LAW AND LITERATURE:
THEORY AND EVIDENCE ON EMPATHY AND GUILE

Daniel L. Chen[*]

Abstract Legal theorists have suggested that literature stimulates empathy and affects moral judgment and decision-making. I present a model to formalize the potential effects of empathy on third parties. Empathy is modeled as having two components–sympathy (the decision-maker's reference point about what the third party deserves) and emotional theory of mind (anticipating the emotions of another in reaction to certain actions). I study the causal effect with a data entry experiment. Workers enter text whose content is randomized to relate to empathy, guile, or a control. Workers then take the Reading the Mind in the Eyes Test (RMET) and participate in a simple economic game. On average, workers exposed to empathy become less deceptive towards third parties. The result is stronger when workers are nearly indifferent. These results are robust to a variety of controls and model specifications.

Keywords: Normative Commitments, Other-Regarding Preferences, Empathy, Deception, Guile

JEL codes: D64, D03, K00

[*] Daniel L. Chen, daniel.chen@iast.fr, Toulouse School of Economics, Institute for Advanced Study in Toulouse, University of Toulouse Capitole, Toulouse, France; dchen@law.harvard.edu, LWP, Harvard Law School. First draft: May 2009. Current draft: December 2017. Most recent version at: http://users.nber.org/∼dlchen/papers/Law_and_Literature.pdf. I thank Aroha Bahuguna and Yutong Li for research assistance. Work on this project was conducted while I received financial support from the European Research Council (Grant No. 614708), Swiss National Science Foundation (Grant Nos. 100018-152678 and 106014-150820), Agence Nationale de la Recherche, John M. Olin Foundation, Ewing Marion Kauffman Foundation, and Templeton Foundation (Grant No. 22420).

1 Introduction

In today's global world, modern democratic societies face the challenges of immigration and increasing multiculturalism; but we see in response to this also a rise of nationalism. Pro-social attitudes among group members alone will not help us meet the social and economic challenges of the future. While people in many cultures may consider themselves to be competent moral judges and devoted moral agents – and will thus expect other people to agree with their moral judgments – they may disagree with each other about what constitutes the morally good thing to do in various circumstances. Legal philosophers emphasize the importance of "recognition" of and "respect" for other people. In "Beyond toleration to equal respect," Martha Nussbaum writes that "When people 'tolerate' others, they give them grudging acceptance, but they don't think of them as equals, with fully equal rights" (Nussbaum 2008). Intrinsically moral motivation and action, recognition of others (including social and cultural strangers) as equals, and the universality of moral commitments go hand in hand (Darwall 2006). Experimental psychologists, cognitive scientists and moral philosophers have suggested possible factors that direct our morally relevant choices and behavior: moral beliefs and reasoning, moral emotions (especially empathy) (Hoffman 2001), moral values or virtues, and various kinds of moral intuitions. Ample evidence indicates that emotions affect decision-making (Engelmann and Hare 2017). Emotion, specifically, facial expressions with varying levels of positive affect are perceived and categorized differently across cultures, and differences in perception of emotion are said to partly account for cultural agreements and disagreements (Engelmann and Pogosyan 2013).

The importance of empathy in legal institutions has been highlighted in recent years by President Obama, who emphasized Justice Sonia Sotomayor's empathy as a criterion for nominating her to the US Supreme Court. The ensuing public discussion about empathy as a judicial criterion revolved around whether understanding different cultural perspectives was necessary for upholding justice. In philosophical arguments, Nussbaum advocates the reading of novels as a building-block of social justice. She argues that stories play a role in developing moral imagination. In Poetic Justice, Nussbaum wrote that judicial decisions informed by "the literary imagination" are likely to be sounder and wiser than judgments reached by other means (Nussbaum 1996).[1] If reading literature is an exercise in empathy, then literature which perpetuates essentialist, colonial, racist, sexist and otherwise dehumanizing depictions does the opposite (McRobie 2014).

I use a revealed preference experiment where subjects make moral judgments after reading a short piece of literature that is randomized. I present a reference-point dependent model of social behavior

[1] The role of empathy has been highlighted in debates about the effects of market forces as a specific channel to explain the impact of competition on moral decision-making (Chen 2016). For instance, Simmel (1955) noted that competition fosters empathy between competitors and third parties Hirschman (1982).

where individuals maximize a three-term utility function: a consumption utility term and two "social" terms. One social term captures a preference for "just desert" (others getting what we think they deserve) and the other term a preference for the satisfaction of other's expectations (i.e., them getting what we think they think they deserve). The legal literature typically does not distinguish between two basic kinds of empathy. The first component of empathy has a focus on motivational impact, that is, empathy as triggering helping behavior (what I call "sympathy" – a shift in the decision-maker's reference point). Sympathy will increase pro-social behavior. The second social term measures the target individuals' anger or gratitude from the interaction which is determined by a value function derived from prospect theory. The second component of empathy is modeled as "emotional theory of mind" (eToM). eToM is whether individuals anticipate the *emotions* of another in reaction to certain actions. This component of empathy has a focus on understanding and normative judgment (empathy as a way to understand how another person feels and as involved in judging the moral adequacy of another person's emotional and behavioral responses to given circumstances). Understanding the other's reference point is modeled as a shift in what the decision-maker perceives as the agent's reference point. Empathy can increase or decrease pro-social behavior.

eToM can be distinguished from sToM–strategic ToM, which is anticipating the *actions* of another–and *actions* are the topic of usual study by economists. In combination with sToM, eToM could be important in explaining heterogeneous behavior in experiments, and perhaps outcomes outside the lab. eToM might be culture-specific, and a low degree of eToM (e.g., toleration, but not recognition-respect, or indifference) towards people of different cultures is likely to cause tensions and a lack of mutual respect and sympathy. The challenge is to measure eToM in relevant situations, based on measures of actual emotions, and where we know how those emotions translate into actual decisions. This paper uses survey instruments that are paradigm-setting for economics – a game involving deception (Gneezy 2005)[2], for psychology – Reading the Eyes in the Mind test (RMET) (Baron-Cohen et al. 2001)[3], and for law – reading literature (Nussbaum 1996). While there may be limitations to this, it is important to note that the state of the art in examining the "Law and Literature" thesis articulated by jurists, legal theorists, legal philosophers is observational data, natural experiments (e.g., the random assignment of judges making different legal precedents on free speech cases), or artifactual experiments (the random assignment of news summaries of legal cases randomized to be liberal or conservative and then measuring the preferences using attitudinal surveys) (Chen and Yeh 2016). The ideal experiment would randomize judges to treatments that amplify or suppress empathy and observe their real world decisions over third parties (like defendants). The

[2] It has over 1000 google scholar citations since 2005.

[3] It has over 3000 google scholar citations since 2001.

findings that judges decide cases differently and sentence more harshly after economics training can be due to shifts in reference points (Ash et al. 2017). Whether empathy and emotion affect moral judgment might inform whether we want algorithms instead of prosecutors (Amaranto et al. 2017).

The reason that empathy can increase or decrease pro-social behavior is because the agent is loss-averse at his reference point. This renders a concave cost of deviating from the optimal decision. As the decision-maker's beliefs about the agent's reference point increases, the decision-maker may decide "what-the-hell" and jump to a lower optimum (Chen 2017c). Popular terms like the "What the Hell Effect" (Ariely 2012; Baumeister and Heatherton 1996) capture this behavioral tendency—the cognitive cost of deviating downwards in a large scale is not much larger than in a small scale. A number of theoretical papers show that the curvature of bliss-point deviations has important implications for decision making in social and political settings. For instance, as discussed by Osborne (1995) and shown by Kamada and Kojima (2014), concavity drives polarization in political platforms. This is since voters do not perceive a difference between a policy slightly away from their bliss point and a policy far away. Hence, unless a candidate adheres very closely to a group of voters' preferences, these voters will not vote at all, implying that in a polarized electorate political platforms will be polarized when ideological costs are concave but not otherwise. A formalization of moral decision-making that is deontological ("first, do no harm") would be lexicographic[4], and an approximation would be a concave cost of deviating from the moral bliss point. Chen et al. (2016c) shows that concavity in deviating from ideological bliss points can lead judges to cave-in – dissent less often – than other judges despite having the least say in shaping court decisions.

The basic idea is quite simple: we may treat people fairly because we do not want to anger them, and since individuals are "prospectors," we must learn about their expectations in order to avoid their wrath. Experimental moral psychology highlights these two distinct ways in which humans react to moral dilemmas.

> "Evidence from observations of great apes suggests that our common ancestors lived intensely social lives guided by emotions such as empathy, anger, gratitude, jealousy, joy, love and a sense of fairness, and all in the apparent absence of moral *reasoning*. Thus, from an evolutionary standpoint, it would seem strange if human behavior were not driven in part by domain-specific social-emotional dispositions. At the same time, however, humans appear to possess a domain-general capacity for sophisticated abstract reasoning, and it would be surprising as well if this capacity played no role in human moral judgement." (Greene et al., 2004)

[4] For an economic model and test of Kant's categorical imperative, see Chen and Schonger (2016; 2017).

