Uncalculating cooperation is used to signal trustworthiness

Jillian J. Jordan, Moshe Hoffman, Martin A. Nowak, and David G. Rand

Humans frequently cooperate without carefully weighing the costs and benefits. As a result, people may wind up cooperating when it is not worthwhile to do so. Why risk making costly mistakes? Here, we present experimental evidence that reputation concerns provide an answer: people cooperate in an uncalculating way to signal their trustworthiness to observers. We present two economic game experiments in which uncalculating versus calculating decision-making is operationalized by either a subject’s choice of whether to reveal the precise costs of cooperating (Exp. 1) or the time a subject spends considering these costs (Exp. 2). In both experiments, we find that participants are more likely to engage in uncalculating cooperation when their decision-making process is observable to others. Furthermore, we confirm that people who engage in uncalculating cooperation are perceived as, and actually are, more trustworthy than people who cooperate in a calculating way. Taken together, these data provide the first empirical evidence, to our knowledge, that uncalculating cooperation is used to signal trustworthiness, and is not merely an efficient decision-making strategy that reduces cognitive costs. Our results thus help to explain a range of puzzling behaviors, such as extreme altruism, the use of ethical principles, and romantic love.

Humans are exceptional in their willingness to incur personal costs to benefit others, and a great deal of work across the social and natural sciences has sought to understand this exceptionally cooperative behavior (1–16). A central explanation that has emerged is reciprocity: often, the future benefits of cooperation outweigh the present costs, and so it is in your long-run self-interest to cooperate with others (17–21). However, there are also many contexts in which the future benefits are not sufficient to outweigh the immediate costs of cooperating: humans sometimes face the opportunity to cooperate in anonymous settings, or with strangers, or to make huge sacrifices and receive only moderate benefits in return. Thus, theories of reciprocity predict that when given the opportunity to cooperate, people should calculate the costs and benefits, and cooperate only when doing so is worthwhile. In other words, people should constantly be computing in every decision whether cooperating is worth it.

Despite this clear theoretical prediction, however, people often appear to cooperate without calculating the costs and benefits. Friends frequently grant requests to help each other without inquiring about how much time or effort will be involved, and avoid precisely tracking favors (22–24). Intimate relationships often foster strong prosocial emotions, such as devotion and love, that encourage extreme cooperative behavior that is insensitive to costs or contexts (25). People impulsively decide to help strangers in emergencies (26), and there are rich traditions of adhering to ethical principles (27) or religious teachings (28–31) that prescribe rigid guidelines for when cooperation is obligatory, regardless of the costs and benefits to the actor. These diverse examples likely evoke a broad range of proximate psychologies, ranging from intuitive and emotional processes to explicit conscious decisions not to calculate (that may themselves be the result of calculation). However, these various proximate mechanisms all lead to cooperative behavior that is not conditional on the precise cost of cooperating in a specific situation or context—what we term “uncalculating cooperation.”

When considered in the light of reciprocity, uncalculating cooperation is therefore a puzzling phenomenon that makes individuals liable to cooperate in contexts in which they would have been better off defecting. Why should people put themselves at risk of giving too much and receiving too little?

One possible explanation is that people engage in uncalculating cooperation in contexts in which they are willing to pay even the maximum possible cost of cooperation, so calculating is unnecessary. Or, relatedly, so long as cooperation is typically worthwhile, cooperating without calculating the costs can be an efficient “heuristic”: it usually leads to the right decision, and avoids costs associated with calculating (e.g., cognitive costs of deliberation, or time and effort involved in gathering relevant information) (32–34), a proposal that is supported by empirical evidence (e.g., ref. 35). Here, however, we provide the first experimental evidence, to our knowledge, that uncalculating cooperation is more than just an efficient way to make cooperative decisions.

Specifically, we demonstrate that uncalculating cooperation is motivated by reputation concerns: People use uncalculating cooperation to signal their trustworthiness to observers. This hypothesis builds on evidence that people who calculate when presented with the opportunity to behave morally are perceived as less prosocial (27, 36), even when they do ultimately wind up making the “right” decision (37, 38). Calculating behavior is seen as a sign of doubt or uncertainty (37, 39, 40), whereas prosocial decisions that are quick, impulsive, or emotional are seen as reflecting genuine moral goodness (38, 41). As a result of this social cost of
observers perceive uncalculating cooperation as a reliable signal and trust uncalculating cooperators with more money. Finally, we show that this perception is valid: People who reach the same cooperative decisions in an uncalculating way are actually more trustworthy.

In our experiments, we use a two-stage incentivized economic game (Fig. 1). Both experiments are almost identical in design, except for how they operationalize uncalculating versus calculating decision-making. In the first stage of both experiments [the Helping Game (HG)], player A decides whether to pay a cost to benefit a recipient and can make this decision in a way that is calculating or uncalculating (see following for details on how “calculating” is operationalized in each study).

In the second stage of both experiments [the Trust Game (TG)], player B (who was not involved in the HG) receives an endowment and decides how much to send to player A. Any money sent is tripled by the experimenter. Player A then decides what percentage of this tripled amount (if any) to return to player B (without knowing how much he or she has been sent by player B). Critically, then, what is payoff-maximizing for player B depends on the behavior of player A. It is payoff-maximizing for player B to send her entire endowment to player A if player A is trustworthy (i.e., will return a substantial amount, more than a third); conversely, if player A is untrustworthy and returns less than a third, it is payoff-maximizing for player B to send nothing. In contrast, it is never payoff-maximizing for player A to return anything to player B. Thus, the amount player B sends to player A reflects player B’s trust of player A, and the amount player A returns to player B reflects player A’s trustworthiness.

In the process observable condition, player B can condition trust (using the strategy method) both on player A’s stage 1 decision (did player A cooperate?) and decision-making process (did player A make this decision in a calculating or uncalculating way?). In the process hidden condition, player B can condition trust only on player A’s stage 1 decision, and not player A’s decision-making process.

In Exp. 1 (n = 735 player A–player B pairs), we operationalize calculating behavior with a looking choice, loosely inspired by a recent formal model of “cooperating without looking” in an “envelope game” (42, 43). (For further discussion of the relationship between our experiments and this model, see Discussion). In this experiment, we tell player A they have the opportunity to pay a cost to help another player, but that by default they do not know how much that cost is; they only know that it is some fraction of their starting endowment. Player A’s decision-making behavior is measured by whether or not player A chooses to look at the cost (calculating behavior) or to look at the cost before deciding (calculating behavior). Thus, in Exp. 1, calculating behavior is measured by whether or not player A chooses to look at the cost. Then, in the process observable condition, but not the process hidden condition, player B can condition stage 2 trust, via the strategy method, on whether player A looked at the cost.

In Exp. 2 (n = 737 player A–player B pairs), we operationalize calculating behavior with decision times, inspired by experimental evidence that quick cooperative choices are perceived to be more prosocial (37, 39, 40). In this experiment, we tell player A they have the opportunity to pay a cost to help another player, and that initially they do not know how large that cost is, they only know that it is some fraction of their starting endowment. Then, right before they make their decisions, we reveal the exact cost of helping to player A and measure the time they spend deciding, with shorter times reflecting less calculating behavior. Thus, in Exp. 2, calculating behavior is measured by the number of seconds spent making the helping decision (decision times were natural-log transformed because they were highly right-skewed, as in ref. 44). Then, in the process observable condition, but not the process hidden condition, player B can condition stage 2 trust, via the strategy method, on whether player A decided quickly (specifically, on whether player A was faster or slower than the median decision time). We note that player A deciding quickly could reflect either the use of intuitive cognition (45) or a lack of decision conflict (i.e., an “easy choice”...
Uncalculating cooperation is used as a signal of trustworthiness. Player As are more likely to engage in uncalculating behavior in the HG (stage 1) when they know their decision process will be observed by a subsequent partner in the TG (stage 2). (A) Results from study 1 (n = 735), in which we plot proportion of player As choosing to look in the HG. (B) Results from study 2 (n = 737), in which we plot natural-log transformed average decision times for player As in the HG. Error bars indicate ±1 SEM.

Results

Uncalculating Cooperation Is Perceived as a Signal. Our key prediction is that people use uncalculating cooperation for reputational benefits: player A should be more likely to decide in an uncalculating way (and thus risk making a suboptimal decision) in the process observable condition, where doing so can confer reputational benefits, than in the process hidden condition, where doing so cannot. Indeed, we confirm this prediction in both experiments. In Exp. 1, 70% of player As looked at the cost of helping in the process hidden condition compared with only 59% in the process observable condition (n = 735; P = 0.002) (Fig. 2.4). In Exp. 2, the mean player A decision time was 12.50 s (2.28 log-seconds) in the process hidden condition compared with 10.26 s (2.17 log-seconds) in the process observable condition (n = 737; P = 0.014) (Fig. 2B). Thus, we confirm our key prediction: Across two experiments, subjects behaved in a less calculating manner when their reputations were at stake. In Exp. 1, they looked less at the cost of helping when their looking choice was observable, and in Exp. 2, they decided faster when their decision time was observable.

Uncalculating Cooperation Actually Is a Signal. Finally, we show that trusting uncalculating cooperators is, in fact, reasonable. Across both conditions, player As who reach a cooperative decision in an uncalculating way return more to player Bs than player As who reach a cooperative decision in a calculating way. In Exp. 1, player As who helped without looking at the cost returned an average of 50% of the amount they were sent to player B compared with 41% among player As who helped after looking (n = 361; P < 0.001) (Fig. 3, A). In Exp. 2, player Bs sent an average of 22% of their endowments to player As who decided not to help relatively quickly (decision time below the median) compared with 27% to player As who decided not to help relatively slowly (decision time above the median; n = 365; P < 0.001) (Fig. 3B, Right). Thus, across two experiments, we confirmed our prediction that uncalculating cooperation was perceived positively, but uncalculating defection was not.
controlling for general comprehension speed): coefficient

uncalculating cooperators (but not defectors) really do behave in a more trustworthy way than calculating cooperators.

Our key result, that people engage in less calculating behavior when their decision process is observable, provides the first evidence, to our knowledge, that people use uncalculating cooperation for reputational benefits, and not merely as a useful way to reduce the nonsocial costs of calculating (32–35). Although a theory of uncalculating cooperation as merely an efficient decision-making strategy can explain our second and third results (individuals who cooperate across contexts to reduce the nonsocial costs of calculating will end up cooperating more, and thus should be perceived as, and should actually be, more trustworthy), it cannot explain why uncalculating decision-making should decrease when it is not observable. Based only on decision-making efficiency, acting in an uncalculating way should be equally valuable, regardless of who is watching. Thus, the fact that uncalculating decision-making is sensitive to observability suggests it represents a costly strategy that risks making a suboptimal choice, but has the benefit of signaling trustworthiness. This result has important implications for our understanding of the function of uncalculating cooperation, implicating reputational motives. It also suggests boundary conditions for when uncalculating cooperation should be observed. For example, when uncalculating cooperation serves as a reputation strategy, it should be particularly likely when trust is relatively important, as opposed to when the maximum cost of cooperation is relatively small or when the cognitive and temporal costs of calculating are relatively large.

Of course, our reputation-based account is not mutually exclusive with the idea that people are sometimes uncalculating cooperators because they are willing to pay even the maximum cost or because calculating has nonsocial costs; these accounts may help explain why some subjects cooperated in an uncalculating manner even when their decision process was hidden. However, we note that because calculating whether your decision process is observable may itself be observable to others, people may also engage in “meta uncalculating cooperation” (i.e., uncalculating cooperation that is itself uncalculated, and not conditional on whether one’s decision process is observable), which may also help explain uncalculating cooperation in the process hidden condition, and make one especially trustworthy to others.

It is important to note that in the process observable conditions of both experiments, we explicitly inform player A that player B can condition their trust on looking choices/decision times, and thus that there are possible reputation consequences of calculating; conversely, in the process hidden conditions, we explain that there are no possible reputation consequences of calculating. Critically, however, this information about reputational consequences needs not be presented so explicitly to obtain our key result. In the SI Appendix, we present a subtler version of Exp. 1, in which we refrain from directly telling player A what player B can condition their trust on (but instead convey this information indirectly via screenshots of the study from player B’s perspective). This additional study replicates our key finding that player A is less calculating when they know their decision process is observable. These results suggest that in real-world contexts of interest, when subjects are aware that their decision processes are observable (via a range of real-world observability cues, which may typically not be explicit), they are likely to act on this information by making cooperative decisions in a less calculating way. See SI Appendix for details.

Our second result (that people preferentially trust uncalculating cooperators) demonstrates that engaging in uncalculating decision-making can be worth the costs: Uncalculating cooperators receive reputational benefits in the form of increased trust. These findings add to a growing body of evidence showing that people attend to whether decisions are made in a calculating or uncalculating way. In one prior study, for example, subjects judged characters more positively if they made prosocial decisions without hesitation, because their decisions were perceived as more certain (37), which fits with

Fig. 4. Uncalculating cooperation actually is a signal of trustworthiness. Across both conditions, player A who engaged in uncalculating cooperation are more trustworthy than player A who engaged in calculating cooperation, but uncalculating behavior does not predict trustworthiness among player A who chose not to help. (A) Results from study 1 (n = 735), in which we plot average proportion returned by player A in the TG, averaged across both conditions. (B) Results from study 2 (n = 737), in which we plot predicted proportion returned by player A in the TG, based on a regression model taking data from both conditions and including natural-log transformed comprehension speed as a control variable. Predictions are generated for a subject with a helping decision time that is either 1 SD above or below the mean. Error bars indicate ±1 SEM.
evidence that decision time is seen as reflecting doubt across a range of social contexts (39, 40) [and evidence that decision conflict does indeed drive decision times (46, 47)]. Other studies have also found that prosocial decisions that are motivated by emotion (38) or made impulsively (41) are seen as reflecting genuine altruistic motives, that deontological decision-makers are perceived as more trustworthy (27), and individuals who decline to reveal the exact payoffs of cooperating are predicted to behave more prosocially (36). We build on this research by using incentivized economic games to demonstrate that subjects show more “revealed” trust of people who help (i) without looking at the cost or (ii) relatively quickly.

