Blind Spots:
The Effects of Information and Stakes on Fairness Bias and Dispersion

James Konow
Department of Economics
Loyola Marymount University
One LMU Drive, Suite 4200
Los Angeles, CA 90045-2659
Telephone: (310) 338-2956
Fax: (310) 338-1950
Email: jkonow@lmu.edu

I wish to thank John Jost and three anonymous referees for helpful comments and suggestions on this version. I am also grateful to the Ludwig Boltzmann Institute for financial support and to Wulf Gaertner, Reinhard Neck, Joe Oppenheimer, Lars Schwettmann, Christian Seidl and participants of the meetings of the Society for Social Choice and Welfare and the Public Choice Society for valuable input on earlier versions of this paper. Any remaining shortcomings are, of course, my own.
Abstract

Mounting empirical research provides evidence of fairness bias and its economic and social effects, where fairness bias refers here to a deviation of claims from unbiased justice due to a personal stake. A far less appreciated issue is dispersion of fairness views and claims, which is also important for its effects on disagreements, empirical analysis and philosophical theories. This study undertakes a systematic analysis of the effects on fairness bias and dispersion of two variables: stakes and information. Most philosophical and social science analyses related to justice and bias associate heightened bias with increased information and, conversely, impartiality with the elimination of certain information. Less attention has been paid to the opposing impact of information, which is to supply the facts needed to achieve justice more reliably. An important open question is whether, on balance, increased information helps agents to achieve fairer outcomes or whether biased use of such information contributes to less fair outcomes. This study focuses on a set of previously reported experiments that share certain features and subjects them to a new analysis. The results of this analysis suggest that, although information is often used in a self-serving way, increased information can, under certain conditions, contribute to fairness claims becoming less biased and less dispersed, both for stakeholders as well as impartial spectators.

Keywords: Fairness, justice, equity, bias, dispersion, impartial spectator
“No fair!” – My son Alex, whenever something does not suit him.

A common perception is that justice is an imprecise, amorphous or perhaps even nonexistent concept, in large part because of the opportunistic and self-serving manner in which it is often employed. The approximately 26,000 hits on a Google search for the statement “fairness is in the eye of the beholder” attest to the popularity of this sentiment. Even some justice scholars express similar views. For example, Pillutla and Murnighan (2003) conclude from the results of certain bargaining experiments that “people base their behaviors on their personal conclusions about fairness, rather than on an objective standard” (pg. 254). If true, this proposition should trouble not only social scientists, who rely on patterned and converging evidence on fairness preferences (distinct from self-interest) for explanation, policy and prediction, but might also cause advocates of prescriptive justice theories to despair. Although normative theories of justice in philosophy, economics and legal studies typically involve reasoning from some ideal conditions, the relevance of such theorizing is surely undermined if actual views of justice are inexorably biased, idiosyncratic, disperse and lacking any real substance independent of other motives such as self-interest.

This paper undertakes a comparative analysis of evidence from experimental studies on how variations in stakes and in information affect two properties of fairness. One property is the fairness bias that individual stakeholders exhibit and that has been documented in Messick and Sentis (1983) and in a flurry of more recent investigations (see Babcock and Loewenstein, 1997, and Handgraaf, Van Dijk and De Cremer, 2003, for surveys). Another is the dispersion of fairness views and claims, which can be traced to biased use of fairness by stakeholders to promote their own disparate and conflicting interests as well as to other causes (e.g., unbiased but “noisy” perceptions of what is fair). This is a far less appreciated but, in this author’s view, no less important dimension of the problem. Indeed, dispersion would be an important issue even in the absence of any fairness bias: for example, suppose two individuals with opposing stakes bargain over a fixed pie but have claims drawn from identical distributions (in terms of both mean and variance). More disperse (i.e., higher variance) distributions will generate a greater
impasse rate since average differences in claims will be larger, and, as Babcock et al. (1995) have demonstrated, that leads to a greater incidence of disagreements. The additional presence of a fairness bias exacerbates this problem by further distancing parties with opposing stakes.

Both unbiasedness and convergence (to use these terms as antonyms of bias and dispersion, respectively) are important both to descriptive and prescriptive justice research. Confidence in the findings of empirical social science research relates directly to observed patterns of concurrence by the very nature of the statistical analysis they apply. Prescriptive claims regarding justice in philosophy, law and economics also gain legitimacy if consistent with impartial views that command a consensus, e.g., Rawls’s (1971) construct rests not only on a particular concept of impartiality but also on the presumed consensus among people about what is just. Moreover, these properties also bear on bargaining (as alluded to above) and on social policies. Policy recommendations regarding, for example, affirmative action, income taxation and education, derive their ethical force from principles that can be defended based on impartiality, including impartial justice, and on a high level of agreement on its exigencies. The current study finds a clear pattern of effects of stakes and information on bias and dispersion that can inform the practice of social scientists, philosophers and policy makers and that suggests means to temper bias and reduce dispersion.

The study of fairness bias is still developing, as is the accompanying terminology, so it is important to clarify terms and concepts that are employed in this paper. I proceed from the assumption that impartial (or unbiased) justice does exist. Later in the paper, I will be more specific about this claim of its existence and about how it might be identified. But it seems necessary to take as first principles in any study of justice the existence of unbiased justice, unless one wishes to argue that there is no justice independent of, say, self-interest, which implies there is no such thing as justice at all: justice is merely a word people sometimes (mysteriously) choose to attach to self-interested acts. Then, fairness bias refers to a deviation of claims from unbiased justice due to a personal stake. As used here, it does not preclude unbiased fairness; indeed, it implies a deviation that, on average, reflects the combined forces of both
unbiased fairness and self-interest. This bias can be decomposed into at least two parts. First, there is the effect of the unadulterated self-interest of agents, which I will call a *self-centered bias*. This is the part of the fairness bias that people acknowledge to themselves as unfair but act on, anyway, out of self-interest. Note that self-recognition of this unfairness does not rule out the possibility that people might misrepresent the fairness of their actions to others for personal gain. Second, there is the *self-serving bias* (also called egocentric bias), which involves self-deception. This is the part of the fairness bias that is due to agents adjusting their beliefs about what is fair to reduce the unpleasantness of blatantly unfair behavior.

On the latter point, Festinger (1957) proposed that people engage in self-deception to reduce dissonant cognitions, in this context, the often diverging demands of unbiased justice and self-interest, and much subsequent research appears to support this claim. For example, 88% of Americans believe they are above average drivers (Svensson, 1981). Even the educated succumb to this bias: 94% of college professors in one survey said they perform above average work (Cross, 1977). When people are stakeholders, they tend to view their fair share of a reward as being greater than their fair share in the view of others (e.g., Messick and Sentis, 1983), and they reveal a willingness to act on biased views of fairness, including in experiments with monetary rewards (e.g., Babcock et al., 1995, Konow, 2000) and in the field as real employers and employees (Babcock, Wang and Loewenstein, 1996). Such moral “blind spots” are implicated in a wide range of social and economic problems, including armed conflicts, labor strikes, legislative logjams and personal disputes. In addition, a necessary first step in designing policies to rectify injustice is agreement on what justice demands, and self-serving biases prove a potentially formidable obstacle to identifying such goals.

Various studies indicate that stakes and information work in tandem to produce fairness biases. For instance, Babcock and Loewenstein (1997) report both laboratory and field studies

---

1 Indeed, the study of Peters, van den Bos and Bobocel (2004) suggests that there is even a kind of second order self-serving bias that they call the *moral superiority effect*: in situations of overcompensation, people expect their own inequity aversion to be more acute than that of others.
demonstrating that fairness biases are sensitive to the stakes and information of individuals and result in bargaining impasses that carry significant economic costs. With regard first to stakes, the “blind justice” we will explore in this paper is akin to the juridical model: impartiality is associated with “impartial spectators” who have no stakes in the allocation being evaluated but are aware of stakes and stakeholders, indeed, in that model, ideally, they know all the relevant particulars. Note that this differs from some common notions of “blindness” in the justice literature, which conceive of justice as being “blind” with respect to stakes. For example, from a social contract approach, Rawls posits a hypothetical state from which agents do not possess any information related to their stakes that could anyway bias their judgments. In the real world, of course, such conditions are rarely, if ever, fulfilled. In the current study, therefore, we focus on studies in which agents are informed about stakes and on the effects on fairness bias and dispersion of being in the role of either stakeholder or impartial spectator.

Turning now to information, the main concern of this paper, as just stated, is not for information about stakes but rather for other variables. Numerous studies suggest that information, i.e., the quantity and quality of real or perceived facts, provides the raw material for fairness biases, specifically, through the self-serving bias component. Babcock et al. (1995) and Loewenstein et al. (1993), for example, conducted contextually rich bargaining experiments and found egocentric processing of information: adversarial parties cited more arguments in their own favor and attached greater importance to those arguments. In contrast to the use above, this is a different and unfavorable type of “blindness”: information appears to feed self-deception and to produce moral “blind spots” that impair an agent’s capacity to perceive fairness impartially. Less attention, however, has been paid to the opposing impact of information: knowledge about facts relevant to justice provides the means to achieve justice more reliably. Another goal of this paper is to address the important open question of whether, on balance, increased information helps agents to achieve fairer outcomes or whether biased processing and use of such information contributes to less fair outcomes. Moreover, do the effects of information on bias and dispersion differ for stakeholders and spectators?
Although there is a substantial literature documenting fairness biases, to my knowledge, such a systematic analysis of the effects of both stakes and information on both bias and dispersion has been lacking. In particular, the new and more detailed data analysis in this paper focuses on evidence from nine previously reported experiments involving real monetary stakes in which the subjects’ interests conflict. Data from these studies are presented in a common format and undergo a new analysis for bias and dispersion. The results suggest a trade-off between self-interest and fairness. Nevertheless, parties often employ fairness concepts opportunistically, using the one that serves their interests. Dispersion of views and disagreements, therefore, tend to increase with conflicts of interests and with the availability of competing concepts. Overall, however, providing parties with information relevant to justice tends to bring about outcomes that are less biased and less dispersed, at least where bargaining is symmetric. Studies that manipulate stakes or the timing of information about stakes suggest that the self-serving bias can be traced partly (if not mostly) to the egocentric processing of information to bolster beliefs that it is fair to be unfair to others. Collectively, these studies suggest that views of justice lose their bias when people are disassociated from their interests while becoming informed of the relevant specifics. This provides a means for justice scholars to identify the effects of unbiased justice while supplying a foundation for normative and philosophical inquiry.

The following section reviews the foundational literature from studies of fairness biases. Then I examine the effects on bias and dispersion when stakes vary. This is followed by an analysis of the effects of information on bias and dispersion, and the final section concludes.

BACKGROUND

The results of numerous survey studies suggest that views of justice are related to the personal interests, experiences and characteristics of individuals. Many of these have employed vignettes, or hypothetical scenarios, and have elicited the preferences of respondents over alternate allocations. Gaertner, Jungeilges and Neck (2001), for example, investigated the preferences of students in several countries for the distribution of economic resources. In this survey, respondents made pair-wise choices between distributions \((x, y)\) that differed with respect
to the degree they satisfied basic needs, addressed a utilitarian concern for the welfare of largest number, encouraged economic growth or infringed on human rights. The authors relate many of the patterns of cross-country differences in judgments to the circumstances of individuals in those nations. For instance, students in several economically depressed formerly socialist countries exhibited a greater concern for economic growth and promoting more advantaged persons than for helping the needy or protecting human rights. Alluding to the Rawlsian impartiality construct, they conclude that it is “obvious that people do not put themselves under a ‘veil of ignorance’ but consciously take into account their personal interests when making judgments” (pg. 963).