Furthermore, as Nussbaum (2017) notes, many believe that anger is central to justice, that it is impossible to care for justice without anger at injustice, and that anger should be encouraged as part of a transformative process. Many view that the law ought to punish aggressors in a manner that embodies the spirit of justified anger. And it is also very widely believed that successful challenges against great injustice need anger to make progress. In the animal world, there is considerable evidence that fighting among social animals occurs where expectations (such as who should get some piece of food) are unclear:

> "Baboons ordinarily forage like flocks of birds, fanning out in a search for small vegetable items that are picked off the ground and eaten quickly. The troop members seldom challenge each one another under these circumstances. But when a clump of grass shoots is discovered in elephant dung, or a small animal is killed, the baboons threaten one another and may even fight over the food."[5]

As some psychologists have defined it, empathy is a non-voluntary, subconscious process of affective understanding of other's mental states or events. Empathy was a necessity in order for humans to be able to communicate effectively with each other (De Waal 2008). As communication between humans became more complex and individuals were confronted with competing interests, deception may have evolved as a means of exploiting empathy. A suggestive piece of evidence is that a basic form of deception, which sometimes even occurs subconsciously, is mimicking another, which encourages empathy in the individual. Such mimicry can even be seen in infants (emotional contagion, facial mimicking) (see e.g. Sonnby-Borgström 2002; Lakin and Chartrand 2003) and courts (vocal mimicry) (see e.g. Chen et al. 2015). In bargaining situations, empathy may be antecedent to guile – when someone knows what the other person wants and intentionally deceives him or her (e.g., exploit take-it-or-leave-it offers to their advantage). Some studies find a positive relationship between reflection and pro-sociality (Corgnet et al. 2015; Lohse 2016; Capraro and Cococcioni 2016) while others find a negative relationship (Ponti and Rodriguez-Lara 2015; Cueva et al. 2016) or no relationship (Hauge et al. 2015; Tinghög et al. 2013; Verkoeijen and Bouwmeester 2014). Arruñada et al. (2015) finds that strategic reasoning has no significant relationship with pro-sociality, Bayer and Renou (2016) and Dittrich and Leipold (2014) find that strategic people are more selfish. DeAngelo and McCannon (2017) finds that ability to process social and emotional cognition is negatively correlated with cooperation in prisoner dilemma games. In the model I present in this paper, whether empathy renders less pro-social actions depends on the degree of perceived loss-aversion of the agent. If the agent is sufficiently loss averse, then a decision-maker may cave-in when she perceives a higher reference point for the agent. To be sure, in most regions of the parameter space, empathy increases pro-social action.

I conduct three experiments on MTurk (the original and two replications for a total of 862 subjects)

[5] *Sociobiology*, Wilson (2000), p249.

in which first, subjects are copying a short paragraph whose content depends on the condition (guile, empathy or control). This method has been used in other legal settings. For example, to test whether exposure to a liberal or conservative judicial decision leads to shifts in moral attitudes towards the litigated subject, several experiments randomize the final paragraph of a data entry task to be a recent news report about a liberal or conservative legal decision, which is found to affect moral attitudes on abortion (Chen et al. 2017b) and obscenity (Chen and Yeh 2014). The data entry task renders the reading of the passage more likely, and this is the same methodology employed here.

Next, subjects have to do an eye expressions test, which means that they look at different photographs of eyes and have to indicate which of four kinds of feelings or emotions those eyes transmit. The survey instrument is the Reading the Mind in the Eyes (RMET) test (Baron-Cohen et al. 2001). The survey asks subjects to look at a small portion of the face (a pair of eyes) and pick one of four possible emotions that that person is feeling. Faces are of central importance for social communication as a window into the mental states of other people via gaze direction, which indicates focus and shifts of attention, and expression, which reveal emotional states. The biological significance of facial cues is highlighted by evidence from developmental and cross-cultural psychology, as well as cognitive neuroscience. For example, even 9 minutes after birth, infants show attentional preferences for faces over similar objects (Johnson et al. 1991) that develop into identity discriminatory abilities within as little as 3 months (Kelly et al. 2005, 2007). RMET is widely used. However, I do not find that being exposed to different types of text affects RMET, on average. Several other papers since this project began in 2009 have also found mixed evidence on whether exposure to literature (with longer texts than mine) affect RMET. Kidd and Castano (2013) found that fiction of 3 pages in length impacted RMET scores, but Djikic et al. (2013) found 10 pages of text did not affect RMET, Panero et al. (2016), Panero et al. (2017), and Samur et al. (2017) were unable to replicate Kidd and Castano's experiments. However, subjects who entered the data entry paragraph accurately do have significant differences in RMET scores. Subjects who enter the paragraphs more accurately have higher RMET scores by 9 points on a scale of 36.[6] This is consistent with these subjects paying more attention to the study. However, RMET can be criticized for using Western emotions and faces. Even within the US, perceptions about masculinity based on voice recordings differ by geography and demographic group of the perceiver (Chen et al. 2016a, 2017a). Indeed, it appears that subjects' demographic characteristics are predictive of RMET scores. It is possible, indeed likely, that the perceived emotion being expressed is just as subjective as perceptions of a masculine voice. Thus, there is a need to develop a measure that transcends cultural groups and is easy for others to use.

The test for deception poses a simple two-player game where the second player makes a decision, option

[6] This table is removed at the suggestion of the editor.

A or option B, that determines the payoffs for both players, but the first player knows the payoffs for both options and has the moral decision to tell the second player which payoff is better for the second person. Varying the difference in payoffs identifies how much people are willing to deceive. The task is made morally difficult to the extent that sending the true message, i.e. which of the two options gives the higher payoff to the other player, coincides with the option in which the subject who sends the message receives a lower payoff. This may create an incentive to lie. Different payoff pairs with more or less equal advantageous or disadvantageous payoff differences for the subject are tested. A final treatment (from the original Gneezy study) tests how the decision of a car seller is viewed who wants to sell a lemon and does not want to tell the potential buyer of a problem the buyer will eventually face. There is the option to indicate how fair or unfair the hiding of information is viewed in two scenarios that depend on the cost of the repair. In the experiment, subjects exposed to the empathy treatment are 9 percentage points less likely to deceive and this effect is coming primarily when subjects are nearly indifferent between outcomes, where the effect size becomes 18 percentage points and increases to 24 percentage points with the inclusion of controls. This is sizeable relative to the control group baseline of 54% choosing to send the true message.

2 Model

Like Koszegi and Rabin (2006), I posit a reference-dependent kind of utility. To distinguish between the two forms of empathy, I adapt a reference-point dependent model of social behavior where individuals maximize a three-term utility function: a consumption utility term and two “social” terms that can be affected by emotions. One social term captures a preference for desert (individuals getting what we think they deserve) and the other term a preference for the satisfaction of other’s expectations (i.e., them getting what they think they deserve) (or more precisely, the individuals’ perceptions of others’ expectations). The second social term involves beliefs about the other individual’s loss aversion with respect to a reference point.

One way to model differences in judges’ decision-making is through shifts in their reference points about what is the just and fair decision, given the circumstances. A decision-maker might be a judge or prosecutor who has to determine what is the fair sentencing decision or sentence to charge. Being too harsh or too lenient, from their perspective, is undesirable. The first social term is simply maximization. As the *Stanford Encyclopedia of Philosophy* (Alexander and Moore 2012) puts it, “For the consequentialist, if one’s act is not morally demanded, it is morally wrong and forbidden.”

On the other hand, a defendant experiences gains with sentencing leniency, and experiences losses from being treated unfairly. The second social term is in the spirit of prospect theory.[7] These reference points may shift consciously or unconsciously.[8] Kahneman (1992) defines reference points as being characterized by abrupt changes in the valuation of gains and losses and of acceptable or reprehensible behavior, whereas in contrast, norm theory is applied to the treatment of mixed feelings about outcomes to which multiple reference points are relevant. The proposed utility function formalizes this treatment.

Given insights from prospect theory (Kahneman and Tversky 1979) and the insights from the sociology literature on comparison theory, formulating other-regarding preferences in terms of reference points seems well motivated. Shaw et al. (2011)'s finding that combining incentives with social comparisons was most effective in incentivizing individuals in a field experiment is supportive of the relevance of reference points.[9] The model is introduced and motivated in Chen (2017c) and applied to understand why judicial decisions are more lenient when sentencing defendants on their birthday in the U.S. and in France (if the defendant is present at the trial) (Chen and Philippe 2017). The innovation here is to use the model to study empathy.

Suppose there are two individuals playing a dictator game, Proposer and Receiver. Proposer has to decide on a transfer $x \in \{0, w\}$ to give to the receiver. Since the decision-maker's choice is to lie or tell the truth, the action space is binary. I present the slightly more general formulation, $x \in [0, w]$, to show how the curvature of deviating from a bliss point drives the intuition for why empathy can decrease pro-social behavior.

The utility of the giver ("Gabriel" in Figure 1) from making a transfer is given by Equation 1.

$$U_g(x) = u(w - x) + \alpha \cdot u_g(x - x_g^{RP}) + \beta \cdot v_r(x - x_r^{RP}) \tag{1}$$

Where w is her initial wealth, x_g^{RP} is her subjective reference point of what is the just transfer in this situation and x_r^{RP} is her best estimate or belief about what the Receiver ("Randy")'s reference point is for the game. Gabriel's belief about what Randy deserves (i.e. her belief about his reference point),

[7] For example, a judge may think a defendant deserves 10 months sentence length, so 12 months or 8 months would be worse according to the first social term of "just deserts". A defendant may think he deserves 6 months sentence length, so 7 months is a loss, whereas 4 months is a gain. The positive utility from the second social term can increase for gains relative to this reference point.

[8] A large collection of findings on the malleability of moral reasoning by judges have been documented in U.S. federal circuit judges (Chen 2017b; **?**), federal district judges (Chen 2017a; Barry et al. 2016), immigration judges (Chen et al. 2016d), sentencing judges (Chen and Prescott 2016), and juvenile judges (Eren and Mocan 2016). Some of these findings can be attributed to snap judgments whether from analysis of the first three seconds of oral arguments (Chen et al. 2016a; Chen et al. 2017a) or from early predictability of judicial decisions based on race or nationality (Chen et al. 2016b; Chen and Eagel 2016). In asylum decisions, the defendant reference points may play a role, and when individuals feel they are not being treated justly or fairly, the perceived legitimacy of legal institutions is affected (Chen 2017d).

[9] I use expectation or reference point in a very general sense and not in the strict mathematical sense; an expectation might mean that the other person follows a certain custom or norm. Outside the lab, conceptions of human rights may also hinge on the context, for example, on rights pertaining to sexual harassment and violence (Chen 2005; Chen and Sethi 2016) or for repugnance norms (Chen 2015a). The malleability of injunctive norms to formal institutions is suggestive of the relevance of reference points (Chen 2015b; Chen and Lind 2016; Chen 2016).

irrespective of the costs that would be incurred in actually making this transfer, is captured by x_g^{RP}. Randy's actual beliefs about what he deserves are x_r^{RP}.

The model assumptions are:

- The consumption utility term is increasing and is strictly concave: $u'() > 0$, $u''() \leq 0$
- The desert utility function is concave and is maximized at the desert point: $x_g^{RP} = argmax_x u_g(x, x_g^{RP})$

- α and β are assumed to be positive.
- The $v_r()$ function is concave in gains and is zero when the Receiver receives exactly what Giver believes he believes he should receive: $v_r(x_r^{RP}) = 0$
- If $x > x_r^{RP}$, then $v_r(x - x_r^{RP}) > 0$ and $v_r''() \leq 0$, otherwise $v_r(x - x_r^{RP}) \leq 0$, $v_r''() \geq 0$.