Likewise, our third result (that uncalkulating cooperators really are more trustworthy) confirms that it can be beneficial to trust uncalkulating cooperators: they really do return more money in the TG, suggesting that uncalkulating cooperation serves as an honest signal of trustworthiness. This work builds on the finding that intuitive decisions are typically more cooperative (35) by showing that intuitive cooperation in one decision (compared with more calculated cooperation) predicts trustworthiness in a future decision.

Relatelt, our second and third results are particularly powerful, given that the target of uncalkulating cooperation in the HG was a different person than the target of the trustworthiness in the TG. Uncalkulating cooperation is often an important signal within dyadic relationships (e.g., a willingness to help, regardless of the costs, is a key quality of a loyal friend or romantic partner); thus, we might expect even stronger results if the recipient from the HG was the first mover in the TG. However, our results provide evidence that, to some extent, subjects expect uncalkulating cooperation to predict prosociality across decisions and interaction partners (at least in a lab experiment).

Importantly, we found that uncalkulating cooperation, but not uncalkulating defection, is perceived as, and actually is, a positive signal of trustworthiness. Directionally, uncalkulating defection is perceived as and actually is a negative signal, and these effects reach significance in some analyses (“perceived as”: Exp. 2, supplemental experiment (SI Appendix), and pooled datasets, and marginally significant in Exp. 1; “actually is”: marginally significant in supplemental experiment; see SI Appendix for details]. This reversal is consistent with evidence that quick decisions can be perceived positively or negatively, depending on the nature of the decision (42, 43), and demonstrates that it is specifically uncalkulating cooperation that is seen positively, rather than uncalkulating decision-making generally being seen as desirable. An interesting question for future research is why uncalkulating cooperation is perceived more positively (relative to calculating cooperation) than uncalkulating defection is perceived negatively (relative to calculating defection). One possibility is that subjects perceived not helping as more diagnostic than helping (48), perhaps because helping could have been motivated by a desire to elicit trust, and thus were more attentive to helpers’ decision processes as a result.

Another important question is how uncalkulating cooperation remains an honest signal of trustworthiness. What stops people from using uncalkulating cooperation to elicit trust from others, but then behaving exploitively? One possibility comes from a model of “cooperating without looking” in the “envelope game” (42, 43). In this game, uncalkulating cooperation prevents an individual from learning whether, in the current situation, defection would earn a higher payoff than cooperation. As a result, uncalkulating individuals are precommitted to not knowing when defection is worthwhile, and thus will best respond to the information they have by reliably cooperating across contexts (so long as, on average, cooperation pays for them).

Another possibility is that uncalkulating cooperation remains an honest indicator of trustworthiness via costly signaling (49–52). For individuals who face incentives to be trustworthy (i.e., who typically find cooperation advantageous), agreeing to cooperate without calculating is not very costly: In any given situation, it is likely that a cost–benefit analysis would support cooperation. In contrast, for individuals who face incentives to be exploitative (i.e., who rarely find cooperation advantageous), agreeing to cooperate without calculating is costlier, and it is more likely that a cost–benefit analysis would favor defection. Thus, exploitative individuals may not find uncalkulating cooperation worthwhile, even when factoring in the increased trust it elicits, keeping uncalkulating cooperation an honest signal.

Critically, our experiments did not explicitly build in either of these mechanisms for keeping signals honest: Nothing about our game structure could stop player A from engaging in uncalkulating cooperation in the HG and then returning nothing in the TG. If we had exactly recreated the envelope game or a costly signaling model in the laboratory, purely rational subjects (without any psychological predisposition toward treating uncalkulating cooperation as an honest signal of trustworthiness) would use uncalkulating cooperation to signal trustworthiness and would trust uncalkulating cooperators as a way to maximize their payoffs, and thus our results could merely reflect strategic reasoning in a novel game. Instead, we created a game setup in which there is not actually an “honest signaling” equilibrium, such that positive results would point to psychological predispositions regarding uncalkulating cooperation (thus, for example, Exp. 1 uses a “looking choice” measure inspired by the envelope game of refs. 42 and 43, but does not formally conform to its structure). Our findings therefore suggest that human psychology has been shaped by daily life contexts in which uncalkulating cooperation honestly signals trustworthiness, and that this psychology “spills over” to situations in which it does not actually make “rational” sense. Future research should further investigate the ultimate mechanisms responsible for keeping signals honest, and thus creating this psychology.

A final important future direction is investigating the proximate motivations that underlie uncalkulating cooperation: How often is the choice to be uncalkulating itself calculated and strategic (53)? And when uncalkulating cooperation is nonstrategic, when is it deliberate (e.g., somebody consciously applying an unconditional ethical principle) versus automatic (e.g., somebody being blinded by love in an intimate relationship (25) or spontaneously helping in an emergency (26))? Future research should investigate the proximate psychologies at play in our experiments and in daily life.

In sum, humans frequently cooperate without calculating the costs, despite the fact that doing so forfeits their ability to condition cooperation on whether or not it is worthwhile. Here, we provide the first empirical evidence, to our knowledge, that people preferentially engage in uncalkulating cooperation when their reputations are at stake. Our results thus suggest that people use uncalkulating cooperation to signal trustworthiness, and that this psychology “spills over” to situations ranging from extreme altruism to ethical principles to romantic love. If I help without thinking today, you know you can rely on me, no matter what tomorrow brings.

Materials and Methods
In both studies, we recruited pairs of subjects (player A and player B) online through Amazon Mechanical Turk to play a two-stage incentivized, anonymous economic game with no deception. We received approval from the Yale University Human Subjects Committee and gained informed consent as required by the Yale Institutional Review Board when subjects accepted our task on Amazon. In stage 1, player A participates in a HG and player B does not participate. In the HG, player A starts with 20 cents, whereas a recipient from the HG is informed that she is informed that c is 10 cents. In stage 2, player A receives the exact value of c (in which case he is informed that c is 10 cents). In Exp. 2, player A is always informed as to the exact value of c (10 cents), and then immediately after, makes his or her helping decision on the same screen. The time spent on this screen is recorded.
In stage 2, both player A and player B participate in a TG. In the TG, player B starts with 30 cents and chooses how much, if anything, to send to player A. Player A then decides how much, if anything, to return to player B. In the process observable condition, player B can condition his sending decision on player A’s HG decision process (whether player A revealed the cost of c in Exp. 1, and player A’s decision time in Exp. 2) and on player A’s HG decision (whether player A helped the recipient). In the process hidden condition, player B can only condition his sending decision on player A’s HG decision. We use the strategy method for both players [i.e., player B decides how much to send to a player A, who engaged in all possible combinations of HG decision processes and/or decisions (depending on condition), without knowing what player A actually did, and player A decides what percentage of the amount player B will send to return, without knowing how much player B actually sent].

We ask subjects comprehension questions to assess their understanding of the incentive structure of both phases of the game. In our primary analyses, we report results from all subjects, but all of our results are robust to the incentive structure of both phases of the game. In our primary analyses, we use logistic regressions when predicting HG looking decisions (which are binary) and linear regressions when predicting HG decision times (which are continuous; see TG sending and returning decisions (which are continuous). We use robust SEs in all regressions. Player Bs make multiple TG sending decisions (because they condition their sending on different possible player A HG behaviors); we analyze these data by treating each sending decision as an observation and clustering robust SEs on subject to account for the nonindependence of repeated observations from the same subject. In our analyses of HG decision times, we natural-log transform times (because they are highly right-skewed) and control for general comprehension speed when taking decision time as an independent variable (because variance in helping decision time is likely to reflect both time spent considering the cost of helping and general comprehension ability). We operationalize general comprehension speed as the natural-log transformed sum of the time the subject spent on the two screens involving comprehension questions (about the two stages of our game).

ACKNOWLEDGMENTS. We thank Adam Bear for helpful comments on the manuscript. We gratefully acknowledge the John Templeton Foundation for financial support.

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Player As</td>
<td>34</td>
</tr>
<tr>
<td>Player Bs</td>
<td>37</td>
</tr>
<tr>
<td>Supplemental experiment</td>
<td>40</td>
</tr>
<tr>
<td>Player As</td>
<td>40</td>
</tr>
<tr>
<td>Player Bs</td>
<td>43</td>
</tr>
<tr>
<td>References</td>
<td>43</td>
</tr>
</tbody>
</table>
Methods

Experimental design overview

In both experiments, subjects participate in a two-stage, incentivized, anonymous economic game in one of two roles: Player A or Player B. In Stage 1, Player A participates in a Helping Game (HG) and Player B does not participate. In the HG, Player A starts with 20 cents, while a passive Recipient (who is neither Player A nor Player B) starts with nothing. Player A is informed that she will have a choice over whether to help the Recipient by sacrificing some fraction of her endowment $c$ to deliver $2c$ to the recipient. Initially, Player A is not told the exact value of $c$, and knows only that it is between 0 and 20 cents. In Experiment 1, before making her helping decision, Player A chooses whether or not to reveal the exact value of $c$ (in which case she is informed that $c$ is 10 cents). In Experiment 2, Player A is always informed as to the exact value of $c$ (10 cents), and then immediately after, makes her helping decision on the same screen. The time spent on this screen is recorded.

In Stage 2, Both Player A and Player B participate in a Trust Game (TG). In the TG, Player B starts with 30 cents and chooses how much, if anything, to send to Player A. Any amount sent is tripled. Player A then chooses how much, if anything, to return to Player B. In the process observable condition, Player B can condition his sending decision both on (i) Player A’s HG decision process (whether Player A revealed the cost of $c$ in Experiment 1, and Player A’s decision time in Experiment 2, which is categorized as relatively fast or relatively slow) and (ii) Player A’s HG decision (whether Player A helped the Recipient). In the process hidden condition, Player B can only condition his sending decision on Player A’s HG decision. We employ the strategy method for both players (i.e. Player B decides how much to send to a Player A who engaged in all possible combinations of HG decision processes and/or decisions (depending on condition) without knowing what Player A actually did, and Player A decides what percentage of the amount Player B will send to return, without knowing how much Player B actually sent).

Subject recruitment method

We recruit all subjects online through Amazon Mechanical Turk. Our subjects receive a show-up fee of 50 cents immediately upon completing the study, and then are paid an additional bonus once data for all participants has been collected. After all data is collected, we form random pairs of Player A and Player B subjects, and then bonuses are determined based on subjects’ decisions, and the decisions of the players they are matched with. In our analyses, we start with all responses to our online surveys, and then exclude any subjects who (i) did not complete all measures within the survey or (ii) did not have a unique IP address (to exclude additional responses from subjects who had already completed the survey previously).

In Experiment 1, we initially planned to request $N = 200$ Player As and Player Bs per condition from Amazon Turk ($N = 400$ Player A-Player B pairs in total). After this initial collection, the comparison between rates of Player A “looking choices” in the process observable versus process hidden condition was marginally significant (logistic regression coefficient $= - .366$, $N = 398$, $p = .080$). We subsequently doubled the sample size (requesting another $N = 400$ Player A-Player B pairs from Amazon Turk); however, because of a technical issue, only a total of $N = 375$ additional Player As were requested (while the full $N = 400$ Player Bs were successfully requested). Then, to achieve equal numbers of Player As and Player Bs (so that we could match pairs evenly), we excluded extra Player Bs. This procedure resulted in a final
sample of N=735 Player A-Player B pairs. In this final sample, the comparison between rates of Player A “looking choices” in process observable versus process hidden condition was significant (logistic regression coefficient = -.486, N = 735, p = .002).

To address potential concerns arising from having increased our sample size because the result was not initially significant, we then applied Sagarin, Ambler, and Lee’s (2014) method to calculate adjusted type-I error rates for augmented data sets that are based on “peeking” at marginally significant results (1). We calculated that for the result reported above, if we determined statistical significance in the full augmented sample by asking if p<.05 (i.e. used the standard alpha value of .05 to evaluate the full augmented sample’s significance), the actual type-I error rate ($p_{augmented}$) would range from .0502 to .0505. We report a range, because the actual type-I error rate depends on the maximum p-value observed in the initial collection for which we would have augmented the dataset rather than declaring the initial results non-significant ($p_{max}$); this could range from a “best-case scenario” of .08 (in which case the type-I error rate would be .0502) to a “worst-case scenario” of 1 (in which case the type-I error rate would be .0505).

We also calculated the range of alpha values that we would have had to use to evaluate the full augmented sample’s significance ($p_{crit}$) to maintain an actual type-I error rate of .05, and found a range of .0299 to .0437. Again, this is a range based on the range of possible $p_{max}$ values; thus, in the “best-case scenario” of $p_{max} = .08$, we would have had to determine statistical significance in the full augmented sample by asking if p<.0437, and in the “worst-case scenario” of $p_{max} = 1$, we would have had to determine statistical significance in the full augmented sample by asking if p<.0299. And indeed, the relevant p-value from the full augmented sample (p = .002) is well below the “worst-case scenario” threshold of .0299.

Next, in Experiment 2, we planned to recruit N = 400 pairs of Player As and Player Bs per condition (N = 800 Player A-Player B pairs total), to achieve an equivalent sample size to Experiment 1. After excluding any subjects who did not complete all measures within the survey or have a unique IP address, and then excluding extra responses from the role (Player As or Player Bs) for which we had more responses, we reached a final sample of N=737 Player A-Player B pairs in Experiment 2.

**Procedure for Player As**

**Experiment 1**

In Experiment 1, Player As first participate in the HG, and then participate in the TG. These subjects first read that they will participate in two different interactive games with other MTurk workers. They then read instructions describing the rules of the HG, and view a decision tree illustrating their choices in the game (instructions use neutral language to label the different players and games; full instructions shown in the “experimental instructions” section at the end of this document). After subjects read the instructions, they answer three comprehension questions to ensure that they understand that helping is costly for them and beneficial for the recipient, and that they have two choices: whether to find out the exact cost of helping, and whether to help the recipient.

