Allocators Are More Prosocial When Affected Agents Can Visually Eavesdrop

Stephanie W. Wang
University of Pittsburgh

Colin F. Camerer
California Institute of Technology

Abstract:

In these experiments, participants made binary choices in single-chooser "dictator" games choosing distributions for themselves and others. . All payoffs are initially hidden and can be clicked open using a mouse. To study the effect of social image on attention and choices, we used a novel screensharing technique: One of the participants receiving the chooser's allocation can observe the chooser's clicks, so they can see if the chooser is looking up what the impact will be on their own allocation (but the observers cannot observe the chooser's choices). This change in observability increases the possible impact of social image concerns on expressed social preferences. It increases the time choosers spend looking at the potential payoffs to the observer and makes their choices less selfish. This finding goes against the hypothesis of "willful ignorance" and suggests other behavioral influences.

Keywords: Social preferences, Social image, Observability, Attention
JEL Classification: C92, D91

Support was provided by the Betty and Gordon Moore Foundation (CFC, SW), a Tamagawa GCOE grant (CFC), and the Tianqiao and Chrissy Chen Center for Social and Decision Neuroscience. Thanks to Walter Yuan and Chris Crabbe for rapid helpful programming, to Antonio Rangel for active collaboration early in this project, and to the referees and Editor.

Corresponding author: Wang, swwang@pitt.edu
Camerer, camerer@hss.caltech.edu

I. Introduction

In Plato's Republic, a character called Glaucon tells of a myth about a ring that can turn a person invisible. Plato's Glaucon insists that in general even a good person could not resist the temptation to behave at least a bit antisocially when cloaked in invisibility (as described fictionally in Baker, 1995). Put the other way around, knowing that we are observed puts us on our best behavior. As Plato writes (Jowett, 1941):

> "If you could imagine anyone obtaining this power of becoming invisible, and never doing any wrong or touching what was another's, he would be thought by the lookers-on to be a most wretched idiot, although they would praise him to one another's faces, and keep up appearances with one another from a fear that they too might suffer injustice. (2.359a–2.360d)

Glaucon's claim that people behave less badly when being observed is a hypothesis about one way that human nature deviates from simple selfishness. While the model of pure material self-interest will always prove to be a useful benchmark, social and natural scientists are now establishing a model of human nature that is richer, but also predictive. Such useful models include concerns for the distribution of rewards to other people (Bolton and Ockenfels, 1999; Fehr and Schmidt, 1999; Charness and Rabin, 2002) and reciprocity (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006). In these models, whether a person is observed by others does not make any difference.

However, in other models, people sometimes help others to increase their own "social image" (e.g., "manners" in Camerer and Thaler, 1995; Rotemberg, 2008, 2014; Benabou and Tirole, 2006; Andreoni and Bernheim, 2009). In these models, people derive utility when others believe they adhere to a norm of pro-sociality or are genuinely altruistic.

Some evidence from psychology and neuroscience indicates that being observed can influence behavior, including prosociality (e.g. Guerin, 1986). For example, an early small-sample study showed that people washed their hands in bathrooms twice as often (80% vs 40%) when they knew somebody was watching them (Munger and Harris, 1989). This kind of behavioral change in response to public observation indicates that some hand-washing is done to appear like a hygienic person, rather than to actually be hygienic.

In our study, participants choose allocations of money for themselves and others. All the monetary payoffs are hidden in boxes on a computer screen. When a cursor moves into the box area participants can see the monetary payoffs (a method called Mouselab). In the Baseline condition, other participants don't see the payoffs the decision maker is looking at. In the social image (SI) condition, an observer whose payoffs will be determined by the decision maker's choice gets to see-- through screen-sharing-- exactly what payoffs the decision maker is looking at. In both cases, the affected other participant does not observe the actual choices made by the active participant until the end of the experiment.

The goal of our experiment is to contribute new evidence of observability effects. While other studies have experimentally changed social observation (see the next section), our study is the first to vary actual social observation of a person's attention and decision-making process— visual eavesdropping-- by a social observer whose payoff is affected by the person being observed. This design can establish how social observation affects both attention and choice at the same time. Changing what other people know or see about a person's decision-making process, and measuring whether the person's decisions change, is a method that can discriminate social image models from distributional or reciprocity models most clearly. Specifically, this paradigm measures three interesting empirical properties of observability and social image: (1) Are people more prosocial when they are observed? (2) Does their visual attention to payoffs change when they are observed? And (3) Do people deliberately control social attention when they are observed, and in which direction? That is, do they look *more* at others' payoffs to improve social image, or do they look *less* to create deniability or "moral wiggle room"? Learning more about the answers to these questions is important for economics and social science for three reasons.

