**Bias, Trust, and Trustworthiness: An Experimental Study of Post Justice System Outcomes**

**Jason Ralston**

**Jason A. Aimone**

**Charles North**

*Baylor University*

**&**

**Lucas Rentschler**

*Utah State University*

**Abstract:** The criminal justice system imposes long-term spillover costs for the convicted and acquitted in the form of reduced employment opportunities and high rates of recidivism, possibly a byproduct of poor employment opportunities. This paper examines discriminatory behavior of investors and employers when they are given the opportunity to condition on their trustee's/worker's criminal record. Similar to the real world, our experiment shows that employers and investors do exhibit discriminatory behavior toward those with criminal conviction and those with criminal acquittals. However, no basis exists for the statistical discrimination, since reciprocator behavior is found to not depend on criminal record, while true innocence or guilt is found to play a large role in a subject's reciprocity. While access to true innocence and guiltiness would prove invaluable to investors and employers, this is unobtainable in the real world. Furthermore, equating convictions with true guiltiness is worrisome, given truly innocent subject's willingness to plead guilty.

**JEL codes:** K40, D90, D91, C91

## 1. Introduction

The American criminal justice system is known to be costly to all parties involved. Though many of these costs are upfront (hiring counsel, potential punitive costs, etc.), some spillover costs may linger for a substantial amount of time. These spillover effects, on the defendant's side, come in the form of reduced employment opportunities and recidivism (Alper el al., 2018; Raphael 2007; Couloute & Kopf 2018). The full magnitude of these costs is difficult to calculate, but are potentially greater than any of the upfront costs.

What are the economic consequences of plea bargains and convictions to truly innocent and truly guilty individuals? We will use a post-conviction economic game to examine the interaction of participants who have confessed to misconduct or been found guilty at trial (both when the participant was truly guilty and truly innocent) with a new counterpart who discovers that the agent has "a record." Specifically, we explore the behavior of such agents in trust games (of the investment game style of (Berg, Dickhaut, & McCabe, 1995) and wage/effort games (of the gift exchange style of (Charness, Attribution and Reciprocity in an Experimental Labor Market, 2004)) played before and after having accepted plea bargains or discovering the outcome of their trials. The laboratory is an ideal location for such a study, for as experimenters we can perfectly distinguish between false and accurate convictions and confessions.

We study defendant's spillover effects of involvement in the criminal justice system in an experimental setting, allowing for clean identification of those who are truly guilty and truly innocent. Additionally, it presents us the opportunity to causally identify the effects of our experimental criminal justice system. We find evidence that subject's trust, trustworthiness, and equity concerns are affected by their interaction with the criminal justice system. We also find evidence that the effect of the criminal justice system is heterogenous across a person's true innocence/guilt status. Specifically, bias against subjects accused, but found not guilty, increases after interaction with the criminal justice system. We also find evidence that reciprocity decisions change in a trust game environment post criminal justice system, with truly guilty subjects showing higher prevalence of betrayal behavior and truly innocent subjects reciprocating significantly more after the criminal justice system resolves. When studying the effects of the criminal justice system through the lens of the of the gift exchange, we find that the truly innocent are responsible for increases in payoff equity between themselves and their employer and *decrease* inequity that favors themselves at the expense of their employer partner. In general, the criminal justice system acts to exacerbate a-social behaviors in the truly guilty and promotes pro-social behaviors in the truly innocent.

Our study is not the first in the field of experimental economics to tack the actors in the criminal justice system. The many behaviors produced by judges (Guthrie et al., 2007; Rachlinski et al., 2009; Rachlinski et al., 2011), juries (Bornstein & Greene, 2011; Minson & Mueller, 2012), attorneys (Bushway et al., 2014; Marselli et al., 2014) have been studied in both field and laboratory experiments, both in framed and neutral environments. A more comprehensive survey of the experimental law literature can be found in Charnes & DeAngelo (2017). However, this study is unique in that it does not target decisions made during the criminal justice process. Rather, we target biases and behaviors that occur after the criminal justice system reaches its conclusion.

In this sense, our paper is related to a long line of experimental and behavioral research in trust and inequity aversion. The strand of literature in trust begun by Berg et al. (1995) examines how people

make investment decisions when their return is entirely at the discretion of another person. In general this literature finds that people are predisposed to trusting others, even when the Nash equilibrium of a one shot game would predict an absence of trust. This willingness to trust is thought to be deep rooted and related to oxytocin levels, which modulate the amount of trust a person exhibits. Our experiment allows us to study the dynamics of trust as it is changed by the criminal justice system, a process that can induce participants to think more critically about the nature of trust.

Our study is also informed by the inequity aversion literature. Games, such as the dictator game (Engel 2011), ultimatum game (Oosterbeek et al. 2004), and trust game (Berg et al. 1995) have been purposed to study subjects' equity preferences. In each paradigm it is generally observed that subjects split endowments or payments such that each player receives around 50% of the total payoff. When the design permits, subjects are often seen rejecting proposed allocations of money where they receive less than 30% of the total payoff. Though inequity aversion has been studied in the context of the gift exchange game (Abeler et al., 2010), we are the first to measure inequity aversion from the perspective of the worker.

**2. Design**

The current paper is part of a larger experimental study of the criminal justice system and its actors. We examine the behaviors of jurors, prosecutors, and defendants with each examination taking place within its own separate, but related experiment. This paper is not directly related to these studies but is closely related. To summarize briefly, the experiments cover the following topics:

1. Juries: The jury experiment examined how jury members made guilty/not guilty verdicts when presented with varying levels of crime severity and varying strengths of evidence of innocence and guilt. From this experiment, we collected juror verdicts, which could be used at the end of an experimental criminal justice session.
2. Prosecutors: The prosecutor experiment studied how prosecutors handled plea bargaining under different incentive schemes when presented with varying amounts of evidence of guilt and crime severity. From this experiment, we collected how prosecutors would act when presented with details about a potential defendant's case.
3. Defendant Plea Bargaining: We study the defendant side of plea bargaining, studying how defendants decide to accept or reject plea offers from a prosecutor when a rejection involves heading to trial. During the course of the experiment, we are able to record subject's true innocence or true guilt, which is known only to the defendant and to no one else. At the end of the criminal justice system, an outcome (never accused, acquitted, found guilty of a small crime, found guilty of a medium crime, or found guilty of a large crime) is realized. These two measures – the true innocence/guilt of a subject and their final criminal justice outcome are used extensively in the design and in the analysis of this paper.