Next, we inform subjects that they will participate in the next game, the TG, with another player (Player B) who did not participate in the HG. We provide instructions describing the rules of the TG. Then, subjects answer three more comprehension questions to ensure that they understand that (i) Player B’s payoff-maximizing decision depends on how much they decide to
return to Player B; (ii) their payoff-maximizing decision is always to return nothing; and (iii) Player B did not participate in the HG.

Next, we tell subjects about the information that Player Bs can base their TG sending decisions on. In the process hidden condition, we explain that Player Bs can base their TG sending decisions on subjects’ HG helping decisions (but not their HG looking decisions). In the process observable condition, we explain that Player Bs can base their TG sending decisions on subjects’ HG helping and looking decision. Then, subjects answer another comprehension question to ensure that they understand what information Player Bs can base their TG sending decisions on.

Next, subjects make their looking decisions. Before deciding, they are reminded that (i) all they currently know is that helping entails paying somewhere between 0 and 20 cents to deliver twice as much to the recipient; (ii) the other players are real; and (iii) they really do have a chance to find out the precise cost of helping. Player As in the process observable condition, but not the process hidden condition, are also reminded that (iv) Player Bs really will get to base their TG sending decisions on whether they decide to help. They then make a binary looking decision, and afterwards, subjects who decide to look are informed that the precise cost of helping is 10 cents.

Next, subjects make their helping decisions. Before deciding, they are reminded (i) of the parameters of the helping decision (in which the cost is labeled as 10 cents only for subjects who decided to look); (ii) that the other players are real; (iii) that they really do have a chance to help the recipient; and (iv) that Player Bs really will get to base their TG sending decisions on whether they decide to help. They then make a binary helping decision.

Next, subjects make their TG returning decisions. Before deciding, they are reminded (i) of the parameters of the returning decision (in which they are told to decide what percentage to return of the amount Player B will send); (ii) that the other players are real; and (iii) that they really do have a chance to decide how much to return. They make a continuous returning decision, indicating the percentage (in 10-cent increments) they would like to return. Finally, Player As receive a post-experimental questionnaire.

**Experiment 2**

In Experiment 2, the procedure for Player As is identical to Experiment 1, with the following exceptions.

When we describe the HG, subjects are not told that they will have a choice to find out the precise cost of helping, but rather that they will find out what it is later. Because subjects only make one decision, they are not shown a decision tree or asked a comprehension question about the two decisions they will make. Instead, they are asked a comprehension question to ensure that they understand that they do not yet know the precise cost of helping, but they will find out later.

When we describe the information that Player Bs can base their TG sending decisions on, we explain to all subjects that we will measure and record the amount of time they spend a screen that (i) reveals the precise cost of helping, and (ii) asks subjects to make their helping decisions. Thus, all subjects are informed that their looking time will be monitored and recorded. Then, subjects in the process hidden condition are told that Player Bs can base their TG sending decisions on subjects’ HG helping decisions (but not how long they take to make this decision), while subjects in the process observable condition are told that Player Bs can base their TG
sending decisions on subjects’ HG helping decisions *and* how long they take to make these decisions.

Then, before subjects make their helping decisions, we warn them that on the next screen, we will tell them what the precise cost of helping is, ask them to decide if they want to help, and record how long they spend on the screen. On the next screen, we inform all subjects that the precise cost of helping is 10 cents, and then present them with the same binary helping decision as in Experiment 1. We record the number of seconds the subject spends on the page before clicking to proceed.

**Procedure for Player Bs**

*Experiment 1*

In Experiment 1, the procedure for Player Bs is identical to the procedure for Player As, with the following exceptions. Player Bs do not participate in the HG, but do participate in the TG. When describing the HG, we thus tell subjects that they will not participate, but that we want them to understand the game. When describing the TG, we tell subjects that they will play with a Player A who previously played the HG.

When subjects make their TG decisions, we explain, according to condition, what information they can base their sending decisions on. Thus, we tell subjects in the Process hidden condition that they can only base their sending on whether Player A helped in the HG, and we tell subjects in the Process observable condition that they can base their sending both on whether Player A helped, and whether Player A looked at the precise cost of helping. We then explain that they will make multiple sending decisions that correspond to different possible Player A behaviors in the HG, and that the decision that matches the behavior of the actual Player A they are matched with will determine their earnings (the “strategy method”). Then, subjects decide how much to send (in 5-cent increments) to each possible type of Player A (helpers and non-helpers in the Process hidden condition; helpers who looked, helpers who did not look, non-helpers who looked, and non-helpers who did not look in the Process observable condition) in a random order, with each decision on a separate page.

*Experiment 2*

In Experiment 2, the procedure for Player Bs is analogous to Experiment 1 (i.e. identical to the procedure for Experiment 2 Player As, with exceptions analogous to the ones described above). Thus, in the Process observable condition, we explain to subjects that they can condition their TG sending on the amount of time Player A spent deciding whether to help. To implement this, we explain to all subjects (in both conditions) that half of Player As decided whether to help in less than 2.05 seconds, while the other half took 2.05 seconds or longer. Then, in the Process observable condition only, we explain that subjects can condition their sending on whether Player A “took a long time to decide” (defined as 2.05 seconds or longer).

**Procedure for HG Recipients**

Because HG Recipients do not make any decisions, we do not require them to read the rules or answer comprehension questions. Instead, for every Player A who chooses to help the recipient, we give a bonus of 20 cents to a randomly selected Mturk worker (picked from our database of Mturk worker IDs who have previously participated in our studies) who thus functions as the Recipient.

**Procedure for matching Player As and Player Bs**
After recruiting and collecting data from all Player As and Player Bs, we randomly match pairs of Player As and Player Bs who are in the same condition, implement their strategy method decisions, and pay them bonuses according to these decisions.

**Results**

**General analysis approach**

We divide our results into three sections, in which we answer the questions (i) Is uncalculated cooperation used as a signal of trustworthiness? (ii) Is uncalculated cooperation perceived as a signal of trustworthiness? and (iii) Is uncalculated cooperation actually a signal of trustworthiness? In (i) we analyze the effect of our observability manipulation on whether Player A behaves in an uncalculated manner (in Experiment 1, Player A’s choice to look at the precise cost of helping, and in Experiment 2, the amount of time Player A spends deciding whether to help). In (ii) we analyze the way that (in the process observable condition) Player B conditions trust on whether Player A behaves in an uncalculated manner, as well as whether Player A helps. In (iii) we analyze the way that (in both conditions) Player A’s trustworthiness is predicted by whether she behaves in an uncalculated manner, as well as whether she helps.

In our analyses, we use Stata/MP 13.1 for Mac. We use logistic regressions when predicting HG looking decisions (which are binary), and linear regressions when predicting HG decision times, as well as TG sending and returning decisions (which are continuous). We use robust standard errors in all regressions.

In our analyses of HG decision times, we natural-log transform times (because they are highly right-skewed), and control for general comprehension speed when taking decision time as an independent variable (because variance in helping decision time is likely to reflect both time spent considering the cost of helping, and general compression ability). We operationalize general comprehension speed as the natural-log transformed sum of the time the subject spent on the two screens involving comprehension questions (about the two stages of our game).

In all our analyses of TG sending and returning, for ease of interpretation, we take the percentage of the endowment sent / returned (0-100) as our dependent variable (rather than the absolute number of cents). In our analyses of TG sending, because Player Bs make multiple decisions (as they condition their sending on different possible Player A HG behaviors), we analyze each decision as an observation, and cluster robust standard errors on subject to account for the non-independence of repeated observations from the same subject.

We report results from all subjects, but all of our results are robust to restricting to subjects who answered all comprehension questions correctly, and this restriction increases effect sizes for most of our effects. In all regression tables, we report results from all subjects, as well as from subjects who answered all comprehension questions correctly (columns with the heading “Perf comp”).

**Comprehension questions**

*Experiment 1*

In Experiment 1, we presented N = 735 Player As with 7 comprehension questions. On average, they answered 5.44 questions correctly (78% correct), and 29% of individuals answered all questions correctly. We also presented N = 735 Player Bs with 6 comprehension questions. On average, they answered 4.89 questions correctly (81% correct), and 41% of individuals answered all questions correctly.

*Experiment 2*
In Experiment 2, we presented N = 737 Player As with 7 comprehension questions. On average, they answered 5.89 questions correctly (84% correct), and 41% of individuals answered all questions correctly. We also presented N = 737 Player Bs with 6 comprehension questions. On average, they answered 4.99 questions correctly (83% correct), and 42% of individuals answered all questions correctly.

In the main text, we include all subjects in our analyses. However, all of our results hold qualitatively when excluding subjects who answered one or more comprehension questions incorrectly; see regression tables for results from models that restrict to perfect comprehenders.

**Primary Analyses**

*Uncalculated cooperation is used as a signal*

We first investigate our prediction that people use uncalculated cooperation for reputational benefits: Player A should be more likely to decide in an uncalculated way in the process observable condition (when doing so can confer reputational benefits) than in the process hidden condition (when doing so cannot).

In Experiment 1, we measure uncalculated behavior as the decision not to look at the precise cost of helping. We find that 70% of Player As decided to look at the cost of helping in the process hidden condition, as compared to 59% in the process observable condition. We confirm that this is significant with a logistic regression predicting the decision to look (1 = look, 0 = do not look) as a function of a process observable condition dummy. We find a significant negative effect of the process observable condition dummy (coeff = -0.49, p = .002, N=735, Table 1 Columns 1-2).

<table>
<thead>
<tr>
<th></th>
<th>All subjects</th>
<th>Perf comp</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Process observable condition</td>
<td>-0.486**</td>
<td>-0.368</td>
</tr>
<tr>
<td></td>
<td>(0.156)</td>
<td>(0.299)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.850***</td>
<td>1.018***</td>
</tr>
<tr>
<td></td>
<td>(0.113)</td>
<td>(0.214)</td>
</tr>
<tr>
<td>Observations</td>
<td>735</td>
<td>215</td>
</tr>
<tr>
<td>Subjects</td>
<td>735</td>
<td>215</td>
</tr>
<tr>
<td>Pseudo R-squared</td>
<td>0.010</td>
<td>0.006</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses

*** p<0.001, ** p<0.01,  * p<0.05

**Table 1.** Player A’s HG decision to look at the precise cost of helping (1 = Look, 0 = Do not look) as a function of condition (1 = Process observable, 0 = Process hidden) in Experiment 1. Column 1 shows results from all subjects; Column 2 shows results from subjects who showed perfect comprehension.

In Experiment 2, we measure uncalculated behavior as the amount of time spent deciding whether to help, after the precise cost of helping is revealed. We find that decision times are highly right-skewed (M = 11.39 seconds, SD = 18.54, skewedness = 21.63, kurktoisis = 539.01), and so we natural log transform times; after this transformation, we find that times are much less skewed (M = 2.23 log-seconds, SD = 0.57, skewedness = 0.43, kurktoisis = 6.49). We find that
the mean transformed decision time of Player As was 2.28 log-seconds in the *process hidden* condition, as compared to 2.17 log-seconds the *process observable* condition. We confirm that this is significant with a linear regression predicting log-transformed decision time as a function of a process observable condition dummy. We find a significant negative effect of the process observable condition (coeff = -0.103, p = .014, N = 737, Table 2 Columns 1-2).

<table>
<thead>
<tr>
<th></th>
<th>All subjects (1)</th>
<th>Perf comp (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process observable condition</td>
<td>-0.103*</td>
<td>-0.198**</td>
</tr>
<tr>
<td></td>
<td>(0.0418)</td>
<td>(0.0586)</td>
</tr>
<tr>
<td>Constant</td>
<td>2.277***</td>
<td>2.263***</td>
</tr>
<tr>
<td></td>
<td>(0.0306)</td>
<td>(0.0411)</td>
</tr>
<tr>
<td>Observations</td>
<td>737</td>
<td>303</td>
</tr>
<tr>
<td>Subjects</td>
<td>737</td>
<td>303</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.008</td>
<td>0.037</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

| Table 2. Player A’s HG helping decision time (in log-seconds) as a function of condition (1 = Process observable, 0 = Process hidden) in Experiment 2. Column 1 shows results from all subjects; Column 2 shows results from subjects who showed perfect comprehension.

Thus, we confirm our first prediction: across two experiments, subjects behaved in a less calculated manner when their reputation was at stake. In Experiment 1, they were less likely to look at the cost of helping when their looking was observable, and in Experiment 2, they decided faster when their decision time was observable.

*Uncalculated cooperation is perceived as a signal*

We next investigate our prediction that uncalculated cooperation does, in fact, confer reputational benefits: uncalculated cooperators should be perceived as more trustworthy than calculated cooperators. Furthermore, we test the prediction that the reputational benefit of uncalculated behavior should be specific to uncalculated cooperation, and should not apply to uncalculated defection.

We begin by investigating the first half of this prediction: that uncalculated cooperation confers reputational benefits. Specifically, we test the prediction that in the process observable condition, Player Bs should send more to Player As who reached a cooperative decision in an uncalculated way than Player As who reached a cooperative decision in a calculated way. We restrict our analyses to the process observable condition because this was the only condition in which Player Bs were allowed to condition their trust on Player A’s decision process.

In Experiment 1, we find that Player Bs sent an average of 55% of their endowments to Player As who helped without looking at the cost, as compared to 49% to Player As who helped after looking. We confirm that this is significant with a linear regression predicting percentage of endowment sent to helpers as a function of looking (clustering on subject to account for the fact that there are two decisions for each subject). We find a significant negative effect of looking (coeff = -5.67, p < .001, N = 361, Table 3 Columns 1-2).
In Experiment 2, we find that Player Bs sent an average of 60% of their endowments to Player As who helped relatively quickly (decision time below the median), as compared to 50% to Player As who helped relatively slowly (decision time above the median). We confirm that this is significant with a linear regression predicting percentage of endowment sent to helpers as a function of fast vs. slow decision time (again, clustering on subject). We find a significant negative effect of slow decisions (coeff = -9.84, p < .001, N = 365, Table 4 Columns 1-2).