First, in recent decades a lively debate has arisen in economics about which models of social preference are best for various empirical purposes (Camerer, 2003; Cooper and Kagel, 2016).

Social image modeling is newer and more complex than the earlier alternatives. Knowing more about the importance and properties of social image effects is essential to advance the debate about whether the social image models are a workable improvement over simpler ones (Rotemberg, 2014).

Second, if social image concerns are important, and observability therefore matters, then institutional details of how social information is organized and broadcast will affect choices (and presumably welfare). One class of effects is how observation of actions enhances social image. An obvious example is how charities use publicity in acknowledging donors-- naming buildings, or putting their names on a plaque that is prominently displayed-- to advertise their generosity. Other studies indicate that social observation changes voting, education, and consumption decisions (DellaVigna et al., 2017; Bursztyn, and Jensen, 2015; Bursztyn et al., 2017). Social signaling has also been used to motivate prosocial choices such as deworming (Karing and Naguib, 2018).

A second class of effects is how observation of the information used by decision makers reassures economic agents that their interests are being considered (buttressed by the decision makers' desire for a good social image). For example, in business and management, law, and politics there are often rules about disclosure, transparency, and how the impact on people should be considered in decisions. These institutions include: Public corporate shareholder meetings, the right of the accused to physically face one's accuser in court, freedoms of the press, "sunshine laws", the Freedom of Information Act, widespread reliance on face-to-face personal interviews in selection and hiring, etc. These rules all promote the right of a person who is affected by decisions to get a chance to socially observe a decision maker— by directly confronting decision makers, or by seeing their paper trail. Some of these rules are required by law to mitigate agency problems, but others are voluntary actions designed to enhance image (e.g., WalMart donating goods to Hurricane Katrina victims). Thus, the institutions and practices designed to benefit observers presumably do so in part because they discipline and change what those being observed do.

Many studies in social psychology and organizational behavior suggest that workers will more readily accept unpleasant outcomes-- such as layoffs and pay cuts-- if they believe their interests are considered during a “procedurally just” process. (Lind and Tyler, 1988; Taylor et al., 1995). A related empirical example is that “apology laws”—which exclude apologies as evidence of guilt in malpractice cases-- reduce litigation (Ho and Liu, 2011a,b). These studies suggest that if observers see that a participant is looking at their payoffs, those observers will more readily accept selfish actions (which could lead participants to act more selfishly).

Third, technology has rapidly made social observation easier (e.g., through Facebook and other social media). That is, the information that people pay attention to and the options they weigh may have become more apparent to others while their final choices often remain private and unknown to anyone else. This boom has made image-cultivation a bigger business and changed the modern workplace. Understanding the underlying motivation for both observation and image are necessary to understand the development and welfare consequences of these technologies, the market activity they catalyze, and the new management challenges and opportunities they have created.

II. Background

Previous social image studies fall into three categories: (1) Presentation of abstract cues of observation; (2) Actual observation by disinterested people; and (3) deliberate image management by participants (through endogenously choosing what others know about their choice set and choice).

Abstract cues: Several experiments show pro-sociality is increased by indirect cues to being watched or heard (“surveillance cues”, in evolutionary psychology language). Cues in the form of stylized or abstract human eyes, or a vaguely human-like robot, have shown modest effects on prosociality in dictator games (Haley and Fessler, 2005; Burnham and Hare, 2007; Rigdon et al., 2009), in a workplace pay-what-you-want setting (Bateson et al., 2006), and in moral condemnation of crimes and villains in vignettes (Bourrat et al., 2011). However, there have been many failed replications. For example, Fehr and Schneider (2010) found no effect in a trust game. Most effects also appear to be interactions between a trait or state variable and surveillance cues.

Effects have been found when the eyes are in 3D rather than 2D (Krátký et al., 2016), and when eyes look directly compared to being averted (consistent with heightened biological activity from direct eye contact) (Manesi et al., 2016). There is also substantial evidence of effects from short exposure (<1min) and no effect at longer exposures (several minutes) (Sparks and Barclay, 2013).

A field experiment on voting measured the effect of sending a postcard encouraging voting ("Do your civic duty and vote!") with three background images—eyes, an American flag, or a palm tree. The eyes increased turnout by 1-2%, but the effect was only significant among Republican voters (Panagopoulos and van der Linden, 2016)

As a whole, these results demonstrate the intriguing possibility that even simple observation cues *could* have an effect, but the effect is not well-established (and could be inflated by publication bias if non-effects are either unsubmitted or rejected).

