, predict the self-interest bias? Yes. As shown in Table 3, there was no main effect of Mach upon proposal self-interest ($b = 1.25$, $\beta = .05$, $p = .25$). However, Mach did interact with coalition strength ($b = .12$, $\beta = .08$, $p = .045$). Simple slopes analysis indicated that, among high Machs (1 *SD* above the mean), coalition strength positively predicted proposal self-interest ($b = .11$, $\beta = .13$, $p = .045$). In contrast, among low Machs (1 *SD* below the mean), coalition strength was unrelated to proposal self-interest ($p = .57$). The observed pattern, shown in Figure 1, suggests that coalition strength only has an effect among high Machs, a finding that is inconsistent with the self-deception hypothesis and consistent with the strategizing hypothesis.

The interaction between Mach and coalition strength did not apply to moral self-interest ($p = .93$). However, Mach did interact with role to predict moral self-interest ($b = 1.25$, $\beta = .28$, $p = .002$; Table 4). Simple slopes indicated that Mach was positively associated with moral self-interest among Typists ($b = .74$, $\beta = .17$, $p = .004$), but not among Checkers ($p = .12$). The

significant effect of Mach among Typists may provide support for the strategizing hypothesis, but a substantive interpretation of Mach's interaction with role may be unwarranted. The observed interaction may be an artifact resulting from the fact that high Machs have higher levels of relative equity endorsement regardless of role ($b = .66$, $\beta = .14$, $p = .001$; Table 8, Model 1). If high Machs were strategically changing their moral judgment based on their role (more so than the average individual), then Mach should moderate the effect of role on relative equity endorsement, which it does not ($p = .54$). The fact that role predicted relative equity endorsement regardless of Mach is consistent with the self-deception hypothesis.

Did explicit influence, by itself or interacting with coalition strength, predict the self-interest bias? Yes. Consistent with the strategizing hypothesis, individuals higher on explicit influence made more self-interested proposals ($b = 1.24$, $\beta = .10$, $p = .02$; Table 3). However, the effect of explicit influence on proposal self-interest was not moderated by coalition strength ($p = .16$). Individuals higher on explicit influence also exhibited more moral self-interest ($b = .33$, $\beta = .16$, $p < .001$; Table 3), which is again consistent with the strategizing hypothesis. The effect of explicit influence on moral self-interest was not moderated by coalition strength ($p = .71$). The fact that the effect of explicit influence was not moderated by coalition strength in the case of proposal or moral self-interest suggests that explicit influence is not strictly measuring the degree to which individuals were influenced by relative coalition size: if individuals were in fact influenced, then coalition strength in combination with explicit influence should have an effect on bias, which is not the case. Consequently, explicit influence appears to be a marker for an individual difference in how one behaves in resource allocation situations. Individuals high on explicit influence engage in more self-interested behavior, and their reporting that they

considered relative coalition size when making their decisions may indicate a cognitive

attunement to strategic information and payoff information more broadly.

To examine whether the self-interest bias was operative among individuals both low and

high on explicit influence (1 *SD* below and above the mean, respectively), the interaction

between explicit influence and role predicting relative equity endorsement ($b = .69$, $\beta = .32$, $p <$

.001; Table 8, Model 1) was broken down into simple slopes. Role significantly predicted

relative equity endorsement among individuals high on explicit influence ($b = 2.31$, $\beta = .86$, $p <$

.001), while this relationship was marginally significant among individuals low on explicit

influence ($b = .60$, $\beta = .22$, $p = .09$). This pattern, shown in Figure 2, is consistent with the

strategizing hypothesis and provides weak evidence for the self-deception hypothesis.

Exploratory Analyses – The Effects of Time, Gender and Taking Action

Exploratory: Did the self-interest bias in proposal behavior change over time? Yes. As

shown in Figure 3, Proposals tended to become more self-interested in later periods of play ($b =$

2.44, $\beta = .20$, $p < .001$; Table 3). Although the observed temporal trend was unanticipated,

exploratory models were created to examine whether the effects of the variables of interest

changed over time. As shown in Table 5, several exploratory interactions between period of play

and other variables emerged.

There was a marginally significant three-way interaction between coalition strength,

period and gender ($b = -.13$, $\beta = -.19$, $p = .07$). Simple slopes indicated that among men in the

first period of play, coalition strength had a marginally significant positive impact on proposal

self-interest ($b = .18$, $\beta = .20$, $p = .08$). Among men in the fifth period of play, coalition strength

was unrelated to proposal self-interest ($p = .45$). Among women, coalition strength was not

related to proposal self-interest in either the first period of play ($p = .44$) or the fifth period of

play ($p = .26$). The three-way interaction provides marginal, qualified support for the coalitional theory of morality.

There was also a three-way interaction between period, Mach, and role ($b = -4.29$, $\beta = -.21$, $p = .03$). Changing the role reference group revealed that the two-way interaction between period and Mach was significant among Checkers ($b = 4.21$, $\beta = .20$, $p = .02$), but not among Typists ($p = .93$). Simple slopes indicated that, among high Machs (1 *SD* above the mean) in the Checker role, proposal self-interest increased in later periods of play ($b = 4.06$, $\beta = .33$, $p = .008$). In contrast, among low Machs (1 *SD* below the mean) in the Checker role, proposal self-interest was not related to period of play ($p = .52$).

Finally, the effect of period was moderated by explicit influence ($b = 1.16$, $\beta = .12$, $p = .08$). Simple slopes indicated that, among individuals high on explicit influence (1 *SD* above the mean), proposal self-interest increased in later periods ($b = 3.00$, $\beta = .24$, $p = .01$). In contrast, among individuals low on explicit influence (1 *SD* below the mean), proposal self-interest was not related to period of play ($p = .93$). The implications of the observed temporal trends are attended to in the discussion.