Thus, we confirmed our prediction that uncalculated cooperators are trusted more than calculated cooperators. In Experiment 1, subjects were less trusting of individuals who checked the cost before helping, and in Experiment 2, they were less trusting of individuals who considered the cost for a long time before helping.

We next investigate the second half of our second prediction: that the positive effect of uncalculated behavior on trust should be specific to uncalculated cooperation.

In Experiment 1, we first conduct a linear regression predicting endowment sent (to either helpers or non-helpers) as a function of helping, looking, and their interaction (again, clustering on subject; in this regression, there are four decisions for each subject). We find a significant negative interaction (coeff = -7.73, p<.001, N = 361, Table 3 Columns 5-6), indicating that looking is perceived more negatively when Player A decided to help than when Player A decided not to help. We then investigate the effect of looking when Player A decided not to help. We find that Player Bs sent an average of 19% of their endowments to Player As who chose not to help without looking at the cost, as compared to 21% to Player As who chose not to help after looking. A linear regression predicting percentage of endowment sent to non-helpers as a function of looking (clustering on subject) demonstrates that this difference is marginally significant (coeff = 2.06, p = .080, N = 361, Table 3 Columns 3-4).

In Experiment 2, we conduct a linear regression predicting endowment sent (to either helpers or non-helpers) as a function of helping, decision speed, and their interaction (clustering on subject). We find a significant negative interaction between helping and slow decisions (coeff = -14.74, p<.001, N = 365, Table 4 Columns 5-6), indicating that deciding slowly is perceived more negatively when Player A decided to help than when Player A decided not to help. We then investigate the effect of decision speed when Player A decided not to help. We find that Player Bs sent an average of 22% of their endowments to Player As who decided relatively quickly not to help, as compared to 27% to Player As who decided relatively slowly not to help. A linear regression predicting percentage of endowment sent to non-helpers as a function of fast vs. slow decision time demonstrates that this difference is significant (coeff = 4.90, p < .001, N = 365, Table 4 Columns 3-4).

Thus, across both of our experiments, we confirmed our prediction that uncalculated cooperation was perceived positively, but uncalculated defection was not.
<table>
<thead>
<tr>
<th></th>
<th>Helpers</th>
<th>Non helpers</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All subjects</td>
<td>Perf comp</td>
<td>All subjects</td>
</tr>
<tr>
<td>Help</td>
<td></td>
<td></td>
<td>35.72***</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(1.945)</td>
</tr>
<tr>
<td>Look at cost</td>
<td>-5.669***</td>
<td>-9.363***</td>
<td>2.059</td>
</tr>
<tr>
<td></td>
<td>(1.227)</td>
<td>(2.047)</td>
<td>(1.174)</td>
</tr>
<tr>
<td>Help X Look at cost</td>
<td>-7.729***</td>
<td>-12.19***</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(1.670)</td>
<td>(2.728)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>54.78***</td>
<td>59.79***</td>
<td>19.07***</td>
</tr>
<tr>
<td></td>
<td>(1.950)</td>
<td>(2.862)</td>
<td>(1.462)</td>
</tr>
<tr>
<td>Observations</td>
<td>722</td>
<td>314</td>
<td>722</td>
</tr>
<tr>
<td>Subjects</td>
<td>361</td>
<td>157</td>
<td>361</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.006</td>
<td>0.017</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

**Table 3.** Player B’s TG sending as a function of Player A’s HG choice to look at the precise cost of helping in the process observable condition of Experiment 1. Columns 1-2 show results from decisions about Player As who chose to help in the HG; Columns 3-4 show results from decisions about Player As who chose not to help in the HG; Columns 5-6 show results from all decisions and also includes a term for whether Player A chose to help, and the interaction between Player A’s helping and looking choices. Columns 1, 3, and 5 show results from all subjects; Columns 2, 4, and 6 show results from subjects who showed perfect comprehension. Standard errors are clustered on subject to account for repeated observations.
Table 4. Player B’s TG sending as a function of Player A’s helping decision time (slower or faster than the median) in the process observable condition of Experiment 2. Columns 1-2 show results from decisions about Player As who chose to help in the HG; Columns 3-4 show results from decisions about Player As who chose not to help in the HG; Columns 5-6 show results from all decisions and also includes a term for whether Player A chose to help, and the interaction between Player A’s helping and decision time. Columns 1, 3, and 5 show results from all subjects; Columns 2, 4, and 6 show results from subjects who showed perfect comprehension. Standard errors are clustered on subject to account for repeated observations.

Uncalculated cooperation actually is a signal

Finally, we investigate our prediction that uncalculated cooperators is reasonable: uncalculated cooperators should actually be more trustworthy than calculated cooperators. Furthermore, we test the prediction that the effect of uncalculated behavior should be specific to uncalculated cooperation, and should not apply to uncalculated defection. Here, we investigate both experimental conditions (because even though Player B could not observe Player A’s decision process in the process hidden condition, we still have data on Player A’s decision process in both conditions and can thus correlate it with trustworthiness in both conditions).

We begin by investigating the first half of this prediction: that uncalculated cooperators should actually be more trustworthy than calculated cooperators.

In Experiment 1, we find that Player As who helped without looking at the cost returned an average of 50% of the amount they were sent to Player B, as compared to 41% among Player As who helped after looking. We confirm that this is significant with a linear regression predicting percentage returned by helpers as a function of looking. We find a significant negative effect of looking (coeff = -8.90, p < .001, N = 595, Table 5 Columns 1-2).
In Experiment 2, we investigate the effect of Player A’s continuous decision time on Player A’s trustworthiness. However, because this analysis is correlational, there is a concern that an individual’s helping decision time reflects not only the extent to which she considered the cost of helping (i.e. calculated decision-making) but also her general comprehension speed. Thus, in our analyses of the effect of helping decision time on trustworthiness, we also control for general comprehension speed. As discussed in the methods section, we operationalize general comprehension speed as the natural long-transformed time spent reading the comprehension questions (sum of time spent on the two comprehension question pages, for the two different stages of the game). We note that we do not include this control in our analyses in which helping decision time is the dependent variable, because this analysis is not correlational (in it, we investigate the effect of our manipulated observability condition on helping decision time).

Thus, we conduct a linear regression predicting percentage returned by helpers as a function of log-transformed helping decision time, and log-transformed general comprehension speed. We find a significant negative effect of helping decision time on the amount returned to Player B (coef = -4.61, p=.021, N = 624, Table 6 Columns 1-2), and a marginally significant positive effect of general comprehension speed (coef = 2.15, p = .090, N = 624, Table 6 Columns 1-2). We note that we visualize the effect of helping decision time among helpers in main text Figure 4, left bars by calculating the predicted percentage returned, from this regression model, by a helping Player A with a general comprehension speed that is the overall sample mean of 8.85 log-seconds, and a helping decision time that is either 1 standard deviation (0.57 log-seconds) above or below the overall sample mean of 2.23 log seconds.

Thus, across two experiments, we confirmed our prediction that uncalculated cooperators are more trustworthy than calculated cooperators. In Experiment 1, subjects were more trustworthy if they helped without looking at the cost, and in Experiment 2, they were more trustworthy if they helped without considering the cost for a long time.

We next investigate the second half of our second prediction: that the positive effect of uncalculated behavior on trustworthiness should be specific to uncalculated cooperation.

In Experiment 1, we first conduct a linear regression predicting percentage returned (by either helpers or non-helpers) as a function of helping, looking, and their interaction. We find a significant negative interaction (coef = -10.56, p = .031, N = 735, Table 5 Columns 5-6), indicating that looking is a stronger predictor of untrustworthiness when Player A decided to help than when Player A decided not to help. We then investigate the effect of looking when Player A decided not to help. We find that Player As who decided not to help without looking at the cost returned an average of 18% of the amount they were sent to Player B, as compared to 20% among Player As who decided not to help after looking. However, a linear regression predicting percentage returned by non-helpers as a function of looking demonstrates that this difference is not significant (coef = 1.67, p = .718, N = 140, Table 5 Columns 3-4).

In Experiment 2, we first conduct a linear regression predicting percentage returned (by either helpers or non-helpers) as a function of helping, log-transformed helping decision time, their interaction, and log-transformed general comprehension speed. We find a significant negative interaction (coef = -7.69, p = .019, N = 737, Table 6 Columns 5-6), indicating that helping decision time is a stronger predictor of untrustworthiness when Player A decided to help than when Player A decided not to help. We then investigate the effect of helping decision time on returning among Player As who decided not to help (again, controlling for general comprehension speed). We find a non-significant positive effect of helping decision time (coef = 2.54, p = .486, N = 113, Table 6 Columns 3-4) and non-significant positive effect of general
comprehension speed (coeff = 2.77, p = .178, N = 113, Table 6 Columns 3-4). We note that we visualize the effect of helping decision time among non-helpers main text Figure 4, right bars by calculating the predicted percentage returned, from this regression model, by a non-helping Player A with a general comprehension speed that is the overall sample mean of 8.85 log-seconds, and a helping decision time that is either 1 standard deviation (0.57 log-seconds) above or below the overall sample mean of 2.23 log seconds.

Thus, across two experiments, we confirmed our prediction that uncalculated decision-making only predicted trustworthiness when Player A helped.

<table>
<thead>
<tr>
<th></th>
<th>Helpers</th>
<th></th>
<th>Non-helpers</th>
<th></th>
<th>Both</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All subjects</td>
<td>Perf comp</td>
<td>All subjects</td>
<td>Perf comp</td>
<td>All subjects</td>
<td>Perf comp</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>Help</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>31.30***</td>
<td>43.12***</td>
</tr>
<tr>
<td>Look at cost</td>
<td>-8.896***</td>
<td>-14.57***</td>
<td>1.667</td>
<td>7.822</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.702)</td>
<td>(2.884)</td>
<td>(4.601)</td>
<td>(5.724)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Help X Look at cost</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-10.56*</td>
<td>-22.40***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(4.887)</td>
<td>(6.309)</td>
</tr>
<tr>
<td>Constant</td>
<td>49.51***</td>
<td>50.89***</td>
<td>18.21***</td>
<td>7.778</td>
<td>18.21***</td>
<td>7.778</td>
</tr>
<tr>
<td></td>
<td>(1.255)</td>
<td>(2.174)</td>
<td>(3.791)</td>
<td>(4.217)</td>
<td>(3.774)</td>
<td>(4.129)</td>
</tr>
<tr>
<td>Observations</td>
<td>595</td>
<td>181</td>
<td>140</td>
<td>34</td>
<td>735</td>
<td>215</td>
</tr>
<tr>
<td>Subjects</td>
<td>595</td>
<td>181</td>
<td>140</td>
<td>34</td>
<td>735</td>
<td>215</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.038</td>
<td>0.105</td>
<td>0.001</td>
<td>0.038</td>
<td>0.180</td>
<td>0.272</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Table 5. Player A’s TG returning as a function of Player A’s HG choice to look at the precise cost of helping in both conditions of Experiment 1. Columns 1-2 show results from Player As who chose to help in the HG; Columns 3-4 show results from Player As who chose not to help in the HG; Columns 5-6 show results from all Player As and also includes a term for whether Player A chose to help, and the interaction between Player A’s helping and looking choices. Columns 1, 3, and 5 show results from all subjects; Columns 2, 4, and 6 show results from subjects who showed perfect comprehension.
Table 6. Player A’s TG returning as a function of Player A’s HG helping decision time, and a control for general comprehension speed, in both conditions of Experiment 1. Columns 1-2 show results from Player As who chose to help in the HG; Columns 3-4 show results from Player As who chose not to help in the HG; Columns 5-6 show results from all Player As and also includes a term for whether Player A chose to help, and the interaction between Player A’s helping and looking choices. Columns 1, 3, and 5 show results from all subjects; Columns 2, 4, and 6 show results from subjects who showed perfect comprehension.

**Supplemental Experiment**

*Experimental design overview*

In our Supplemental experiment, our goal was to test whether we could obtain our key result, whereby Player As are less likely to calculate when their decision processes are observable, without such explicit reminders of the possible reputation consequences of calculation. To this end, we conducted a replication of Experiment 1 with minor alterations. For Player As, our aim was to convey to subjects whether their decision process was observable less explicitly, using screenshots of the study from Player B’s perspective. Then, after Player As finished all of their economic game decisions, we explicitly measured their comprehension of which aspects of their decision processes were observable. For Player Bs, the experiment was identical to Experiment 1.
Subject recruitment

We planned to recruit N = 400 pairs of Player As and Player Bs per condition (N = 800 Player A-Player B pairs total) via Mturk, as in Experiment 1. After excluding any subjects who did not complete all measures within the survey or have a unique IP address, and then excluding extra responses from the role (Player As or Player Bs) for which we had more responses, we reached a final sample of N=805 Player A-Player B pairs.

Procedure

The procedure for Player As has the following changes, relative to Experiment 1. First, when presenting the initial TG instructions, we insert a sentence explaining to Player As that Player B will be a new player who did not participate in the HG, but will learn information about how Player A behaved in the HG. Our goal was to convey that Player B TG decisions would be conditional on Player A HG behavior, without specifying which aspects of HG behavior they would be conditional on.

Second, on the next page, instead of explicitly providing this information as in Experiment 1, we convey it indirectly by providing screenshots of the study from Player B’s perspective. To keep our intentions subtle, we frame the screenshots as providing general logistical information about how Player Bs will make their decisions. Our instructions (note that Player Bs were referred to as “Senders” to subjects) read: “As we explained, in the second game, the Sender will decide how much money to send to you. But how will this work? The Sender will decide using the screens shown below. Each question will be shown on a different screen, and the screens will be presented in a random order. The Sender will use the sliding scale to decide how much to send, with the 5 cent increments as anchor points. Please look carefully at the screens as you may be asked a comprehension question about them later.”