The complete descriptions of the experimental designs of the above studies, can be found in Amoine et al. (2019), Ralston et al. (forthcoming), and Ralston et al. (forthcoming) respectively. Aside from places where the prior experiments directly inform aspects of the current study's design or analysis, we will refrain from fully describing the aspects of these experiments for the sake of brevity.

The current study made use of an experiment comprised of multiple games and tasks. The various orderings in which our experimental defendants proceeded through the tasks is displayed in Figure 1.

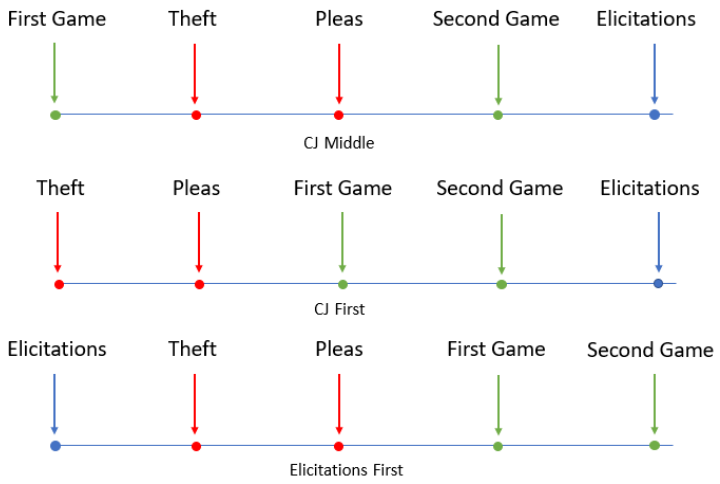


*Figure 1: Different orderings of tasks and games in the defendant experiment*

As can be seen from Figure 1, along with the theft game, plea-bargaining game, and elicitation tasks, subjects also partook in one of two games: a gift exchange game or a trust game. This paper concerns itself primarily with behaviors occurring in these gift exchange and trust games. Subjects always played only one of these two types of games in a session and they always played it twice. In some sessions, subjects played one game before the theft and plea-bargaining games and one game after. In other sessions, subjects played both games after the theft and plea-bargaining games. When we perform pre-post analyses, our post group is comprised only of observations of those in the second economic game for each ordering.

In the first of the two games, subjects were always split into two groups: those who were allowed to condition on the criminal status of their partner and those who could not. Those who could not condition on the criminal status of their partner instead could condition on the strategy of their partner – that is to say that those who could not condition of the criminal status of their partner engaged in a strategy method elicitation for the game they were playing.

In the second of the two games, players changed roles. Those who were able to condition on criminal status in the first game engaged in the strategy method elicitation in the second game, and those who engaged in the strategy method elicitation in the first game instead conditioned on criminal status in the second. Subjects did not keep the same partners from the first game to the second and matching was done randomly. Partners were also different from partners in the theft game. A more detailed explanation of how each game was conducted is described below.

**2.1 Gift Exchange Games**

The gift exchange implemented in our study is similar to that found in Charness et al. (2004). In it, subjects form pairs, with one filling the role of an employer and the other a worker. Both the employer and the worker simultaneous make decisions. The employer chooses a fixed payment to be given to the worker, in increments of 10 dimes, between 0 and 120, while workers select an amount of costly effort to return to give to their employer. A worker always had the option of rejecting the employers wage

offer, resulting in a payoff of $0 for both players. Payoff and cost functions for each player can be found below.

$$\pi_{employer} = (\mathbf{120} - \mathbf{wage}) * \mathbf{effort}$$

$$\pi_{employee} = \mathbf{wage} - \mathbf{cost\ of\ effort} - \mathbf{20}$$

| **Effort Level** | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Effort Cost** | 0 dimes | 1 dime | 2 dimes | 4 dimes | 6 dimes | 8 dimes | 10 dimes | 12 dimes | 15 dimes | 18 dimes |

As mentioned before, half of our subjects could condition on criminal status in each game. The statuses that could be conditioned on were Never Accused, Acquitted, Guilty of a Small Crime, Guilty of a Medium Crime, and Guilty of a Large Crime. Those that could condition on criminal status were *always* in the role of the employer. Therefore, all subjects that were employers in the first round were workers in the second round. Workers were never able to condition on criminal status. This design decision was made to reflect the occurrence in naturally existing labor markets that employers can ask and discriminate against employees on the basis of criminal history, the reverse is rarely true. An example screen for subjects in each roll are provided in the supplemental information.

**2.2 Trust Games**

The trust game we use in our study is similar to that of Berg et al (1995). In it, two subjects are matched to form a pair and each are given an endowment for 10 quarters. One subject plays first and must decide on a number between 0 and all ten of their quarters to give to their partner. Whatever the first mover sends is multiplied by a factor of 3 before it is given to their partner. The first mover's partner then chooses to send nothing or some portion of the multiplied sum of money back to the first mover. The second player cannot give any of their endowment back to the first player.

As noted previously, half of our subjects always conditioned on criminal justice outcomes (Never Accused, Acquitted, Guilty of a Small Crime, Guilty of a Medium Crime, and Guilty of a Large Crime.) In the case of the Trust game, the first movers always were the ones who could condition on criminal status and their partners never could. The same rationale as used in the gift exchange game applies here. Investors (whether an employer or someone investing in an entrepreneur) can ask and condition upon the criminal justice status of those with whom they are interacting. An example of each subjects' screen can be found in the appendix.

**Results**

**3.1 Bias against guilty and acquitted**

Before evaluating changes in behavior resulting from exposure to our experimental criminal justice system, we present basic results of bias against those who are guilty either by plea or by a guilty verdict rendered at trial. We pool decisions made when conditioning on the guilt status of those who were never accused and those who were acquitted together (the "not guilty" group) and compare them against the decisions regarding anyone who was determined to be "guilty" of any crime (the "guilty" group). These comparisons can be seen in Figure 2. Evident is the clear bias against those who have committed a crime, which is significant across both games (MW, $p<0.01$ in each).