Exploratory: Did gender, by itself or interacting with coalition strength, predict self-interest bias in moral judgment? Yes. As shown in Table 6, there was no main effect of gender upon moral self-interest ($b = -.29$, $\beta = -.11$, $p = .25$). However, as shown in Figure 4, gender did interact with coalition strength ($b = .03$, $\beta = .23$, $p = .02$). Changing the gender reference group revealed that there was a marginal positive simple effect of coalition strength upon moral self-interest among men ($b = .02$, $\beta = .15$, $p = .07$), but not among women ($p = .20$). The effect of coalition strength upon the moral judgments of men provides marginal, qualified support for the coalitional theory of morality.

Exploratory: Did being a proposer predict self-interest bias in moral judgment? Yes. Compared to individuals who did not make a proposal, proposers exhibited more moral self-interest ($b = .81$, $\beta = .31$, $p = .01$; Table 6).

To examine whether the self-interest bias was operative among both proposers and non-proposers, the interaction between proposing and role predicting relative equity endorsement ($b = 1.53$, $\beta = .57$, $p = .02$; Table 8, Model 2) was broken down into simple slopes. Changing the proposer reference group revealed that there was a simple effect of role upon relative equity endorsement among proposers ($b = 1.81$, $\beta = .68$, $p < .001$), but not among non-proposers ($p = .65$), suggesting that making a proposal was a necessary condition for the moral self-interest bias to occur. Although not included in the model in Table 6, there was a marginally significant interaction between role and proposing in predicting moral self-interest ($p = .08$), such that the effect of proposing was stronger among Checkers, a trend which is evident in Figure 5.

Exploratory: Did the predicted effects of coalition strength upon proposal behavior emerge under conditions of post-hoc interest? Because of the moderators identified above, proposal data was re-analyzed to see if the predicted effects of coalition strength emerged under select conditions. Generally, re-analyses did not substantively change the findings reported thus far.

The first re-analysis strategy was to exclude proposals that fell outside the plausibly fair range (50-75) and re-run the confirmatory model (Table 3). Excluding data outside this range may be justified if individuals who proposed outside the plausible range were engaged in a decision-making process qualitatively different from offering what they think is morally acceptable (e.g. extreme proposals may have been driven by explicit payoff maximization). When proposals within the plausibly fair range were analyzed separately ($N = 423$), the effect of

coalition strength on proposal self-interest remained non-significant ($p = .67$). The interaction between coalition strength and Mach was reduced to marginal significance ($p = .08$). Thus, exclusively analyzing plausibly fair proposals did not substantively change the results.

The second re-analysis strategy was to re-run the confirmatory model with an extreme groups approach and only retain data from sessions with a coalition strength split of 75%/25% or greater ($N = 161$), so as to focus on cases where the manipulation of coalition strength was strongest. The extreme groups analysis found no main effect of coalition strength ($p = .51$), no interaction between coalition strength and period ($p = .52$), and no three-way interaction between coalition strength, period and gender ($p = .85$). To be thorough, a follow-up test was conducted with period centered on the first period of play, and this test found no effect of coalition strength during the first period ($p = .41$). Thus, the extreme groups analysis did not substantively change the results. Treating coalition strength as a categorical variable was not possible in the extreme groups analysis because role was perfectly confounded with having an extreme coalitional majority.

The third re-analysis strategy was to examine simple effects when moderators were at extreme values. In particular, the three-way interaction between period, gender and coalition strength (Table 5), which suggested that the effect of coalition might be greater in earlier rounds of play, was further explored. Prior to adding gender to the exploratory model, the interaction between period and coalition strength was non-significant ($p = .87$). Nevertheless, a follow-up analysis was conducted where period was centered on the first period of play. In the first period of play, the effect of coalition strength remained non-significant ($p = .68$). Next, to follow up on the marginally significant simple effect of coalition strength among men in the first period reported above, the three-way interaction was interpreted from a different point of view that

allowed for coalition strength to be centered at its maximum observed value of 83. When

coalition strength was at its maximum and the game was in the first period, simple slopes

indicated that the proposals of men were marginally more self-interested than those of women (b

$= 7.71$, $\beta = .50$, $p = .06$). However, when coalition strength was at its minimum of 17 and the

game was in the first period, the proposals of men and women did not differ ($p = .13$). This

conditional gender effect supports the possibility that the behavior of men was sensitive to

coalition strength.

DISCUSSION

Random assignment to a task-specific role caused self-interested changes in moral judgments about how to divide a reward, consistent with the findings of DeScioli et al. (2014) and a broader literature about the role of self-interest in moral judgment (Brown-Iannuzzi et al., 2015; Petersen et al., 2013; Rustichini & Villeval, 2014). Indeed, the effect of role upon moral judgment was larger than that of individual difference variables predicting relative equity endorsement, including political conservatism and trait Machiavellianism (Table 8, Model 1). Consistent with the premise that there are plausibility constraints on what constitutes fairness, proposals between the poles of equality and equity (i.e. 50 and 75) accounted for 85.8% of all proposals. However, Checker proposers ($M = 53.01$) and Typist proposers ($M = 58.76$) preferred different regions of this plausible range.

The primary aim of the current study was to test the coalitional theory of morality. The central predictions – that the self-interest bias in moral behavior and/or judgment would increase as a function of coalition strength – were not supported in confirmatory analyses. However, qualified support for the predictions did emerge when potential moderators were examined. Three separate interactions suggested that coalition strength can affect the self-interest bias. First, among high Machs but not low Machs, greater coalition strength led to more self-interested proposals about how to divide the task bonus. Second, among men but not women, coalition strength had a marginally significant association with proposal self-interest in the first period of play, but this association declined in later periods. Third, among men but not women, coalition strength led to more self-interested moral judgments about how the task bonus should be divided. These interactions provide hints about the critical conditions under which coalitional effects on

morality might be observed.