Human observers: Several studies have found that the presence of actual human observers influences prosociality (broadly interpreted). Izuma et al. (2010) shows heightened activation in well-established reward regions (ventral striatum), and more giving, when participants are watched while giving to charity; autistic adults show a diminished observation effect (Izuma et al., 2011). Coricelli et al. (2010) find that having your picture displayed to others deters an experimental version of tax evasion (and arousal, measured by skin conductance response, increases more during an experimental "audit" when your picture is displayed).

An effect of human observers on anti-social cheating was measured by a natural experiment in Italy, explicitly randomizing adult examiners to classrooms, as Italian 2nd and 5th graders took tests (Bertoni et al., 2013). The examiners watched for cheating and transmitted answers when the students were done. In classes with examiners, test scores were about 5-8% lower. The effect is probably due to reduced cheating, because students did not report feeling more anxious.

Endogenous image management: Experimental participants sometimes exhibit a preference to limit whether or not they are socially observed when they are making prosocial choices. Limiting social observation can create deniability or "moral wiggle room" to then act selfishly without the emotional burden (perhaps guilt) of being watched. In a modified dictator game, Lazear et al. (2012) and Dana et al. (2006) found that many people deliberately opted for a choice which had no scope for prosociality but gave them less money, in order to avoid a choice in which they could earn more but another person would know they had acted selfishly.

DellaVigna et al. (2012) found that some households, when notified that a charity would be visiting the next day, did not answer the door at the scheduled visiting time, presumably to avoid observability by the charity solicitor of the household's explicit decision to give nothing. Dillenberger and Sadowski (2011) and Benabou and Tirole (2006) provide formal models of these preferences.

Two other field studies looked at reactions of shoppers to solicitors asking for charitable donations outside of stores. One was the familiar red kettle of Salvation Army at Christmas holiday times. Not surprisingly, to anyone who has shopped and seen solicitors, about a third of shoppers avoided the likely guilt from not giving, by walking an average of 70 feet extra to go through a door with no solicitor (Andreoni et al., 2017). In a similar experiment in Alaska, when the temperature dropped below 0°F, the avoidance disappeared, indicating that avoidance is sensitive to its price (Trachtman et al., 2015).

All these studies are important for showing how observation can matter. However, they are generally limited on one of two dimensions: (1) Basic internal validity— i.e., we cannot be certain that the experimental treatment was applied as intended (i.e., that participants knew they were being observed). In the abstract cue studies, participants were typically not asked if they noticed the eyes or the robot picture (and their attention was not measured in any subtler way). (2) Do observers have a stake in the outcome? Observability effects might work differently— presumably more strongly— if the observers care about how the participant treats them.[1]

---

[1] Studies by Dana et al. (2006) and Lazear et al. (2012) are important exceptions because they have strong internal validity and do have financially- interested observers. They clearly show

Our design is different along both of these dimensions. The decision makers know that the observers see exactly what they themselves see (as if the observers are standing over their shoulder looking at the screen). They also know that the observers' payoffs will depend on what they choose. We measure the decision-making participants' attention to their own and others' payoffs directly.

III Framework and Hypotheses

First, we show a highly stylized model. This is a language to express some hypotheses about the influence of social observation on lookups and choice. Our experiments were not well-designed to estimate parametric features of this structural model, so we will not do so. However, other papers do so (e.g., Bruhin et al., 2019).

Player roles and allocations: We analyze the case of two players, the chooser C and the observer O. When there is an observer and a passive third player, the analysis is complicated by whether the observer's utility depends on what the passive third player gets, and what the chooser C believes about that utility. We did not design those treatments to test any theory; they are purely exploratory to generate data that may provoke future theorizing.

Player C chooses between two allocations denoted A_S, A_F. The allocation payoff pairs are denoted $(\pi_C(A), \pi_O(A))$. In the design, these are payoff vectors such as (400,400) and (300, 600). Without loss of generality (given further modelling assumptions below), the payoff allocations can be normalized by subtracting the lowest role-specific payoff from each of the two allocations, which creates A_S=(100,0) and A_F=(0,200) for the previous example. We can further normalize these to A_S=(1,0) and A_F=(0,y), where y is the ratio of the payoff advantage to O from the allocation A_F over A_S, relative to the advantage to C of A_S over A_F. The example (400,400) and (300, 600) allocations then are transformed to (1,0) and (0, 2).

---

that some participants have a preference for avoiding social observation and choosing selfishly. Our design goes further by exogenous manipulation of observation, implementing direct observation of information gathering through screen-sharing, and recording visual attention.