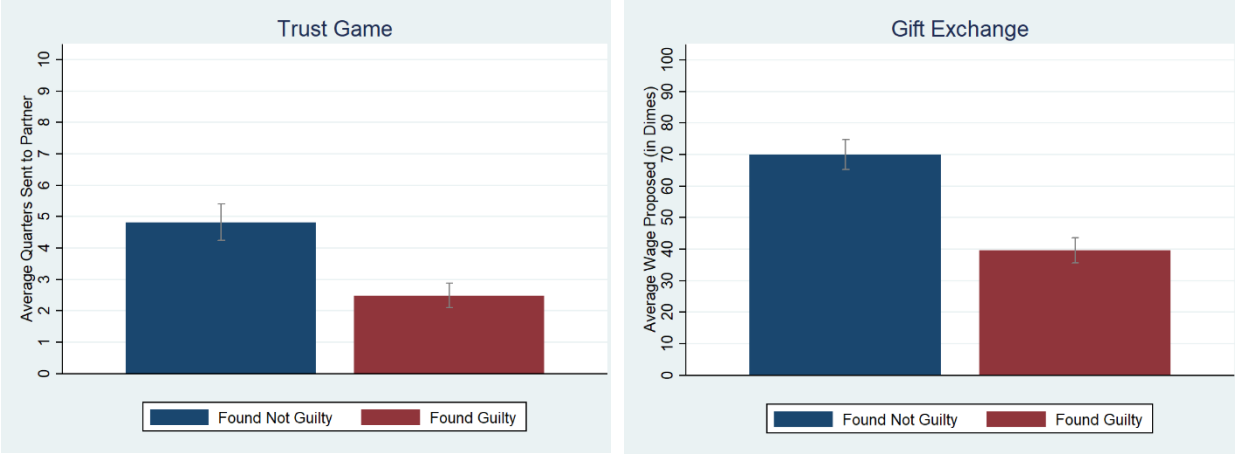


*Figure 2:Guilty bias across Trust and Gift Exchange games*

In the following analysis we check for deeper changes in behavior in our various economic games pre and post theft and plea-bargaining games. Our analysis of pre and post always compares a game that was played before the theft and plea-bargaining games to the second game played after the theft and plea-bargaining games.

We use a probit estimation to identify the effect the experimental criminal justice system has on different types of biases subjects may possess. A participant is labeled as having bias against a group if they either invested (in trust game) or offered a wage (gift exchange) that was strictly greater in a partner who had never been accused of a crime compared to the comparison group. To be clearer on the bias we differently explore in the left column a bias between someone never accused of a crime and someone who was found guilty of a crime (using average investment or wage for small/medium/large crime) and in the right column a bias between someone never accused of a crime and someone who was acquitted of a crime. From our previous analysis of plea-bargaining behavior, we know that truly innocent and truly guilty subjects are different, and we interact the effect of the experimental criminal justice system with true innocence/true guilty to study heterogenous effects. The results of these probit models are found in Table 1. A graphical comparison of bias follows in Figure 3.

| | **Probit** | |
|---|---|---|
| | Anti-Guilty Bias | Anti-Acquitted Bias |
| Truly Innocent (After CJ) | 0.129 | -0.476 |
| | (0.437) | (0.342) |
| Found Guilty (After CJ) | 0.119 | -0.181 |
| | (0.380) | (0.336) |
| Trust Game | -0.624 | 0.0131 |
| | (0.477) | (0.361) |
| After CJ Exposure | -0.291 | **0.997**** |
| | (0.497) | (0.413) |
| Trust Game (After CJ) | 0.986 | 0.0966 |
| | (0.633) | (0.485) |
| Baylor | 0.820* | 0.650** |

| | | |
|---|---|---|
| | (0.454) | (0.308) |
| Constant | 1.329*** | -0.180 |
| | (0.373) | (0.254) |
| Observations | 125 | 125 |

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

*Table 1:Probit estimation of effect of experimental criminal justice system on bias*

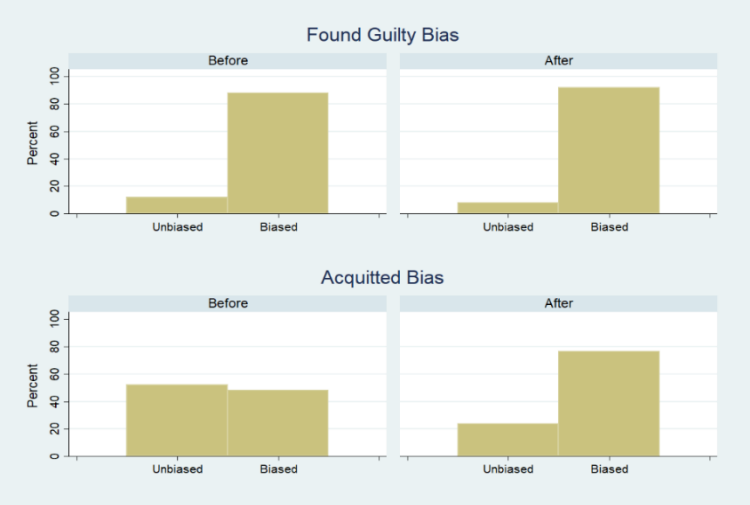


*Figure 3: Bias Pre and Post Criminal Justice*

*Finding 1: Bias against the acquitted increases post experimental criminal justice system*

Figure 4 shows that following exposure to the criminal justice system, those who can condition on criminal justice outcomes exhibit significantly more bias toward those who were accused but acquitted ($p<0.05$). The bias against the convicted does not change.

**3.2 Reciprocation**

The above biases exhibited by our investors/employers in our trust/gift-exchange games raise the following question: are investors/employers investing optimally in their partners conditional on their criminal justice status? That is, are investors/employers correctly anticipating what their partners would do for the different amounts of investments/wages they could be given?

To answer this question, we start by examining amounts returned to the investor (trust game) and effort given to the employer (gift exchange) conditional on the trustee's/worker's criminal justice status. Figure 5 shows the average return and effort level selected by trustees/workers conditional on the amount invested in them/wage paid to them. Figure 5 further decomposes these measures over the three criminal justice outcomes of interest – never accused, acquitted, and found guilty.