Self-Deception and Strategizing

       Turning to the mechanisms underlying the self-interest bias, explicit influence predicted greater bias in both proposal behavior and moral judgment, while the effects of Mach were qualified by coalition strength, in the case of proposal behavior, and role, in the case of moral judgment. However, the apparent increased moral self-interest of high Machs relative to low Machs in the Typist role is likely an artifact resulting from the tendency of Machs to endorse equity regardless of role (Table 8). Altogether, the effects of explicit influence and Mach suggest that individual differences in strategic motives and cognition affect the self-interest bias, providing support for the strategizing hypothesis. The effect of role on moral judgment was not moderated by Mach, but was moderated by explicit influence, such that the effect of role was reduced to marginal significance among individuals low on explicit influence. These patterns suggest that Mach is not a necessary condition for the moral self-interest bias to occur while explicit consideration of strategic information may be, a picture that provides mixed support for the self-deception hypothesis. The relation between self-interested behavior and attention to strategic information is consistent with past findings suggesting that a reflective cognitive style is associated with more rational decision making in cooperative games, which is in turn associated with more behavioral sensitivity to the specific contingencies of the cooperation context (Boone et al., 2010; Peysakhovich & Rand, 2016; Rand, Greene, & Nowak, 2012a). The current study extends previous findings by suggesting that strategic thinking not only leads to more self-serving economic behavior, but also to concomitant shifts in moral attitudes.

Agency and the Self-Interest Bias

Another potential necessary condition for the self-interest bias to emerge is moral agency. In the current study, moral judgments were only biased by role among individuals who took action as a proposer. The importance of personal action in guiding attitude change is readily predicted by cognitive dissonance theory (Aronson, 1969). Similarly, an intuitionist perspective suggests that explicit moral beliefs may arise as post-hoc justifications (Haidt, 2001). Functionally, attitude change could serve several purposes. First, proposers are at risk of being accused of acting out of self-interest, and self-deceptive attitude shifts may help maintain plausible deniability about the role of self-interest (Chakroff, Thomas, Haque, & Young, 2015). On this view, non-proposers face no condemnation risk and therefore have no incentive to shift their attitudes. Second, maintaining a consistent self-presentational story—i.e. maintaining consonance between one's behaviors and public beliefs—is an essential part of preserving one's credibility as a communicator and therefore one's ability to use messages to attain social influence (Kurzban & Aktipis, 2007). Third, the coalitional perspective suggests that action preparation may be accompanied by attitudinal shifts in morality that ultimately cause the actor to emit coordination messages that recruit coalitional support for one's preferred course of action (DeScioli & Kurzban, 2013). Finally, it is possible that the powerlessness of non-proposers explains the absence of the self-interest bias among them. If individuals have no power, then self-interested advocacy may have no material probability of succeeding, and worse, it may entail negative repercussions from those who do have power, such that deferring to the apparent moral consensus is the safest option (Trivers, 2011). Consistent with this possibility, non-proposer Checkers endorsed equity over equality (see Figure 5), which reflects the average (i.e. normative) trend in actual proposals ($M = 56.51$). The current study cannot confirm or reject any

of the above explanations for the effect of action on moral judgment, but future investigations of the effect of taking action, messaging an audience, or possessing power could shed light on the social challenges faced by moral agents.

Limitations and Implications for the Coalitional Theory of Morality

The absence of a main effect of coalition strength upon self-interest bias in moral judgment is surprising. In general, perception is known to track affordances available to the individual in the present situation (Proffitt, 2006; Witt & Proffitt, 2008), and this affordance tracking principle is known to extend to the social domain (Cesario & Navarrete, 2014; Fessler & Holbrook, 2013; Schnall, Harber, Stefanucci, & Proffitt, 2008). Indeed, studies of moral hypocrisy show that increasing power—i.e. increasing affordances to impose one's interests on others—is associated with increasingly self-serving morality (Lammers, Stapel, & Galinsky, 2010[1]; Rustichini & Villeval, 2014). In the current study, one explanation for the failure to find a main effect of coalition strength may be that individuals did not perceive role size as granting increased affordances for passing a proposal. There is reason to think that the manipulation of role size failed to serve its intended function as a manipulation of affordances for coordinated action.

First, because 93.5% of all proposals were accepted, participants may have learned that coalition strength was simply irrelevant to voting outcomes. Of the proposals between the plausible poles of equality and equity (i.e. between 50 and 75), 98.8% were accepted. Consistent with a learning hypothesis, proposals became more self-interested in later periods of play, suggesting that as participants gained experience in the game, they noticed and took advantage of the reluctance to reject proposals. This interpretation is also supported by the fact the effect

---

[1] *Psychological Science* concluded that this paper, which had Diederik Stapel as an author, contained no evidence of fraud and should not be retracted ("Retraction of 'The secret life of emotions'," 2012).

of time was greater among high Machs (in the Checker role) and individuals high on explicit influence. In a very broad sense, the observed temporal trend is consistent with the central notion underlying the coalitional theory of morality, which is that moral behavior tracks social affordances. If the current study failed to make coalition strength relevant, then the coalitional theory was not adequately tested. The marginal effect of coalition strength upon proposal behavior in the first period among men, for whom coalitional information might be salient, suggests that coalition strength may have initially been perceived as relevant to voting outcomes but that this perception was not borne out by ensuing experiences in the game and thus abandoned.

More generally, participants may not have perceived their roles as socially meaningful entities, as they may neither have received cues that role members were coordinating as a coalition, nor signs that their coalitions actually possessed strength. Like any other psychological effects, group processes have boundary conditions and causal mediators, and in the current study, the mere assigning of roles may have been insufficient to realize group effects. For example, in cooperation contexts with minimal groups, identification with the in-group is a necessary condition for group-biased cooperation to occur, as is common knowledge of group identity between interaction partners (Yamagishi, Mifune, Liu, & Pauling, 2008). Common knowledge of group membership moderates in-group biases in trusting behavior as well (Platow, Foddy, Yamagishi, Lim, & Chow, 2012). These studies also show that psychological expectations that in-group members will behave in reciprocating and generous ways are important mediating processes underlying group effects, consistent with the idea that groups may only affect behavior insofar as they are perceived as socially functional units of organization (Yamagishi & Kiyonari, 2000). In the current study, the coalition strength manipulation may have been unsuccessful

because role identity was perceived as purely nominal, with limited implications for coalitional functioning.