The above study drew on student respondents and asked them which distribution “should” be chosen. Another set of studies examined non-student populations and used explicit fairness terminology. Kahneman, Knetsch and Thaler (1986) presented randomly sampled telephone respondents in Toronto and Vancouver with hypothetical scenarios and asked them to evaluate whether or not the transaction terms in each of those questions were fair. Gorman and Kehr (1992) posed most of these same questions (22, counting contrasting versions of some scenarios) to a sample of business executives. Although the majority response (acceptable or unfair) in the two samples was the same for most of the questions, the percentage of respondents judging transaction terms unfair differed significantly across the two groups of respondents for 82% of the questions. In every one of those cases, this difference was associated with the business executives favoring the interests of the firm over those of the consumer or worker, relative to the more general population.

The study of Schokkaert and Capeau (1991) provides more thorough evidence on subject attributes that permits interpretations of personal bias. They elicited views of a representative sample of the adult Belgian working population on the fair compensation to workers in several hypothetical scenarios and collected information on various socio-economic characteristics of the respondents. Also, their survey contained five to seven quantitative response categories per question representing progressive differences in distributions as opposed to the more limited
qualitative response formats of the Gaertner, Jungeilges and Neck (distribution $x$ or $y$) and
Gorman and Kehr (acceptable or unfair) studies. Certain patterns emerged, such as the tendency
of blue-collar workers to reward direct contributions and of older people to defend seniority
rights. Schokkaert and Capeau conclude that “There are clear indications that the self-interest of
the respondents influences their answers” (pg. 342), although they do not find this effect large.

The previously cited studies examined distributive preferences of subjects over the
hypothetical allocations of others and permit one to associate these with personal characteristics
of respondents. Those results suggest that subjects respond at least somewhat vicariously,
identifying with one party or another. Messick and Sentis (1983), on the other hand, directly
made subjects hypothetical stakeholders and examined their biases. Specifically, student subjects
were instructed to “imagine” that they worked a different number of hours (7 or 10) grading
exams for a professor. The fairness wording was explicit, and the response format continuous.
These studies reveal significant biases in the pay deemed fair. In particular, the use of fairness
rules is self-serving: subjects presumed to work 7 hours were more than twice as likely to
employ equal pay than those working 10 hours, whereas those who worked more hours were
more likely to apply proportionality of pay to hours worked. This self-serving use of equality and
proportionality is similar to that predicted by Komorita and Chertkoff (1973) for coalitions.

In the above studies, respondents were asked to judge hypothetical scenarios with
hypothetical stakes. An important issue is whether these results generalize to real stakeholders in
the real world. As Messick and Sentis ask, “Could it be that the egocentric fairness bias is a
phenomenon that is manifested only when people are filling out questionnaires about
hypothetical situations and hypothetical payoffs?” (pg. 76). The study of Babcock, Wang and
Loewenstein (1996) helps shed light on this matter as well as on the question of whether even
experienced negotiators fall victim to the egocentric bias. They surveyed union presidents and
school board presidents in Pennsylvania regarding what school districts they would consider
comparable for purposes of teacher salary negotiations. The average salary in districts chosen by
the unions was significantly greater than the average in districts chosen by school boards. This
difference was also positively correlated with the variance in teacher salaries in neighboring districts, implying larger variation in comparison values provides more opportunity for egocentric evaluations. Moreover, districts in which differences in the two parties’ estimates were greater were much more likely to have experienced strikes in the past, which suggests the bias leads to disagreements with real social and economic costs.

The Babcock, Wang and Loewenstein study helps address the concern of Messick and Sentis about hypothetical situations. But the decisions of the participants in that study did not affect any real payoffs, which is one reason paid experiments have figured prominently in much of the fairness literature. One design, called the ultimatum game, has generated an especially large volume of experimental research on the topic. In particular, the publication of the first experimental results of the ultimatum game by Güth, Schmittberger and Schwarze in 1982 was probably the single most important impetus to stimulating (belatedly) the interest of large numbers of economists in fairness and social preferences, in general.

The ultimatum game is a simple bargaining game that proceeds as follows. The first decision maker, or the first mover, is given a fixed sum of money, say $10, and is anonymously matched with a counterpart, the second mover. The first mover proposes a division of the sum to the second, e.g., $7 to the first and $3 to second. If the second mover agrees, the money is divided as proposed, whereas if the second rejects the proposal, both receive nothing. The ultimatum game possesses several properties that are very useful for evaluating the behavioral predictions of alternate theories. It is the simplest of bargaining games, since there is only one round of decision-making and only one decision per person, yet it is truly a game since the players interact strategically. Consider the predicted outcome if players are rational and self-interested. The relevant game theoretic concept here is subgame perfect equilibrium (Selten, 1965), which solves optimal moves beginning with the final one(s) and working backward to the first (a process known as backward induction). For the ultimatum game, we begin with the second mover, who would accept the smallest offer, say, 1 cent (being rational, self-interested and not at all concerned with fairness), since 1 cent is more than nothing if he rejects. Indeed, the
second mover is indifferent about whether to accept nothing and might agree to that, as well. Knowing this, the first mover (also being rational and self-interested) makes the minimal offer to the second.

In fact, Güth, Schmittberger and Schwarze (1982) found that many first movers made generous offers and many second movers rejected small offers. In a recent book, Camerer (2003) summarizes numerous experimental results on bargaining, including ultimatum games. He reports that first movers in that experiment offer an average of 40% of the stakes and that second movers reject offers of less than 20% about one-half of the time. Although these findings have been widely attributed to a taste for fairness, some contributors have stressed the importance of self-interest. Pillutla and Murnighan (1995, 2003), for example, argue that ostensibly “generous” offers are motivated largely by a self-interested desire by first movers to avoid rejection by second movers. Others have voiced this same suspicion, beginning with Güth et al. (1982).

Nevertheless, self-interest does not explain the consistent rejection of meager offers by many second movers. Moreover, it does not account for the results of another experiment, the so-called dictator game. In this experiment, the first mover, called the dictator, is given a sum of money. The dictator may share any amount, or nothing, with an anonymous counterpart, whereby the counterpart has no recourse. Of course, a self-interested first mover should give nothing, but Camerer reports that dictators share on average around 20-25% of their money. Nevertheless, Dana, Weber and Kuang (2004) are skeptical that transfers in the standard dictator game reflect a true preference for fairness. They employed a binary version of the dictator game in which dictators could choose either allocation A, which always gave them $6, or allocation B, which always gave them $5. There were two possibilities for the payoffs of their counterparts: under one, A was the fairer choice, whereas under the other, B was fairer. Initially, dictators did not know which set of counterpart payoffs was in effect, but they could choose for that information to be revealed at no cost to them. Interestingly, 44% of dictators chose for the information about counterparts’ payoffs not to be revealed. There was also an increase in the percentage of choices that maximized the dictator’s payoff (A) in comparison to a baseline.
treatment where the information about counterparts’ payoffs is not initially hidden. Dana, Weber and Kuang conclude that this finding conflicts with a preference for fairness since dictators could costlessly choose to receive information that would help them achieve fairer outcomes. Instead, dictators prefer to remain ignorant and to take, on average, a larger payoff for themselves.

Actually, this finding does not necessarily conflict with a preference for fairness. In fact, given that fairness preferences are usually formalized as inequity aversion, it is very much consistent with it. Fehr and Schmidt (1999), for example, propose a simple linear preference model of agents who are averse to any deviation from the fair allocation, which, in their model, is taken to be an equal split. They apply this model to reconcile a number of ostensibly contradictory experimental findings. Konow (2005) proposes a model of preferences in a more general (nonlinear) functional form, whereby the fair amount is also variable and not restricted to equality. This model is then used to explain the results of dictator experiments, including how variations in fairness account for differences in dictator transfers across treatments. In the case of both models, though, inequity aversion never makes agents better off – at best, it leaves them the same as agents with no inequity aversion, and usually it makes them worse off. Thus, it is quite reasonable to expect that dictators in the Dana, Weber and Kuang experiments would prefer to remain ignorant of possible inequities caused by their actions, except, perhaps, for those subjects whose choices are more strongly directed by moral principles than by a concern for their own happiness. Despite some healthy skepticism about fairness, though, most researchers in this field would probably agree with Fehr and Schmidt who, citing both experimental and field studies, declare that “By now we have substantial evidence suggesting that fairness motives affect the behavior of many people” (pg. 817).

---

2 This inference is also consistent with another finding in Konow (2005). In that study, one treatment was a standard dictator game where dictators shared $10 with anonymous counterparts, whereas another was a control treatment in which dictators were given $10 but were prevented from sharing any of this with their counterparts. The change transient affect of dictators was assessed in each treatment using mood measures. The average mood change was more favorable for dictators in the control treatment, who had no opportunity to share with their counterparts. Among dictators in the standard treatment, though, fair dictators were happier than unfair dictators. Together with the Dana, Weber and Kuang study, this suggests that people are, on average, happier not knowing about or being able to act on fairness. Nevertheless, these results also demonstrate that when people are able to act and understand the consequences of their actions, many, in fact, will choose to be fair and are happier than those who are not fair.
The experimental conditions of ultimatum and dictator games have been varied to examine possible effects on subject behavior. Camerer (2003) identifies five categories of variables, including methodological, demographic and descriptive variables. But he concludes that two variables have the biggest effects: cultural and structural variables. With respect to the former, behavioral differences have been attributed to cultural differences in subjects. For example, in a widely cited study, Henrich et al. (2001) conducted the ultimatum game in 15 small scale societies, i.e., with villagers, farmers and herders in developing countries. They found that mean first mover offers in the ultimatum game ranged from 26% for a Peruvian tribe to 58% for an Indonesian group, although modal offers for most groups were in the 40-50% range. These results are striking, but, as Camerer points out, cross-cultural comparisons raise a number of methodological questions. In particular, I believe we do not yet know what experimental results tell us about the underlying preferences of people who are unfamiliar with experiments. As Henrich et al. write, “A plausible interpretation of our subjects’ behaviors is that, when faced with a novel situation (the experiment), they looked for analogues in their daily experience … and then acted in a way appropriate for the analogous situation” (pg. 76). Thus, it is possible that the results reflect cultural differences how the experiment is interpreted rather than cultural differences in the underlying social preferences.

Finally, structural variables include subject identity, entitlement, competitive pressure, information and intentions. Of his five categories of variables, Camerer (2003) believes that “structural variables are the most useful to study because they connect simple games to richer economic structures (e.g., adding competition) and also provide the most direct clues to the psychology underlying social preferences” (pp. 75-76). An important variable in the studies examined in the following sections is the entitlement, i.e., the outcome to which an agent has a right according to unbiased fairness. In particular, in this paper, I define unbiased fairness as do Handgraaf, Van Dijk and De Cremer (2003) using equity theory: fair outcomes (i.e., entitlements) are proportional to inputs, and when no design feature indicates any inequality of inputs (as is the case in many experiments), fair outcomes are equal. I extend this second part
slightly to state that fair *expected* outcomes are equal, when actual outcomes are probabilistic rather than certain. The studies considered in this paper include ones with both equal and unequal entitlements, according to this definition.

Fairness as equality is a more complex and subtle concept than it first appears, so it is worthwhile to clarify how it is used here. I have argued elsewhere (e.g., Konow, 2003) that, although equality in some sense is related to justice, equality of outcomes is not a general principle of justice. The claim is that in a world of perfectly informed agents with unlimited cognitive abilities, equality might sometimes emerge as a *special case*, but never as a *general principle*, of justice. For example, applying equity theory, fair outcomes are proportional to inputs; if inputs happen to be equal, so will be fair outcomes, but proportionality, not equality, is the general principle. Nevertheless, the argument is that, in the absence of information to evaluate fairness, equal outcomes are fair. Using again the equity principle as an example, if there is no information about the inputs of agents, equal outcomes are deemed fair.³ This finds substantial support from numerous contextually simple experimental studies, including from ultimatum games in which modal first mover offers are equal splits, e.g., Forsythe et al. (1994), Roth et al. (1991), and the Hoffman et al. (1994) replications of Forsythe et al. When, however, subjects are provided with information relevant to fairness, more departures from equality are typically observed. For example, Hoffman et al. (1994) found that dictators shared significantly less with counterparts if they first earned the right to be dictator by winning a contest. Later in the paper we will explore the possibility that such behavior is magnified by fairness biases.