Below these instructions, in the process hidden condition, we show the two Player B decision screens from this condition (which allow Player B to condition trust only on Player A helping). In contrast, in the process observable condition, we show the four Player B decision screens from this condition (which allow Player B to condition trust on both Player A looking and helping). Importantly, while in Experiment 1 Player As also had to complete a comprehension question on this page about what Player B could base their sending on, in the Supplemental Experiment no such comprehension question is included on this page (to avoid making this information explicit).

Third, when Player As make their looking decisions, we eliminate the reminder that there are possible reputation consequences of this decision. In Experiment 1, we reminded Player As in both conditions that they really did have a chance to find out the precise cost of helping, and then reminded Player As in the process observable condition that Player Bs would really get to base their TG decisions on whether they decided to do so. We remove both of these reminders in our Supplemental experiment.

Finally, after Player As complete all of their economic game decisions (whether to look and whether to help in the HG, and how much to return in the TG), we ask them an explicit comprehension question about what information Player Bs could base their trust on in the TG. This question is identical to the one that was asked before Player As made their decisions in Experiment 1.
Player A results

Comprehension

We begin by investigating comprehension among Player As. We presented N = 805 Player As with 6 pre-decision comprehension questions, and then our key post-decision comprehension question (about what Player Bs could condition their trust on). On average, Player As answered 4.62 pre-decision questions correctly (77% correct), and 31% of individuals answered all pre-decision questions correctly.

On our key post-decision comprehension question, 57% of subjects correctly understood what information Player Bs could condition their trust on (where chance = 25%). Rates of comprehension on this question do not differ significantly by condition (logistic regression coeff = 0.02, p = .871, N = 805): 56.86% of subjects answered correctly in process hidden, as compared to 57.43% in process observable.

Critically, however, the rate of comprehension on this question in our Supplementary experiment is significantly lower than the rate of comprehension on the same question in Experiment 1, where 83% of subjects answered correctly (logistic regression coeff = 1.32, p < .001, N = 1540). This demonstrates that presenting information about observability less explicitly in our Supplementary experiment made it less likely that subjects comprehended it, which was consistent with our goal of being subtler.

Uncalculating cooperation is used as a signal

Next, we turn to our key prediction that Player As are less likely to look when they know their looking decision will be seen by Player B (process observable condition), compared to when it will not be seen by Player B (process hidden condition). Of course, we would not expect to observe this difference among the large fraction of Player As who did not correctly understand what information Player Bs could condition their trust on (and thus did not correctly understand whether their decision process was observable). And indeed, a logistic regression predicting the decision to look as a function of a process observable condition dummy, a non-comprehender dummy (indicating whether subjects failed our critical post-decision comprehension question), and their interaction (N = 805, Table 7 Column 1), reveals a significant interaction between the process observable condition and being a non-comprehender (coeff = 0.83, p = .005): the effect of the process observable condition was significantly different among comprehenders versus non-comprehenders.

Specifically, there was a large significant effect of the process observable condition among the N = 460 comprehenders (coeff = -0.60, p = .002), reflecting that (as predicted), comprehenders looked 68% of the time in the process hidden condition, but only 54% of the time in the process observable condition. Among the N = 345 non-comprehenders, in contrast, there was no significant effect of the process observable condition (coeff = 0.23, p = .295; we note that this non-significant trend in the opposite direction is consistent with the fact that many non-comprehenders’ answers reflected that they thought they were in the opposite condition to the one they were in). And naturally, collapsing across comprehension (i.e. analyzing all subjects together) produces a result that is still in the expected direction but substantially attenuated by the non-comprehenders (63% looking in the process hidden condition, 57% looking in the process observable condition; logistic regression coeff = -0.24, p = .101, N = 805).

Thus, our results suggest that we do replicate our key result among subjects who correctly understood the game structure; and we note that this restriction does not violate random assignment, because rates of understanding did not differ by condition (see above).
We found that, quite naturally, comprehension played a critical role in observing our predicted effect within in the Supplemental experiment. Is the same true of Experiment 1? Indeed, conducting the same regression described above using the data from Experiment 1 (N = 735, Table 7 Column 2) also reveals a significant interaction between the process observable condition and being a non-comprehender (coeff = 1.32, p = .002), reflecting that there was a significant effect of the process observable condition among the N = 612 comprehenders (coeff = -0.70, p < .001; 72% looking in process hidden vs 56% looking in process observable), but not among the N = 123 non-comprehenders (coeff = 0.62, p = .118; again, this directional reversal is consistent with the fact that many non-comprehenders thought they were in the opposite condition). Thus, while more subjects comprehended the game structure when it was described to them explicitly, subjects who did comprehend the game structure behaved almost identically regardless of whether it was described explicitly or not.

We provide further evidence for the equivalency of the effect across the explicit and implicit designs by directly comparing the effect of observability on looking among comprehenders in the two designs. We conduct a logistic regression predicting looking as a function of the process observable condition, an Experiment 1 dummy, and their interaction among comprehending subjects in both Experiment 1 and the Supplemental experiment. Indeed, we find no significant interaction (coeff = -0.10, p = .701, N = 1072), confirming that subjects who understood the game structure were not significantly more responsive to our manipulation when it was presented explicitly, as compared to when it was presented more subtly.

<table>
<thead>
<tr>
<th></th>
<th>Supp. Exp</th>
<th>Exp 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Process observable</td>
<td>-0.600**</td>
<td>-0.700***</td>
</tr>
<tr>
<td>condition</td>
<td>(0.194)</td>
<td>(0.172)</td>
</tr>
<tr>
<td>Non comprehender</td>
<td>-0.529*</td>
<td>-0.753*</td>
</tr>
<tr>
<td></td>
<td>(0.209)</td>
<td>(0.333)</td>
</tr>
<tr>
<td>Observable X Non-comp</td>
<td>0.830**</td>
<td>1.315**</td>
</tr>
<tr>
<td></td>
<td>(0.293)</td>
<td>(0.429)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.773***</td>
<td>0.944***</td>
</tr>
<tr>
<td></td>
<td>(0.143)</td>
<td>(0.122)</td>
</tr>
<tr>
<td>Observations</td>
<td>805</td>
<td>735</td>
</tr>
<tr>
<td>Subjects</td>
<td>805</td>
<td>735</td>
</tr>
<tr>
<td>Pseudo R-squared</td>
<td>0.010</td>
<td>0.020</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

**Table 7.** Player A’s HG decision to look at the precise cost of helping (1 = Look, 0 = Do not look) as a function of condition (1 = Process observable, 0 = Process hidden), non-comprehension (1 = Non comprehender, 0 = Comprehender), and their interaction, by experiment. Column 1 shows results from the Supplemental experiment; Column 2 shows results from Experiment 1.
Uncalculating cooperation actually is a signal

Next, we seek to replicate our finding that uncalculating cooperation actually is a signal in our Supplemental experiment, using the same analysis approach as in Experiment 1. We thus again investigate both experimental conditions, and begin by testing the effect of looking among helpers. We find the same pattern as in Experiment 1: Player As who helped without looking at the cost return an average of 51% of the amount they were sent to Player B, as compared to 44% among Player As who helped after looking. We confirm that this difference is significant with a linear regression predicting percentage returned by helpers as a function of looking, which finds a significant negative effect (coeff = -7.91, p < .001, N = 674, Table 8 Columns 1-2).

Next, we test our prediction that looking is a more negative predictor of trustworthiness among helpers than non-helpers. We conduct a linear regression predicting percentage returned (by either helpers or non-helpers) as a function of helping, looking, and their interaction. As in Experiment 1, we find a significant negative interaction (coeff = -15.64, p < .001, N = 805, Table 8 Columns 5-6), indicating that looking is a stronger predictor of untrustworthiness when Player A decided to help than when Player A decided not to help. We then investigate the effect of looking when Player A decided not to help. We find the same pattern as in Experiment 1: Player As who decided not to help without looking at the cost return an average of 17% of the amount they were sent to Player B, as compared to 25% among Player As who decided not to help after looking. However, while this difference was not significant in Experiment 1, a linear regression predicting percentage returned by non-helpers as a function of looking demonstrates that it is marginally significant in our Supplemental experiment (coeff = 7.72, p = .065, N = 131, Table 8 Columns 3-4). In sum, then, we replicate all of our findings from Experiment 1, and find a marginally significant reversal of the effect of looking among non-helpers (which was only observed directionally but not significantly in Experiment 1).

<table>
<thead>
<tr>
<th></th>
<th>Helpers</th>
<th>Non-helpers</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All subjects</td>
<td>Perf comp</td>
<td>All subjects</td>
</tr>
<tr>
<td>Help</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.702)</td>
<td>(3.186)</td>
<td>(4.145)</td>
</tr>
<tr>
<td>Help X Look at cost</td>
<td>-15.64***</td>
<td>-15.98</td>
<td>-15.64***</td>
</tr>
<tr>
<td>Constant</td>
<td>51.44***</td>
<td>50.16***</td>
<td>16.98***</td>
</tr>
<tr>
<td></td>
<td>(1.336)</td>
<td>(2.426)</td>
<td>(3.107)</td>
</tr>
<tr>
<td>Observations</td>
<td>674</td>
<td>177</td>
<td>131</td>
</tr>
<tr>
<td>Subjects</td>
<td>674</td>
<td>177</td>
<td>131</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.031</td>
<td>0.049</td>
<td>0.026</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05
Table 8. Player A’s TG returning as a function of Player A’s HG choice to look at the precise cost of helping in both conditions of our Supplemental experiment. Columns 1-2 show results from Player As who chose to help in the HG; Columns 3-4 show results from Player As who chose not to help in the HG; Columns 5-6 show results from all Player As and also includes a term for whether Player A chose to help, and the interaction between Player A’s helping and looking choices. Columns 1, 3, and 5 show results from all subjects; Columns 2, 4, and 6 show results from subjects who showed perfect comprehension. We note that “perfect comprehension” is defined here as answering all 6 pre-decision questions correctly, as well as the key post-decision question.

Player B results

We next turn to investigating Player Bs. Because the procedure for Player Bs was identical to Experiment 1, this is a direct replication, and we use the exact same analysis approach as in Experiment 1.

Comprehension

We begin with comprehension. We presented N = 805 Player Bs with 6 comprehension questions. On average, they answered 4.60 questions correctly (77% correct), and 31% of individuals answered all questions correctly.

Uncalculated cooperation is perceived as a signal

We next seek to replicate our finding that uncalculating cooperation is perceived as a signal. We thus again investigate the process observable condition, and begin by testing the effect of looking on helpers. We find the same pattern as in Experiment 1: Player Bs send an average of 57% of their endowments to Player As who helped without looking at the cost, as compared to 52% to Player As who helped after looking. We confirm that this is significant with a linear regression predicting percentage of endowment sent to helpers as a function of looking (clustering on subject), which finds a significant negative effect (coeff = -4.15, p < .001, N = 404, Table 9 Columns 1-2).

Next, we test our prediction that looking is a more negative predictor of trust among helpers than non helpers. We conduct a linear regression predicting endowment sent (to either helpers or non-helpers) as a function of helping, looking, and their interaction (again, clustering on subject; in this regression, there are four decisions for each subject). We find a significant negative interaction (coeff = -7.63, p<.001, N = 404, Table 9 Columns 5-6), indicating that looking is perceived more negatively when Player A decided to help than when Player A decided not to help. We then investigate the effect of looking when Player A decided not to help. We find the same pattern as in Experiment 1: Player Bs send an average of 24% of their endowments to Player As who chose not to help without looking at the cost, as compared to 27% to Player As who chose not to help after looking. However, while this difference was only marginally significant in Experiment 1, a linear regression predicting percentage of endowment sent to non-helpers as a function of looking (clustering on subject) demonstrates that it is significant in our Supplemental experiment (coeff = 3.48, p = .005, N = 404, Table 9 Columns 3-4). In sum, then, we replicate all of our findings from Experiment 1, and find a significant reversal of the effect of looking among non-helpers.
<table>
<thead>
<tr>
<th></th>
<th>Helpers</th>
<th>Non-helpers</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Perf comp</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td>subjects</td>
<td></td>
<td>subjects</td>
</tr>
<tr>
<td>Help</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Look at cost</td>
<td>-4.150***</td>
<td>-6.990**</td>
<td>3.482**</td>
</tr>
<tr>
<td></td>
<td>(1.094)</td>
<td>(2.193)</td>
<td>(1.221)</td>
</tr>
<tr>
<td>Help X Look at cost</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>56.51***</td>
<td>65.30***</td>
<td>23.99***</td>
</tr>
<tr>
<td></td>
<td>(1.774)</td>
<td>(2.997)</td>
<td>(1.560)</td>
</tr>
<tr>
<td>Observations</td>
<td>808</td>
<td>268</td>
<td>808</td>
</tr>
<tr>
<td>Subjects</td>
<td>404</td>
<td>134</td>
<td>404</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.003</td>
<td>0.010</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Table 9. Player B’s TG sending as a function of Player A’s HG choice to look at the precise cost of helping in the process observable condition of our Supplemental Experiment. Columns 1-2 show results from decisions about Player As who chose to help in the HG; Columns 3-4 show results from decisions about Player As who chose not to help in the HG; Columns 5-6 show results from all decisions and also includes a term for whether Player A chose to help, and the interaction between Player A’s helping and looking choices. Columns 1, 3, and 5 show results from all subjects; Columns 2, 4, and 6 show results from subjects who showed perfect comprehension. Standard errors are clustered on subject to account for repeated observations.
Is uncalculating decision-making a stronger signal when decisions are cooperative than uncooperative? A comparison using pooled datasets

An important finding in our experiments is that uncalculating cooperation but not uncalculating defection is perceived as, and actually is, a positive signal of trustworthiness. Directionally, uncalculating defection is perceived as and actually is a negative signal, and these effects reach significance in some but not all analyses (see below for details). This reversal is interesting because it is consistent with evidence that quick decisions can be perceived positively or negatively, depending on the nature of the decision (2), and demonstrates that it is specifically uncalculating cooperation that is seen positively, rather than uncalculating decision-making generally being seen as desirable.