The moderators of coalition strength observed in the current study suggest the possibility that because the situation was only weakly organized in coalitional terms, effects emerged exclusively among individuals who are especially attuned to coalitional information. Machiavellian individuals, who are higher on social astuteness, higher on competitive motivation, and more likely to engage in opportunistic behavior, increased their proposal self-interest as a function of coalition strength (Dahling et al., 2008). Moral self-interest and potentially proposal self-interest in early periods were affected by coalition strength among men, who are more prone than women to engage in intergroup competition (McDonald, Navarrete, & Van Vugt, 2012).

In the current study, communication among coalition members was strictly limited in the interest of experimental control, and such restrictions may have prevented the activation of the social psychological processes necessary for producing coalitional behavior. The social cognitive details of how individuals experienced their roles are critical for the coalitional theory of morality. Moral coalitions are predicted to form when individuals have common knowledge of shared interests, and coalitions are predicted to increase the self-interest bias when coalitions are perceived as capable of coordinated action—and thus formidable. These essential conditions may not have been met in the current study, as anonymous proposing and voting procedures prevented coalition members from emitting any cues about their coordination intentions. While participants may have reasonably assumed that fellow coalition members had similar goals vis-à-vis the bonus division, common knowledge of shared interests was absent, which is critical for coordination (Thomas et al., 2014). Importantly, if the current study failed to manipulate

perceived affordances for coordinated action, then the study did not adequately test the coalitional theory of morality. In future investigations, a more adequate test would involve communication within roles, which could allow coalition members to establish consensus about what proposals are acceptable (i.e. establish a coalitional platform) and to provide mutual assurances about voting in a coordinated fashion.

Conclusion

The current findings clarify the social conditions under which moral phenomena might become organized in terms of coalitions and suggest that individual differences in strategic and coalitional psychology impact the manifestation of self-interest bias in the moral domain. Additionally, the current study raises the possibility that agency is a boundary condition for the self-interest bias. Future research focusing on the social risks and opportunities faced by moral agents has the potential to further elucidate the functional causes and consequences of the self-interest bias.

APPENDIX

APPENDIX

Table 1:

Sample sizes across different levels of coalition strength and conditions

Coalition Strength	Sample Size		Condition/Role	Sample Size
16.67	8		Proposed/Typist	300
20.00	17		Proposed/Checker	193
25.00	16		Never Proposed/Typist	45
33.33	37		Never Proposed/Checker	47
40.00	75			
50.00	113			
60.00	116			
66.67	67			
75.00	46			
80.00	71			
83.33	32			

Table 2:

Descriptive statistics and correlations among study variables

	Mean (SD)	1	2	3	4	5	6	7
1. Role	.59							
2. Coalition Strength	56.82 (17.49)	.47**						
3. Explicit Influence	2.57 (1.23)	-.02	-.02					
4. Machiavellianism	2.50 (.58)	-.02	.02	.15**				
5. Proposed	.83	.09*	-.20**		-.01			
6. Period	2.57 (1.26)	-.09[at]	.22[a] **			.12**		
7. Proposal Self-Interest	54.15 (15.52)	.37[a] **	.23[a] **	.12[a] *	.08[a]		.14[a] **	
8. Moral Self-Interest	.69 (2.59)	.00	-.02	.16**	.08[t]	.11*		.26[a] **

Note. Role: 0 = Checker; 1 = Typist. Proposed: 0 = No; 1 = Yes. The [a] superscript indicates the correlations are solely based on observations of proposers (N = 493). [t]$p < .10$. * $p < .05$. ** $p < .01$.

Table 3:

Confirmatory linear model predicting proposal self-interest

	Proposal Self-Interest					
	B	B	*SE (b)*	*df*	*t*	*p*
Intercept	46.86	-.47	1.20			
Role	12.66	.82	1.51	415	8.43	$< .001$
Coalition Strength	.04	.05	.04	415	.95	.35
Machiavellianism	1.25	.05	1.09	415	1.15	.25
Explicit Influence	1.24	.10	.53	415	2.36	.02
Period	2.44	.20	.55	415	4.46	$< .001$
Coalition Strength × Machiavellianism	.12	.08	.06	415	2.01	.045

Note. Role: 0 = Checker; 1 = Typist.

Table 4:

Confirmatory linear model predicting moral self-interest

	Moral Self-Interest					
	B	B	*SE (b)*	*df*	*t*	*p*
Intercept	.55	-.06	.19			
Role	.25	.10	.26	493	.96	.34
Coalition Strength	-.01	-.04	.01	493	-.76	.45
Machiavellianism	-.51	-.11	.32	493	-1.57	.12
Explicit Influence	.33	.16	.09	493	3.51	< .001
Role × Machiavellianism	1.25	.28	.41	493	3.06	.002

Note. Role: 0 = Checker; 1 = Typist.

Table 5:

Exploratory linear model predicting proposal self-interest

	Proposal Self-Interest					
	b	β	*SE (b)*	*df*	*t*	*p*
Intercept	46.30	-.51	1.22			
Role	12.28	.79	1.51	405	8.15	< .001
Coalition Strength	.02	.02	.05	405	.38	.70
Machiavellianism	-.17	-.01	1.86	405	-.09	.93
Explicit Influence	.99	.08	.53	405	1.86	.06
Period	1.55	.13	1.07	405	1.45	.15
Gender	2.32	.15	1.45	405	1.61	.11
Period × Explicit Influence	1.16	.12	.43	405	2.70	.007
Role × Machiavellianism	1.07	.04	1.07	405	.47	.64
Period × Role	1.12	.09	1.13	405	.92	.36
Period × Machiavellianism	4.21	.20	1.75	405	2.40	.02
Period × Role × Machiavellianism	-4.29	-.21	1.96	405	-2.20	.03
Coalition Strength × Gender	.03	.04	.08	405	.43	.67
Period × Gender	.73	.06	1.19	405	.61	.54
Period × Coalition Strength	.05	.07	.04	405	1.15	.25
Period × Coalition Strength × Gender	-.13	-.19	.07	405	-1.81	.07

Note. Role: 0 = Checker; 1 = Typist. Gender: 0 = Female; 1 = Male.