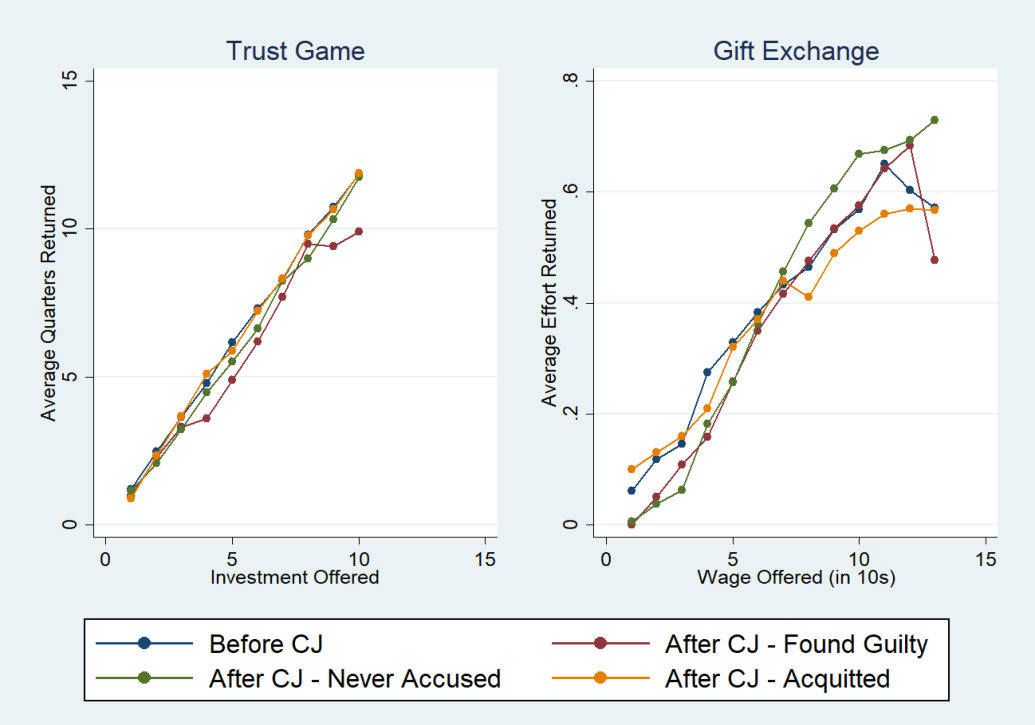


*Figure 4: Second Mover Behavior by Criminal Justice Status*

Figure 4 shows that a subject's reciprocal behavior varies little conditional on criminal justice outcome. Figure 4 also allows for us to test if investors/employers are making investment/wage offers in accordance with how their trustees/workers would reciprocate. That is to say, are investor/employers "rational" to discriminate against acquitted/guilty partners in a payoff maximizing sense? From our graphs, we can calculate the corresponding payoffs investors/employers would have received had they offered a particular investment/wage. From these potential payoffs, we can see what the payoff maximizing investment/wage offer would have been and for each type of partner they might be paired with (never accused, acquitted, or found guilty).

Given that the reciprocation behavior does not vary much across the different CJ statuses, it should come as no surprise that investors/employers should not be treating their potential partners much differently. Indeed, in the trust game, investors should be investing 10 quarters in never accused/acquitted subjects and 9 quarters into those found guilty in order to maximize their payoff. In the gift exchange game, employers should be offering all 3 types of partners 60 dimes in order to maximize their payoffs. However, when we test if our investors'/employers' behavior differs from payoff maximizing behavior using Mann-Whitney signed-rank tests, we find that investors always significantly under invest in their partners in the trust game (*p<0.001, MW signed-rank test*). In the gift exchange game, employers offer too high a wage to those who have never been accused (*p=0.0189, MW signed-rank test*), too low a wage to those who have been found guilty (*p<0.001, MW signed-rank test*), but offer the optimal wage to those who have been acquitted (*p=0.3922, MW signed-rank test*). This indicates that while investors/employers display a bias against the acquitted and those found guilty, the bias is neither warranted nor optimal.

Next, we examine coarser measures of reciprocity. In the trust game for each decision a subject makes, we calculate whether they reciprocated trust placed in them by their investor counterpart. If they returned less than their investor invested in them, we label that a betrayal ("B."). If they returned exactly what was invested, we label them a minimal reciprocator ("M.R."). If they return more than what was invested in them, we label them a reciprocator ("R."). Again, we check for differences in reciprocation between games played before the criminal justice games and after. These differences are displayed in Figure 5.

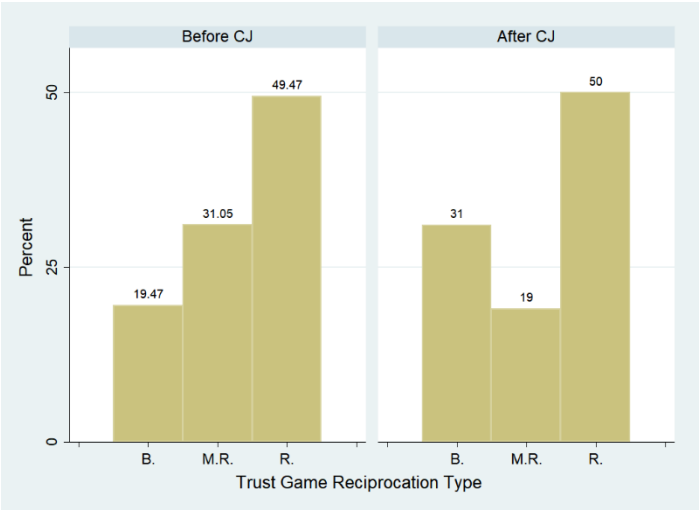


*Figure 5:Difference in rate of return pre and post criminal justice system*

It may be that our investors may be exhibiting their bias as a response to these coarser measures of reciprocity. Figure 6 displays the breakdown of reciprocity between the three CJ outcome classes – never accused, acquitted, and found guilty.