These assumptions generate some very simple comparative statics results for the effect of empathy on pro-social decisions. For ease of notation, let $x_r^{RP} = x_r$ and $x_g^{RP} = x_g$. For the basic results, let each second-derivative be negative so that $u''(w - x^*) + \alpha \cdot u_g''(x^* - x_g) + \beta \cdot v_r''(x^* - x_r) < 0$. After the results are presented in the figure, this assumption is relaxed and discussed.

Lemma 1: The optimal transfer x^* is strictly increasing in what the Giver believes the Receiver deserves, or: $\delta x^* / \delta x_g > 0$. Thus, the first component of empathy that focuses on motivational impact (as triggering helping behavior, what I call "sympathy") increases pro-social behavior. Sympathy increases the deservingness of the recipient, which increases pro-sociality.

Lemma 2: The optimal transfer x^* is strictly increasing in what the Giver believes the Receiver believes he deserves, or: $\delta x^* / \delta x_r > 0$.

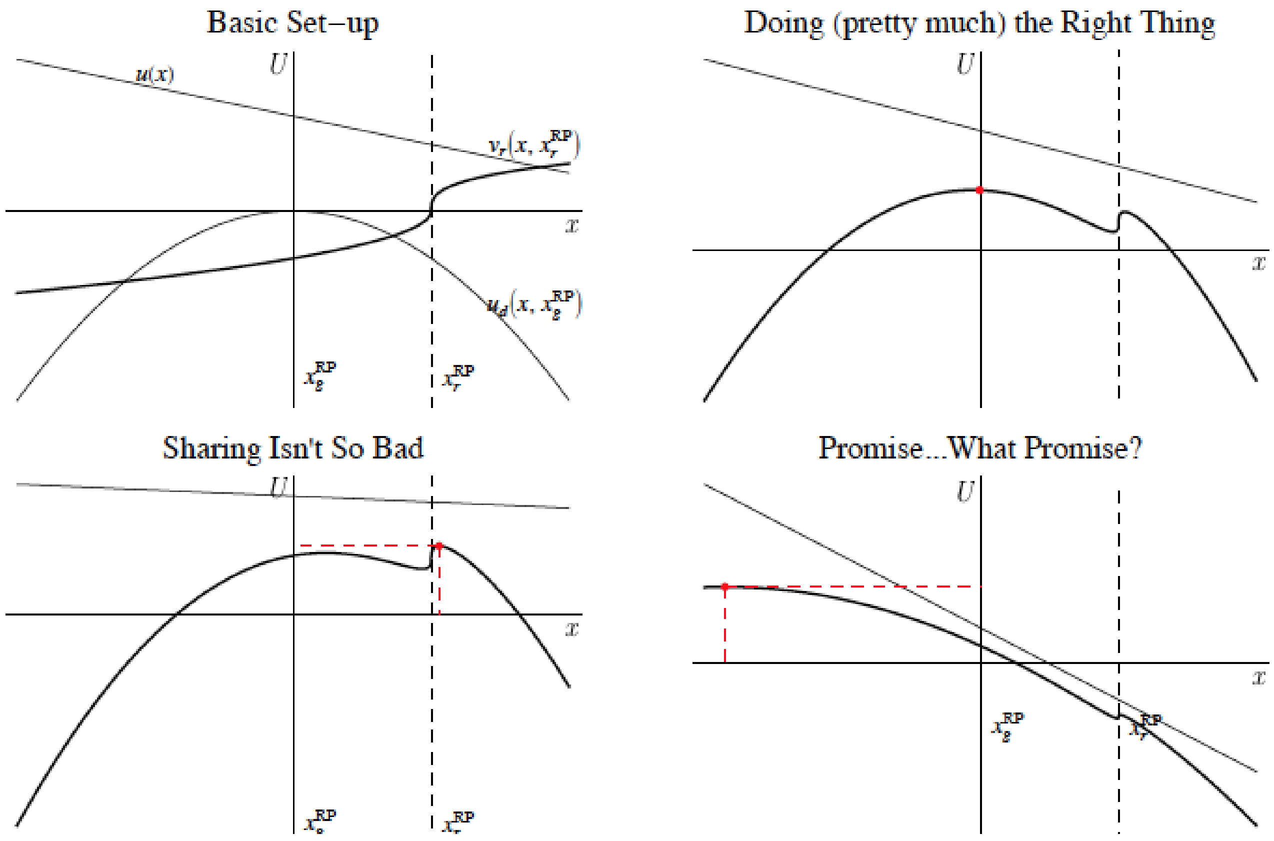
These lemmas are illustrated in Figure 1.

The figure shows that the Receiver is loss averse relative to his reference point. If $x < x_r^{RP}$, then $v_r(x - x_r^{RP}) \leq 0$, $v_r''() \geq 0$, and the proof is slightly more complicated.

Loss-Averse Receiver:

If the second derivative of $v_r()$ is positive enough, then $u''(w - x^*) + \alpha \cdot u_g''(x^* - x_g) + \beta \cdot v_r''(x^* - x_r) > 0$ can lead to the opposite conclusion for Lemma 1. The intuition can be illustrated by supposing the Giver begins in the lower-left quadrant of the figure ("Sharing Isn't So Bad"), and has made a decision at the Recipient's reference point. Now, the optimal decision can decrease as the Recipient's reference point increases. This is because the Giver has a concave cost of deviating from the optimal decision. As the Recipient's reference point increases (such that the shape emulates the upper-right quadrant of the figure and $v_r()$ shifts rightward), the Giver may decide "what-the-hell" and jump to a lower optimum rather than placate the Recipient. This is consistent with situations when the Recipient asks or expects

Figure 1.— Sympathy and Empathy



Basic Set–up

Doing (pretty much) the Right Thing

Sharing Isn't So Bad

Promise...What Promise?

Note: The upper-left panel shows all the component utility terms as well as Gabriel's (the giver) desert reference point for Randy (the receiver) and Randy's own reference point. In the remaining three panels, Gabriel's composite utility $U()$ is shown as well as her utility-maximzing transfer (a red dot). In the upper-right panel, the desert term dominates and Gabriel transfers what she believes Randy deserves according to her own estimate. From Randy's perspective, this action looks like a punishment and he is angry. In the lower-left panel, the slop of the consumption utility is smaller and hence Gabriel transfers a "gift" (from her perspective) to Randy, which he finds to be slightly in the domain of gains. In the lower-right panel, the slope of the consumption utility is just too steep and Gabriel cannot resist cheating Randy and giving him less than even she thinks he deserves.

too much, leading the Giver to ignore the Recipient's expectations. In the upper-right quadrant of the figure, the Giver is simply following what the Giver thinks is the right thing to do. What happens is that the marginal cost of meeting the Recipient's expectations increases (as the Recipient's reference point increases) with pressure from the desert and consumption utility terms to do what the Giver thinks is just.

In Lemma 2, if the Giver's reference point decreases, the optimal decision may jump to a lower optimum. This is because the numerator, $v_r''() \geq 0$, so if the Giver is loss averse enough and $u''(w-x^*) + \alpha \cdot u_g''(x^* - x_g) + \beta \cdot v_r''(x^* - x_r) > 0$, the original prediction holds. The intuition can be seen by supposing the Giver begins in the lower-left quadrant of the figure. As the Giver's reference point decreases and $u_g(x, x_g^{RP})$ shifts leftward, the Giver's utility at x_r^{RP} decreases more sharply than the utility for some decision less than x_r^{RP} (such that the shape emulates the upper-right quadrant of the figure). That is, the marginal cost of meeting the Recipient's expectations increases with pressure from the desert term. Thus, if the Recipient is sufficiently loss averse, decreasing the reference point of the Giver leads the Giver to decide "what-the-hell" and jump to a lower optimum, what the Giver thinks is right and just.

Another way to model empathy is increased certainty about the Receiver's reference point. Certainty about reference points renders a steeper $v_r()$. The cost of deviating from the bliss point becomes more concave if the decision-maker is at the Receiver's reference point, which predicts a null effect (or weakly positive one). In contrast, indifference (lack of recognition-respect) is not knowing where is the reference point, which makes the loss aversion at the reference point become less concave. As the concave cost of deviating from the bliss point decreases, the decision-maker can jump to a less pro-social outcome – or be more susceptible to behavioral biases. One seemingly counter-intuitive implication of the model is that we might actually punish more harshly those people who are socially "close", and do so when indifferent to the Receiver's outcome. Judges sentence defendants more harshly when they share the same first initial and this effect is larger when the defendants are classified as "Negro" by the New Orleans District Attorneys office (Chen and Prescott 2016). Throughout, I assume α and β to be positive. This can be modified for more complex analysis of group conflict or bias or heterogenous types. Movement towards one reference point or the other can derive from shifting the weights on different terms in the utility function.

3 Methodology

The methodology is similar to what the author has developed in other studies (Chen 2016; Chen and Horton 2014). I recruit workers through Amazon Mechanical Turk, where tasks are often done multiple times by different workers for quality-control purposes. Amazon Mechanical Turk ensures the same person does not do the same task more than once by preventing unique worker IDs from accepting the same task. It also prevents users from generating multiple worker IDs by using e-mail addresses, IP addresses, and in some cases, bank accounts. These measures prevent workers from entering the experiment more than once. Hundreds of thousands of jobs are posted each day. The behavior of subjects on MTurk is comparable to the behavior of subjects in a laboratory and may be comparable to subjects in a real labor market (Barankay 2010).

MTurk is designed to recruit a large number of workers in a short amount of time. Through an interface, registered users perform tasks posted by buyers for money. The tasks are generally simple to do for humans yet difficult for computers. Common tasks include captioning photographs, extracting data from scanned documents, and transcribing audio clips. A buyer controls the features, the contract terms of the tasks, and the number of times the buyer wants a task completed. Workers, who are identified to buyers only by a string of letters and numbers, can inspect tasks and the offered terms before deciding whether to complete them. Buyers can require workers to have certain qualifications, but the default is that workers can accept a task immediately and begin work. Once workers submit their work, buyers can approve or reject their submission. If the buyer approves the work, MTurk pays the worker with escrow funds provided by the buyer; the worker is paid nothing if the buyer rejects his work.