Of course, there are many important contributions to the literature on the topics addressed in this paper, which, in the interests of brevity, cannot be treated here. The studies reviewed in this section are offered as exemplars of various methods and findings. To put the discussion in sharper focus, a more select set of studies are analyzed in greater detail in the following sections.

³ This differs from Messick’s (1993) treatment of equality as a “decision heuristic.” This is an important use of equality but is quite different from the current one. As a decision heuristic, equality helps to cope with situations that are *too complex*, e.g., agents face too much information to process all of it effectively and, instead, employ a simple rule. In the current instance, on the other hand, equality is a means of dealing with situations that are *too simple*: agents do not possess enough information to calculate entitlements using general principles.
These were chosen because, unlike other studies, they satisfy all of five criteria that are useful given the specific topics of this paper and the goal of maintaining comparability across studies in terms of experimental methods and presentation of results. First, they all involve salient rewards, i.e., the stakes are not hypothetical (or paid to a random fraction of subjects) but rather involve real monetary rewards, and individual payoffs depend in an important way on subject decisions. Although studies with and without real stakes often lead to similar conclusions, some results are affected by the presence of real stakes, such as the average level of unfairness. The monetary stakes also permit an unambiguous, interpersonally observable measure of different motives. Second, the stakes in each of the studies are fixed, which helps to focus the attention on equity issues, as opposed to preferences over the size of surplus. Third, the stakes are allocated between only two parties. This limits the role of more complicated strategic considerations that might confound inferences about justice. Fourth, the formats for decisions are continuous, or nearly continuous. Thus, allocations can be clearly expressed and interpreted on the unit interval, which permits comparative analysis of bias and dispersion across studies. In this paper, bias is measured by differences in mean allocations. Dispersion is evaluated by differences in standard deviations. Finally, stakes and/or information are varied across treatments in a way that permits analysis of their effects on both means and standard deviations. In addition to the more detailed evidence from these studies, corroborative results are also cited from various additional studies.

**STAKES AND FAIRNESS**

This section explores questions of how the willingness of agents to act on fairness varies with stakes. Does it matter whether the stakes are real or hypothetical? What effect does the size of stakes have? How do variations in the dictates of self-interest affect claims on stakes? For the contextually simple experiments in this section, which lack information about inputs, the working definition of fair shares corresponds to equal splits.

**Effects of Variation in Stakes on Claims**

Forsythe, Horowitz, Savin and Sefton (1994) conducted three treatments of the dictator game in a between subjects design: 1. Hypothetical $5 Dictator, 2. Actual $5 Dictator and 3.
Actual $10 Dictator. In the first treatment, dictators were asked how much of $5 they wished to share with their counterpart, but no one was actually paid the chosen amounts. The second treatment was similar, but dictator decisions resulted in actual payments and subjects were told that. The third treatment was the same as the second, except dictators were given $10 to divide.

The results are summarized in the first rubric of Table 1. Results reported in this paper are normalized: the means and standard deviations are expressed as a fraction of the total stakes. These values refer to claims by the party that is typically advantaged in the sense of being the party that, in most treatments of the experiment, is able to claim more than one-half. Here, of course, that is the dictator, who is the first mover (indeed, in this experiment, the only mover). We see that, when the stakes are not salient and subject decisions do not result in real payments, dictators indicate on average that they will take 61.7% of the (hypothetical) $5. The superscript 0 on the mean of .617 indicates that this mean differs from equal splits of .5 at the 5 percent level of significance. Thus, the result from treatment 1 shows that subjects yield, to some degree, to self-interest and are willing to acknowledge their unfairness, even when there is no material benefit to doing so. Nevertheless, hypothetical dictator decisions underestimate the degree of unfairness exhibited by real stakeholders, as the second treatment demonstrates. When the $5 payments are real, dictators choose to keep 77.8% of the stakes. This differs significantly from .5, again denoted by the superscript 0 on the mean, and from the mean in treatment 1, as indicated by the superscript 1 on the mean of .778 (as well as by the superscript 2 on the mean of .617 in the first treatment). Raising the stakes from $5 to $10, however, has no significant effect on dictator claims, given the mean of .767 in the third treatment. This mean differs significantly from .5 and from the mean in the first treatment, but not from the mean in the second treatment. In addition, all of the means reported in this and all other tables in the paper differ significantly from the claim of 1 that corresponds to pure self-interest. These results suggest that, on average, people trade off self-interest with fairness and do not choose completely self-interested or completely fair allocations. In addition, hypothetical dictator decisions underestimate the magnitude of unfairness, but $5 and $10 stakes produce similar claims.
The next column in Table 1 lists the standard deviations of the claims. These do not differ significantly based on the stakes, suggesting that dispersion is not affected by the magnitude of stakes involved. The final two columns show the percentage of claims that can be categorized as equal (.5) or self-interested (1). Using real versus hypothetical stakes causes the percentage of equal, or fair, claims to drop significantly from 48% to around 20% in treatments 2 and 3. Using real $5 stakes causes the percentage of completely self-interested claims to rise significantly from 13% to 36%. The increase in completely self-interested claims from 13% with hypothetical stakes to 21% using $10 stakes is not significant. If, however, one includes claims of .9, which are close to pure self-interest, then these claims rise to 37% in treatment 3 whereas the other percentages stay the same. Thus, the percentage of approximately self-interested claims with $10 stakes is significantly greater than that with hypothetical stakes but the same as with $5 stakes.

Forsythe et al. also explored the effects of stakes on the ultimatum game. Because of the structure of the ultimatum game, the first mover has more power than the second, but, as reported in the previous section, empirical tests suggest this is less than the absolute power in the dictator game: first movers typically propose positive amounts, and second movers often reject small positive offers. Moreover, the rejection rate usually rises, as the size of the offer falls. Although these results depart from the predictions of pure self-interest, fairness alone cannot account for them either. A purely fair-minded second mover would reject any proposal other than equal splits, even if the inequality favored him or her, and a fair-minded first mover would only propose equal splits. Thus, this “fairness hypothesis” also produces extreme predictions.

Forsythe et al. conducted ultimatum games in a between subjects design for the same three levels of stakes as their dictator games. The results are summarized in the second rubric of Table 1. The mean first mover claims do not differ significantly from one another in the ultimatum games (characterized by the absence of superscripts 4, 5 and 6 on those means), indicating, in contrast to the dictator games, that hypothetical stakes yield the same results as actual ones. Comparing claims across the dictator and ultimatum games, however, one does observe differences. Mean claims and the percentage of self-interested claims are significantly
less in the paid ultimatum games (5 and 6) than in the paid dictator games (2 and 3), indicating that subjects do respond to the strategic differences in the two games. Subjects do not exhibit such a difference, however, when stakes are merely hypothetical (i.e., 1 versus 4). The closeness of mean claims in the ultimatum game is related to the high percentage of equal claims in that case, in fact, hypothetical first movers appear to underestimate their tendency with real stakes to offer equal splits. Mostly, differences in stakes do not matter for the ultimatum game, although the jump from zero to $10 stakes does significantly raise the percentage of equal split claims and reduce the variance in claims. Mean claims, however, are not significantly affected by the level of already positive stakes.

Pure self-interest and pure fairness produce extreme predictions for the dictator and ultimatum games. A more reasonable assumption about motivations is that agents care about both their own interests as well as fairness. This explains the tendency of second movers to reject small offers and of first movers to propose generous splits that, nonetheless, often favor themselves. In the ultimatum game, first mover fairness is reinforced by two considerations: their preference for fairness as well as their self-interested concern that overly selfish claims will be rejected by fair second movers. Thus, it is not surprising that first mover claims are fairer in the ultimatum game than in the dictator game.

The results of the Forsythe et al. experiments suggest that even small stakes yield results equivalent to those with larger stakes. It might be the case, however, that the stakes typical of most experiments are still insufficient to be indicative of behavior when significant economic rewards are at stake. Hoffman, McCabe and Smith (1996) decided to sink some serious money into examining this question using the ultimatum game with stakes of $10 in one treatment and of $100 in another. The results of their study are reported in the bottom rubric of Table 1. Mean claims by first movers do not differ significantly when stakes are raised from $10 to $100. In fact, these means are not significantly different from those in the paid Forsythe et al. ultimatum games. The $10 and $100 ultimatum games of Hoffman et al. are also indistinguishable in terms of the percentage of equal split claims and self-interested claims. The standard deviation is
significantly greater with $100 stakes, but even this disappears if a single outlier is eliminated: one observation actually corresponds to pure self-interest with the first mover claiming the whole $100 and the second mover consenting. Collectively, the findings from the studies above imply Results 1 and 2.

**Result 1:** For a given incentive structure, even small real stakes generate claims equivalent to those on large real stakes.

**Result 2:** Strategically advantaged agents claim, on average, more than the fair share but less than their “self-interest share,” i.e., their optimal claim if all agents were self-interested and rational.

Result 1 is reassuring regarding our choice to analyze results based on shares as opposed to absolute levels. Result 2 suggests a trade-off between fairness and self-interest, which we will explore further in the following section.

**Effects of Variation in Self-Interest Shares on Claims**

Differences in the claims of first movers in dictator versus ultimatum games in the Forsythe et al. study suggest that the willingness of people to act on fairness is related to their strategically optimal shares as prescribed by self-interest. The bargaining experiments of Spiegel, Currie, Sonnenschein and Sen (1994) provide more thorough evidence on this relationship and on the trade-off between fairness and self-interest. Among the experiments reported in that paper is one involving two rounds of bargaining over $6 stakes. In the first round, the first mover (denoted F) proposes a division of the $6. If the second mover (denoted S) agrees, the money is divided as proposed. If S rejects the proposal, however, the stakes shrink, and now S can make a proposal to F. If F accepts, the shrunken stakes are divided as proposed, and if not, both players receive nothing.

Table 2 summarizes the stakes and results for this experiment. In Game 1, the first round stakes are $6, but if S rejects the proposal of F, the second round stakes shrink to $0. That is, Game 1 reduces to an ultimatum game, and, as previously discussed, a self-interested F claims the whole amount (or that less 1 cent), which is denoted in the “Self-Interest Share” column of
Table 2 as 1.00. In Game 2 the second round share shrinks to $1. One can solve the optimal claim of the first mover from this game by backward induction, i.e., beginning with the last possible move. In the second round, the stakes are $1, and S proposes a division of this to F. The final move is F’s choice to accept or reject this proposal. Being self-interested, F will accept 1 cent and, possibly, nothing. Knowing this, S will propose 1 cent or nothing. The second round is essentially an ultimatum game with $1 stakes where S is the first mover. Thus, S can claim $1 in the second round, leaving nothing for F. Working back to the first round, F knows S can claim $1 in the second round and, therefore, offers $1 to S in the first round so that F can secure $5 in the first round rather than nothing in the second. Thus, F’s optimal claim in the first round is $5, or .833 of the $6 stakes, leaving $1 for S. Games 3 through 6 progressively increase the second round stakes, which S can claim in that round. Solving similarly by backward induction, one can find F’s optimal first round claims, which, as we see in the Self-Interest Shares column, decline as the second round stakes increase.

Table 2 illustrates that the relationship of actual claims to self-interest shares is (weakly) monotonic, viz., as the first mover’s optimal self-interest shares decrease, claims by the first mover fall. The first mover responds, then, to deteriorating bargaining power by reducing claims, consistent with self-interest, leading to this result:

**Result 3:** Actual claims are moderated by fairness but also respond to changes in self-interest shares, suggesting a well-behaved trade-off between fairness and self-interest.