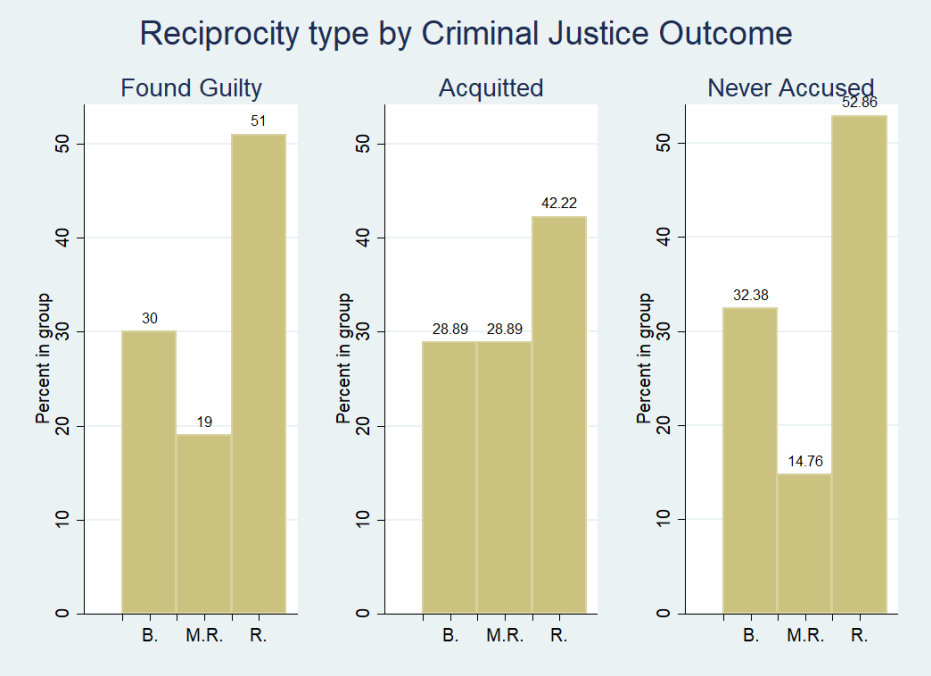


*Figure 6: Return rates by criminal justice status*

Figure 6 reveals that, even by coarser measures, behavior does not differ significantly between the three criminal justice outcome types. Using Mann-Whitney tests, we find that the only statistically significant difference lies in the number of minimal reciprocators there are among acquitted and among the never accused (*p=0.0043*). All other differences are insignificant at the 0.05 level. Given that the three types behave in statistically indistinguishable ways, it seems that the bias exhibited by investors is unwarranted.

Next, we explore what might be driving the large shift in reciprocity pre and post criminal justice system. Figure 5 shows that exposure to the criminal justice system is changing reciprocator behavior, thus we explore some other dimensions which might shift reciprocator behavior.

*Finding 2: The criminal justice system increases extreme reciprocity rates, both high and low. The truly guilty increase rates of betrayal and the truly innocent decrease the rates of betrayal. Criminal justice status insignificantly impacts reciprocity.*

Figure 5 shows that post criminal justice system, behavior becomes more extreme than prior to exposure to the criminal justice system. The percentage of minimal reciprocators drops significantly (from 31.05 to 19%, Mann-Whitney, $p<0.01$) and are seen reallocating themselves in the distribution to the betrayer category (from 19.47% to 31%, Mann-Whitney, $p < 0.01$).

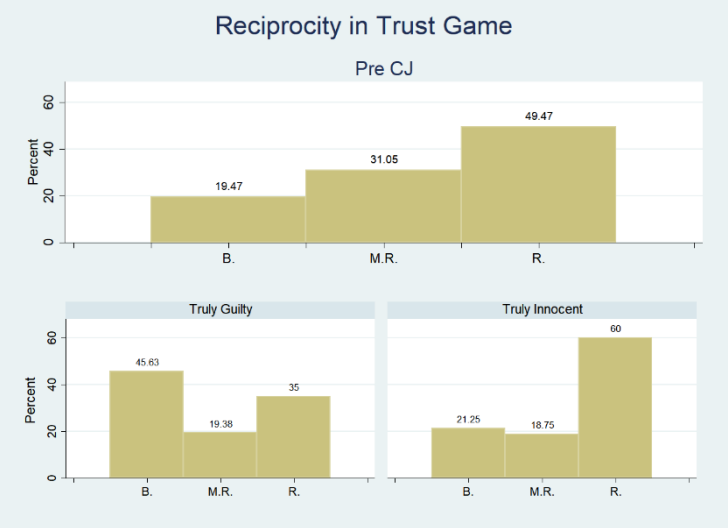


*Figure 7:Changes in return rate by true guilt/innocence*

When we separate our post-criminal justice group into those that were truly innocent and those that were truly guilty of the crime they made decisions about in the criminal justice system, differences in true innocence generate large differences in reciprocity, shown in Figure 7. Here we can see that the truly guilty are responsible for the shift into betrayers. Both the increase from 19.47% to 45.63% in betrayers and decrease in minimal reciprocators from 31.05% to 19.38% compared to the before group is significant (Mann-Whitney tests, $p<0.01$ and $p=0.01$ respectively). The truly innocent are seen to be responsible for an increase in reciprocators (from 49.47% to 60%, $p=0.03$).

### 3.3 Inequity Aversion

Next, we turn to a metric related to reciprocity, inequity aversion. In the context of the trust game, this metric measures one's sensitivity to payoff inequality between the two economic actors. The metric is simply the final payoff of the trustee minus the final payoff of the investor. If the difference is less than -2, we say that the trustee has pro-investor preference ("P.I"). If the difference is greater than 2, we label the trustee as pro-trustee ("P.T."). If the difference is between -2 and 2, then we say the trustee has a preference for equity ("P.E."). Figure 8 displays difference in the distribution of inequity preferences between pre and post criminal justice system trust games.