However, while we consistently found that uncalculating cooperation was perceived as, and actually was, significantly better than calculating cooperation, we did not consistently find that calculating defection was significantly better than uncalculating defection. This raises the interesting possibility that calculating decision-making is a stronger signal in the context of cooperative than uncooperative decisions. In this section we attempt to better understand whether, in our experiments, (i) calculating defection is indeed significantly better than uncalculating defection and (ii) calculating decision-making is indeed a significantly stronger signal when decisions are cooperative, by pooling our datasets.

Analysis of perceived signaling value

We begin by investigating the perceived signaling value of calculating decision-making in the context of cooperation and defection. As a reminder, we found a significant positive effect of uncalculating cooperation (relative to calculating cooperation) on perceived trustworthiness within Experiment 1, Experiment 2, and our Supplemental experiment. In contrast, we found a significant negative effect of uncalculating defection (relative to calculating defection) in Experiment 2 and our Supplemental experiment, while finding a marginally significant effect in Experiment 1.

In this analysis, we pool all three experiments (Experiment 1, Experiment 2, and Supplemental experiment), as they all can be analyzed the same way (by comparing sending to uncalculating versus calculating helpers, and then by comparing sending to calculating versus uncalculating non-helpers, within the process observable condition; this allows us to estimate the positive effects of a “desirable” decision process (i.e. uncalculating decisions for helpers, and calculating decisions for non-helpers) in the context of cooperative / uncooperative decisions, respectively).

First, we compute the pooled positive effect of uncalculating cooperation, as compared to calculating cooperation, on sending. We find that on average, across our three experiments, Player Bs send an average of 57% of their endowments to uncalculating cooperators, as compared to 51% to calculating cooperators. A linear regression (clustering on subject) reveals that this difference is significant (coeff = 6.47, p < .001, N = 1130, Table 10 Columns 1-2).

Next, we compute the pooled positive effect of calculating defection, as compared to uncalculating defection, on sending. We find that on average, across our three experiments, Player Bs send an average of 25% of their endowments to calculating defectors, as compared to 22% to uncalculating defectors. A linear regression (clustering on subject) reveals that this difference is significant (coeff = 3.49, p < .001, N = 1130, Table 10 Columns 3-4).
Thus, these analyses demonstrate that when pooling our data, we find significant effects of both uncalculating cooperation, and calculating defection: both are perceived as positive signals of trustworthiness. However, the effect of uncalculating cooperation is close to twice as large in magnitude as the effect of calculating defection. A linear regression (clustering on subject) reveals that this difference is significant (coeff = 2.99, p = .003, N = 1130, Table 10 Columns 5-6). In sum, then, our pooled dataset reveals that both uncalculating cooperation and calculating defection are perceived as signals, but the effect of uncalculating cooperation is significantly larger.

<table>
<thead>
<tr>
<th></th>
<th>Helpers</th>
<th>Non-helper</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Perf comp</td>
<td>All</td>
</tr>
<tr>
<td>Help</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Help</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Desirable&quot; decision process</td>
<td>6.472***</td>
<td>10.21***</td>
<td>3.487***</td>
</tr>
<tr>
<td>Help X Desirable process</td>
<td>(0.703)</td>
<td>(1.247)</td>
<td>(0.728)</td>
</tr>
<tr>
<td>Constant</td>
<td>50.71***</td>
<td>53.17***</td>
<td>21.69***</td>
</tr>
<tr>
<td></td>
<td>(1.062)</td>
<td>(1.713)</td>
<td>(0.909)</td>
</tr>
<tr>
<td>Observations</td>
<td>2,260</td>
<td>888</td>
<td>2,260</td>
</tr>
<tr>
<td>Subjects</td>
<td>1130</td>
<td>444</td>
<td>1130</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.008</td>
<td>0.019</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Table 10. Player B’s TG sending as a function of whether Player A used a “desirable” decision process in the HG (i.e. uncalculating decisions for helpers, or calculating decisions for non-helpers) in the process observable condition of our pooled dataset (Exp 1, Exp 2, and Supp exp). Columns 1-2 show results from decisions about Player As who chose to help in the HG; Columns 3-4 show results from decisions about Player As who chose not to help in the HG; Columns 5-6 show results from all decisions and also includes a term for whether Player A chose to help, and the interaction between Player A’s helping choice and decision process. Columns 1, 3, and 5 show results from all subjects; Columns 2, 4, and 6 show results from subjects who showed perfect comprehension. Standard errors are clustered on subject to account for repeated observations.
Analysis of actual signaling value

We next investigate the actual signaling value of calculating decision-making in the context of cooperation and defection. As a reminder, we found a significant positive effect of uncalculating cooperation (relative to calculating cooperation) on actual trustworthiness within Experiment 1, Experiment 2, and our Supplemental experiment. In contrast, we found a marginally significant negative effect of uncalculating defection (relative to calculating defection) in our Supplemental experiment, while finding non-significant effects in same direction in Experiments 1 and 2.

In this analysis, we pool only Experiment 1 and our Supplemental experiment, because only they can be analyzed the same way (by comparing returning by uncalculating versus calculating helpers, and then by comparing returning by uncalculating versus calculating defectors, across both conditions; this allows us to estimate the positive effects of a “desirable” decision process (i.e. uncalculating decisions for helpers, and calculating decisions for non-helpers) in the context of cooperative / uncooperative decisions, respectively). Because Study 2 involves a continuous variable (helping decision time) as the measure of uncalculating versus calculating decision processes, it cannot be analyzed in this format and thus is not included in our pooled analysis.

First, we compute the pooled positive effect of uncalculating cooperation, as compared to calculating cooperation, on returning. We find that on average, across our two experiments, Player As who were uncalculating cooperators return an average of 51% in the TG, as compared to 42% among calculating cooperators. A linear regression reveals that this difference is significant (coeff = 8.47, p < .001, N = 1269, Table 11 Columns 1-2).

Second, we compute the pooled positive effect of calculating defection, as compared to uncalculating defection, on returning. We find that on average, across our two experiments, Player As who were calculating defectors return an average of 22% in the TG, as compared to 18% among uncalculating defectors. A linear regression reveals that this difference is not significant (coeff = 4.48, p = .146, N = 271, Table 11 Columns 3-4), although we note that the sample size of defectors is substantially smaller than the sample size of cooperators.

Thus, these analyses demonstrate that when pooling our data, we find significant effects of uncalculating cooperation, but not calculating defection. We also find that the effect of uncalculating cooperation is close to twice as large in magnitude as the effect of calculating defection. However, a linear regression (clustering on subject) reveals that this difference is not significant (coeff = 3.99, p = .226, N = 1540, Table 12 Columns 5-6).

In sum, then, our pooled dataset reveals that uncalculating cooperation actually is a significant signal of trustworthiness, whereas the effect of uncalculating defection on untrustworthiness does not reach significance. However, the difference between the effects of uncalculating cooperation and uncalculating defection also does not reach significance. We note that these non-conclusive results likely reflect that, relative to our “perceived as” analysis (where sending is our dependent variable), our “actually is” analysis (where returning is our dependent variable) has substantially less power. The reason for this is that it does not involve a fully balanced within-subjects dataset (whereby every subject provides an observation for uncalculating and calculating helping and non-helping): each subject either was calculating or uncalculating, and helped or did not help.
<table>
<thead>
<tr>
<th></th>
<th>Helpers</th>
<th>Non-helpers</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Perf comp</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td>subjects</td>
<td></td>
<td>subjects</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Help</td>
<td></td>
<td></td>
<td>24.55***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2.535)</td>
</tr>
<tr>
<td></td>
<td>(1.213)</td>
<td>(2.154)</td>
<td>(3.073)</td>
</tr>
<tr>
<td>Help X Desirable process</td>
<td></td>
<td></td>
<td>3.993</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(3.297)</td>
</tr>
<tr>
<td>Constant</td>
<td>42.12***</td>
<td>38.16***</td>
<td>17.56***</td>
</tr>
<tr>
<td></td>
<td>(0.779)</td>
<td>(1.399)</td>
<td>(2.418)</td>
</tr>
<tr>
<td>Observations</td>
<td>1,269</td>
<td>358</td>
<td>271</td>
</tr>
<tr>
<td>Subjects</td>
<td>1,269</td>
<td>358</td>
<td>271</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.035</td>
<td>0.075</td>
<td>0.008</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Table 11. Player A’s TG returning as a function of whether Player A used a “desirable” decision process in the HG (i.e. uncalculating decisions for helpers, or calculating decisions for non-helpers) in both conditions of our pooled dataset (Exp 1 and Supp exp). Columns 1-2 show results from decisions about Player As who chose to help in the HG; Columns 3-4 show results from decisions about Player As who chose not to help in the HG; Columns 5-6 show results from all decisions and also includes a term for whether Player A chose to help, and the interaction between Player A’s helping choice and decision process. Columns 1, 3, and 5 show results from all subjects; Columns 2, 4, and 6 show results from subjects who showed perfect comprehension.
Experimental instructions

Experiment 1

Player As

In Experiment 1, we describe the HG to Player As with instructions and comprehension questions (on the same page) in the format illustrated below. We note that Player A and the recipient are are labeled as “Player 1” and “Player 2.”
In addition to the payment you will receive for participating in this HIT, you can earn more money as a bonus, as follows:

You will participate in an interaction with other MTurk workers, involving two different games. You will take part in both games.

The first game has two players. In this game, you will be Player 1, and the other person will be Player 2.

You (Player 1) start with 20 cents, while Player 2 starts with nothing. You then have the chance to pay a cost to help Player 2 gain a bonus. Specifically, if you help, whatever cost you pay, Player 2 will receive twice that much as a bonus.

Importantly, by default, you do not know how much this cost is, only that it is somewhere between 1 and 20 cents (that is, it is some portion of the 20 cents that you start with). If you choose to help Player 2 by paying the cost, you only get to keep whatever remains of the 20 cents (i.e. 20 cents minus the cost). But if you do NOT choose to help, you get to keep all 20 cents.

So, for example, the cost turned out to be 5 cents, helping would entail you sacrificing 5 cents to deliver 10 cents to Player 2 (and keeping the remaining 15 cents). Alternatively, if the cost turned out to be 20 cents, helping would entail sacrificing 20 cents to deliver 40 cents to Player 2 (and keeping nothing).

While you by default do NOT know what the cost is, you can request to find out before deciding whether to help.

So, by default, you must choose whether to help Player 2 without knowing what the cost is. But, you can decide to find out what the cost is before making this choice.

So, your decisions can be represented with this tree:

Please answer the following questions, to make sure you understand this first game. You MUST answer ALL questions correctly to receive your bonus!

Imagine that you are deciding whether or not to help Player 2.
Which decision will result in the highest possible bonus for you?
- You deciding to help
- You deciding NOT to help
- It depends on what the cost of helping is

Imagine that you are deciding whether or not to help Player 2.
Which decision will result in the highest possible bonus for Player 2?
- You deciding to help
- You deciding NOT to help
- It depends on what the cost of helping is

What are the two the choices that you have?
- Whether to find out the cost of helping, and whether to help Player 2
- Whether to help Player 2 if the cost of helping is large, and whether to help Player 2 if the cost of helping is small
- Whether to help Player 2, and how much to help Player 2
Correct answers: 1) You deciding NOT to help; 2) You deciding to help; 3) Whether to find out the cost of helping, and whether to help Player 2.

We subsequently describe the TG with instructions and comprehension questions (on the same page) in the format illustrated below. We note that Player A and Player B are labeled as the “Sender” and “Receiver.”

After finishing the first game, you will play a second game. The second game has two players. You will play the second game with a NEW player who did NOT participate in the first game.

In this game, you will be the “Receiver.” The new player will be the “Sender.”

In this second game:

• The Sender starts with 30 cents.
• The Sender then chooses how many cents, if any, to send to you (the Receiver).
• Any money sent to you is tripled: for every 1 cent the Sender sends you, you will receive 3 cents.
• You then choose how many cents, if any, to return to the Sender. You can return nothing, any intermediate amount, or everything.

If the Sender sends all 30 cents, you will receive 90 cents. If you return half of the 90 cents, you and the Sender will both earn 45 cents, and will both have more than the Sender started with.

But if you return nothing, you will earn 90 cents, while the Sender will earn nothing.

So, the Sender can gain money or lose money by sending you money, depending on how much you return.

Please answer the following questions, to make sure you understand the second game. You MUST answer ALL questions correctly to receive your bonus!

Imagine that the Sender is deciding how much to send to you.

Which decision will result in the Sender earning the highest payoff?

- The Sender deciding to send nothing
- The Sender deciding to send everything
- It depends on how much you decide to return to the Sender

Imagine that you are deciding how much to return to the Sender.

Which decision will result in you earning the highest payoff?

- You deciding to return nothing
- You deciding to return everything
- It depends on how much the Sender sent you

Did the Sender participate in the first game?

- Yes
- No

Correct answers: 1) It depends on how much you decide to return to the Sender; 2) You deciding to return nothing; 3) No.
Next, we inform subjects about the information that Player B can use when deciding how much to send them in the TG, and include an additional comprehension question about this. In the process observable condition, we present the following screen:

In the second game, the Sender will decide how much money to send to you.

As we explained previously, the Sender will be a NEW player who did NOT participate in the first game. However, when deciding how much money to send to you, the Sender can base this decision on how you behaved in the first game.

Specifically, the Sender can base this decision on whether you decided to help Player 2, as well as whether you decided to find out the cost of helping.

Please answer the following question, to make sure you understand. You MUST answer ALL questions correctly to receive your bonus!

When deciding how much to send to you, what information can the Sender base this decision on?