This is also consistent with both the implication of dictator-ultimatum game comparisons above and with the results of Andreoni and Miller (2002), who vary the price of giving in dictator games, and find that the preferences of most agents are well-behaved, i.e., consistent with the standard economic model adapted to incorporate social preferences. Also, as the standard deviations reveal, there is the least dispersion in first mover claims when the self-interest shares of both parties converge to fair shares, summarized as follows:

---

4 Claims in Games 1 and 2 are the same: perhaps the failure of Game 1 claims to be greater reflects the greater salience (to both F and S) of fairness over strategic self-interest in this game given the single round of bargaining.
**Result 4:** Dispersion in claims rises with the degree of conflict between self-interest and fair shares.

Evidence from numerous dictator games implies people differ widely with respect to their relative preference for fairness. Given that fact, one can expect that, when the tension between self-interest and fairness is greater, the dispersion of claims will be greater. In the Spiegel et al. study, the standard deviations of claims are smallest when self-interest and fairness correspond to similar claims (Games 3, 4 and 5) and greatest when they do not (Games 1, 2 and 6).

Nevertheless, only in the case of Game 3 does a majority of first movers claim the self-interest share, and in Games 1 and 6, no claims coincide with pure self-interest.

There is a noticeable break in several statistics in comparing the first three games, in which the first mover is strategically advantaged, with the last three, in which he is not. One is the significant difference in mean claims between the first and second three sets of games. Another is the statistically significant difference in the percentage of equal claims between the two groups. A final pattern is that the first three games do not result in disagreement (except for two pairs in Game 1), whereas the last three games are significantly more likely to end in disagreement. The percentages of equal claims and of disagreements for these games are illustrated in Figure 1. Thus, in Games 1 through 3, first movers tend to claim unequal shares that favor them, whereas in Games 4 through 6, they are significantly more likely to claim equal shares, suggesting this result:

**Result 5:** When multiple allocation rules are salient, many people employ them in a self-serving fashion, choosing the one that yields a greater share for themselves.

This replicates with real stakes the finding of Messick and Sentis that advantaged subjects claim larger shares whereas disadvantaged ones appeal to equality. When the first mover is advantaged, claims of equal splits are unusual, but when self-interest shares call for equal shares or less, around one-half of first movers suddenly claim equal splits.

**INFORMATION AND EGOCENTRIC BIASES**

The results of the previous section suggest that information about stakes, self-interest and
fairness influence fairness biases and the willingness of agents to behave fairly. This section examines evidence on fairness biases from studies that, in contrast to the prior ones, specifically vary information about facts relevant to self-interest and fairness. We first review and analyze the studies of Roth and his collaborators, which reveal how varying information relevant to fairness affects claims and outcomes. Then we examine studies that manipulate stakes or information about stakes. These confirm fairness biases and suggest that self-deception is an important cause as people process information in a self-serving manner.

**Variation in Information Relevant to Self-Interest and Fairness**

Roth and his colleagues Malouf and Murnighan conducted a series of bargaining experiments that are especially helpful for illuminating the relationship between information and fairness. These studies shared a number of common design features: Subjects were paired anonymously, but bargaining was much less structured than in the previously discussed experiments. A distinctive feature of this design was that subjects did not bargain directly over stakes but rather over lottery tickets that determined their probability of winning a prize. That is, a subject with, say, 60% of the tickets had a 60% probability of winning his prize. In almost all treatments, these prizes differed across pairs, e.g., one player had a $3 prize whereas the other player had a $9 prize. These studies, then, precluded the equal outcomes generally considered fair in such experiments. Since no other fairness relevant information was provided in these studies, fair shares were associated with equal expected outcomes, which could be achieved by allocating fewer tickets to the player with the high prize. This would occur, for example, if a $3 player received 75% of the tickets and a $9 player 25%, creating equal expected values of $2.25. Thus, equality in the variable being allocated (tickets) corresponds here to a different allocation from equality in the variable players value (expected earnings). Subjects always knew their own prizes, and information was varied across treatments in these experiments in terms of what

---

5 Specifically, players were given 10 or 12 minutes, depending on the study, to communicate and come to an agreement via computers, or to earn nothing. Any messages were allowed, except those that identified the subject, proposals were binding on the sender, and bargaining ended when a subject accepted a sender’s proposal or when time ran out.
subjects knew about the value of their counterparts’ prizes.

In Roth and Malouf (1979), players bargained over lottery tickets under one of two conditions. In the full information condition, both players knew the values of both prizes. In the partial information condition, each player knew only his or her own prize, but not that of the counterpart. Subjects played four games in random order under just one of the two information conditions. This is summarized in the top two rubrics of Table 3. In Games 1 and 2, both players had equal prizes of $1.00, whereas in Games 3 and 4, one player had a prize of $1.25 (the Low Prize Player) and the other of $3.75 (the High Prize Player). Games 1 and 2 (3 and 4, respectively) differed in inconsequential ways for our purposes, so their results are collapsed in Table 3.\(^6\)

Treatments 1 and 2 in Table 3 deal with Games 1 and 2, where both players had equal $1.00 prizes. Taking the fair division of tickets to be the one that produces equal expected payoffs, this gives each player, of course, one-half of the tickets, indicated in the Fair column (for consistency, all results in Table 3 are expressed for the High Prize Player, although, in this case, there is no difference). The mean share of tickets actually agreed upon does not differ significantly from .5 in either information condition. In fact, the two information conditions also do not differ significantly in terms of mean, variance, percentage of equal and fair shares (which are the same, in this case), or percentage of bargaining encounters that led to disagreement.\(^7\)

These experiments lack the explicit asymmetry in bargaining power present in the previously discussed games, and one finds that a high percentage of outcomes are fair. The increased information in treatment 2 versus treatment 1 appears to have mostly beneficial effects by increasing the fraction of fair agreements and reducing disagreements (although variance also

---

\(^6\) Games 2 and 4 contained the additional restriction that the High Prize Player was not permitted to receive more than 60% of the tickets. Although this constraint would have been binding on subjects pursuing equal expected values if it had been placed on Low Prize Players, it was not so when placed on High Prize Players. Moreover, the only statistically significant difference Roth and Malouf report here is between Games 3 and 4 in the partial information condition.

\(^7\) Note that the Mean, Standard Deviation, % Equal and % Fair are for those pairs that reached an agreement, i.e., the first number in the N column, whereas the % Disagree is out of the total sample, i.e., the second number in the N column.
rises slightly), but these values are already so favorable even in the partial information condition that these improvements are not significant. Since there is no tension between allocating tickets equally and equalizing expected values, however, these treatments are not useful for examining information effects when there are conflicting fairness rules. Therefore, in all of the discussion that follows, I will refer only to the results for the other treatments by Roth and his colleagues in which subjects may appeal to differing rules.

The unequal prizes in treatments 3 and 4 of Roth and Malouf imply that expected payoffs are equalized when the High Prize Player receives 25% of the tickets. When each player only knows his own prize in the partial information condition, however, mean shares do not differ significantly from .5, and 69% of agreements result in equal splits of tickets. If both players are fully informed about prize values, however, there is a significant shift in mean ticket shares toward fair shares, as evidenced in Figure 2. The percentage of agreements with equal shares drops from 69% to only 17% while the percentage with fair shares rises from 0% to 22%, both significantly. On the other hand, increasing information does not appear to reduce dispersion, as the standard deviation rises in treatment 4 versus 3 (and 1 and 2), and the percentage disagreements rises, although not significantly. We will see that precisely these patterns of effects from increasing information repeat themselves in other studies.

Roth, Malouf and Murnighan (1981) built on this work by increasing the average stakes, the number of different prize values, and the number of levels of information. In the high information condition, subjects were able to calculate fair shares, whereas the information in the intermediate and low information conditions did not permit them to do so. Since mean shares in the latter two conditions did not differ significantly, we combine them. The results of this

---

8 Specifically, each subject played four games with differing prizes under just one of the three information conditions. As in Roth and Malouf, players bargained for lottery tickets, but here the tickets were expressed in terms of an intermediate commodity, viz., the prize was a certain number of chips that had a certain monetary value. The number of chips, the value per chip and the total prize (the product of the two previous values) differed across the four games. In every treatment, each player knew his own number of chips, values per chip and total prize, but the information each player was given about his opponent’s prize was varied across conditions. In the high information condition, each player knew the number and value of chips in his opponent’s prize. In the intermediate information condition, each player knew the number of chips in his opponent’s prize, but not the value per chip. In the low information condition, neither player knew the number or value of chips in his opponent’s prize.
experiment are summarized in the middle rubrics of Table 3. Each subject played the high and the low prize position in Games 1 and 3, but since the prizes were the same for these games, these results have been consolidated. Note that Games 2 and 4 of this study require more extreme differences in ticket shares to equalize expected values, as can be seen from the Fair column. Comparing the Low/Intermediate and High information conditions for each treatment, one sees that increasing information causes mean shares to move toward equalizing expected values, variance to increase, percentage equal shares to fall, percentage fair shares to rise (except treatments 3 and 4, which are the same) and percentage disagreements to rise. These are the same patterns observed in the Roth and Malouf study and are generally significant at the 5% level for means and standard deviations.

The procedures of a third study by Roth and Murnighan (1982) were similar to those of Roth and Malouf (1979), with a few crucial differences. The prizes were always equal to $5 and $20, and there were four information conditions with respect to what each subject knew about the other’s prize: 1. Neither player knows his opponent’s prize, 2. Only the player with the $20 prize knows the other’s ($5) prize, 3. Only the $5 player knows the other’s ($20) prize, and 4. Both players know each other’s prize.\textsuperscript{9} Considering the results in the bottom rubric of Table 3 for each of the information treatments progressively from 1 to 4, the mean share moves away from equality and toward fair shares, as the percentage of equal shares falls. Nevertheless, the tests of significance point toward a break between the first and second two sets of treatments. The mean share when only the $20 player knows the other’s prize does not differ significantly from the case in which neither knows. But when the $5 player knows the other’s prize (i.e., that it is $20), there is a significant shift toward equal expected payoffs. Moreover, this is not significantly

\textsuperscript{9} There was one other treatment variable consisting of two conditions pertaining to common knowledge, resulting in a 4 information \times 2 common knowledge between subjects design. Under \textit{common knowledge}, both players are told what information is available to whom. For example, under common knowledge, both the $20 and $5 players in treatment 2 above know that the $20 player knows both prizes but that the $5 player only knows his own prize. Under \textit{non-common knowledge}, it is still the case that the $20 player knows both prizes and the $5 player knows only his own prize, but both are told that the other might or might not know their prize. In any case, mean shares do not differ significantly across common knowledge conditions for any of the four information treatments, so these results are combined.
different from the case in which both know both prizes. This suggests that providing information relevant to fairness does not strongly motivate the advantaged player to act on it, but it does cause the disadvantaged player to work toward fairer outcomes. In fact, the lack of significant differences between treatments 3 and 4 by any measure suggests that fair outcomes in this study are due solely to the efforts of the disadvantaged party to defend its fair share. This is consistent with evidence Güth, Huck and Ockenfels (1996), Güth and van Damme (1998), and Straub and Murnighan (1995) provide from modified ultimatum games that first movers make significantly lower offers when the second movers do not know the size of the stakes and are less likely, therefore, to reject small and unfair shares. In terms of dispersion, the only significant difference in variances is between treatment 1 and all others. Also, the only significant difference in the rate of disagreements is between treatment 1 and one other (treatment 3). The implication of these results is that information that introduces alternate fairness rules (equal expected values) increases dispersion, but additional information has no significant impact on dispersion.

The general conclusions one can draw from the studies of Roth and his colleagues are the following. Increased information tends to reduce bias and sometimes to increase the percentage of fair shares, at least, if the information is sufficient to calculate fair outcomes. On the other hand, additional information seems to have an unfavorable effect on dispersion: standard deviations of shares rise as do rates of disagreement, although the latter is usually not significant. The increased variance is associated with information that introduces conflicting allocation rules.