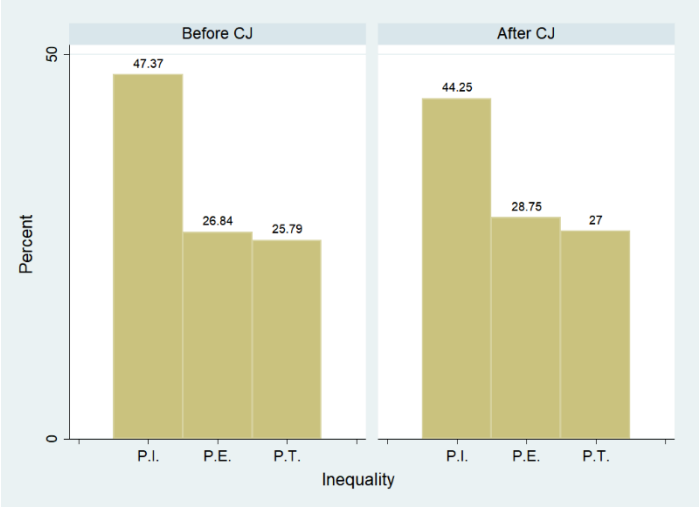


*Figure 8: Inequity preferences before and after the criminal justice system*

These differences are minor and the distribution does not change significantly (Mann-Whitney, $p > 0.1$). However, as per the previous analysis, we break observations of post-CJ preferences into those held by the truly innocent and those held by the truly guilty. These results are displayed in Figure 9.


*Figure 9: Inequity preferences by truly guilty/innocence status*

Figure 9 displays a few patterns that we saw when analyzing reciprocity, namely that the truly guilty become more self-interested (Mann-Whitney, $p = 0.019$) at the expense of becoming marginally less generous to their partners (Mann-Whitney, $p = 0.063$). The truly innocent do not change in any significant manner as a result of the criminal justice system.

We use probit models to measure the effect of the criminal justice system on inequity preferences to take advantage of subject's multiple trustee decisions. These results are found in Table 2. We find some support for our prior results and some results that were not present in the nonparametric analysis. Specifically, these results reveal that the criminal justice system increases more equitable behavior among those who are found guilty by our mock criminal justice system. Table 2 also reveals that the

truly innocent become less self-interested (and, conversely, the truly guilty become more self-interested).

| **Trust Game** | (1) Favors Investor | (2) Favors Trustee | (3) Favors Neither |
|---|---|---|---|
| After CJ | -0.188 | 0.3694 | -0.1749 |
| | (0.249) | (0.283) | (0.173) |
| Truly Innocent | 0.267 | **-0.5373**** | 0.2282 |
| | (0.224) | (0.261) | (0.156) |
| Convicted | -0.217 | -0.1229 | **0.3453**** |
| | (0.194) | (0.283) | (0.158) |
| Obs | 590 | 590 | 590 |
| Number of subjects | 55 | 55 | 55 |

*** p<0.005 or p>0.995, ** p<0.025 or p>0.975, *p<0.05 or p>0.95

Table 2: Probit estimation of inequity preferences in the trust game

Inequity aversion is especially relevant when we consider the gift exchange game. Unlike the trust game, where reciprocity is easily calculable and an intuitive metric to judge decisions by, the gift exchange game's structure does not allow for the same categorization technique. Subjects in the gift exchange must give to each other – there is no outside option to hold their endowment/contribution. Our metric of inequality is the difference between payoffs to employers and payoffs to employees. Figure 6 displays differences in inequality pre and post criminal justice system. Ideally a subject who is concerned only with equality would set their effort to a level that ensured both parties received an equal payoff. However, since this outcome is impossible, we classify subjects as equality concerned when they choose an effort that would result in both partners receiving within 10 dimes of an equal payoff ("P.Eq."). If the employer receives 10 dimes more than what would have been equal, then we say the subject has made a "pro-employer" choice ("P.E."). Similarly, if the worker receives 10 dimes more than what would have been equal, then we say the subject is "pro-worker" ("P.W."). We restrict our attention to choices made in a setting where all three outcomes are possible.


*Figure 10: Inequality pre and post criminal justice system*

*Finding 3: Interaction with the justice system increases (decreases) pro-social (asocial) behavior in the truly innocent and increases (decreases) asocial (pro-social) behavior in the truly guilty.*

Figure 10 shows evidence that differences in inequality are insignificantly different by second movers in the gift exchange game pre-and-post criminal justice exposure (Mann-Whitney, $p>0.1$). However, this may be masking differences we have witnessed between truly innocent and truly guilty subjects as well as between those found guilty and not guilty. We check for differences between truly innocent and truly guilty in Figure 11. Mann-Whitney U-tests reveal that the post criminal justice distributions are insignificantly different from the pre criminal justice in both the and found guilty and found not guilty cases ($p=0.336$ and $p=0.642$, respectively). However, when examining the decisions of those who are truly innocent, we find evidence that the criminal justice system changes behavior. Indeed, truly innocent subjects marginally increase equality preserving behavior (Mann-Whitney, $p=0.093$) and significantly *decrease* pro-worker behavior (Mann-Whitney, $p=0.023$). This is related to a result from the trust game, that truly innocent subjects tend to increase reciprocation to their investor partners.

Similar to the trust game, we use probit models to measure the effects of the criminal justice system on our experimental workers inequity preferences. The results of these probit models are found in Table 3. Controlling for other factors, the level effect of the criminal justice system is more muted. Though point estimates of the effects point in the correct direction, they are statistically insignificant.

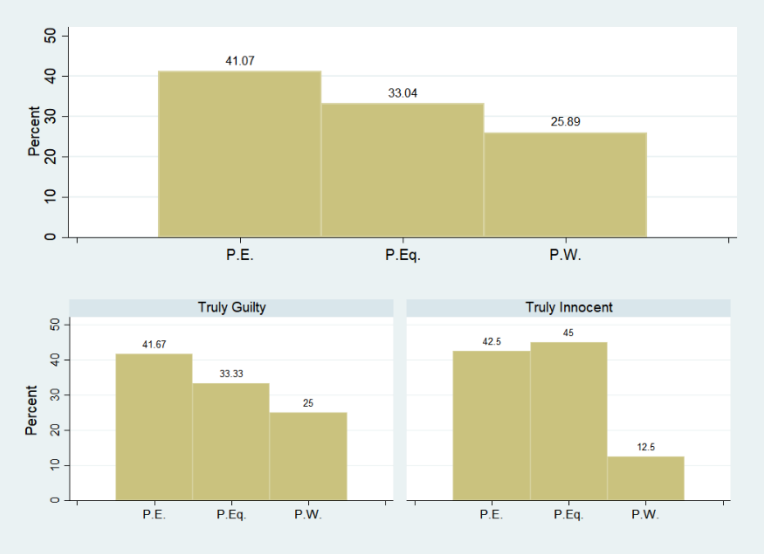


*Figure 11: Heterogenous criminal justice effects by true innocence status*

| **Gift Exchange** | (1) Favors Employer | (2) Favors Employee | (3) Favors Neither |
|---|---|---|---|
| After CJ | -0.055 | 0.073 | 0.004 |
| | (0.320) | (0.354) | (0.285) |
| Truly Innocent | 0.061 | -.5313 | 0.308 |
| | (0.276) | (0.352) | (0.246) |
| Convicted | 0.156 | -0.240 | 0.010 |
| | (0.262) | (0.387) | (0.226) |
| Obs | 264 | 264 | 264 |
| Number of subjects | 69 | 69 | 69 |

Note: Randomized p-values are reported in parentheses. Each value is the proportion of coefficients from 10,000 randomly generated replications that are less than the indicated coefficient.