- Nothing
- Whether you helped Player 2
- Whether you decided to find out the cost of helping
- Whether you helped Player 2, and whether you decided to find out the cost of helping

**Correct answer:** Whether you helped Player 2, and whether you decided to find out the cost of helping.

In the process hidden condition, we present the same screen, with the exception that the bolded sentence reads “Specifically, the Sender can base this decision on whether you decided to help Player 2 (but not on whether you decided to find out the cost of helping).” Thus, the answer to the comprehension question is “Whether you helped Player 2”.

Next, subjects decide whether to find out the cost of helping. In the process observable condition, we show the following screen:

It is now the first game. You will now decide whether or not to find out the cost of helping Player 2.

Right now, all you know is that helping means you paying a cost to help Player 2 gain a bonus twice the size of that cost. You know that the cost is somewhere between 0 and 20 cents, but you do not know exactly how much it is.

**NOTE:** This game involves REAL PEOPLE, and your decisions in this game will determine how much bonus you and the other players actually receive. You really do have the chance to find out the cost of helping, and the Sender really does get to base their sending decision on whether you decide to find out. Once the HIT is over, we will use your decisions from this game to calculate your bonus, and the bonuses of the other players.

What would you like to do?

- I WOULD like to find out the cost of helping
- I would NOT like to find out the cost of helping

In the process hidden condition, we present the same screen, with the exception that the underlined sentence reads “You really do have the chance to find out the cost of helping.”

If subjects decide to look at the cost of helping, they see the following screen:
You decided to find out what the cost of helping is. The cost of helping is **10 cents**.

And if they decide not to look, they see the following screen:

You decided not to find out what the cost of helping is.

Then, all subjects make their helping decisions. If they decided to look at the cost of helping, they see the following screen:

You will now make your Player 1 helping decision. It is your job to decide whether or not to help Player 2.

You have received 20 cents. Helping means paying a cost to help Player 2 gain a bonus twice the size of that cost. You chose to find out what the cost is, so you know that the cost is 10 cents.

**NOTE:** This game involves REAL PEOPLE, and your decisions in this game will determine how much bonus you and the other players actually receive. You really do have the chance to help Player 2, and the Sender really does get to base their sending decision on whether you decide to help. Once the HIT is over, we will use your decisions from this game to calculate your bonus, and the bonuses of the other players.

What would you like to do?
- I WOULD like to pay the cost to help Player 2
- I would NOT like to pay the cost to help Player 2

And if they decided not to look, we present the same screen, with the exception that the last sentence of the second paragraph is replaced with “You chose not to find out what the cost is, so all you know is that the cost is somewhere between 0 and 20 cents. You do not know exactly how much it is.”

All subjects subsequently decide how much to return in the TG on the following page:

It is now your turn to play the second game. The Sender will decide how much to send to you. **You are the Receiver.**

It is now your job to decide how much to return to the Sender.

Specifically, you will decide the percentage of cents you would like to return.

For example, if you decide to return 40 percent of the cents you receive...

- You will return 0 cents if the sender sends 0 cents (and you receive 0 cents)
- You will return 36 cents if the sender sends 30 cents (and you receive 90 cents)

**NOTE:** This game involves REAL PEOPLE, and your decisions in this game will determine how much bonus you and the other players actually receive. You really do have the chance to decide how much to return to the Sender. Once the HIT is over, we will use your decisions from this game to calculate your bonus, and the bonuses of the other players.

What percentage would you like to return?

<table>
<thead>
<tr>
<th>Percentage to return</th>
<th>0%</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
<th>40%</th>
<th>50%</th>
<th>60%</th>
<th>70%</th>
<th>80%</th>
<th>90%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Player Bs**

For Player Bs, we describe the HG with the following instructions and comprehension questions:
In addition to the payment you will receive for participating in this HIT, you can earn more money as a bonus, as follows:

You will participate in an interaction with other MTurk workers, involving two different games.

You will ONLY take part in the second game, and will NOT play in the first game. However, we would like you to read about and understand both games.

The first game has two players: Player 1 and Player 2. In this game:

Player 1 starts with 20 cents, while Player 2 starts with nothing. **Player 1 then has the chance to pay a cost to help Player 2 gain a bonus. Specifically, if Player 1 helps, whatever cost Player 1 pays, Player 2 will receive twice that much as a bonus.**

Importantly, by default, **Player 1 does not know how much this cost is**, only that it is somewhere between 1 and 20 cents (that is, it is some portion of the 20 cents that Player 1 starts with). If Player 1 chooses to help Player 2 by paying the cost, Player 1 only gets to keep the whatever remains of the 20 cents (i.e. 20 cents minus the cost). But if Player 1 does NOT choose to help, Player 1 gets to keep all 20 cents.

So if, for example, the cost turned out to be 5 cents, helping would entail Player 1 sacrificing 5 cents to deliver 10 cents to Player 2 (and keeping the remaining 15 cents). Alternatively, if the cost turned out to be 20 cents, helping would entail sacrificing 20 cents to deliver 40 cents to Player 2 (and keeping nothing).

**While Player 1 by default does NOT know what the cost is, Player 1 can request to find out before deciding whether to help.**

So, by default, Player 1 must choose whether to help Player 2 without knowing what the cost is. But, Player 1 can decide to find out what the cost is before making this choice.

So, Player 1's decisions can be represented with this tree:

Please answer the following questions, to make sure you understand this first game. **You MUST answer ALL questions correctly to receive your bonus!**

Imagine that Player 1 is deciding whether or not to help Player 2.

Which decision will result in the highest possible bonus for **Player 1**?
- Player 1 deciding to help
- Player 1 deciding NOT to help
- It depends on what the cost of helping is

Imagine that Player 1 is deciding whether or not to help Player 2.

Which decision will result in the highest possible bonus for **Player 2**?
- Player 1 deciding to help
- Player 1 deciding NOT to help
- It depends on what the cost of helping is

What are the two the choices that Player 1 has?
- Whether to find out what the cost of helping, and whether to help Player 2
- Whether to help Player 2 if the cost of helping is large, and whether to help Player 2 if the cost of helping is small
- Whether to help Player 2, and how much to help Player 2
Correct answers: 1) Player 1 deciding NOT to help; 2) Player 1 deciding to help; 3) Whether to find out the cost of helping, and whether to help Player 2.

We subsequently describe the TG with the following instructions and comprehension questions:

In the second game, you WILL participate. The second game has two players. You will play this game with another person who ALSO participated in the first game, in the role of Player 1.

In this game, you will be the "Sender." This other person, who was previously Player 1, will be the "Receiver."

In this second game:
• You (the Sender) start with 30 cents.
• You then choose how many cents, if any, to send to the Receiver.
• Any money you send to the Receiver is tripled: for every 1 cent you send, the Receiver will receive 3 cents.
• The Receiver then chooses how many cents, if any, to return to you. The Receiver can return nothing, any intermediate amount, or everything.

If you send all 30 cents, the Receiver will receive 90 cents. If the Receiver returns half of the 90 cents, you and the Receiver will both earn 45 cents, and will both have more than you started with.

But if the Receiver returns nothing, he or she will earn 90 cents, while you will earn nothing.

So, you can gain money or lose money by sending money to the Receiver, depending on how much the Receiver returns.

Please answer the following questions, to make sure you understand the second game. You MUST answer ALL questions correctly to receive your bonus!

Imagine that you are deciding how much to send to the Receiver.

Which decision will result in the highest possible bonus for you?
- You deciding to send nothing
- You deciding to send everything
- It depends on how much the Receiver decides to return to you

Imagine that the Receiver is deciding how much to return to you.

Which decision will result in the highest possible bonus for the Receiver?
- The Receiver deciding to return nothing
- The Receiver deciding to return everything
- It depends on how much you sent to the Receiver

Did the Receiver previously participate in the first game?
- No
- Yes

Correct answers: 1) It depends on how much the Receiver decides to return to you; 2) The Receiver deciding to return nothing; 3) Yes

We subsequently describe the “strategy method” (in which subjects make multiple sending decisions for different possible HG behaviors) in the format illustrated below. Shown below are the instructions from the process observable condition (in which subjects make 4 decisions based on Player A’s HG helping and looking decisions).
It is now your turn to play the second game.

You are the Sender. You have received 30 cents. You will now decide how much, if anything, to send to the Receiver.

Remember, the Receiver in this game previously participated in the first game in the role of Player 1.

You can base your decision on how the Receiver behaved as Player 1 in the first game. Specifically, you can base your decision both on whether the Receiver decided to find out what the cost of helping was, and whether the Receiver decided to help Player 2.

In the first game, the cost of helping turned out to be 10 cents. So, helping entailed Player 1 sacrificing 10 cents to deliver 20 cents to Player 2. But remember, Player 1 may or may not have known what the cost of helping was, depending on if Player 1 decided to find out.

Now, in the second game, you will make four different choices, one for each of the possible Player 1 decisions (i.e. to help without finding out the cost of helping, to help after finding out the cost of helping, to not help without finding out the cost of helping, and to not help after finding out the cost of helping). Your choice that goes with the decision the Receiver actually made in the first game will count for your bonus. But because you do not know what the Receiver chose in the first game, you do not know which of your decisions will actually be used to calculate your bonus. So, all of your decisions are important.

For every 1 cent you send to the Receiver, the Receiver gets 3 cents. Then the Receiver decides how much, if anything, to return to you.

NOTE: The other players are REAL, and your decisions will determine how much bonus you and the Receiver actually receive. Once the HIT is over, you will be told whether the Receiver decided to find out how much helping cost, and whether the Receiver helped, in the first game. Then, we will use your decisions and the Receiver's decision to calculate each of your bonuses.

Please advance the screen to make your first decision. You will make each decision on a separate page.

In the process hidden condition, we present the same screen, with the following exceptions:

- The second underlined sentence reads “You can base your decision on how the Receiver behaved as Player 1 in the first game. Specifically, you can base your decision on whether the Receiver decided to help Player 2.”
- The paragraph with the bolded text reads “Now, in the second game, you will make two different choices, one for each of the possible Player 1 decisions (i.e. to help, and to not help). Your choice that goes with the decision the Receiver actually made in the first game will count for your bonus. But because you do not know what the Receiver chose in the first game, you do not know which of your decisions will actually be used to calculate your bonus. So, both of your decisions are important.”
- The second sentence of the last paragraph reads “Once the HIT is over, you will be told whether the Receiver helped in the first game.”

Subjects subsequently make their sending decisions on decision pages in the format illustrated below. Subjects make one decision per page in a randomized order; show is the decision screen from the process observable condition about a Player A who did not look at the cost of helping and did help.

<table>
<thead>
<tr>
<th>Cents to send</th>
<th>0</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>How many cents would you like to send to a Receiver who, as Player 1 in the first game...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-Chose TO help Player 1, WITHOUT finding out the cost of helping?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The decisions about Player As who did not help replace the word “TO” with “NOT to”; the decisions about Player As who did look replace the word “WITHOUT” with “AFTER”. Subjects in the process hidden condition decide only on the basis of helping; decisions screens are thus modified to remove the text after the comma.

Importantly, we note that these decision screens had a typo, which we detected after completing data collection. They refer to Player 1 helping “Player 1” instead of “Player 2”. This error was constant across conditions and applied to all decision pages. We do not think that this typo affects the interpretation of the results, for two reasons. First, our results from Experiments 1 and 2 (shown in Figure 3 of the main text) clearly demonstrate that Player Bs treated helping as a very strong signal of trustworthiness: they sent much more when the decision screen read “Chose TO help Player 1” than when it read “Chose NOT to help Player 1.” This strongly suggests that subjects either did not notice the typo, or understood that our intended meaning was helping Player 2: their behavior demonstrates that they perceived helping as prosocial, even though it would not be if they had taken our typo literally, and interpreted “helping” as directed at the self. (And, it is reasonable that subjects could understand our intended meaning, given that we had previously described the HG as a situation where Player 1 could help Player 2, not Player 1). Second, and perhaps more importantly, we corrected this typo before running our Supplemental Experiment, which was otherwise identical to Experiment 1, and we replicated our results there.

**Experiment 2**

*Player As*

In Experiment 2, we describe the HG to Player As with the following instructions and comprehension questions:
In addition to the payment you will receive for participating in this HIT, you can earn more money as a bonus, as follows:

You will participate in an interaction with other MTurk workers, involving two different games. You will take part in both games.

The first game has two players. In this game, you will be Player 1, and the other person will be Player 2.

You (Player 1) start with 20 cents, while Player 2 starts with nothing. **You then have the chance to pay a cost to help Player 2 gain a bonus.** Specifically, if you help, whatever cost you pay, Player 2 will receive twice that much as a bonus.

Importantly, right now, you do not know how much this cost is, only that it is somewhere between 1 and 20 cents (that is, it is some portion of the 20 cents that you start with). If you choose to help Player 2 by paying the cost, you only get to keep whatever remains of the 20 cents (i.e. 20 cents minus the cost). But if you do NOT choose to help, you get to keep all 20 cents.

So if, for example, the cost turned out to be 5 cents, helping would entail you sacrificing 5 cents to deliver 10 cents to Player 2 (and keeping the remaining 15 cents). Alternatively, if the cost turned out to be 20 cents, helping would entail sacrificing 20 cents to deliver 40 cents to Player 2 (and keeping nothing).

While you currently do not know what the cost is, we will tell you before you have to decide whether to help.

Please answer the following questions, to make sure you understand this first game. You MUST answer ALL questions correctly to receive your bonus!

Imagine that you are deciding whether or not to help Player 2.

Which decision will result in the highest possible bonus for you?

- You deciding to help
- You deciding NOT to help
- It depends on what the cost of helping is

Imagine that you are deciding whether or not to help Player 2.

Which decision will result in the highest possible bonus for Player 2?

- You deciding to help
- You deciding NOT to help
- It depends on what the cost of helping is

If you decide to help Player 2, what is the cost to you?