The foregoing result on dispersion, however, is misleading: if parties genuinely value fair outcomes, here fair expected shares of earnings, then shares of tickets are merely an intermediate means of achieving the more important goal. In that case, dispersion should be evaluated in terms of the earnings, or rather expected earnings, that the bargaining process produces. These studies do not report the actual payoffs resulting for the randomly chosen winners, but one can easily calculate the distribution of expected payoffs from the product for each player of his share of tickets and his prize. The results of this exercise for each of the treatments are reported in Table 4. One sees that the mean expected payoff to the High Prize Player always moves toward
equality as information rises, consistent with the pattern of change in the shares of tickets. Calculating the expected payoffs to both players, the standard deviation always falls, sometimes significantly, with increased information. Figure 3 illustrates these results for treatments 3 and 4 of Roth and Malouf.

Although greater information is typically associated with an (insignificant) increase in the rate of disagreement in the studies above, it can also reduce disagreements, as demonstrated by the study of Camerer and Loewenstein (1993). They conducted ultimatum games with five levels of stakes: $1, $3, $5, $7 and $9. In the certain amount treatment, first movers indicated how much they would offer, and second movers indicated the minimum offers they were willing to accept, in both cases for each of the five stakes. Subjects were matched, one decision was randomly chosen for payment, and if the first mover’s offer was no less than the second mover’s minimum acceptable offer, the offer was paid out; otherwise both received nothing. The uncertain amount treatment, conducted with a separate group of subjects, was the same, except that only first movers knew the stakes and made five separate decisions for each case; the second movers stated a single minimum acceptable offer without knowing the stakes. In the certain amount condition, the means and standard deviations of first mover offers (calculated as shares of the total stakes) were very similar across the five stakes as were means and standard deviations of the second mover minimum acceptable offers. Moreover, the disagreement rates for all levels of stakes were very close to the overall average of 15%. In the uncertain amount condition, the means and standard deviations of first mover offers were again quite similar to one another as well as to those in the certain amount condition. But, since second movers did not know what level of stakes their minimum accept offer applied to, higher rates of disagreement resulted that, across all stakes, averaged 39%.

Thus, we conclude the following. In the face of countervailing bargaining power, the overall effect of increased information is to reduce bias in the claims settled (Table 3, Figure 2) as well as to reduce bias and dispersion in the outcomes, or expected outcomes, that people value (Table 4, Figure 3). Although information can supply fodder for conflicting claims, it is
beneficial (especially in the hands of the disadvantaged person) if it is relevant to fairness, i.e., information that is necessary to calculate fair allocations according to accepted norms. Result 6 summarizes this.

**Result 6:** Under symmetric bargaining, providing information relevant to fairness results in fairer mean expected outcomes, a higher percentage of fair expected outcomes and greater convergence (i.e., less disperse expected outcomes). Disagreement rates can be favorably or unfavorably affected, depending on the structure of bargaining.

**Egocentric Processing of Information**

We have seen evidence from numerous studies that information affects the fairness bias, which in turn can affect claims, allocations and disagreement rates. We turn now to studies that explore the link between information and this bias and that point toward a self-serving bias (i.e., self-deception) that is related to egocentric processing of information. These studies also provide a more realistic set of conditions than most laboratory experiments, including contextual richness and entitlements that, as in most real world contexts, do not generally correspond to equal splits.

Babcock, Loewenstein, Issacharoff and Camerer (1995) created a contextually rich laboratory experiment, in which each subject was first assigned to the role of plaintiff or defendant in a legal dispute. Subjects read 27 pages of testimony abstracted from an actual case and were informed that a judge had read the same materials and had decided how much, if anything, to award to the plaintiff. Then, they provided their judgments of 1) what they considered a fair settlement, and 2) their best estimate of what the judge’s award would be. Finally, the parties had 30 minutes to reach an agreement (the division of the experimental stakes of $10 were in proportion to the agreement), or the judge’s decision was imposed, less legal costs. This describes the condition with *ex ante stakes*, where each subject knew his role in the

---

10 This innovative approach to analyzing bias in bargaining made a break with two literatures. Previous theoretical research in law and economics suggested that cases fail to settle out of court because of random errors in litigants’ estimates of the value of going to trial, which implies that increasing information to the parties should reduce disputes. Previous experimental research on bargaining, on the other hand, usually strived for contextual simplicity, often reducing or eliminating information about specifics of the persons and conditions involved.

11 Subjects were paid a bonus at the end of the experiment for accurately estimating the latter value within a range. The plaintiff was suing for $100,000, and the settlement could lie anywhere between that and $0, so the defendant’s
dispute at the start. There was a second condition with ex post stakes, which was identical, except that subjects were not informed whether they were the plaintiff or the defendant until just before beginning negotiations, but after they had read the court materials and given their estimates of the fair settlement and the judge’s award. The idea is that egocentric processing of information begins when roles are known, and that manipulation of the timing of that information produces stable biased beliefs, when roles are known ex ante, and stable unbiased beliefs, when roles are revealed ex post. Thus, Babcock et al. predicted that the ex ante treatment, in comparison to the ex post treatment, would result in more biased judgments of the fair settlement and the judge’s award and lead to a higher rate of disagreement.

The results of this experiment are presented at the top of Table 5. The high stakeholder refers to the advantaged player, which is the defendant in this experiment. This is consistent with the judge’s award of .306 to the plaintiff, which leaves .694 for the defendant. One sees dramatic evidence of bias in the results for the condition with ex ante stakes. The high stakeholder has significantly higher estimates of what he should receive according to the fair judgment (.791) and judge’s award (.761) than the low stakeholder (.593 and .575, respectively). In the ex post stakes condition, these two estimates fall significantly for the high stakeholder (to .690 and .638) and rise significantly for the low stakeholder (to .753 and .707). The differences between high and low stakeholder estimates are significantly greater in the ex ante condition than in the ex post condition. Moreover, the ex post condition eliminates any bias in fairness judgments at conventional levels of significance. Finally, in comparing the ex ante and ex post agreements, the difference in agreed shares is borderline significant and the difference in disagreement rates is highly significant.

All of the Babcock et al. predictions are borne out: parties in the ex ante condition have more biased fairness judgments and are much less likely to reach a settlement. In addition, differences in fairness judgments were a significant (directly related) predictor of the time parties

\[\text{initial experimental endowment of $10 was allocated proportionately at a ratio of } 1:10,000.\]
took to settle and the probability of impasse. The results of this study are summarized below:

**Result 7:** *The self-serving bias of stakeholders is due to egocentric processing of information. This bias is an important and significant predictor of bargaining impasse.*

It is reasonable to ask whether the powerful findings of this seminal study are robust with respect to changes in experimental procedures.

Gächter and Riedl (2005) constructed a bargaining experiment, similar in several ways to Babcock et al., in which subjects were randomly and anonymously paired and asked to place themselves in the role of one of two department heads in a firm. The subjects were then told that, historically, the department head who scored better on a general knowledge quiz also performed better as head and, therefore, received twice the salary of the lower scoring head. Next, subjects took the quiz, were informed which of the two bargaining partners scored higher, and responded to a question about what they considered to be the fair distribution according to a neutral arbitrator. They then had 15 minutes to reach an agreement over 205 Austrian shillings (about US $18) through free-form bargaining or receive zero payoffs, if no agreement was reached. This was their condition with *ex ante stakes.* The *ex post stakes* treatment was identical, except that the fairness question was posed right after reading the instructions but before subjects knew whether they had scored higher on the test, indeed, before they even took the test.

Gächter and Riedl found that fairness judgments correlate significantly with a number of behaviors, including opening offers, bargaining duration, concessionary behavior and agreed shares. There was also an important *entitlement effect;* both parties responded strongly to the two-to-one salary frame of the experiment, which suggests the winner of the quiz was entitled to two-thirds of the earnings. The bottom of Table 5 summarizes this result, where the high stakeholder in this study was the winner of the quiz. In the *ex ante stakes* treatment, the fair share to the winner according to both high and low stakeholders was close to two-thirds. Nevertheless, as this table shows, the results on fairness bias and disagreement rates from this experiment are not as large or significant as in the Babcock et al. study. In the *ex ante* treatment, high stakeholders considered their fair share to be greater than did low stakeholders, but this
difference (.024) is not significant at conventional levels (p=.08). The difference in fair judgments between high and low stakeholders in the ex ante treatment is significantly different, however, from that in the ex post treatment. Gächter and Riedl also reported opening offers in the negotiations and found that these differed significantly for high and low stakeholders in both treatments. High stakeholders lowered and low stakeholders raised their opening offers slightly in the ex post treatment in comparison to the ex ante treatment, although these differences are not significant. Thus, self-interest is still at play in this study but not so prominently because of a self-serving bias. Finally, the disagreement rate is no higher (indeed, is lower) in the ex ante condition than in the ex post condition, and the authors found no evidence of a direct effect of fairness judgments on disagreement rates.

The Gächter and Riedl experiment provides more modest evidence of an egocentric bias and its impact on bargaining than the Babcock et al. study. In comparing the two studies, two important procedural issues stand out that might explain these differences. First, Babcock et al. presented their subjects with a much more lengthy list of facts than Gächter and Riedl. This provided a rich basis for choosing, interpreting, weighting and recalling facts in a self-serving fashion. As Dunning, Meyerowitz and Holzberg (1989) report, self-serving evaluations rise with an increase in the number of criteria at one’s disposal. In an earlier study employing mostly the same procedures as the Babcock et al. ex ante treatment, Loewenstein, Issacharoff, Camerer and Babcock (1993) asked parties after negotiation to list and rate the importance of arguments favoring themselves and their partners. Parties were significantly more likely to recall more arguments favoring themselves and to rate self-serving arguments as more important. Second, Gächter and Riedl explicitly presented subjects with a statement of fair shares, whereas Babcock et al. did not. As Gächter and Riedl themselves conjecture (pg. 19), this precedent apparently conveyed a strong sense of entitlement, even to the disadvantaged low stakeholders, who usually judged their own fair share as less than one-half. Comparing the various studies, we come to the following conclusion:
Result 8: The self-serving bias of stakeholders is enhanced by more ambiguous facts and is diminished by an independently determined estimate of fair shares.

These studies represent significant contributions to the literature on fairness biases and stimulate a number of unresolved questions. They rely on self-reports of fairness beliefs to measure self-serving biases, but these do not directly determine any actual allocations. In addition, their context is bargaining, which is appropriate for the types of situations their authors wish to examine, but the question remains whether such biases arise even when strategic interaction is not involved. For the purpose of reducing self-serving biases, withholding information about roles, although a useful mechanism for identifying these biases experimentally, does not represent an effective means of alleviating them in most real world situations where parties know their stakes, as these authors acknowledge. Finally, all of the foregoing experimental studies analyzed the behavior of individuals who were in the position of stakeholder. The last two studies addressed bias by manipulating the timing of information about roles. Another approach to identifying biased and unbiased beliefs about fairness is to manipulate the roles themselves.

In Konow (2000), I sought to address these issues, to generate additional evidence and to formulate a theory of these and other findings about justice. The theory integrates several concepts from social psychology. One is that fair outcomes are proportional to contributions, which is the main tenet of equity theory (e.g., Homans, 1958, Adams, 1965, and Walster, Walster and Berscheid, 1973). Another is attribution theory (e.g., Fritz Heider 1958, Julian Rotter 1966, and Bernard Weiner and Andy Kukla 1970), which asserts that, in evaluating contributions to outcomes, agents are only held responsible for factors they can control. The paper claims that the relevant justice principle in this context is the “accountability principle,” which merges these concepts. The accountability principle simply states that fair allocations are proportional to the contributions agents control (called “discretionary” variables) but do not adjust for factors they cannot influence (called “exogenous” variables). For example, suppose one worker is twice as productive as another and, therefore, earns twice as much. This allocation of earnings is fair if
the productivity differences are completely discretionary, e.g., if the former works twice as hard as the latter. It is not fair, however, if these differences are entirely exogenous, e.g., if both work equally hard but the latter is encumbered by poor work conditions; then, fair earnings are equal.