MTurk also allows the buyer to implement randomization although randomization is not intrinsic to the platform.[10] Buyers usually post tasks directly, but they can also host tasks on an external site. I use this external hosting method; I post a single placeholder task containing a description of the work and the link for workers to follow if they wish to participate. The subjects are then randomized, via stratification in the order in which they arrived at the job, to one of the several treatments.

The experimental task was a real-work task (though not very taxing): subjects were shown a scanned image of text and asked to do a transcription. In this experiment I investigate the effect of exposing workers to transcribe a paragraph with a specific content. The workers are told that the data entry is to ensure that they are human and not computers.

> **Empathy:** For the 60 soldiers of the Army's 601st Area Support Medical Company who deployed Saturday evening from Fort Bragg to Afghanistan, a holiday weekend of Thanksgiving ended with tears of uncertainty and worry. Deployment is difficult both for those who go to do their duty and for loved ones left behind. Soldier Zachary Hogan left when his wife was pregnant and got back when the child was about four months old. His son is about 18 months now, and it is even more difficult to leave his family now with the holidays right around the corner. The soldiers of the 601st are expected to be gone for about a year.
>
> **Guile:** The defense in the Glenn Agliotti trial on Wednesday lashed restaurateur Alexis Christopher for omitting vital information about a "clandestine" meeting he arranged from his statement to the prosecution. Advocate Laurence Hodes SC slammed Christopher for failing to mention that he arranged a "clandestine" meeting with Agliotti's ex-wife Vivian and slain mining magnate Brett Kebble's head of security, Clinton

[10] It was not intrinsic to the platform in 2009 when this project began.

> Nassif. Nassif was fingered by three State witnesses as having arranged the "muscle" to kill Kebble. Christopher raised his voice, shook his head and waved off Hodes' barrage of questions over his role in the meeting. During the meeting, Nassif told Vivian to give her ex-husband a message.
>
> **Control:** Virginia Department of Transportation documents show the state has a big number of aging bridges. The Daily Press reported Sunday that the agency found more than 80 structurally deficient bridges in the Hampton Roads area alone. Statewide, the agency says 9 percent of bridges and culverts – or about 1,800 of them – are deficient. Mal Kerley, the agency's chief engineer, says the bridges aren't dangerous. He says the agency would close any dangerous bridge. Kerley says many bridges built decades ago are starting to show their age, but must compete with other needs for money. Kerley estimates it would cost $4 billion to bring all Virginia bridges up to standards.

This task is an omnibus treatment. It is challenging to empirically assess the claims of legal scholarship on the effect of literature. Even with random assignment of judges allowing different types of literature (Chen and Yeh 2016), it is difficult to isolate the specific channel through which these decisions (much less which literatures, or which aspect of the literature) affects societal attitudes.

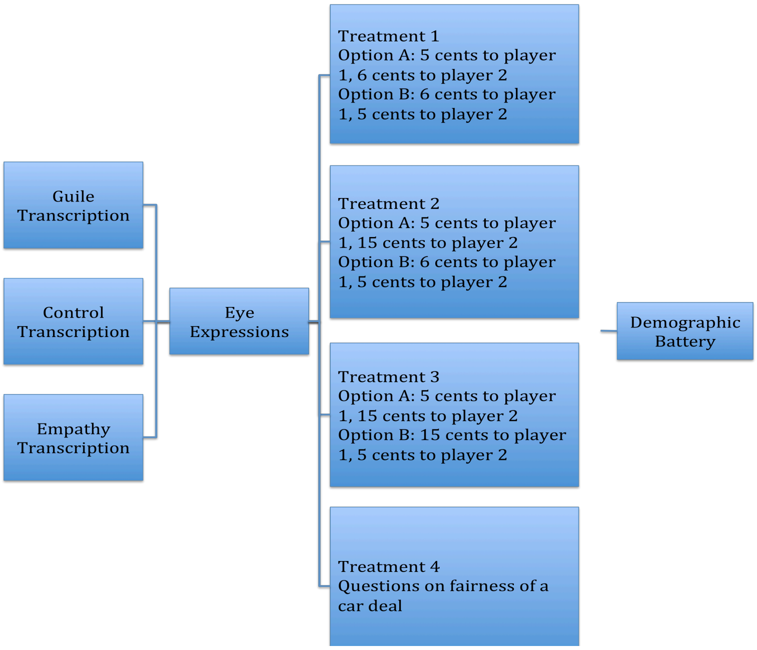
The workers are then given a set of 36 eye expressions and asked to gauge the feelings portrayed. The scores (eyescore) from these expressions are used to assess if the text was successful in inducing empathy. These scores are calculated such that those who are correct in identifying the expression get a score of 1. Hence getting all expressions correct will fetch 36 points. Higher points are supposed to reflect higher empathy.

Next, I reveal a second tier of treatments to assess the effect of the text on empathy for fellow workers (see Figure 2). In this case, I use the two-player communication game (Gneezy 2005) where Player 1 has private information and Player 2 takes an action. Player 2's action may or may not be based on the message Player 1 chooses to send. Payoffs to the players depend on the action chosen and not on the message sent. The two monetary distributions are A and B. Workers are informed about the monetary consequences of each option and they can choose to send one of the two messages to the other worker.

> Message 1: "Option A will earn you more money than Option B."
>
> Message 2: "Option B will earn you more money than Option A."

The treatments are constructed such that Option A always gives Player 1 less than Option B and the reverse for Player 2. Hence, Message 1 is true and Message 2 is false. The actual payoffs used in the original experiment are presented in Appendix Table 1. In one treatment, which I label as "Nearly Indifferent" to facilitate reading, the incentives to lie are almost non-existent with allocations (5, 6) vs. (6, 5). In a second treatment, which I label as "Lying Hurts Others", the allocations are (5, 15) vs. (6, 5). The difference between the truth and the lie is magnified for Player 2. In a third treatment, which I label as

Figure 2: Experimental Design


Guile Transcription
Control Transcription
Empathy Transcription
Eye Expressions
Treatment 1
Option A: 5 cents to player 1, 6 cents to player 2
Option B: 6 cents to player 1, 5 cents to player 2
Treatment 2
Option A: 5 cents to player 1, 15 cents to player 2
Option B: 6 cents to player 1, 5 cents to player 2
Treatment 3
Option A: 5 cents to player 1, 15 cents to player 2
Option B: 15 cents to player 1, 5 cents to player 2
Treatment 4
Questions on fairness of a car deal
Demographic Battery

"Advantageous to Lie", the allocations are (5, 15) vs. (15, 5). Lying is strongly advantageous for Player 1.

In the communication game, subjects are randomly assigned to the role of Player 1 or Player 2. The original experiment is programmed in Surveygizmo and the replication is programmed in oTree (Chen et al. 2016e).[11] The first survey ran in 2011 and the replications ran in 2017.[12] In the third launch, the stakes were raised by a factor of four. In addition, there was a participation fee in the form of payment for data entry, which was 20 cents in launches 1 and 2 and 80 cents in launch 3.

In addition, a few workers are administered a fourth treatment, or Treatment 4 where participants judged the fairness of lying. In this treatment, workers were given the following scenarios (Gneezy 2005):

> "Mr. Johnson is about to close a deal and sell his car for $1,200. The engine's oil pump does not work well, and Mr. Johnson knows that if the buyer learns about this, he will have to reduce the price by $250 (the cost of fixing the pump). If Mr. Johnson doesn't tell the buyer, the engine will overheat on the first hot day, resulting in damages of $250 for the buyer. Being winter, the only way the buyer can learn about this now is if Mr. Johnson were to tell him. Otherwise, the buyer will learn about it only on the next hot day. Mr. Johnson chose not to tell the buyer about the problems with the oil pump. In your opinion, Mr. Johnson's behavior is: *Completely Fair; Fair; Unfair; Very Unfair.*"

[11] The experiment is available here: https://glacial-bayou-38696.herokuapp.com/demo/.

[12] In the first survey, 447 completed the RMET exam and 415 completed the entire survey. There are 148 subjects and 267 subjects in 2017 replications. Attrition was also low in 2017.

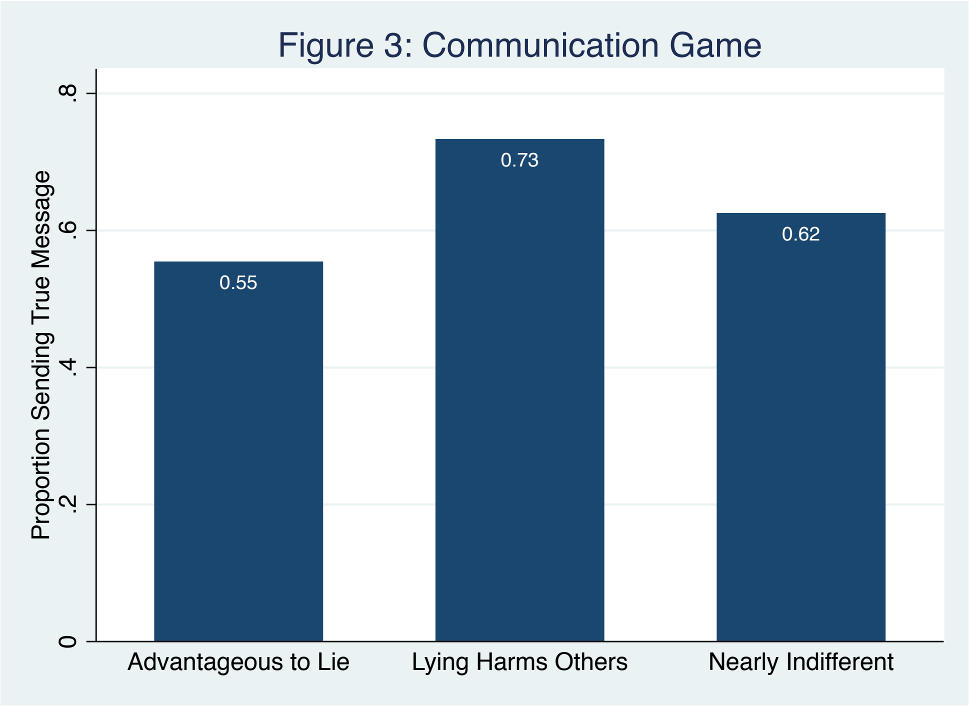


Figure 3: Communication Game

> “What would your answer be if the cost of fixing the damage for the buyer incase Mr. Johnson does not tell him is $1,000 instead of $250? Mr. Johnson’s behavior is: *Completely Fair; Fair; Unfair; Very Unfair.*”

In the two scenarios, there is no difference in the seller’s payoffs but the buyer’s cost increases from $250 to $1000. In the tables, this is labeled as “Very Unfair to Not Disclose .. Minor Harm” and “Major Harm” to facilitate reading the results.