- 0 cents
- 10 cents
- 20 cents
- I do not know yet, but I will find out later

**Correct answers:** 1) You deciding NOT to help; 2) You deciding to help; 3) I do not know yet, but I will find out later

We subsequently describe the TG as in Experiment 1. Next, we inform subjects (i) that we will measure and record their time on the decision screen, and (ii) about the information that Player B can use when deciding how much to send them in the TG. We also include an additional
comprehension question about (ii). In the process observable condition, we present the following screen:

Soon, you will play the first game and decide whether to help Player 2. On the decision screen, you will be told the cost of helping Player 2, and you will get to use this information to make your choice. We will measure and record how long you spend on this screen. In other words, we will take note of how long it takes you to decide whether to help Player 2.

Then, in the second game, the Sender will decide how much money to send to you. As we explained previously, the Sender will be a NEW player who did NOT participate in the first game. However, when deciding how much money to send to you, the Sender can base this decision on how you behaved in the first game.

Specifically, the Sender can base this decision on whether you decided to help Player 2, as well as how long it took you to decide whether to help Player 2.

Please answer the following question, to make sure you understand. You MUST answer ALL questions correctly to receive your bonus!

When deciding how much to send to you, what information can the Sender base this decision on?

- Nothing
- Whether you helped Player 2
- How long it took you to decide whether to help Player 2
- Whether you helped Player 2, and how long it took you to decide whether to help Player 2

Correct answer: Whether you helped Player 2, and how long it took you to decide whether to help Player 2

In the process hidden condition, we present the same screen, with the exception that the bolded sentence reads “Specifically, the Sender can base this on whether you decided to help Player 2 (but not on how long it took you to decide whether to help Player 2).” Thus, the answer to the comprehension question is “Whether you helped Player 2”.

Next, we warn subjects that the next screen will be the decision screen, in which the cost of helping is revealed and time spent on the page is recorded. In the process observable condition, we present the following screen:

On the next screen, you will decide whether to help Player 2 in the first game. We will tell you how much it costs to help, and then you will decide whether to help.

NOTE: These games involve REAL PEOPLE, and your decisions will determine how much bonus you and the other players actually receive. In the second game, the Sender really does get to base their sending decision on whether you decide to help in the first game, and on how long it takes you to decide whether to help. Once the HIT is over, we will use your decisions from this game to calculate your bonus, and the bonuses of the other players.

Please proceed to the next screen when you are ready to make your decision. We will record how long you spend on this screen.

In the process hidden condition, we present the same screen, with the exception that the underlined sentence reads “In the second game, the Sender really does get to base their sending decision on whether you decide to help in the first game.”

Next, all subjects find out the cost of helping and make their helping decision on the following page:
You will now make your Player 1 helping decision.

You have received 20 cents. The cost of helping is 10 cents. So, helping means sacrificing 10 cents to deliver a bonus of 20 cents to Player 2.

Would you like to help Player 2?

- YES
- NO

Finally, all subjects decide how much to return in the TG as in Study 1.

Player Bs
For Player Bs, we describe the HG with the following instructions and comprehension questions:
In addition to the payment you will receive for participating in this HIT, you can earn more money as a bonus, as follows:

You will participate in an interaction with other MTurk workers, involving two different games.

You will ONLY take part in the second game, and will NOT play in the first game. However, we would like you to read about and understand both games.

The first game has two players: Player 1 and Player 2. In this game:

Player 1 starts with 20 cents, while Player 2 starts with nothing. Player 1 then has the chance to pay a cost to help Player 2 gain a bonus. Specifically, if Player 1 helps, whatever cost Player 1 pays, Player 2 will receive twice that much as a bonus.

Importantly, initially, Player 1 does not know how much this cost is, only that it is somewhere between 1 and 20 cents (that is, it is some portion of the 20 cents that Player 1 starts with).

However, right before we ask Player 1 to decide whether to help Player 2, we tell Player 1 that the cost is 10 cents. Thus, we reveal that helping entails Player 1 sacrificing 10 cents to deliver 20 cents to Player 2. After we reveal this information to Player 1, we measure and record how long Player 1 takes to decide whether or not to help Player 2.

Please answer the following questions, to make sure you understand this first game. You MUST answer ALL questions correctly to receive your bonus!

Imagine that Player 1 is deciding whether or not to help Player 2.

Which decision will result in the highest possible bonus for Player 1?

- Player 1 deciding to help
- Player 1 deciding NOT to help
- It depends on what the cost of helping is

Imagine that Player 1 is deciding whether or not to help Player 2.

Which decision will result in the highest possible bonus for Player 2?

- Player 1 deciding to help
- Player 1 deciding NOT to help
- It depends on what the cost of helping is

What does Player 1 know about the cost of helping Player 2?

- Player 1 never knows the exact cost of helping (but does know that the cost is between 0 and 20 cents)
- Player 1 always knows that the exact cost of helping is 10 cents
- Player 1 initially does not know the exact cost of helping, but finds out that it is 10 cents right before deciding whether to help Player 2

Correct answers: 1) Player 1 deciding NOT to help; 2) Player 1 deciding to help; 3) Player 1 initially knows the exact cost of helping, but finds out that it is 10 cents right before deciding whether to help Player 2

We subsequently describe the TG as in Experiment 1. Next, we describe the strategy method and provide information about Player A decision times in the format illustrated below. Shown below are the instructions from the process observable condition (in which subjects make 4 decisions based on Player A’s HG helping and looking decisions).
It is now your turn to play the second game.

**You are the Sender.** You have received 30 cents. You will now decide how much, if anything, to send to the Receiver.

Remember, the Receiver in this game previously participated in the first game in the role of Player 1. As Player 1, the Receiver's job was to decide whether to sacrifice money to help Player 2. Recall that Player 1 found out that the cost of helping was 10 cents right before deciding whether to help Player 2, and then we measured and recorded how long Player 1 took to decide.

In the first game, half of Player 1s took 2.05 seconds or longer to decide whether to help. We will refer to these Player 1s as "taking a long time to decide". The other half of Player 1s were faster, and took less than 2.05 seconds to decide. We will refer to these Player 1s as "not taking a long time to decide".

Now, in the second game, you can base your sending decision on how the Receiver behaved as Player 1 in the first game. Specifically, you can base your decision on whether the Receiver decided to help Player 2, and how long the Receiver took to decide.

You will make **four different choices**, one for each of the possible Player 1 actions (i.e. helping without taking a long time to decide, helping after taking a long time to decide, not helping without taking a long time to decide, and not helping after taking a long time to decide). Your choice that goes with the action the Receiver actually took in the first game will count for your bonus. But because you do not know how the Receiver acted in the first game, you do not know which of your choices will actually be used to calculate your bonus. **So, all of your choices are important.**

For every 1 cent you send to the Receiver, the Receiver gets 3 cents. Then the Receiver decides how much, if anything, to return to you.

**NOTE:** The other players are REAL, and your choices will determine how much bonus you and the Receiver actually receive. Once the HIT is over, you will be told whether the Receiver helped, and how long the Receiver took to decide whether to help, in the first game. Then, we will use your choices and the Receiver's action to calculate each of your bonuses.

Please advance the screen to make your first decision. You will make each decision on a separate page.

In the process hidden condition, we present the same screen, with the following exceptions:

- The second underlined sentence reads “Now, in the second game, you can base your sending decision on how the Receiver behaved as Player 1 in the first game. Specifically, you can base your decision on whether the Receiver decided to help Player 2.”
- The paragraph with the bolded text reads “You will make **two different choices**, one for each of the possible Player 1 actions (i.e. helping, and not helping). Your choice that goes with the action the Receiver actually took in the first game will count for your bonus. But because you do not know how the Receiver acted in the first game, you do not know which of your choices will actually be used to calculate your bonus. **So, both of your choices are important.**”
- The second sentence of the last paragraph reads “Once the HIT is over, you will be told whether the Receiver helped in the first game.”

Subjects subsequently make their sending decisions as in Study 1, with the exception that we describe looking time instead of looking decisions. Below is the decision screen from the process observable condition about a Player A who decided quickly and *did* help.
The decisions about Player As who did not help replace the word “TO” with “NOT to”; the decisions about Player As who decided slowly replace the word “WITHOUT” with “AFTER”. Subjects in the process hidden condition decide only on the basis of helping; decisions screens are thus modified to remove the text after the comma.

**Supplemental experiment**

*Player As*

From the perspective of Player As, the Supplemental experiment was identical to Experiment 1, with the following exceptions:

First, the TG instructions were modified to explain that Player B will be a new player who did not participate in the HG, but will learn information about how Player A behaved in the HG. They thus read as follows (with the same comprehension questions as in Experiment 1):

> After finishing the first game, you will play a second game. The second game has two players. **You will play the second game with a NEW player who did NOT participate in the first game.**

> In this game, you will be the ”Receiver.” The new player will be the ”Sender.”

> While the Sender is a NEW player who was not a part of the first game, the Sender WILL learn information about how you behaved in the first game before playing the second game.

In this second game:

- The Sender starts with 30 cents.
- The Sender then chooses how many cents, if any, to send to you (the Receiver).
- Any money sent to you is tripled: for every 1 cent the Sender sends you, you will receive 3 cents.
- You then choose how many cents, if any, to return to the Sender. You can return nothing, any intermediate amount, or everything.

If the Sender sends all 30 cents, you will receive 90 cents. If you return half of the 90 cents, you and the Sender will both earn 45 cents, and will both have more than the Sender started with.

But if you return nothing, you will earn 90 cents, while the Sender will earn nothing.

So, the Sender can gain money or lose money by sending you money, depending on how much you return.

Please answer the following questions, to make sure you understand the second game. **You MUST answer ALL questions correctly to receive your bonus**!

Next, the subsequent page conveyed information to subjects about their condition less explicitly, via screenshots of the study from the perspective of Player B. Below is this page in the “process observable condition”: 
As we explained, in the second game, the Sender will decide how much money to send to you. But how will this work?

The Sender will decide using the screens shown below. Each question will be shown on a different screen, and the screens will be presented in a random order. The Sender will use the sliding scale to decide how much to send, with the 5 cent increments as anchor points.

Please look carefully at the screens as you may be asked a comprehension question about them later.

<table>
<thead>
<tr>
<th>Question</th>
<th>Cents to send</th>
</tr>
</thead>
<tbody>
<tr>
<td>How many cents would you like to send to a Receiver who, as Player 1 in the first game...</td>
<td></td>
</tr>
<tr>
<td>- Chose TO help Player 2, AFTER finding out the cost of helping?</td>
<td>0 5 10 15 20 25 30</td>
</tr>
<tr>
<td>How many cents would you like to send to a Receiver who, as Player 1 in the first game...</td>
<td></td>
</tr>
<tr>
<td>- Chose TO help Player 2, WITHOUT finding out the cost of helping?</td>
<td>0 5 10 15 20 25 30</td>
</tr>
<tr>
<td>How many cents would you like to send to a Receiver who, as Player 1 in the first game...</td>
<td></td>
</tr>
<tr>
<td>- Chose NOT to help Player 2, AFTER finding out the cost of helping?</td>
<td>0 5 10 15 20 25 30</td>
</tr>
<tr>
<td>How many cents would you like to send to a Receiver who, as Player 1 in the first game...</td>
<td></td>
</tr>
<tr>
<td>- Chose NOT to help Player 2, WITHOUT finding out the cost of helping?</td>
<td>0 5 10 15 20 25 30</td>
</tr>
</tbody>
</table>

And below is this page in the “process hidden” condition:
As we explained, in the second game, the Sender will decide how much money to send to you. But how will this work?

The Sender will decide using the screens shown below. Each question will be shown on a different screen, and the screens will be presented in a random order. The Sender will use the sliding scale to decide how much to send, with the 5 cent increments as anchor points.

Please look carefully at the screens as you may be asked a comprehension question about them later.

| How many cents would you like to send to a Receiver who, as Player 1 in the first game... |
|---------------------------------|---|---|---|---|---|---|
| -Chose TO help Player 2?       |
| 0 5 10 15 20 25 30             |

| How many cents would you like to send to a Receiver who, as Player 1 in the first game... |
|---------------------------------|---|---|---|---|---|---|
| -Chose NOT to help Player 2?   |
| 0 5 10 15 20 25 30             |

This page did not include a comprehension question, as it did in Experiment 1.

Next, the subsequent looking choice page was modified to remove the explicit reminder in the “process observable” condition that Player B could condition TG sending on Player A’s looking choice. For consistency, we also removed the explicit reminder in both conditions that subjects “really do have the chance” to find out the precise cost of helping. Thus, in both conditions, it looked as follows:

It is now the first game. You will now decide whether or not to find out the cost of helping Player 2.

Right now, all you know is that helping means you paying a cost to help Player 2 gain a bonus twice the size of that cost. You know that the cost is somewhere between 1 and 20 cents, but you do not know exactly how much it is.

**NOTE:** This game involves REAL PEOPLE, and your decisions in this game will determine how much bonus you and the other players actually receive. Once the HIT is over, we will use your decisions from this game to calculate your bonus, and the bonuses of the other players.

What would you like to do?
- I WOULD like to find out the cost of helping
- I would NOT like to find out the cost of helping

Then, subjects made their helping decisions and their TG returning decisions. These were identical to Experiment 1, except that, for consistency, we again removed the explicit reminder in both conditions that subjects “really do have the chance” to help / decide how much to return. Thus, the paragraph starting with “**NOTE:**” was identical to the analogous paragraph shown above, for both of these decisions screens.

Then, after subjects make their TG returning decisions, we include the explicit comprehension
question about condition (that was presented *before* looking decisions, on the condition assignment page, in Experiment 1). Thus, this question is presented on its own page, as follows:

Thank you for making your decisions. We would now like you to answer one more question about the game, to make sure you understood.

In the second game, when deciding how much to send to you, what information can the Sender base their decision on?

☐ Nothing
☐ Whether you helped Player 2
☐ Whether you decided to find out the cost of helping
☐ Whether you helped Player 2, and whether you decided to find out the cost of helping

Then, the study concludes as in Experiment 1.

*Player Bs*

From the perspective of Player Bs, the Supplemental Experiment is identical to Experiment 1, with the exception that the typo (described in the section showing instructions from Experiment 1) was corrected on TG decision screens.

**References**