Table 6:

Exploratory linear model predicting moral self-interest

	Moral Self-Interest					
	B	β	*SE (b)*	*df*	*t*	*p*
Intercept	.05	-.25	.30			
Role	.09	.03	.26	489	.34	.73
Coalition Strength	-.01	-.08	.01	489	-1.30	.20
Gender	-.29	-.11	.25	489	-1.16	.25
Proposed	.81	.31	.32	489	2.54	.01
Explicit Influence	.33	.16	.09	489	3.48	.001
Machiavellianism	-.44	-.10	.32	489	-1.38	.17
Role × Machiavellianism	1.18	.26	.41	489	2.92	.004
Coalition Strength × Gender	.03	.23	.01	489	2.38	.02

Note. Role: 0 = Checker; 1 = Typist. Proposed: 0 = No; 1 = Yes. Gender: 0 = Female; 1 = Male.

Table 7:

Linear model predicting proposal behavior

	Proposal					
	b	β	*SE (b)*	*df*	*t*	*p*
Intercept	53.26	-.22	1.16			
Role	5.78	.39	1.45	400.21	3.98	< .001
Coalition Strength	.00	.00	.04	438.94	-.07	.94
Period	-1.23	-.11	.94	454.46	-1.31	.19
Period × Role	3.74	.32	1.10	460.54	2.95	.001

Note. Role: 0 = Checker; 1 = Typist.

Table 8:

Linear models predicting endorsement of equity over equality

	Relative Equity Endorsement											
	Model 1						Model 2					
	b	β	*SE* (*b*)	*df*	*t*	*p*	*b*	β	*SE* (*b*)	*df*	*t*	*p*
Intercept	-.63	-.28	.20				.44	.12	.41			
Role	1.46	.54	.26	337.22	5.60	<.001	.28	.10	.61	439.38	.46	.65
Coalition Strength	-.01	-.06	.01	413.01	-1.19	.24	-.01	-.09	.01	417.75	-1.70	.09
Machiavellianism	.66	.14	.21	427.86	3.20	.001	.66	.14	.21	428.06	3.21	.001
Explicit Influence	-.26	-.12	.15	436.17	-1.72	.09	-.21	-.10	.15	435.55	-1.37	.17
Conservatism	.14	.08	.08	426.50	1.82	.07	.14	.08	.08	426.99	1.77	.08
Role × Explicit Influence	.69	.32	.20	434.60	3.53	<.001	.64	.30	.20	434.22	3.29	.001
Proposed	--						-1.39	-.52	.46	431.52	-3.00	.003
Role × Proposed	--						1.53	.57	.65	439.88	2.36	.02

Note. Role: 0 = Checker; 1 = Typist. Proposed: 0 = No; 1 = Yes.

Figure 1:

Proposal self-interest as a function of coalition strength and Machiavellianism
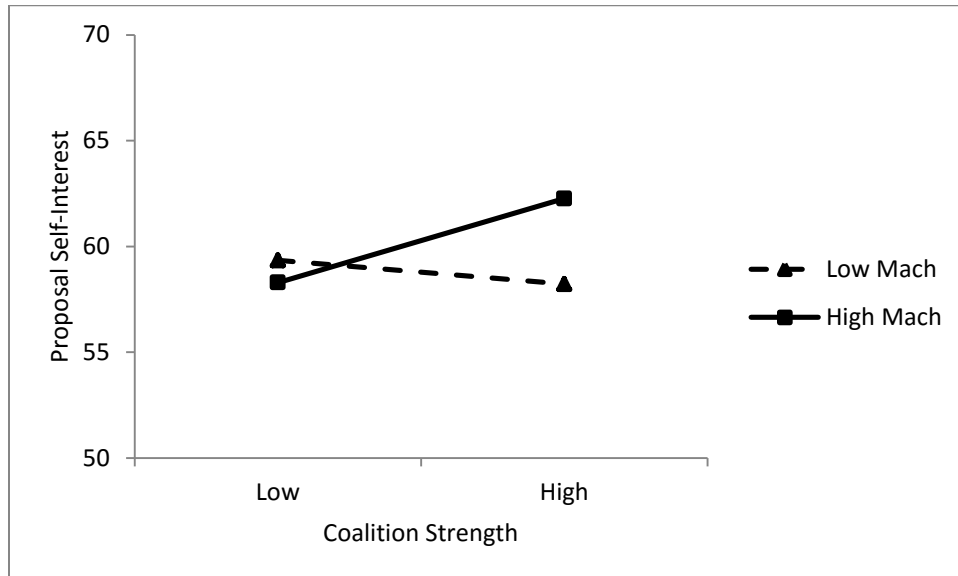
Note. The intercepts for Typists were used to obtain the predicted values.

Figure 2:

Relative equity endorsement by role and explicit influence
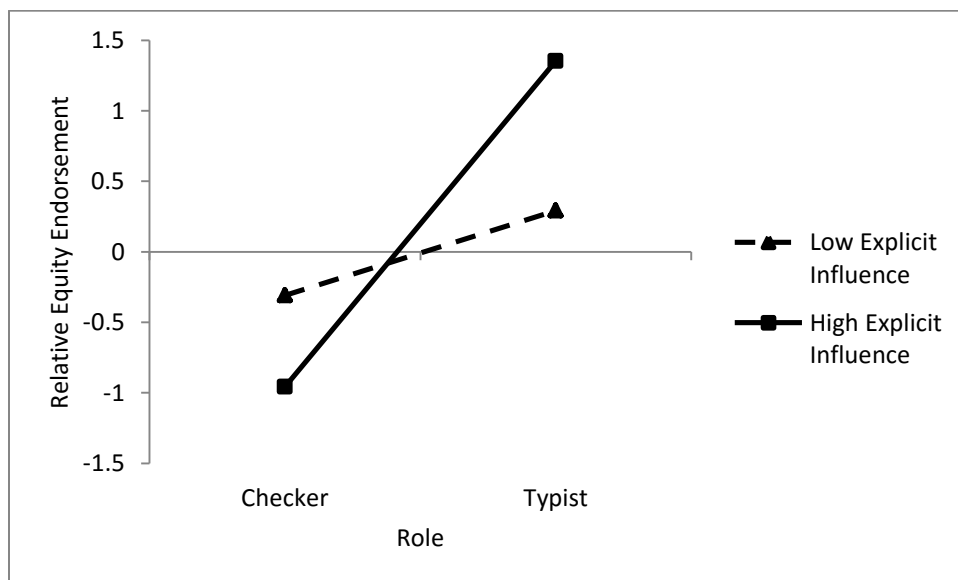
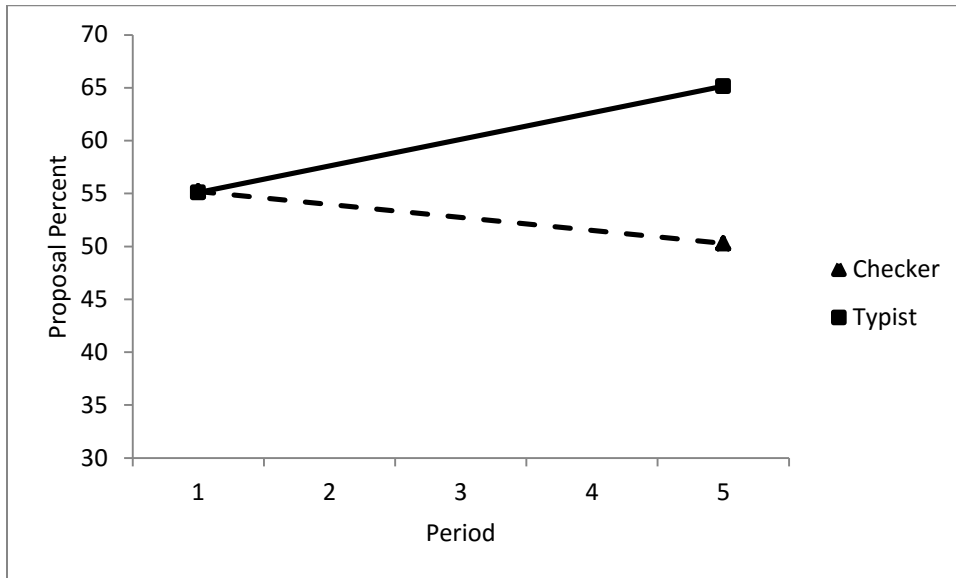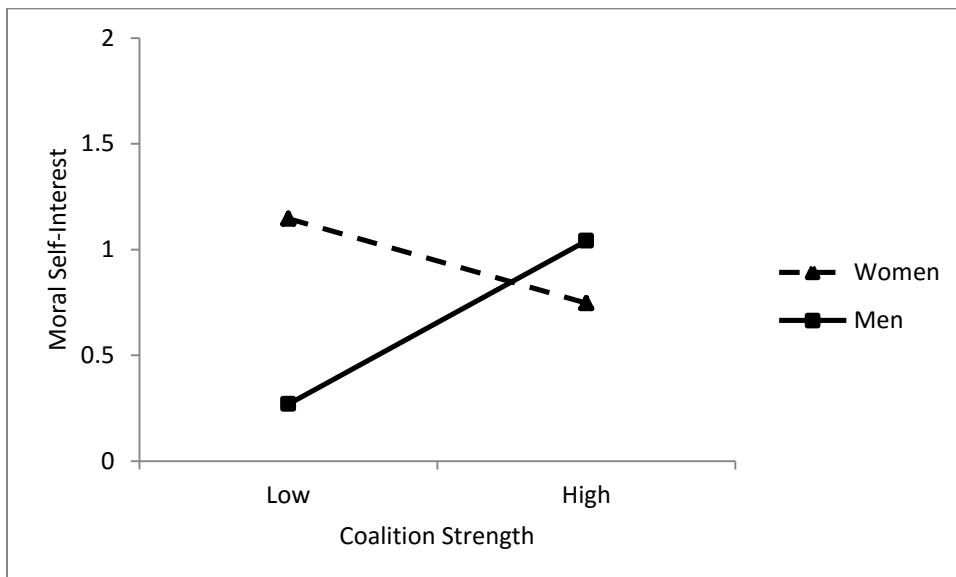Figure 3:
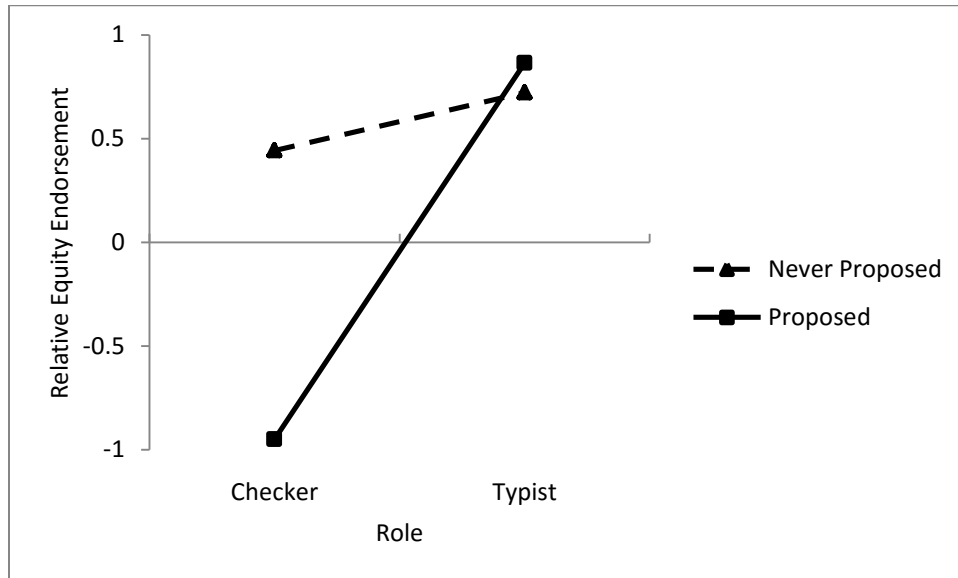
Proposal behavior over time by role

Figure 4:

Moral self-interest as a function of coalition strength and gender

Note. The intercepts for Typists were used to obtain the predicted values.