Following these questions, I ask for demographic characteristics including gender, age, religion (categories in the subsequent regressions are Christian, Hindu, Muslim, Atheists and the omitted category is Other), and frequency of religious attendance (never, once a year, once a month, once a week, or multiple times a week; coded as 1-5 respectively).

After work has been completed, according to the original expiry date listed on MTurk, bonuses are calculated and workers are notified of their earnings. The analysis examines the effect of treatment on outcomes, be it eyescore, sending Message 1 (the truthful response), or deeming Mr. Johnson’s behavior to be very unfair.

4 Results

Table 1 presents summary statistics of the raw data for the pooled data and Table 2 presents the results in a regression. From the raw data, Table 1 shows that the overall results are intuitive and along the lines of the Gneezy study. For example, in Treatment 1 (Nearly Indifferent), when the incentive to lie is small, the proportion of people sending the true message is 62% (Figure 3 displays the mean).

In Treatment 2 (Lying Hurts Others), the loss to player 2 was increased to 10 times the gain to player 1 from lying. In this case, the proportion of people sending the truth increase dramatically to 73%. In Treatment 3 (Advantageous to Lie), where the gains to the sender from lying and the loss to player 2 were both large, the proportion of people sending the true message is 55%.

The results are in line with the Gneezy study. In Gneezy's interpretation, the differences across Treatments 1-3 reflect the effects of consequences on behavior: if subjects perceive the gains by lying to be small and the loss to the other party to be disproportionately higher, subjects are less likely to lie.

Turning to Treatment 4, in the first question (fairness of not disclosing minor harm), 51% chose very unfair. In the second question (fairness of not disclosing major harm), the proportion of people choosing very unfair increased to 64%. To be sure, the question is a hypothetical scenario and thus warrants a grain of salt (people may report decisions that are more "moral" than when decisions are actually carried out and treatment effects can be larger[13]). However, the fact that the size of the damages affects the share of subjects reporting that it is very unfair for the car salesman to hide a significant product deficiency is validation that subjects do pay attention in the study.

Table 1 displays remaining summary statistics. As seen in Column 13, the average age in the sample is 31 years. Males comprise 52% of the sample. 32% are Christian, 15% are Atheists, 29% are Hindus and 7% are Muslims. The average religious attendance is between once a year and once a month. Notably, demographic characteristics are balanced across treatment groups, and this is consistent with the randomization of workers across treatment and non-differential attrition across groups.

Table 2 reports the results of the experiment. The empirical specification examines the effect of treatment on outcomes, be it eyescore, sending Message 1 (the truthful response), or deeming Mr. Johnson's behavior to be very unfair:

$$(2) \qquad Outcome_i = \beta_0 + \beta_1 Treatment_i + \beta_2 X_i + \varepsilon_i$$

$Treatment_i$ represents the treatment group for individual i and X_i represents individual demographic characteristics.

Table 2 Panel A reports the results for the full sample. Column 2 reports the pooled results for all treatments in the communication game.[14] It shows that individuals in the empathy treatment are 9 percentage points more likely to select the truthful message and this effect is statistically significant at the 10% level. Interestingly, this effect occurs when subjects are most indifferent to lying or saying the truth. As seen in Column 3, this effect is largely due to Treatment 1 ("Indifferent"), where there are small differences in payoffs for the two players. The effect increases to 18 percentage points and

[13] For formal proof, see Chen and Schonger (2016; 2017).

[14] The number of subjects in Columns 2 and 6 add up to 830 instead of 862 because in the first survey, 447 completed the RMET exam and 415 completed the entire survey, resulting in 32 fewer subjects.

Table 1: Summary Statistics

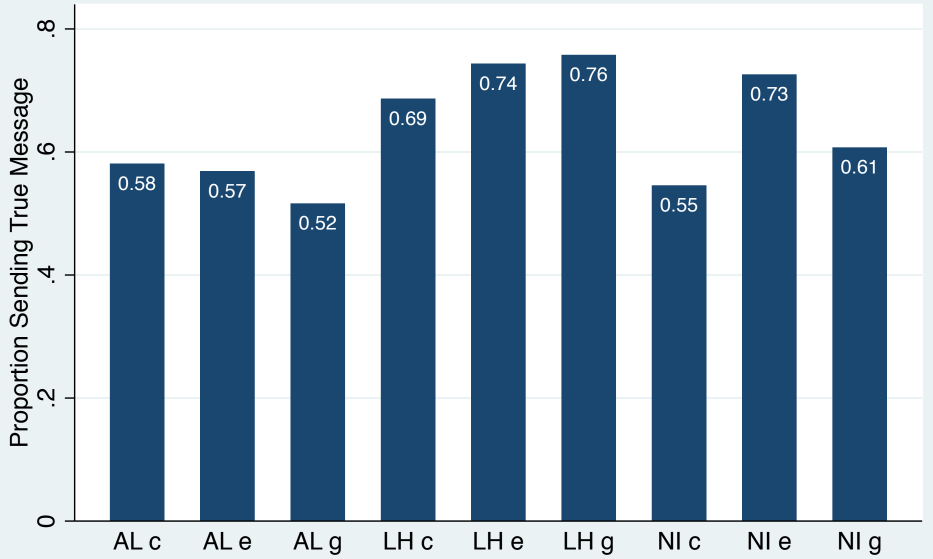
	Nearly Indifferent			Lying Hurts Others			Advantageous to Lie			Non-Disclosure in Sales			
Treatment:	(5,6) vs. (6,5)			(5,15) vs. (6,5)			(5,15) vs. (15, 5)			Vignette			
Prime:	Control	Empathy	Guile	Control	Empathy	Guile	Control	Empathy	Guile	Control	Empathy	Guile	Total
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Eyescore	21.09	22.05	21.50	19.30	21	21.50	22.02	21.67	21.26	22.79	21.72	21.76	21.53
	(7.390)	(6.622)	(7.437)	(8.836)	(7.258)	(7.272)	(6.899)	(7.608)	(7.231)	(6.372)	(6.967)	(7.269)	(7.222)
True Message	0.545	0.726	0.607	0.686	0.743	0.758	0.581	0.569	0.516				0.640
	(0.502)	(0.450)	(0.493)	(0.469)	(0.440)	(0.432)	(0.497)	(0.500)	(0.504)				(0.480)
Very Unfair 1										0.447	0.552	0.533	0.514
(Minor Harm)										(0.500)	(0.500)	(0.502)	(0.501)
Very Unfair 2										0.682	0.629	0.611	0.639
(Major Harm)										(0.468)	(0.486)	(0.490)	(0.481)
Accuracy	0.812	0.889	0.724	0.768	0.821	0.662	0.905	0.796	0.800	0.941	0.872	0.787	0.820
	(0.394)	(0.317)	(0.451)	(0.426)	(0.386)	(0.477)	(0.296)	(0.407)	(0.403)	(0.237)	(0.336)	(0.411)	(0.384)
Age	31.20	31.07	30.36	29.11	29.70	30.93	32.91	32.40	31.23	32.19	32.75	31.56	31.40
	(10.34)	(8.948)	(9.081)	(10.01)	(8.011)	(8.316)	(12.38)	(10.16)	(8.827)	(10.40)	(10.08)	(10.13)	(9.792)
Male	0.435	0.571	0.534	0.500	0.462	0.559	0.413	0.407	0.646	0.553	0.541	0.543	0.517
	(0.499)	(0.499)	(0.503)	(0.505)	(0.502)	(0.500)	(0.496)	(0.496)	(0.482)	(0.500)	(0.501)	(0.501)	(0.500)
Christian	0.290	0.222	0.379	0.321	0.372	0.221	0.317	0.315	0.277	0.435	0.330	0.351	0.324
	(0.457)	(0.419)	(0.489)	(0.471)	(0.486)	(0.418)	(0.469)	(0.469)	(0.451)	(0.499)	(0.472)	(0.480)	(0.468)
Hindu	0.348	0.222	0.207	0.304	0.231	0.397	0.286	0.241	0.354	0.247	0.312	0.330	0.292
	(0.480)	(0.419)	(0.409)	(0.464)	(0.424)	(0.493)	(0.455)	(0.432)	(0.482)	(0.434)	(0.465)	(0.473)	(0.455)
Muslim	0.0580	0.159	0.103	0.0536	0.0513	0.0735	0.0952	0.0556	0.0615	0.0353	0.0642	0.0426	0.0684
	(0.235)	(0.368)	(0.307)	(0.227)	(0.222)	(0.263)	(0.296)	(0.231)	(0.242)	(0.186)	(0.246)	(0.203)	(0.253)
Atheist	0.130	0.238	0.138	0.0714	0.115	0.147	0.143	0.148	0.169	0.176	0.138	0.128	0.145
	(0.339)	(0.429)	(0.348)	(0.260)	(0.322)	(0.357)	(0.353)	(0.359)	(0.378)	(0.383)	(0.346)	(0.335)	(0.352)
Religious	2.420	2.397	2.414	2.214	2.218	2.338	2.397	2.407	2.277	2.412	2.459	2.564	2.386
Services	(1.675)	(1.498)	(1.590)	(1.637)	(1.543)	(1.599)	(1.632)	(1.608)	(1.452)	(1.383)	(1.398)	(1.583)	(1.534)
Launch #2	0.116	0.175	0.155	0.107	0.154	0.132	0.222	0.111	0.123	0.224	0.220	0.234	0.172
	(0.323)	(0.383)	(0.365)	(0.312)	(0.363)	(0.341)	(0.419)	(0.317)	(0.331)	(0.419)	(0.416)	(0.426)	(0.377)
Launch #3	0.275	0.222	0.328	0.214	0.256	0.294	0.222	0.370	0.292	0.435	0.349	0.372	0.310
	(0.450)	(0.419)	(0.473)	(0.414)	(0.439)	(0.459)	(0.419)	(0.487)	(0.458)	(0.499)	(0.479)	(0.486)	(0.463)
Observations	69	63	58	56	78	68	63	54	65	85	109	94	862