*** p<0.005 or p>0.995, ** p<0.025 or p>0.975, *p<0.05 or p>0.95

*Table 3: Probit estimations of inequity preferences in the gift exchange game*

**4. Conclusion**

In our exploration of trust games and gift exchange games, occurring either before or after the criminal justice system, as would be expected, we observed bias against those who have been found guilty of a theft, by both investors and by employers (in their wage offers). While we observed that such bias against those convicted of crimes is unchanged if the investor or employer has been through the criminal justice environment themselves, we find that there is an increased bias against those who have been accused of crimes but who were acquitted. This indicates that even if a truly innocent person is acquitted/found not guilty of a crime they may still suffer costs and labor market biases from simply being falsely accused. However, when we investigate reciprocators behavior using both fine and coarse measures, we find that behavior over all criminal justice outcome groups is similar. This indicates that

bias exhibited against any criminal justice outcome group is unwarranted from a purely self-interested perspective, and that employers should exhibit similar amounts of trust across all groups.

We see that after having gone through the criminal justice system betrayal rates stay approximately the same with a marginal increase in rates of individuals reciprocating trust. However, our deeper analysis shows that underlying the flat rate of change of betrayers we see a significant change of types. Those who are truly guilty of crimes experience a significant increase in the propensity to betray trust, while those who are truly innocent of the crime, they were accused of shift to being significantly less like to betray trust.

It seems that, while employers understand that different types of workers will reciprocate in different manners (truly innocent are more like than truly guilty to reciprocate), our employers make the mistake of interpreting the convicted as truly guilty and those who are never accused as truly innocent. This is especially worrisome, since there is evidence in the real world and from our own experiments that shows that the truly innocent are likely to plead guilty in order to receive a more lenient sentencing than they would face at trial. In Ralston, et al. (2019), we find that innocent people are likely to plead guilty to crimes that they did not commit. In the real world, the number of convictions that result from guilty pleas account for 97% of all convictions (PPI, 2019). These criminal convicts are not necessarily guilty of serious crimes, either, and may plead guilty to minor infractions.

Overall, we see that those truly innocent appear to adopt a significantly more prosocial character in the face of the bias they face in wake of exposure to the criminal justice system, while the truly guilty appear to fall into more a-social behaviors. We see this time and again in our analyses, with regard to inequity concerns in both games we use as well as in reciprocity rates in the trust game. The criminal justice system acts as a primer for both types of subjects, and pushes them in opposite directions. As a result, repeated exposure could be expected to produce a large divergence in behaviors depending on the subject's true innocence or guilt, with those the truly innocent becoming more pro-social on average and those truly guilty becoming more asocial.

## References


Abeler, J., Altmann, S., Kube, S., & Wibral, M. (2010). Gift Exchange and Workers' Fairness Concerns: When Equality is Unfair. *Journal of the European Economic Association*.

Aimone, J., North, C., & Rentschler, L. (2019). Priming the jury by asking for Donations: An empirical and experimental study. *Journal of Economic Behavior and Organization*.

Alper, M., Durose, M., & Markman, J. (2018). 2018 Update on Prison Recidivism: A 9-Year Follow-up Period (2005 - 2014). *U.S. Department of Justice*.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior*.

Bornstein, B., & Greene, E. (2011). Jury Decision Making: Implications For and From Psychology. *Current Directions in Psychological Science*.

Bushway, S., Redlich, A., & Norris, R. (2014). An Explicit Test of Plea Bargaining In The "Shadow Of The Trial". *Criminology*.

Charness, G. (2004). Attribution and Reciprocity in an Experimental Labor Market. *Journal of Labor Economics*.

Charness, G., & DeAngelo, G. (Forthcoming). Law and Economics in the Laboratory.

Couloute, L., & Kopf, D. (2018). Out of Prison & Out of Work: Unemployment among formerly incarcerated people. *Prison Policy Initiative*.

Engel, C. (2011). Dictator games: a meta study. *Experimental Economics*.

Fehr, E., & Fischbacher, U. (2002). Why Social Preferences Matter - The Impact of Non-Selfish Motives On Competition, Cooperation And Incentives. *The Economic Journal*.

Guthrie , C., Rachlinski, J., & Wistrich, A. (2007). Blinking on the Bench: How Judges Decide Cases. *Cornell Law Review*.

Minson, J., & Mueller, J. (2012). The Cost of Collaboration: Why Joint Decision Making Exacerbates Rejection of Outside Information. *Psychological Science*.

Oosterbeek, H., Sloof, R., & van de Kuilen, G. (2004). Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental Economics*.

Rachlinski, J., Johnson, S., Wistrich, A., & Guthrie, C. (2009). Does Unconscious Racial Bias Affect Trial Judges? *Notre Dame Law Review*.

Ralston, J., Aimone, J., North, C., & Rentschler, L. (Forthcoming). False confessions: An experimental laboratory study of the innocence problem .

Ralston, J., Aimone, J., North, C., & Rentschler, L. (Forthcoming). Prosecutor Plea Bargaining and Conviction Rate Structure: Evidence from an Experiment.

Raphael, S. (2018). The employment prospects of ex-offenders. In C. Heinrich, & J. Scholz, *Social Policy Approaches that Promote Self-Sufficiency and Financial Independence Among the Poor.* New York: Russell Sage Foundation.

## Appendix

### A1: Alternate Analyses

One may consider an alternative set of analyses to run when we consider truly innocent and truly guilty subjects, namely that truly guilty and truly innocent subjects constitute static characteristics of the subjects and that changes in behavior pre and post criminal justice system should be measured within the respective group. To that end, we provide an alternative set of analyses which take these considerations into account and note that our results remain relatively unchanged.