Figure 5:

Relative equity endorsement by role and proposing

REFERENCES

REFERENCES

Aronson, E. (1969). The Theory of Cognitive Dissonance: A Current Perspective. *Advances in Experimental Social Psychology*, *4*(C), 1–34. http://doi.org/10.1016/S0065-2601(08)60075-1

Bereczkei, T., Birkas, B., & Kerekes, Z. (2010). The presence of others, prosocial traits, Machiavellianism : A personality × situation approach. *Social Psychology*, *41*(4), 238–245. http://doi.org/10.1027/1864-9335/a000032

Boehm, C. (1999). *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*. Harvard University Press. Retrieved from https://books.google.com/books/about/HIERARCHY_IN_THE_FOREST.html?id=ICGAA AAAMAAJ&pgis=1

Boone, C., Declerck, C., & Kiyonari, T. (2010). Inducing Cooperative Behavior among Proselfs versus Prosocials: The Moderating Role of Incentives and Trust. *Journal of Conflict Resolution*, *54*(5), 799–824. http://doi.org/10.1177/0022002710372329

Brown-Iannuzzi, J. L., Lundberg, K. B., Kay, A. C., & Payne, B. K. (2015). Subjective status shapes political preferences. *Psychological Science*, *26*(1), 15–26. http://doi.org/10.1177/0956797614553947

Cesario, J., & Navarrete, C. D. (2014). Perceptual Bias in Threat Distance: The Critical Roles of In-Group Support and Target Evaluations in Defensive Threat Regulation. *Social Psychological and Personality Science*, *5*(1), 12–17. http://doi.org/10.1177/1948550613485605

Chakroff, A., Thomas, K. A., Haque, O. S., & Young, L. (2015). An indecent proposal: the dual functions of indirect speech. *Cognitive Science*, *39*(1), 199–211. http://doi.org/10.1111/cogs.12145

Charness, G., Rigotti, L., & Rustichini, A. (2007). Individual Behavior and Group Membership. *American Economic Review*, *97*(4), 1340–1352. http://doi.org/10.1257/aer.97.4.1340

Dahling, J. J., Whitaker, B. G., & Levy, P. E. (2008). The Development and Validation of a New Machiavellianism Scale. *Journal of Management*, *35*(2), 219–257. http://doi.org/10.1177/0149206308318618

DeScioli, P., & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological Bulletin*, *139*(2), 477–496. http://doi.org/10.1037/a0029065

DeScioli, P., Massenkoff, M., Shaw, A., Petersen, M. B., & Kurzban, R. (2014). Equity or equality? Moral judgments follow the money. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1797), 20142112. http://doi.org/10.1098/rspb.2014.2112

Ferris, G. R. (2005). Development and Validation of the Political Skill Inventory. *Journal of*

Management, *31*(1), 126–152. http://doi.org/10.1177/0149206304271386

Fessler, D., & Holbrook, C. (2013). Friends shrink foes: the presence of comrades decreases the envisioned physical formidability of an opponent. *Psychological Science*, *24*, 797–802. http://doi.org/10.1177/0956797612461508

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178. http://doi.org/10.1007/s10683-006-9159-4

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in Simple Bargaining Experiments. *Games and Economic Behavior*, *6*(3), 347–369. http://doi.org/10.1006/game.1994.1021

Goodwin, G. P., & Darley, J. M. (2008). The psychology of meta-ethics: exploring objectivism. *Cognition*, *106*(3), 1339–66. http://doi.org/10.1016/j.cognition.2007.06.007

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814–34. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11699120

Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, *65*(4), 613–28. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8229648

Johnson, S. D., & Tamney, J. B. (2001). Social traditionalism and economic conservatism: two conservative political ideologies in the United States. *The Journal of Social Psychology*, *141*(2), 233–43. http://doi.org/10.1080/00224540109600549

Kurzban, R., & Aktipis, C. A. (2007). Modularity and the social mind: are psychologists too selfish? *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc*, *11*(2), 131–149. http://doi.org/10.1177/1088868306294906

Lammers, J., Stapel, D. A., & Galinsky, A. D. (2010). Power Increases Hypocrisy: Moralizing in Reasoning, Immorality in Behavior. *Psychological Science*, *21*(5), 737–744. http://doi.org/10.1177/0956797610368810

McDonald, M. M., Navarrete, C. D., & Van Vugt, M. (2012). Evolution and the psychology of intergroup conflict: the male warrior hypothesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1589), 670–9. http://doi.org/10.1098/rstb.2011.0301

Paxton, J. M., Ungar, L., & Greene, J. D. Reflection and reasoning in moral judgment. *Cognitive Science*, *36*(1), 163–77. http://doi.org/10.1111/j.1551-6709.2011.01210.x

Petersen, M. B., Sznycer, D., Sell, A., Cosmides, L., & Tooby, J. (2013). The ancestral logic of politics: upper-body strength regulates men's assertion of self-interest over economic redistribution. *Psychological Science*, *24*(7), 1098–103. http://doi.org/10.1177/0956797612466415

Peysakhovich, A., & Rand, D. G. (2016). Habits of Virtue: Creating Norms of Cooperation and Defection in the Laboratory. *Management Science*, *62*(3), 631–647. http://doi.org/10.1287/mnsc.2015.2168

Pillutla, M. M., & Murnighan, J. K. (1996). Unfairness, Anger, and Spite: Emotional Rejections of Ultimatum Offers. *Organizational Behavior and Human Decision Processes*, *68*(3), 208–224. http://doi.org/10.1006/obhd.1996.0100

Platow, M. J., Foddy, M., Yamagishi, T., Lim, L., & Chow, A. (2012). Two experimental tests of trust in in-group strangers: The moderating role of common knowledge of group membership. *European Journal of Social Psychology*, *42*(1), 30–35. http://doi.org/10.1002/ejsp.852

Price, M. E., Brown, S., Dukes, A., & Kang, J. (2015). Bodily attractiveness and egalitarianism are negatively related in males. *Evolutionary Psychology : An International Journal of Evolutionary Approaches to Psychology and Behavior*, *13*(1), 140–66. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25664560