Table 2: Main Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A	Eyescore	True Message	True Message	True Message	True Message	Very Unfair to Not Disclose	
Full Sample		Full	Indifferent	Harming	Advantageous	Minor Harm	Major Harm
Guile	0.0646	0.0327	0.0617	0.0713	-0.0645	0.0863	-0.0712
	(0.612)	(0.0504)	(0.0876)	(0.0830)	(0.0899)	(0.0757)	(0.0729)
Empathy	0.130	0.0921*	0.180**	0.0570	-0.0120	0.105	-0.0538
	(0.603)	(0.0502)	(0.0853)	(0.0810)	(0.0946)	(0.0730)	(0.0703)
N	862	550	184	191	175	280	280
R-sq	0.000	0.006	0.025	0.004	0.003	0.008	0.004
Panel B							
Non-Indian Workers							
Guile	-0.0380	-2.59e-17	0.158	-0.00664	-0.170	0.132	0.0277
	(0.817)	(0.0650)	(0.106)	(0.112)	(0.116)	(0.0895)	(0.0779)
Empathy	-0.197	0.0605	0.316***	-0.0304	-0.141	0.147*	0.0428
	(0.788)	(0.0626)	(0.106)	(0.102)	(0.116)	(0.0861)	(0.0749)
N	551	339	114	117	108	180	180
R-sq	0.000	0.004	0.074	0.001	0.023	0.019	0.002
Panel C							
Workers with Accurate Entries							
Guile	0.406	0.0815	0.131	0.157*	-0.0405	0.149+	0.00694
	(0.597)	(0.0560)	(0.0987)	(0.0932)	(0.0962)	(0.0803)	(0.0764)
Empathy	0.268	0.118**	0.200**	0.135	-0.0441	0.148**	-0.0349
	(0.569)	(0.0538)	(0.0912)	(0.0865)	(0.101)	(0.0750)	(0.0714)
N	707	453	151	150	152	247	247
R-sq	0.001	0.011	0.032	0.023	0.002	0.020	0.002
Panel D							
Full Sample with Controls							
Guile	0.454	0.0269	0.0635	0.0793	-0.0658	0.136*	-0.0135
	(0.478)	(0.0521)	(0.0908)	(0.0855)	(0.0977)	(0.0724)	(0.0665)
Empathy	0.235	0.0973*	0.240***	0.0662	-0.0603	0.158**	-0.0140
	(0.473)	(0.0523)	(0.0903)	(0.0828)	(0.104)	(0.0700)	(0.0643)
Age	0.114***	-0.00115	0.00188	-0.00638	0.000435	-0.00103	0.00407
	(0.0217)	(0.00241)	(0.00421)	(0.00419)	(0.00430)	(0.00320)	(0.00294)
Male	-1.489***	0.0648	0.0539	0.0725	0.0356	-0.0791	-0.0545
	(0.399)	(0.0442)	(0.0750)	(0.0708)	(0.0883)	(0.0590)	(0.0542)
Christian	-0.125	0.105	0.101	0.200	-0.0728	0.0736	-0.0750
	(0.765)	(0.0858)	(0.154)	(0.125)	(0.175)	(0.110)	(0.101)
Hindu	-4.775***	0.130	0.137	0.220*	-0.0652	-0.307***	-0.418***
	(0.772)	(0.0862)	(0.155)	(0.125)	(0.178)	(0.111)	(0.102)
Muslim	-3.679***	-0.157	-0.205	-0.100	-0.300	-0.159	-0.282*
	(0.997)	(0.107)	(0.177)	(0.171)	(0.226)	(0.160)	(0.147)
Atheist	0.861	0.0652	-0.0363	0.226	-0.0103	-0.0694	-0.0685
	(0.820)	(0.0904)	(0.157)	(0.145)	(0.172)	(0.121)	(0.111)
Services	-0.701***	-0.0309*	-0.0533*	-0.0312	0.00753	-0.0170	-0.0497**
	(0.161)	(0.0175)	(0.0297)	(0.0269)	(0.0365)	(0.0246)	(0.0226)
Launch #2	0.475	0.142**	0.125	0.168*	0.105	0.156**	0.142*
	(0.540)	(0.0612)	(0.107)	(0.0993)	(0.116)	(0.0792)	(0.0728)
Launch #3	-0.276	0.0298	0.155*	-0.0975	0.0412	0.103	-0.0151
	(0.473)	(0.0515)	(0.0887)	(0.0805)	(0.0997)	(0.0740)	(0.0680)
Constant	23.04***	0.562***	0.452**	0.740***	0.575**	0.547***	0.845***
	(1.054)	(0.119)	(0.204)	(0.187)	(0.239)	(0.150)	(0.138)
N	781	509	171	175	163	272	272
R-sq	0.288	0.056	0.127	0.114	0.033	0.160	0.225

Notes: Linear model is presented. Eyescore refers to the score on the Reading-the-Mind-Behind-the-Eyes exam. Standard errors in parentheses: * p<0.10, ** p<0.05, *** p<0.01

Figure 4: Communication Game



AL: Advantageous to Lie, LH: Lying Harms Others, NI: Nearly Indifferent
c: control, e: empathy, g: guile

becomes significant at the 5% level. This difference is also apparent in Figure 4, which breaks out the communication game proportions sending the true message by treatment group.

Turning to the car dealer vignette, treatment has no significant effect. In Panel D, after controls, treatment has an effect relative to control, but there does not appear to be significant differences between the empathy and guile treatments.[15]

As explained earlier, exposure to text does not always affect RMET scores, and RMET can be criticized for using Western emotions and faces. Males RMET scores are 1.67-1.73 points lower than females RMET scores (this is similar to a recent finding by Ridinger and McBride (2015), who explores why males have lower RMET scores). Hindus have 4.7 lower RMET scores, and Muslims have 3.7 points lower RMET scores. Religious service attendance is negatively correlated with RMET scores. In Column 1, there is no significant impact of treatments on RMET score.

In Panel B, in the sample of non-Indians, the empathy treatment increases by 32 percentage points the likelihood to send the truthful message when subjects are nearly indifferent (there are small differences in payoffs for the two players) as can be seen in Column 3. This effect is significant at the 1% level.

Panel C reports the results for the sample of workers who entered the paragraph accurately. Accuracy is measured using the Levenshtein distance – the minimum number of operations needed to transform one string into another. "Operation" is defined as an insertion, deletion or substitution of a single character

[15] The number of subjects in Columns 2 and 6 add up to 830 instead of 862 because in the first survey, 447 completed the RMET exam and 415 completed the entire survey, resulting in 32 fewer subjects.

(Levenshtein 1966). Accuracy is transformed into a dummy variable where the Levenshtein distance greater than 100 is considered inaccurate. I use this cutoff because the data entry was bimodal with some individuals doing their best to enter the entire paragraph while others did not (they would enter a few words, a line, or a free expression). Accuracy increases the likelihood that the subject paid enough attention to be exposed to treatment. On the other hand, causal interpretation is made more challenging because treatment may cause some people to become more accurate. With this caveat in mind, I leave the interpretation of the results to the reader.

Panel D uses the full sample and controls for all available demographic variables, including dummy indicators for the launch.[16] The impact of the empathy treatment on truth telling is significant at the 10% level for the pooled Treatments 1-3 in Column 2 (10 percentage points more likely to say the truth). The largest impact is through Treatment 1, where subjects are nearly indifferent (24 percentage points more likely to say the truth), which is significant at the 1% level. The empathy and guile treatments increase subjects' tendency to think a car salesman hiding information is very unfair when damages are low (14 and 16 percentage points respectively).

I recognize there is debate in the econometrics literature concerning the relative merits of various binary dependent variable models (Angrist and Pischke 2008). I reestimate all baseline results using logit and probit models and estimate similar marginal effects. I also reestimate the RMET regression with a tobit model. The functional form does not affect the results.

5 Discussion

What game do experimental subjects believe they are playing, especially when there is no performance pay? It seems possible that the subject has several potential objectives: Subjects might try to maximize their monetary pay-off within the context of the current game. However, since most of the experiments have fixed participation payments and since tasks are so simple and unambiguous, maximizing pay-offs consists of simply following instructions. Subjects did not always choose the selfish option, and the paragraph data entry is intended to exclude bots.

Subjects might perceive of themselves as playing a game over multiple periods, with "good" work in early periods possibly leading to greater rewards to through repeat business. However, the very narrow-choice set, the small monetary stakes, the absence of promulgated standards etc. might make it very implausible that workers choose to do what they do for strategic, principal-agent reasons.

Subjects might perceive of the experiments, especially those in which payment if fixed and not tied in any meaningful way to performance, to be about "voice." They are given the ability to judge, decide, punish free-riders and reward good behavior and they find exercising this right to have some value.

[16] Some subjects did not report their demographic variables, resulting in fewer observations in this panel.

Further, subjects, who almost certainly perceive of themselves as quasi-employees, might feel a duty to carry out their task, however vague in a manner they think matches their expectations of the right thing to do.

6 Conclusion

Should judges have empathy? Consider a definition of justice as equal treatment before the law and equality based on recognition of difference. We can imagine a set of covariates X that should lead to the same prediction or predictability of outcomes $Y = f(X) + \varepsilon$; the X's should improve predictions. And, there's a set of W's that should not ($y \perp W, var(\varepsilon) \perp W$). We generally think of the W's as immutable and the X's as mutable—as products of choices ($a \to X, a \nrightarrow W$). Though, this view is also changing—W's can also be mutable if they are expressions of one's identity (in the U.S., less so in Europe). In the U.S., public discourse surrounding the role of empathy in legal institutions is suggested by repeated discussions about whether the federal judiciary lacks diversity. Diversity presumably affects the accommodation of protected characteristics W. At the same time, there is increasing use of machine learning in law, whereby algorithms plausibly improve predictions but probably not empathy. Legal scholars theorize that judicial decisions informed by "the literary imagination" and empathy are likely to be sounder and wiser than judgments reached by other means (Nussbaum 1996), but little causal research exists to date.