The paper presents the results of an experiment designed to test this theory. Paired subjects first generated earnings jointly by preparing letters for a mass mailing, after which one randomly assigned subject (the dictator) allocated the joint earnings from the first phase. One treatment variable concerned how earnings were generated in the first phase, and there were two possibilities. In the “discretionary differences” treatment, 50 cents were credited to the pair for each letter prepared, and the percentage of joint earnings attributable to one of the paired subjects varied because of differences across subjects in the number of letters prepared in the allotted time. In the “exogenous differences” treatment, all subjects had enough time to prepare an equal (and limited) number of letters, but the earnings associated with each subject differed because of arbitrary differences in the per letter credits across the two subjects in each pair. Another treatment variable, which was crossed with these two, concerned who allocated the earnings from the first phase. In the Spectator condition, the dictator was a subject from a third group, who did not prepare letters but was paid a fixed fee to allocate the joint earnings to each of two other subjects who did prepare letters. According to the accountability principle, spectators will allocate earnings in proportion to letters produced in the discretionary differences treatment but will allocate earnings equally across a pair in the exogenous differences treatment, since they will disregard the differences in the per letter credits that subjects do not control.

Figure 4 illustrates the results of the discretionary differences treatment. The horizontal axis shows the fraction of earnings contributed by one of the subjects in each pair (here, the fraction of letters produced), and the vertical axis represents the corresponding fraction of earnings allocated by the spectator to that subject. According to the accountability principle, fair

---

12 Specifically, the average credit per pair was still 50 cents, but if one subject’s credit was, say, 75 cents per letter, then the other subject’s credit was 25 cents per letter. Thus, although each subject produced one-half of the letters in his or her pair, monetary contributions differed.
allocations lie on the 45 degree line, and, in fact, the percentage of exactly proportional allocations is 79%. These results provide powerful corroboration of a strong prediction of equity theory for this case. Indeed, this was the first such evidence of the exact proportionality rule, to this author’s knowledge. Figure 5 illustrates the results of the exogenous differences treatment. The fractional earnings contributed by one subject that appear on the horizontal axis here are due not to differences in letters prepared, which were equal in this condition, but to the arbitrary differences in per letter credits. According to the accountability principle, fair allocations disregard these differences and give each subject one-half of his pair’s earnings, i.e., they lie on the horizontal line at 0.5. In fact, 87% of spectators allocated exactly equal splits, strongly supporting the theory.

In the Spectator treatments described above, dictators were third parties with no stake in the allocations they decided. They were free, therefore, to actualize their undistorted fairness preferences. Of course, people often have a stake in an allocation, which introduces a role for self-interest. The theory in the paper stipulates that people are motivated by both fairness and, when they have a stake, self-interest in making allocation decisions. In addition to the two Spectator treatments, therefore, there were two corresponding Stakeholder treatments, in which the dictator was one of the two subjects who prepared letters – there was no third party. Parallel to the Spectator treatments, in the one Stakeholder treatment differences in earnings were discretionary and in the other they were exogenous (here the dictator was always the subject who

---

13 The results illustrated in Figure 4 are for sessions that corrected an experimental procedure that produced an artifact. Uncorrected sessions, which are discussed below, exhibited greater dispersion around the 45 degree line but generated the same proportionality property, according to the results of statistical tests.

14 In most previous work in this area (e.g., Leventhal and Michaels, 1971), significant deviations from equality were taken as favorable to equity theory. As Hook and Cook (1979) point out, however, mere inequality does not per se corroborate the much stronger and more specific proportionality claim of equity theory. Of the many studies Hook and Cook review, only Hook (1978) uses unequal inputs to generate outcomes that differ insignificantly from strictly proportional ones. Nevertheless, the design used by Hook leaves serious questions about the robustness of the results and what they really captured: the subjects were 13-year-old children who were first presented with a series of exercises that repeatedly reinforced distribution in the same exact proportion as the allocation decision subsequently used to test proportionality. More convincing evidence of proportionality has, however, surfaced in subsequent work: using adult subjects, Ellingsen and Johannesson (2002) combine produced earnings, as in Konow (2000), with manipulation of information about roles, as in Babcock et al. (1995), and also find evidence that perceived contributions are proportional to first mover claims in their ex post treatment.

15 In the original paper, Spectator was called “Benevolent” Dictator and Stakeholder “Standard” Dictator.
had been arbitrarily assigned the higher per letter credits). The theory presented in the paper predicts that stakeholders will take an amount equal to or greater than the fair amount, i.e., the proportional amount in the discretionary treatment and equal splits in the exogenous treatment.

Table 6 summarizes the results for the 2 (Spectator vs. Stakeholder) × 2 (Discretionary vs. Exogenous) between-subjects factorial design (ignore, for the moment, the “Stakeholder beliefs” line). In the first treatment, contributions by one group averaged .491, which is the average fair allocation to that group. The actual allocations of third party dictators to that group averaged .494, an insignificant difference. Simply analyzing differences in means, however, does not evaluate the prediction of proportionality, only that allocations are fair, on average. So a paired difference test of contributions to allocations was conducted, and it established that the allocations of spectators do not differ significantly from fair (proportional) allocations (p=.88). Stakeholders in the discretionary differences condition allocated to themselves .644 of the earnings, an amount that differs significantly from the self-interest share of 1.0 and from their fair share according to average contributions of .516, both by tests of differences in means as well as by paired difference tests. Finally, an examination of the standard deviation of actual allocations reveals that variance is significantly lower among spectators, indicating that this kind of impartiality not only eliminates bias but also reduces dispersion.

The results for the exogenous differences condition are illustrated in the bottom rubric of Table 6. In treatment 1, spectators allocated .508 to one group, which differs insignificantly from the fair share of .5. The stakeholders in treatment 2, on the other hand, allocated .592 of the earnings to themselves, significantly less than the self-interest share of 1.0 and more than the fair share of .5. The standard deviation of spectators is also significantly less than that of stakeholders, as in the discretionary differences condition. The results from both conditions,

---

16 In addition, a regression of allocations on earnings indicates a slope of one, intercept of zero and $R^2$ of 0.63, and the fraction of benevolent dictators allocating exactly in proportion to earnings was 46%. These are the results for sessions that were not corrected for an experimental artifact. In the corrected sessions (see Figure 4), the $R^2$ rises to 0.98 and the percentage of exactly proportional allocations to 79%, as noted earlier.

17 As previously mentioned, 87% of spectators split the earnings exactly in halves. Note also that the paired difference test in this case where fair allocations are always .5 is the same as a standard difference of means test.
then, confirm the bias of stakeholders, and the impartiality and convergent views of spectators. Figure 6 illustrates dictator allocations for these two treatments.

An additional goal of this study was to formalize a theory of the sources of unfairness and to identify and quantify as cleanly as possible these sources when real stakes are involved. As summarized in the introduction to the current paper, the fairness bias was decomposed in Konow (2000) into two parts: the self-centered bias due to unadulterated self-interest, or deliberately selfish behavior, that appears to dominate the Gäechter and Riedl study, and the Babcock et al. finding of a self-serving bias due to self-deception, or unfairness owing to the fact that people apparently deceive themselves sometimes into believing it is fair to be unfair. Although many researchers seem to mean the latter when speaking of fairness bias, unfair biases actually have these related but separate origins. This decomposition is an application of Festinger’s theory of cognitive dissonance (1957): when agents experience displeasure from conflicting facts, they can respond by trying to alter the facts or by changing their beliefs about them. In the context of justice, Kay and Jost (2003) point out that people confronted with such situations (e.g., fairness is good, the world is not fair) can respond one of two ways: they can try to rectify or rationalize the injustice. In the context of the dictator experiment, the dictator can allocate some amount to the counterpart and/or nurture self-serving beliefs about what is fair. As Hafer (2000) notes, self-reported beliefs in these cases are suspect for a number of reasons. The dictator experiment, on the other hand, by employing real monetary allocations in a non-strategic environment, provides incentives that are compatible with truthful revelations of one’s beliefs. The stakeholder treatment in the exogenous differences condition contained a second decision designed for this purpose. After choosing their own divisions, dictators were then informed for the first time that they would be making an additional decision. They were also to allocate earnings between two other subjects, whose unequal earnings from the first phase were the same as the unequal ones of the dictator and his/her counterpart. Although fairness calls for equal shares, to the extent the dictators had convinced themselves that it was fair to take more than one-half in their own case, they should now allocate more to their counterpart with higher credits in the new pair. In fact,
standard dictators gave .558 of the earnings to their counterpart in the new pair, in comparison to the .592 they took for themselves in the earlier round. That is, around 60% of their unfairness (allocations to themselves in excess of .5) appears to be due to self-deception, whereas the remaining 40% is unadulterated self-interest.\footnote{In fact, there are several reasons to believe that this is a conservative estimate of the effect of self-deception. One is that this experiment is contextually simple and provides little basis for egocentric information processing. In addition, given the anonymity of this design, some dictators might believe, or choose to believe, that there are not really any counterparts. I found anecdotal evidence of this in responses to post-experimental questionnaires, and Frohlich, Oppenheimer and Moore (2001) found systematic evidence of such doubts in dictator games. In the current experiment, these doubts lower estimates of self-deception by making it cheap to be selfish in the first round and fair in the second round. The fact that such a large share of unfairness can be traced to self-deception despite these concerns underscores both the importance and the difficulty of achieving impartiality as a stakeholder. On the other hand, the spectator treatments of this experiment indicate how, when personal stakes are removed, bias vanishes and a very high level of consensus can be achieved.}

The fairness bias resembles a related phenomenon called “self-enhancement,” or the tendency to rate oneself more positively than is warranted. Paulhus and his collaborators (for example, Paulhus et al., 2003) distinguish two aspects of positive self-enhancement that are similar to the decomposition above of the fairness bias. One is self-presentation (or positive impression management), whereby people exaggerate their positive qualities in a conscious and deliberate fashion (similar to the self-centered bias). The other is self-deception, whereby people actually believe these exaggerations. Paulhus et al. (2003) conclude that there is a growing consensus among researchers about these two sources of self-enhancement and report that their own measure of self-enhancement suggests that it is “primarily self-deceptive” (pg. 900). This type of self-deception constitutes an intersection with “positive illusions,” or the tendency to engage in overly positive self-evaluations as well as, more generally, exaggerated perceptions of control and unrealistic optimism. Taylor and Brown (1988) survey this literature and conclude that positive illusions are actually conducive to good mental health: they promote the ability to care about others, to be happy and to engage productively in work.

It is difficult to detect such a silver lining on the cloud of self-deception in the context of ethical decision-making. When justice and self-interest conflict, the only direction for self-deception to pull is away from justice. Moreover, the results of numerous studies suggest that
self-serving bias is a widespread phenomenon in ethical decision-making. For example, Jost et al. (2003) note the contradiction between egalitarian ideals and the consistent perception that the free market system is fair despite substantial income inequality, and they attribute this in large part to self-deception. Self-deception might be seen in the tendency of many subjects in the Dana, Weber and Kuang study cited earlier to remain ignorant of information that could reveal a conflict between the pursuit of their self-interest and the promotion of fairness. Many contributors to this literature conclude that the self-serving bias is both pernicious and difficult to eradicate. Messick and Bazerman (1996) propose various steps to improve ethical business decision-making and note various psychologically based obstacles, including self-deception, to achieving that goal, a point underscored by the corporate scandals subsequent to their publication. In the wake of those scandals, Tenbrunsel and Messick (2004) go further and assign primary responsibility to self-deception for these ethical meltdowns. They are also pessimistic about the efficacy of traditional ethics training to avert or significantly attenuate such tendencies. In a related line of research, Haidt (2003) argues that moral judgments typically exhibit inertia, since they usually follow from intuition and emotion rather than a deliberative process. If correct, this fact only reinforces the tenacity of self-serving biases.