Price, M. E., Kang, J., Dunn, J., & Hopkins, S. (2011). Muscularity and attractiveness as predictors of human egalitarianism. *Personality and Individual Differences*, *50*(5), 636–640. http://doi.org/10.1016/j.paid.2010.12.009

Proffitt, D. R. (2006). Embodied Perception and the Economy of Action. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *1*(2), 110–22. http://doi.org/10.1111/j.1745-6916.2006.00008.x

Rand, D. G., Greene, J. D., & Nowak, M. a. (2012a). Spontaneous giving and calculated greed. *Nature*, *489*(7416), 427–430. http://doi.org/10.1038/nature11467

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012b). Spontaneous giving and calculated greed. *Nature*, *489*(7416), 427–30. http://doi.org/10.1038/nature11467

Retraction of "The secret life of emotions" Psychological Science, 19, 385-91 and "Emotion elicitor or emotion messenger? Subliminal priming reveals two faces of facial expressions" Psychological Science, 19, 593-603. (2012). *Psychological Science*, *23*(7), 828. http://doi.org/10.1177/0956797612453137

Ritzman, R. L., & Tomaskovic-Devey, D. (1992). Life Chances and Support for Equality and Equity As Normative and Counternormative Distribution Rules. *Social Forces*, *70*(3), 745–763. http://doi.org/10.1093/sf/70.3.745

Rustichini, A., & Villeval, M. C. (2014). Moral hypocrisy, power and social preferences. *Journal of Economic Behavior & Organization*, *107*, 10–24. http://doi.org/10.1016/j.jebo.2014.08.002

Sakalaki, M., Richardson, C., & Thépaut, Y. (2007). Machiavellianism and Economic Opportunism. *Journal of Applied Social Psychology*, *37*(6), 1181–1190. http://doi.org/10.1111/j.1559-1816.2007.00208.x

Schlenker, B. R., Britt, T. W., & Pennington, J. (1996). Impression regulation and management:

Highlights of a theory of self-identification. In *Handbook of Motivation and Cognition* (pp. 118–147). Retrieved from http://psycnet.apa.org/psycinfo/1996-98385-004

Schnall, S., Harber, K. D., Stefanucci, J. K., & Proffitt, D. R. (2008). Social Support and the Perception of Geographical Slant. *Journal of Experimental Social Psychology*, *44*(5), 1246–1255. http://doi.org/10.1016/j.jesp.2008.04.011

Sell, A. N. (2011). The recalibrational theory and violent anger. *Aggression and Violent Behavior*, *16*(5), 381–389. http://doi.org/10.1016/j.avb.2011.04.013

Sell, A., Tooby, J., & Cosmides, L. (2009). Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(35), 15073–15078. http://doi.org/10.1073/pnas.0904312106

Shaw, A. (2013). Beyond "to Share or Not to Share": The Impartiality Account of Fairness. *Current Directions in Psychological Science*, *22*(5), 413–417. http://doi.org/10.1177/0963721413484467

Silk, J. B. (2007). Social components of fitness in primate groups. *Science*, *317*(5843), 1347–1351. http://doi.org/10.1126/science.1140734

Smetana, J. G., & Braeges, J. (1990). The development of toddlers' moral and conventional judgements. *Merrill-Palmer Quarterly*, *36*(JUNE 1990), 329–346.

Thomas, K. a, DeScioli, P., Haque, O. S., & Pinker, S. (2014). The psychology of coordination and common knowledge. *Journal of Personality and Social Psychology*, *107*(10), 657–676. http://doi.org/10.1037/a0037037

Tooby, J., & Cosmides, L. (2010). Groups in mind: The coalitional roots of war and morality. *Human Morality and Sociality: Evolutionary …*, 191–234. Retrieved from http://www.psych.ucsb.edu/research/cep/papers/groups in mind2010.pdf

Tooby, J., Cosmides, L., Sell, A., Lieberman, D., & Sznycer, D. (2008). Internal regulatory variables and the design of human motivation: A computational and evolutionary approach. *Book*, 1–22. http://doi.org/10.1348/135910710X500828

Trivers, R. L. (1971). The Evolution of Reciprocal Altruism. *Chicago Journals*, *46*(1), 35–57. http://doi.org/10.1086/406755

van Yperen, N. W., van den Bos, K., & de Graaff, D. C. (2005). Performance-based pay is fair, particularly when I perform better: differential fairness perceptions of allocators and recipients. *European Journal of Social Psychology*, *35*(6), 741–754. http://doi.org/10.1002/ejsp.273

von Hippel, W., & Trivers, R. (2011). The evolution and psychology of self-deception. *The Behavioral and Brain Sciences*, *34*(1), 1–16; discussion 16–56. http://doi.org/10.1017/S0140525X10001354

Weeden, J., & Kurzban, R. (2014). *The Hidden Agenda of the Political Mind: How Self-Interest Shapes Our Opinions and Why We Won't Admit It*. Princeton University Press. Retrieved

from
https://books.google.com/books/about/The_Hidden_Agenda_of_the_Political_Mind.html?i
d=z73BAwAAQBAJ&pgis=1

Witt, J. K., & Proffitt, D. R. (2008). Action-specific influences on distance perception: a role for motor simulation. *Journal of Experimental Psychology. Human Perception and Performance*, *34*(6), 1479–1492. http://doi.org/10.1037/a0010781

Wrangham, R. W. (1999). Evolution of coalitionary killing. *Yearbook of Physical Anthropology*, *Suppl 29*, 1–30. http://doi.org/10.1002/(SICI)1096-8644(1999)110:29+<1::AID-AJPA2>3.0.CO;2-E

Yamagishi, T., & Kiyonari, T. (2000). The Group as the Container of Generalized Reciprocity. *Social Psychology Quarterly*, *63*(2), 116. http://doi.org/10.2307/2695887

Yamagishi, T., Mifune, N., Liu, J. H., & Pauling, J. (2008). Exchanges of group-based favours: Ingroup bias in the prisoner's dilemma game with minimal groups in Japan and New Zealand. *Asian Journal of Social Psychology*, *11*(3), 196–207. http://doi.org/10.1111/j.1467-839X.2008.00258.x