This paper offers a formal model of empathy in a reference-point dependent model of social behavior where individuals maximize a three-term utility function: a consumption utility term and two "social" terms. One social term captures a preference for desert (others getting what we think they deserve) and the other term a preference for the satisfaction of other's expectations, or to placate them (i.e. them getting what we think they think they deserve). The first component of empathy (what I call "sympathy" – a shift in the decision-maker's reference point) will tend to increase pro-social behavior. The second social term measures the target individuals' anger or gratitude from the interaction which is determined by a value function derived from prospect theory. Empathy's second component is modeled as "emotional theory of mind" (eToM), whether individuals anticipate the *emotions* of another in reaction to certain actions and understand others' reference points. Empathy can increase or decrease pro-social behavior depending on the exact parameters.

In an MTurk experiment, I offered data entry workers different paragraphs for data entry. These paragraphs had content with empathy or its opposite (perpetuating essentialist, colonial, racist, sexist and otherwise dehumanizing depictions) relative to a control group. When workers are exposed to empathy, they became less deceptive towards third parties in an economic deception game. The effects are robust to a variety of controls and sub-samples and are statistically significant at the 1% level.

References

Alexander, Larry, and Michael Moore, 2012, *Stanford Encyclopedia of Philosophy*.

Amaranto, Daniel, Elliott Ash, Daniel L Chen, Lisa Ren, and Caroline Roper, 2017, Algorithms as Prosecutors: Lowering Rearrest Rates Without Disparate Impacts and Identifying Defendant Characteristics âNoisyâto Human Decision-Makers .

Angrist, Joshua D., and Jörn-Steffen Pischke, 2008, *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton University Press).

Ariely, Dan, 2012, *The (Honest) Truth About Dishonesty* (Harper Collins Publishers, New York).

Arruñada, Benito, Marco Casari, and Francesca Pancotto, 2015, Pro-sociality and strategic reasoning in economic decisions, *Frontiers in behavioral neuroscience* 9.

Ash, Elliott, Daniel L. Chen, and Suresh Naidu, 2017, The Impact of Economics on Moral Decision-Making and Legal Thought, Working paper.

Barankay, Iwan, 2010, Rankings and Social Tournaments: Evidence from a Field Experiment, Working paper, University of Pennsylvania, Mimeo.

Baron-Cohen, Simon, Sally Wheelwright, Jacqueline Hill, Yogini Raste, and Ian Plumb, 2001, The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism, *Journal of child psychology and psychiatry* 42, 241–251.

Barry, Nora, Laura Buchanan, Evelina Bakhturina, and Daniel L. Chen, 2016, Events Unrelated to Crime Predict Criminal Sentence Length, Technical report.

Baumeister, Roy F., and Todd F. Heatherton, 1996, Self-Regulation Failure: An Overview, *Psychological Inquiry* 7, 1–15.

Bayer, R-C, and Ludovic Renou, 2016, Logical abilities and behavior in strategic-form games, *Journal of Economic Psychology* 56, 39–59.

Capraro, Valerio, and Giorgia Cococcioni, 2016, Rethinking spontaneous giving: Extreme time pressure and ego-depletion favor self-regarding reactions, *Scientific reports* 6, 27219.

Chen, Daniel, Yosh Halberstam, and Alan Yu, 2017a, Covering: Mutable Characteristics and Perceptions of Voice in the U.S. Supreme Court, *Review of Economic Studies* invited to resubmit, TSE Working Paper No. 16-680.

Chen, Daniel, Yosh Halberstam, and Alan C. L. Yu, 2016a, Perceived Masculinity Predicts U.S. Supreme Court Outcomes, *PLOS ONE* 11, 1–20, e0164324.

Chen, Daniel, Damian Kozbur, and Alan Yu, 2015, Pandering VS. Persuasion? Phonetic Accommodation in the U.S. Supreme Court, Working paper.

Chen, Daniel, and Arnaud Philippe, 2017, Mental Accounting and Social Preferences: Judicial Leniency on Defendant Birthdays .

Chen, Daniel, and J.J. Prescott, 2016, Implicit Egoism in Sentencing Decisions: First Letter Name Effects with Randomly Assigned Defendants, Technical report.

Chen, Daniel L., 2005, Gender Violence and the Price of Virginity: Theory and Evidence of Incomplete Marriage Contracts, Working paper, University of Chicago, Mimeo.

Chen, Daniel L., 2015a, Can Markets Overcome Repugnance? Muslim Trade Reponse to Anti-Muhammad Cartoons, Working paper, ETH Zurich, Mimeo.

Chen, Daniel L., 2015b, Can markets stimulate rights? On the alienability of legal claims, *RAND Journal of Economics* 46, 23–65.

Chen, Daniel L., 2016, Markets, Morality, and Economic Growth: Competition Affects Moral Judgment, TSE Working

Paper No. 16-692.

Chen, Daniel L., 2017a, Mood and the Malleability of Moral Reasoning, TSE Working Paper No. 16-707.

Chen, Daniel L., 2017b, Priming Ideology: Why Presidential Elections Affect U.S. Judges, *Journal of Law and Economics* resubmitted, TSE Working Paper No. 16-681.

Chen, Daniel L., 2017c, Tastes for Desert and Placation: A Reference Point-Dependent Model of Social Preferences, *Research in Experimental Economics* resubmitted, TSE Working Papers No. 16-725.

Chen, Daniel L., 2017d, The Deterrent Effect of the Death Penalty? Evidence from British Commutations During World War I, *American Economic Review* resubmitted, TSE Working Paper No. 16-706.

Chen, Daniel L., Matt Dunn, Rafael Garcia Cano Da Costa, Ben Jakubowki, and Levent Sagun, 2016b, Early Predictability of Asylum Court Decisions, Technical report.

Chen, Daniel L., and Jess Eagel, 2016, Can Machine Learning Help Predict the Outcome of Asylum Adjudications?, Technical report.

Chen, Daniel L., and John J. Horton, 2014, The Wages of Pay Cuts, Revise and Resubmit at Management Information Systems Quarterly, ETH Zurich and New York University.

Chen, Daniel L., Vardges Levonyan, and Susan Yeh, 2017b, Do Policies Affect Preferences? Evidence from Random Variation in Abortion Jurisprudence, *Journal of Political Economy* TSE Working Paper No. 16-723, under review.

Chen, Daniel L., and Jo Thori Lind, 2016, The Political Economy of Beliefs: Why Fiscal and Social Conservatives/Liberals (Sometimes) Come Hand-in-Hand, under review, TSE Working Paper No. 16-722.

Chen, Daniel L., Moti Michaeli, and Daniel Spiro, 2016c, Ideological Perfectionism, TSE Working Paper No. 16-694.

Chen, Daniel L., Tobias J. Moskowitz, and Kelly Shue, 2016d, Decision Making Under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires, *The Quarterly Journal of Economics* 131, 1181–1242.

Chen, Daniel L., and Martin Schonger, 2016, Social Preferences or Sacred Values? Theory and Evidence of Deontological Motivations, *American Economic Journal: Microeconomics* invited to resubmit, TSE Working Paper No. 16-714.

Chen, Daniel L., and Martin Schonger, 2017, A Theory of Experiments: Invariance of Equilibrium to the Strategy Method of Elicitation, TSE Working Paper No. 16-724.

Chen, Daniel L., Martin Schonger, and Chris Wickens, 2016e, oTree—An open-source platform for laboratory, online, and field experiments, *Journal of Behavioral and Experimental Finance* 9, 88 – 97.

Chen, Daniel L., and Jasmin K. Sethi, 2016, Insiders, Outsiders, and Involuntary Unemployment: Sexual Harassment Exacerbates Gender Inequality, invited to resubmit, TSE Working Paper No. 16-687.

Chen, Daniel L., and Susan Yeh, 2014, The Construction of Morals, *Journal of Economic Behavior and Organization* 104, 84–105.

Chen, Daniel L., and Susan Yeh, 2016, How Do Rights Revolutions Occur? Free Speech and the First Amendment, TSE Working Paper No. 16-705.

Corgnet, Brice, Antonio M Espín, and Roberto Hernán-González, 2015, The cognitive basis of social behavior: cognitive reflection overrides antisocial but not always prosocial motives, *Frontiers in behavioral neuroscience* 9.

Cueva, Carlos, Inigo Iturbe-Ormaetxe, Esther Mata-Pérez, Giovanni Ponti, Marcello Sartarelli, Haihan Yu, and Vita Zhukova, 2016, Cognitive (ir) reflection: New experimental evidence, *Journal of Behavioral and Experimental Economics* 64, 81–93.

Darwall, Stephen L, 2006, *The second-person standpoint: Morality, respect, and accountability* (Harvard University Press).

De Waal, Frans BM, 2008, Putting the altruism back into altruism: the evolution of empathy, *Annu. Rev. Psychol.* 59, 279–300.

DeAngelo, Gregory, and Bryan C McCannon, 2017, Theory of Mind predicts cooperative behavior, *Economics Letters* 155, 1–4.

Dittrich, Marcus, and Kristina Leipold, 2014, Gender differences in strategic reasoning .

Djikic, Maja, Keith Oatley, and Mihnea C Moldoveanu, 2013, Reading other minds: Effects of literature on empathy, *Scientific Study of Literature* 3, 28–47.

Engelmann, Jan B, and Todd A Hare, 2017, Question 13: How are emotions integrated into choice?, *The Nature of Emotion* .

Engelmann, Jan B, and Marianna Pogosyan, 2013, Emotion perception across cultures: the role of cognitive mechanisms, *Frontiers in psychology* 4.

Eren, Ozkan, and Naci Mocan, 2016, Emotional Judges and Unlucky Juveniles, Working paper.

Gneezy, Uri, 2005, Deception: The Role of Consequences, *The American Economic Review* 95, 384–394.