Relating this literature to the previously discussed experiments, could it be that the experimental measures of egocentric bias represent, to some degree, self-presentation rather than self-deception? I believe this question can be meaningfully related to subject anonymity. Suppose subjects’ decisions are completely anonymous: neither the experimenter nor other subjects will ever know these decisions. Then, if subjects value truthfulness or fairness at all and doing so comes at no other cost, it is difficult to conceive why subjects would be motivated to misrepresent their beliefs in order to present themselves favorably. No one will ever know that a given decision was actually theirs. Even if they misrepresent themselves simply in order to feel good or consistent, it is difficult to see how they could achieve the desired feeling unless they truly believe their choices. The measures of self-deception are based on fairness estimates in Babcock et al. (1995) and Gächter and Riedl (2005) and on allocation decisions in Konow

37
(2000), which were all collected under single-blind anonymity conditions. That is, subjects did not know the identity of their counterparts, but the experimenter did know the identity of decision makers, and these facts were revealed to subjects. Evidence from several experiments suggests that single-blind conditions generate results that are roughly the same as complete anonymity (for example, Bolton, Katok and Zwick, 1998, and Cason and Mui, 1997). This should be reassuring regarding the validity of these measures. Moreover, the widespread signs of self-serving bias in the political and business world cited, for example, by Jost et al. (2003), Messick and Bazerman (1996) and Tenbrunsel and Messick (2004), reinforce this interpretation of the laboratory findings and indicate the importance of this phenomenon in the real world. The findings of these studies are summarized below:

**Result 9:** The unfairness of stakeholders is due partly, and perhaps mostly, to self-deception, i.e., beliefs about fairness that are egocentrically biased. The views of spectators, on the other hand, are unbiased and converge significantly more than those of stakeholders.

**DISCUSSION AND CONCLUSIONS**

The analysis of this paper points toward the following conclusions. Even small real stakes produce measures of bias and dispersion in claims that are similar to those with large stakes. There is a well-behaved trade-off between fairness and self-interest. Claims become more biased and disperse and disagreements more frequent as the fair claim increasingly conflicts with self-interest and with the introduction of information that provides a basis for competing allocation rules. Biased claims can be traced to two sources: the self-centered bias due to blatant self-interest and the self-serving bias arising from self-deception, which is fed by egocentric processing of information. If bargaining power is balanced, however, the overall effect of increased information is favorable: settled claims are less biased, and outcomes are less biased and less dispersed, although disagreement rates can increase or decrease, depending on the structure of bargaining. Bias and variance can be reduced by addressing the impact of personal stakes, for example, by the timing of information about roles, by an independent statement of fair shares, or by removing stakes altogether and placing individuals in the role of spectator.
These conclusions differ considerably from the Rawlsian “conventional wisdom” in several ways. First, they follow from a model of impartiality that involves a “spectator,” i.e., one who has no stakes in the allocation being considered. As opposed to the Rawlsian model, one can find many real world examples and empirical studies that correspond to, or at least approximate, the conditions of this approach. Second, given this type of impartiality, we find that justice is better understood, and under certain conditions better realized, by adding rather than removing information. A frequent criticism of the Rawlsian model is that removing all information related to stakes arguably entails eliminating information relevant to justice, indeed, this process could reduce the information set to almost nothing. The approach proposed here, on the other hand, operates under stake and information conditions that are often realizable, at least approximately. The current findings raise the prospect of employing empirical methods to identify impartial justice, which could help address important questions of interest in philosophy, law and the social sciences. It suggests a means to rule out biased claims that are erroneously defended based on self-serving manipulation of justice principles and to develop and implement just policies. As the evidence shows, however, stakeholders develop blind spots about justice due to the particularly insidious tendency to nurture self-serving beliefs. It can be difficult to dislodge such biases, because people believe they are acting on principle. In light of this, I conclude with some observations about additional prospects for identifying and applying impartial justice.

I believe there are important potential benefits of targeting impartial justice for descriptive social science research. Consider, as an example, the history of justice research inspired by equity theory. The important insight of equity theory was the concept that fair allocations are proportional to inputs. At the same time, people were seen as self-interested and inclined to make self-serving claims about which variables count as inputs. Equity theorists originally reconciled these facts by allowing people to appeal to any variables as inputs. The problem with this solution is that it produces a theory that predicts everything, but then it predicts nothing: a theory must generate refutable propositions in order to lay claim to any predictive power. This quandary can be resolved, however, by distinguishing the preferences of
implicated stakeholders from those of impartial spectators. To identify the equity motive, one can observe the decisions about the allocations of other parties by spectators, i.e., informed people with no personal stakes. Applying this method, I found, as previously discussed, that fair allocations are proportional to inputs, specifically, to the contributions people control, but not to those they do not control. This provides a basis for distinguishing which inputs are relevant to equity theory. Stakeholders, who are motivated by both equity and self-interest, are predicted to choose allocations between this fair claim and the purely self-interested claim, inclusive, which the data also corroborate. Of course, a large amount of social science research has explored impartial justice, most often though eliciting attitudes in a context removed from circumstances of the respondent. My contention is merely that progress in descriptive justice research would be enhanced by a heightened awareness of fairness biases and by explicitly distinguishing biased and impartial justice in the design of empirical work.

Turning to possible real world applications of these lessons, conflicts can sometimes be addressed by approximating the conditions of impartial spectatorship: one can apply directly the judgments of spectators, i.e., groups of informed individuals with no stake in the matter under consideration. In fact, that seems to be the concept around which an independent judiciary is designed. Judges and juries are supposed to be unbiased adjudicators to whom the relevant facts are presented so that they may decide the just distribution of benefits or burdens. Of course, the costs of legal proceedings prohibit this as a means to resolve all conflicts, and legal recourse is, at best, an imperfect approach. Third party judgments are also not necessarily the best apparatus for all social and political decisions. Alternately, one can think of impartial spectatorship as did Adam Smith (1759): individuals should strive to be aware of and guided by their own impartial spectator. We have seen how fairness can emerge when conflicting parties engage in symmetric bargaining, presumably through the implicit judgment of the impartial spectator. This is reinforced when adversaries receive an independent (and impartial) estimate of fair shares.

When stakes are salient, however, it can be quite challenging to purge the fairness bias. Babcock and Loewenstein (1997) report the results of various interventions aimed at reducing or
eliminating the self-serving bias in their experiments. They found informing subjects of the bias had no effect, either on the discrepancy in fairness assessments and predictions of the judge’s decision or on the likelihood of settlement. Having subjects write an essay arguing the opponent’s case was similarly ineffectual. They did find, however, that informing subjects of the egocentric bias and having them list weaknesses in their own case significantly reduced the differences in predictions of the judge’s decision and lowered the rate of impasse from 36% to only 4%. In a real world setting, Oberholzer-Gee, Bohnet and Frey (1997) describe the siting of nuclear waste facilities, and their findings support the importance of impartiality, fairness and information. They report a willingness, uncharacteristic in such cases, on the part of a Swiss community to accept the siting, which the authors trace to fair, efficient and informed decision-making procedures. The lessons of impartial justice will, I hope, help to clarify further the nature of justice and to produce additional mechanisms for resolving conflicts and realizing just outcomes and processes.
References


## Table 1
### Variation in Stakes

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Salient Stakes</th>
<th>Share of Stakes Claimed by First Mover</th>
<th>Mean</th>
<th>Std Dev</th>
<th>N</th>
<th>% Equal</th>
<th>% Self-Interested</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forsythe, Horowitz, Savin and Sefton (1994)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Hypothetical $5 Dictator</td>
<td>$0</td>
<td></td>
<td>.617</td>
<td>.184</td>
<td>46</td>
<td>.48</td>
<td>.13</td>
</tr>
<tr>
<td>2. Actual $5 Dictator</td>
<td>$5</td>
<td></td>
<td>.778</td>
<td>.204</td>
<td>45</td>
<td>.18</td>
<td>.36</td>
</tr>
<tr>
<td>3. Actual $10 Dictator</td>
<td>$10</td>
<td></td>
<td>.767</td>
<td>.179</td>
<td>24</td>
<td>.21</td>
<td>.21</td>
</tr>
<tr>
<td>4. Hypothetical $5 Ultimatum</td>
<td>$0</td>
<td></td>
<td>.562</td>
<td>.134</td>
<td>47</td>
<td>.43</td>
<td>.04</td>
</tr>
<tr>
<td>5. Actual $5 Ultimatum</td>
<td>$5</td>
<td></td>
<td>.545</td>
<td>.104</td>
<td>43</td>
<td>.53</td>
<td>.00</td>
</tr>
<tr>
<td>6. Actual $10 Ultimatum</td>
<td>$10</td>
<td></td>
<td>.533</td>
<td>.082</td>
<td>24</td>
<td>.71</td>
<td>.00</td>
</tr>
<tr>
<td>Hoffman, McCabe and Smith (1996)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. $10 Ultimatum</td>
<td>$10</td>
<td></td>
<td>.567</td>
<td>.076</td>
<td>24</td>
<td>.50</td>
<td>.00</td>
</tr>
<tr>
<td>2. $100 Ultimatum</td>
<td>$100</td>
<td></td>
<td>.559</td>
<td>.119</td>
<td>27</td>
<td>.52</td>
<td>.04</td>
</tr>
</tbody>
</table>

Note: Superscripted values indicate that the value in question differs at the five percent level of significance from its counterparts in the treatments with the superscripted numbers. For example, in the Hypothetical $5 treatment of Forsythe et al., the superscripts \(^{23}\) on the mean of .617 indicate that this mean differs from the means in treatments 2, 3 and 6 of the same study at the five percent level of significance. A superscript 0 indicates that this mean differs significantly from equal splits of .5. For tests of differences, a t-test is used for differences in means, an F-test of differences in variances for standard deviations, and a z-test of differences in proportions for percentage values, such as % Equal and % Self-Interested.
<table>
<thead>
<tr>
<th>Treatment</th>
<th>Round</th>
<th>Round</th>
<th>Salient Stakes</th>
<th>First Interest Share</th>
<th>Mean</th>
<th>Std Dev</th>
<th>N</th>
<th>% Equal</th>
<th>% Self-Interested</th>
<th>% Disagreements</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Game 1</td>
<td>$6</td>
<td>$0</td>
<td>1.00</td>
<td>.686 .0456</td>
<td>.686</td>
<td>.136</td>
<td>33</td>
<td>.21</td>
<td>.686 .0456</td>
<td>.06</td>
</tr>
<tr>
<td>2. Game 2</td>
<td>$6</td>
<td>$1</td>
<td>.833</td>
<td>.667 .0456</td>
<td>.667</td>
<td>.141</td>
<td>33</td>
<td>.06</td>
<td>.667 .0456</td>
<td>.18</td>
</tr>
<tr>
<td>3. Game 3</td>
<td>$6</td>
<td>$2</td>
<td>.667</td>
<td>.665 .0456</td>
<td>.665</td>
<td>.067</td>
<td>33</td>
<td>.03</td>
<td>.665 .0456</td>
<td>.00</td>
</tr>
<tr>
<td>4. Game 4</td>
<td>$6</td>
<td>$3</td>
<td>.500</td>
<td>.573 .012356</td>
<td>.573</td>
<td>.076</td>
<td>33</td>
<td>.45</td>
<td>.573 .012356</td>
<td>.18</td>
</tr>
<tr>
<td>5. Game 5</td>
<td>$6</td>
<td>$4</td>
<td>.333</td>
<td>.490 .1234</td>
<td>.490</td>
<td>.085</td>
<td>33</td>
<td>.54</td>
<td>.490 .1234</td>
<td>.18</td>
</tr>
<tr>
<td>6. Game 6</td>
<td>$6</td>
<td>$5</td>
<td>.168</td>
<td>.458 .01234</td>
<td>.458</td>
<td>.104</td>
<td>33</td>
<td>.48</td>
<td>.458 .01234</td>
<td>.21</td>
</tr>
</tbody>
</table>