We start with pre and post acquitted bias analysis, analogous to that found in Figure 4. These are presented in Figure 11.

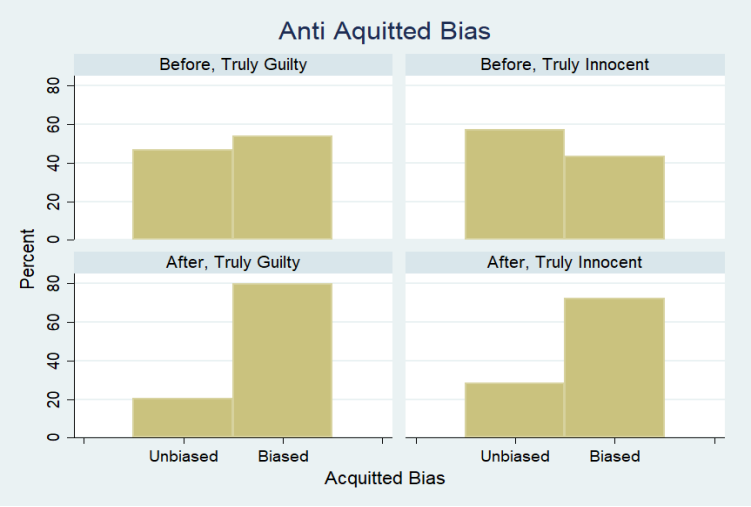


*Figure 11: Pre and post acquitted bias by true innocence/guilt*

As can be seen, both truly guilty and truly innocent subjects become more biased against acquitted subjects post criminal justice system (Mann-Whitney, $p < 0.01$; Mann-Whitney, $p < 0.01$).

We turn next to reciprocity and inequity concerns, using a similar analysis to what was shown in Figure 11 and analogous to Figures 6, 8, and 9 in the main text. Figure 12 compiles the results of these analyses. In general, the same results remain from the original analyses. The only major changes come when we find insignificant changes between pre and post trust game inequity concerns within truly guilty subjects (Mann-Whitney, $p > 0.10$) and insignificant changes between pre and post reciprocity within truly innocent subjects (Mann-Whitney, $p > 0.10$). However, the overall tendency for truly guilty subjects to become more a-social and truly innocent subjects to become more pro-social is still observed in many contexts.

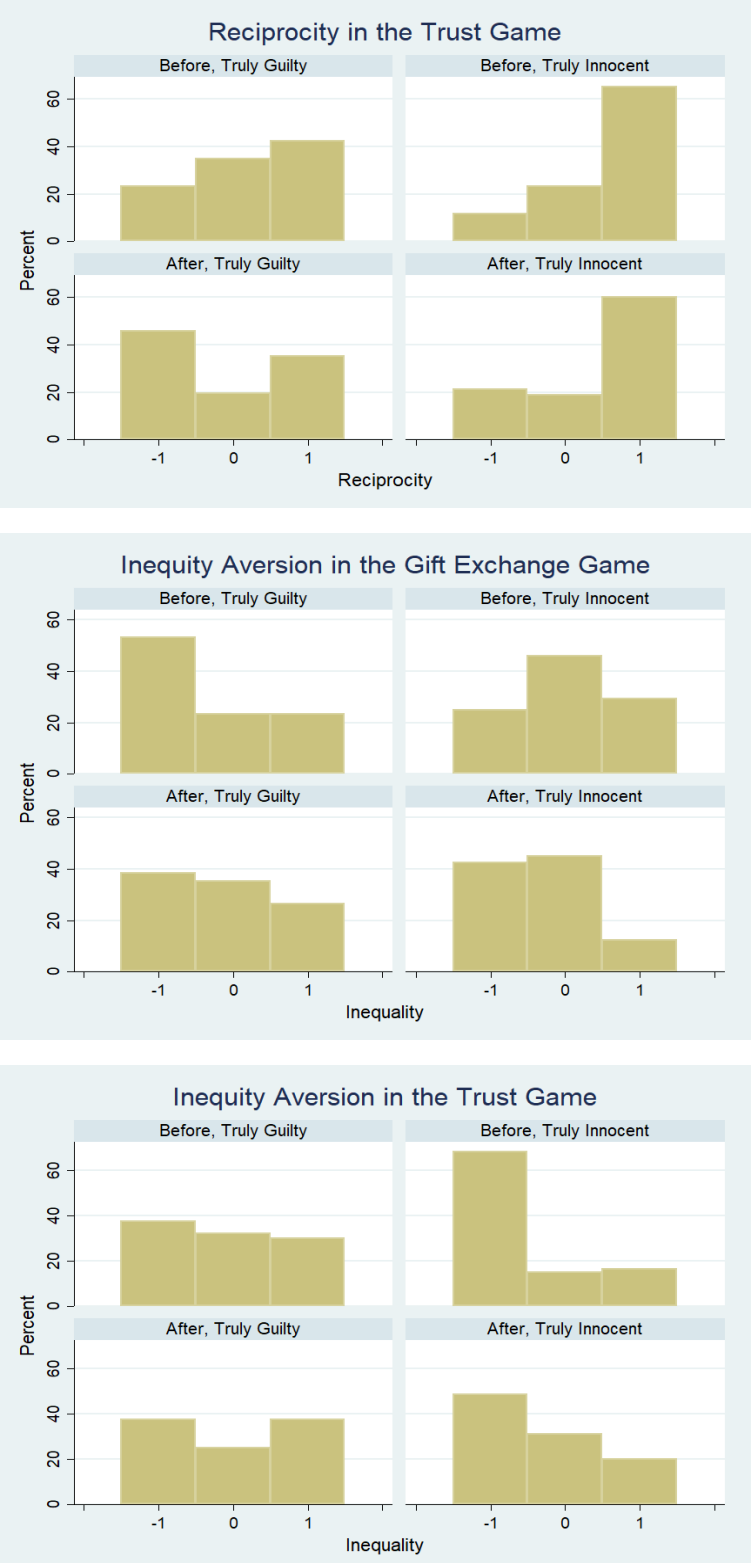


*Figure 3: Pre and post measures by true innocence/guilt*