Greene, Joshua D., Leigh E. Nystrom, Andrew D. Engell, John M. Darley, and Jonathan D. Cohen, 2004, The Neural Bases of Cognitive Conflict and Control in Moral Judgment, *Neuron* 44, 389–400.

Hauge, Karen, Kjell Arne Brekke, Lars-Olof Johansson, Olof Johansson-Stenman, and Henrik Svedsäter, 2015, Keeping others in our mind or in our heart? Distribution games under cognitive load, *Experimental Economics* .

Hirschman, Albert O., 1982, Rival Interpretations of Market Society: Civilizing, Destructive, or Feeble?, *Journal of Economic Literature* 20, 1463–1484.

Hoffman, Martin L., 2001, *Empathy and Moral Development: Implications for Caring and Justice* (Cambridge University Press).

Johnson, Mark H, Suzanne Dziurawiec, Hadyn Ellis, and John Morton, 1991, Newborns' preferential tracking of face-like stimuli and its subsequent decline, *Cognition* 40, 1–19.

Kahneman, Daniel, 1992, Reference points, anchors, norms, and mixed feelings, *Organizational behavior and human decision processes* 51, 296–312.

Kahneman, Daniel, and Amos Tversky, 1979, Prospect Theory: An Analysis of Decision under Risk, *Econometrica* 47, 263–292.

Kamada, Yuichiro, and Fuhito Kojima, 2014, Voter Preferences, Polarization, and Electoral Policies, *American Economic Journal: Microeconomics* 6, 203–236.

Kelly, David J, Shaoying Liu, Liezhong Ge, Paul C Quinn, Alan M Slater, Kang Lee, Qinyao Liu, and Olivier Pascalis, 2007, Cross-race preferences for same-race faces extend beyond the African versus Caucasian contrast in 3-month-old infants, *Infancy* 11, 87–95.

Kelly, Kristen, Arietta Slade, and John F Grienenberger, 2005, Maternal reflective functioning, mother–infant affective communication, and infant attachment: Exploring the link between mental states and observed caregiving behavior in the intergenerational transmission of attachment, *Attachment & human development* 7, 299–311.

Kidd, David Comer, and Emanuele Castano, 2013, Reading literary fiction improves theory of mind, *Science* 342, 377–380.

Koszegi, Botond, and Matthew Rabin, 2006, A Model of Reference-Dependent Preferences, *The Quarterly Journal of Economics* 121, 1133–1165.

Lakin, Jessica L, and Tanya L Chartrand, 2003, Using nonconscious behavioral mimicry to create affiliation and rapport, *Psychological science* 14, 334–339.

Levenshtein, Vladimir I., 1966, Binary Codes Capable of Correcting Deletions, Insertions, and Reversals, in *Soviet Physics-Doklady*, volume 10, 707–710.

Lohse, Johannes, 2016, Smart or selfish–When smart guys finish nice, *Journal of Behavioral and Experimental Economics*

64, 28–40.

McRobie, Heather, 2014, Martha Nussbaum, empathy, and the moral imagination, *Open Democracy* .

Nussbaum, Martha Craven, 1996, Poetic justice: The literary imagination and public life .

Nussbaum, Martha Craven, 2008, Beyond Toleration to Equal Respect, in *Chicago Unbound*, 100.

Nussbaum, Martha Craven, 2017, Jefferson Lecture in the Humanities.

Osborne, Martin J., 1995, Spatial models of political competition under plurality rule: a survey of some explanations of the number of candidates and the positions they take, *The Canadian Journal of Economics* 28, 261–301.

Panero, Maria Eugenia, Deena Skolnick Weisberg, Jessica Black, Thalia R Goldstein, Jennifer L Barnes, Hiram Brownell, and Ellen Winner, 2016, Does reading a single passage of literary fiction really improve theory of mind? An attempt at replication., *Journal of personality and social psychology* 111, e46.

Panero, Maria Eugenia, Deena Skolnick Weisberg, Jessica Black, Thalia R Goldstein, Jennifer L Barnes, Hiram Brownell, and Ellen Winner, 2017, No support for the claim that literary fiction uniquely and immediately improves theory of mind: A reply to Kidd and Castanoâs commentary on Panero et al.(2016). .

Ponti, Giovanni, and Ismael Rodriguez-Lara, 2015, Social preferences and cognitive reflection: evidence from a dictator game experiment, *Frontiers in behavioral neuroscience* 9.

Ridinger, Garret, and Michael McBride, 2015, Money affects theory of mind differently by gender, *PloS one* 10, e0143973.

Samur, Dalya, Mattie Tops, and Sander L Koole, 2017, Does a single session of reading literary fiction prime enhanced mentalising performance? Four replication experiments of Kidd and Castano (2013), *Cognition and Emotion* 1–15.

Shaw, Aaron D., John J. Horton, and Daniel L. Chen, 2011, Designing Incentives for Inexpert Human Raters, in *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, CSCW '11, 275–284 (ACM, New York, NY. USA).

Simmel, Georg, 1955, *Conflict and the Web of Group Affiliations* (Glencoe, IL: The Free Press).

Sonnby-Borgström, Marianne, 2002, Automatic mimicry reactions as related to differences in emotional empathy, *Scandinavian journal of psychology* 43, 433–443.

Tinghög, Gustav, David Andersson, Caroline Bonn, Harald Böttiger, Camilla Josephson, Gustaf Lundgren, Daniel Västfjäll, Michael Kirchler, and Magnus Johannesson, 2013, Intuition and cooperation reconsidered, *Nature* 498, E1–E2.

Verkoeijen, Peter PJL, and Samantha Bouwmeester, 2014, Does intuition cause cooperation?, *PloS one* 9, e96654.

Wilson, Edward O., 2000, *Sociobiology: The New Synthesis*, twenty-fifth anniversary edition (Belknap Press of Harvard University Press, Cambridge, Massachusetts, and London, England).

Appendix Table 1 – Payoffs Used in the Experiment

		Payoff to	
Treatment	Option	Player 1	Player 2
1	A	5	6
	B	6	5
2	A	5	15
	B	6	5
3	A	5	15
	B	15	5

Treatment 1

This is a short experiment in decision making. In this experiment, you will be matched with another worker. Neither of you will ever know the identity of the other. The money that you earn will be paid to you next week, privately and in cash.

Two possible monetary payments are available to you and your counterpart in the experiment. The two payment options are:

Option A: 5 cents to you and 6 cents to the other worker
Option B: 6 cents to you and 5 cents to the other worker

The choice rests with the other worker who will have to choose either Option A or Option B. The only information your counterpart will have is information sent by you in a message. That is, he or she will not know the monetary payments associated with each choice.

We now ask you to choose one of the following two possible messages, which you will send to your counterpart:

Message 1: “Option A will earn you more money than Option B.”

Message 2: “Option B will earn you more money than Option A.”

We will show the other worker your message and ask him or her to choose either A or B. To repeat, your counterpart’s choice will determine the payments in the experiment. However, your counterpart will never know what sums were actually offered in the option not chosen (that is, he or she will never know whether your message was true or not). Moreover, he or she will never know the sums to be paid to you according to the different options.

We will pay the two of you according to the choice made by your counterpart.

I choose to send (please select one option):

Message 1 Message 2

Treatment 2

This is a short experiment in decision making. In this experiment, you will be matched with another worker. Neither of you will ever know the identity of the other. The money that you earn will be paid to you next week, privately and in cash.

Two possible monetary payments are available to you and your counterpart in the experiment. The two payment options are:

Option A: 5 cents to you and 15 cents to the other worker
Option B: 6 cents to you and 5 cents to the other worker

The choice rests with the other worker who will have to choose either Option A or Option B. The only information your counterpart will have is information sent by you in a message. That is, he or she will not know the monetary payments associated with each choice.

We now ask you to choose one of the following two possible messages, which you will send to your counterpart:

Message 1: "Option A will earn you more money than Option B."

Message 2: "Option B will earn you more money than Option A."

We will show the other worker your message and ask him or her to choose either A or B. To repeat, your counterpart's choice will determine the payments in the experiment. However, your counterpart will never know what sums were actually offered in the option not chosen (that is, he or she will never know whether your message was true or not). Moreover, he or she will never know the sums to be paid to you according to the different options.

We will pay the two of you according to the choice made by your counterpart.

I choose to send (please select one option):

Message 1 Message 2

Treatment 3

This is a short experiment in decision making. In this experiment, you will be matched with another worker. Neither of you will ever know the identity of the other. The money that you earn will be paid to you next week, privately and in cash.

Two possible monetary payments are available to you and your counterpart in the experiment. The two payment options are:

Option A: 5 cents to you and 15 cents to the other worker
Option B: 15 cents to you and 5 cents to the other worker

The choice rests with the other worker who will have to choose either Option A or Option B. The only information your counterpart will have is information sent by you in a message. That is, he or she will not know the monetary payments associated with each choice.

We now ask you to choose one of the following two possible messages, which you will send to your counterpart:

Message 1: "Option A will earn you more money than Option B."

Message 2: "Option B will earn you more money than Option A."

We will show the other worker your message and ask him or her to choose either A or B. To repeat, your counterpart's choice will determine the payments in the experiment. However, your counterpart will never know what sums were actually offered in the option not chosen (that is, he or she will never know whether your message was true or not). Moreover, he or she will never know the sums to be paid to you according to the different options.

We will pay the two of you according to the choice made by your counterpart.

I choose to send (please select one option):

Message 1 Message 2

Treatment 4

Mr. Johnson is about to close a deal and sell his car for $1,200. The engine's oil pump does not work well, and Mr. Johnson knows that if the buyer learns about this, he will have to reduce the price by $250 (the cost of fixing the pump). If Mr. Johnson doesn't tell the buyer, the engine will overheat on the first hot day, resulting in damages of $250 for the buyer. Being winter, the only way the buyer can learn about this now is if Mr. Johnson were to tell him. Otherwise, the buyer will learn about it only on the next hot day. Mr. Johnson chose not to tell the buyer about the problems with the oil pump. In your opinion, Mr. Johnson's behavior is:

Completely Fair
Fair
Unfair
Very Unfair

What would your answer be if the cost of fixing the damage for the buyer incase Mr. Johnson does not tell him is $1,000 instead of $250? Mr. Johnson's behavior is:

Completely Fair
Fair
Unfair
Very Unfair