Note: Superscripted values indicate that the value in question differs at the five percent level of significance from its counterparts in the treatments with the superscripted numbers, and a superscript 0 indicates the mean differs significantly from equal splits of .5. For tests of differences, a t-test is used for differences in means, an F-test of differences in variances for standard deviations and a z-test for differences in proportions.
Table 3
Variation in Information
Share of Tickets

<table>
<thead>
<tr>
<th>Treatments</th>
<th>Information</th>
<th>Value of Prizes</th>
<th>Share of Tickets to High Prize Player</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roth and Malouf (1979)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Games 1 &amp; 2</td>
<td>Partial</td>
<td>$1.00 $1.00</td>
<td>.50  .504⁴ .030⁴ 15/16 .80³⁴ .06³⁴</td>
</tr>
<tr>
<td>2. Games 1 &amp; 2</td>
<td>Full</td>
<td>$1.00 $1.00</td>
<td>.50  .491⁴ .039⁴ 21/21 .95³⁴ .95³⁴ .00³⁴</td>
</tr>
<tr>
<td>3. Games 3 &amp; 4</td>
<td>Partial</td>
<td>$1.25 $3.75</td>
<td>.25  .500⁴ .025⁴ 16/16 .69²⁴ .00¹²⁴ .00³⁴</td>
</tr>
<tr>
<td>Roth, Malouf and Murnighan (1981)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Games 1 &amp; 3</td>
<td>Low/Inter</td>
<td>$3.00 $9.00</td>
<td>.25  .509²³⁶ .037²⁴⁶ 38/46 .61²⁴ .00⁶ .17³⁴</td>
</tr>
<tr>
<td>2. Games 1 &amp; 3</td>
<td>High</td>
<td>$3.00 $9.00</td>
<td>.25  .400⁰¹³⁵ .085¹³⁴⁵ 24/30 .29¹ .08⁶ .20³⁴</td>
</tr>
<tr>
<td>3. Game 4</td>
<td>Low/Inter</td>
<td>$3.00 $12.00</td>
<td>.20  .481²³⁶ .043²⁴⁶ 21/23 .52 .00 .09³⁴</td>
</tr>
<tr>
<td>4. Game 4</td>
<td>High</td>
<td>$3.00 $12.00</td>
<td>.20  .219¹³⁴⁵ 11/15 .18¹ .00 .27³⁴</td>
</tr>
<tr>
<td>5. Game 2</td>
<td>Low/Inter</td>
<td>$2.40 $9.60</td>
<td>.20  .494²³⁶ .052²⁴⁶ 21/23 .52 .00 .09³⁴</td>
</tr>
<tr>
<td>6. Game 2</td>
<td>High</td>
<td>$2.40 $9.60</td>
<td>.20  .357⁰¹³⁵ .121¹³⁵ 12/15 .33 .17¹² .20³⁴</td>
</tr>
<tr>
<td>Roth and Murnighan (1982)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Neither knows</td>
<td>Neither</td>
<td>$5.00 $20.00</td>
<td>.20  .482³⁴ .058²³⁴ 56/63 .59²³⁴ .00³⁴ .11³³</td>
</tr>
<tr>
<td>2. $20 player knows</td>
<td>$20 player</td>
<td>$5.00 $20.00</td>
<td>.20  .461³⁴ .108¹ 44/54 .39¹⁴ .02³⁴ .19³⁴</td>
</tr>
<tr>
<td>3. $5 player knows</td>
<td>$5 player</td>
<td>$5.00 $20.00</td>
<td>.20  .363¹²  .135¹ 58/81 .28¹ .21¹² .28³⁴</td>
</tr>
<tr>
<td>4. Both know</td>
<td>Both</td>
<td>$5.00 $20.00</td>
<td>.20  .334¹²  .121¹ 51/65 .20¹² .16¹² .22³⁴</td>
</tr>
</tbody>
</table>

Note: Superscripted values indicate that the value in question differs at the five percent level of significance from its counterparts in the treatments with the superscripted numbers, whereas a 0 indicates that it differs significantly from equal splits. A t-test is used for differences in means, an F-test of differences in variances and a z-test for differences in proportions. The % Disagree is out of all pairs (N: Total) whereas all other figures refer only to pairs that reached an agreement (N: Agree). Two observations have been removed from the Roth et al. (1981) data: one outlier that they also ignored and one inefficient agreement, i.e., not all tickets were allocated.
Table 4
Variation in Information
Expected Payoffs to Pair

<table>
<thead>
<tr>
<th>Treatments</th>
<th>Information</th>
<th>Value of Prizes</th>
<th>Expect Payoffs</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roth and Malouf (1979)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Games 1 &amp; 2</td>
<td>Partial</td>
<td>$1.00</td>
<td>$1.00</td>
<td>.504$^{34}$</td>
</tr>
<tr>
<td>2. Games 1 &amp; 2</td>
<td>Full</td>
<td>$1.00</td>
<td>$1.00</td>
<td>.491$^{34}$</td>
</tr>
<tr>
<td>3. Games 3 &amp; 4</td>
<td>Partial</td>
<td>$1.25</td>
<td>$3.75</td>
<td>.750$^{0124}$</td>
</tr>
<tr>
<td>4. Games 3 &amp; 4</td>
<td>Full</td>
<td>$1.25</td>
<td>$3.75</td>
<td>.594$^{0123}$</td>
</tr>
<tr>
<td>Roth, Malouf and Murnighan (1981)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Games 1 &amp; 3</td>
<td>Low/Inter</td>
<td>$3.00</td>
<td>$9.00</td>
<td>.756$^{02}$</td>
</tr>
<tr>
<td>2. Games 1 &amp; 3</td>
<td>High</td>
<td>$3.00</td>
<td>$9.00</td>
<td>.659$^{0135}$</td>
</tr>
<tr>
<td>3. Game 4</td>
<td>Low/Inter</td>
<td>$3.00</td>
<td>$12.00</td>
<td>.786$^{02}$</td>
</tr>
<tr>
<td>4. Game 4</td>
<td>High</td>
<td>$3.00</td>
<td>$12.00</td>
<td>.709$^{0}$</td>
</tr>
<tr>
<td>5. Game 2</td>
<td>Low/Inter</td>
<td>$2.40</td>
<td>$9.60</td>
<td>.794$^{02}$</td>
</tr>
<tr>
<td>6. Game 2</td>
<td>High</td>
<td>$2.40</td>
<td>$9.60</td>
<td>.671$^{0}$</td>
</tr>
<tr>
<td>Roth and Murnighan (1982)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Neither knows</td>
<td>Neither</td>
<td>$5.00</td>
<td>$20.00</td>
<td>.784$^{034}$</td>
</tr>
<tr>
<td>2. $20 player knows</td>
<td>$20 player</td>
<td>$5.00</td>
<td>$20.00</td>
<td>.762$^{04}$</td>
</tr>
<tr>
<td>3. $5 player knows</td>
<td>$5 player</td>
<td>$5.00</td>
<td>$20.00</td>
<td>.664$^{01}$</td>
</tr>
<tr>
<td>4. Both know</td>
<td>Both</td>
<td>$5.00</td>
<td>$20.00</td>
<td>.644$^{012}$</td>
</tr>
</tbody>
</table>

Note: Superscripted values indicate that the value in question differs at the five percent level of significance from its counterparts in the treatments with the superscripted numbers, and a superscript 0 indicates the mean differs significantly from equal splits of .5. For tests of differences, a t-test is used for differences in means and an F-test of differences in variances for standard deviations.
Table 5
Egocentric Information Processing

<table>
<thead>
<tr>
<th></th>
<th>Ex Ante Stakes</th>
<th>Ex Post Stakes</th>
<th>Differences in Means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>Babcock, Loewenstein, Issacharoff and Camerer (1995, $10 stakes)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fair judgments</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High stakeholder</td>
<td>.791</td>
<td>.189</td>
<td>47</td>
</tr>
<tr>
<td>Low stakeholder</td>
<td>.593</td>
<td>.196</td>
<td>47</td>
</tr>
<tr>
<td>Difference</td>
<td>.198</td>
<td>.231</td>
<td>47</td>
</tr>
<tr>
<td>t-statistic</td>
<td>4.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimates of judge’s award</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High stakeholder</td>
<td>.761</td>
<td>.209</td>
<td>47</td>
</tr>
<tr>
<td>Low stakeholder</td>
<td>.575</td>
<td>.203</td>
<td>47</td>
</tr>
<tr>
<td>Difference</td>
<td>.186</td>
<td>.260</td>
<td>47</td>
</tr>
<tr>
<td>t-statistic</td>
<td>4.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agreements</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agreed shares</td>
<td>.700</td>
<td>.158</td>
<td>34</td>
</tr>
<tr>
<td>Disagreement rate</td>
<td>.28</td>
<td>.44</td>
<td>47</td>
</tr>
<tr>
<td>Gächter and Riedl (2005, $18 stakes)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fair judgments</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High stakeholder</td>
<td>.640</td>
<td>.062</td>
<td>45</td>
</tr>
<tr>
<td>Low stakeholder</td>
<td>.616</td>
<td>.068</td>
<td>45</td>
</tr>
<tr>
<td>Difference</td>
<td>.024</td>
<td>.014</td>
<td>45</td>
</tr>
<tr>
<td>t-statistic</td>
<td>1.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opening offers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High stakeholder</td>
<td>.716</td>
<td>.054</td>
<td>20</td>
</tr>
<tr>
<td>Low stakeholder</td>
<td>.524</td>
<td>.128</td>
<td>25</td>
</tr>
<tr>
<td>Difference</td>
<td>.192</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-statistic</td>
<td>6.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agreements</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agreed shares</td>
<td>.605</td>
<td>.067</td>
<td>37</td>
</tr>
<tr>
<td>Disagreement rate</td>
<td>.16</td>
<td>.37</td>
<td>44</td>
</tr>
</tbody>
</table>

Note: The test statistic for differences in means is the t-statistic and for differences in disagreement rates the z-statistic for differences in proportions.
Table 6
Stakeholders versus Spectators
Konow (2000)

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Fair allocations</th>
<th>Actual allocations</th>
<th>Paired Difference test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean  SD</td>
<td>Mean  SD</td>
<td>t-stat  p-value  N</td>
</tr>
<tr>
<td><strong>Discretionary differences</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Spectator</td>
<td>.491 .113</td>
<td>.494 2 .130 2</td>
<td>-0.156 .88 24</td>
</tr>
<tr>
<td>2. Stakeholder</td>
<td>.516 .171</td>
<td>.644 1 .235 1</td>
<td>2.959 &lt;.01 24</td>
</tr>
<tr>
<td><strong>Exogenous differences</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Spectator</td>
<td>.500 0</td>
<td>.508 23 .056 2</td>
<td>0.723 .48 24</td>
</tr>
<tr>
<td>2. Stakeholder</td>
<td>.500 0</td>
<td>.592 1 .196 13</td>
<td>2.808 &lt;.01 36</td>
</tr>
<tr>
<td>3. Stakeholder beliefs</td>
<td>.500 0</td>
<td>.558 1 .083 2</td>
<td>4.205 &lt;.01 36</td>
</tr>
</tbody>
</table>

Note: The t-statistics result from a two-tail test that the mean actual allocations equal the mean fair allocations. Superscripted means and standard deviations differ at the five percent level of significance from their counterparts in the treatments with the superscripted numbers.

Figure 1. Equal split first mover claims and disagreements in two round bargaining experiments (Spiegel, et al., 1994).
Figure 2. Agreements over lottery tickets in bargaining games (Roth and Malouf, 1979).

Figure 3. Expected payoffs from bargaining games (Roth and Malouf, 1979).
Figure 4. Discretionary/Spectator treatment (Konow, 2000)

Figure 5. Exogenous/Spectator treatment (Konow, 2000)
Figure 6. Dictator allocations in exogenous differences treatments (Konow, 2000).