Player private-information types: The O players do not actively choose. Therefore, for simplicity, we assume they are purely selfish so that $U_O=\pi_O(A)$.[2]

There are two chooser C types, called selfish (S) and prosocial or fair (F). In general $U_C(A)=\pi_C(A)+\alpha\pi_O(A)$. Their utilities put a (weakly positive) weight on the payoffs of O. But S types have $U_S(A)=\pi_C(A)$ (i.e., $\alpha=0$) and F types have $U_F(A)=\pi_C(A)+\alpha\pi_O(A)$ (with $\alpha>0$). The prior proportions of these types are 1-p and p.

Note that F types will choose A_S over A_F iff $1<\alpha y$.

The next part has three steps. In step 1, we describe the general utilities which can arise from social image when O players observe what the C players look up. In step 2, we work through the logic of what happens for all players (C and O) and player types (S and F types of C players) in the baseline condition where there is no observation—that is, the C players can look at whatever payoffs they want, but the O players will not know what was observed. In the most complicated step 3, we describe equilibrium conditions when C players both choose what to observe and what to choose.

Step 1: Judgments of social image and disutility from guilt

We make assumptions on social image and utility such that the Observability (Social Image) treatment could potentially activate a social image judgment by the O players, about the C player's likely type and therefore O's expected payoff, and which also creates potential disutility for C.

In both treatments, S and F type Choosers are assumed to care about their social image. Social image is defined as their beliefs about the difference between what O actually gets and O expects

---

[2] The observer O's could be altruistic, for example, and weight C's payoffs positively. But since the O's do not make any choices, there is no learning based on what the O's observe or do (they aren't observed by C in what they choose to observe or do). This makes it implausible to consider models in which players want to match altruism levels (e.g. Levine, 1998).

to get. Note that this is a socially iterated expectation (C's beliefs about what O believes).

For the two types of Choosers the utilities for S and F types are assumed to be

$U_S(A,\lambda)= \pi_C(A)+ \gamma\ [\pi_O(A)-E_C[E_O[\pi_O(A)|\lambda]]_0$

$U_F(A,\lambda)= \pi_C(A)+ \alpha\pi_O(A) + \gamma\ [\pi_O(A)-E_C[E_O[\pi_O(A)|\lambda]]_0$

An important piece of notation is that $[\pi_O(A)-E_C[E_O[\pi_O(A)|\lambda]]_0$, defines a function which is upward censored at zero (i.e., positive values are set to zero). The idea is that this term measures only negative guilty when the chosen amount $\pi_O(A)$ is less than the Chooser thinks the Observer expects.[3] When the $\pi_O(A)$ is above the iterated expectation it is set to zero.

The first term is the Chooser's personal payoff. The second term, for the F type, is the altruism weight α on the O payoff $\pi_O(A)$. For both types, the last term is a social sensitivity parameter γ times the difference between what O will actually get from choice A, which is $\pi_O(A)$, and C's expectation about what O expects to get. This is the iterated expectation, $E_C[E_O[\pi_O(A)|\lambda]$, given the choice of lookup λ.

Step 2: No observation (Baseline Treatment)

In this Baseline treatment, player O does not know the lookup values $\lambda(A)$. We assume the third term in the specifications does not influence choice because $E_C[E_O[\pi_O(A)|\lambda]]$ cannot vary with the unobserved choice of $\lambda(A)$.

Denote the lookup function $\lambda=1$ if C looked up O's payoffs and $\lambda=0$ if C did not.[4]

---

[3] Notationally, $[\pi_O(A)-E_C[E_O[\pi_O(A)|\lambda]]_0=[\pi_O(A)-E_C[E_O[\pi_O(A)|\lambda]]<0$ if $[\pi_O(A)-E_C[E_O[\pi_O(A)|\lambda]]<0$ and $[\pi_O(A)-E_C[E_O[\pi_O(A)|\lambda]]_0=0$ if $[\pi_O(A)-E_C[E_O[\pi_O(A)|\lambda]]>0$.

[4] Empirically, if C players looked at either of the payoffs $U_O(A_S)$ or $U_O(A_F)$ they also looked at both so we treat O-player payoff lookups as one variable.

Even in the Baseline treatment, the fair-minded type F always looks up ($\lambda(A)=1$) and chooses A_F iff $1<\alpha y$. F types will always look because they need to know if $1<\alpha y$.[5] Intuitively, F is not looking because it affects O's judgments in any way (and hence F's social image concern). Instead, F is looking to find out if the payoff y for player O is sufficiently high to trigger preference for A_F.

The selfish C type doesn't look ($\lambda=1$) because that type does not take into account O's payoff in the selfish utility when social image doesn't matter. She compares $U_S(A_S)=1$ and $U_S(A_F)=0$ and always chooses A_S.

Step 3: Social image: Player O observes what the chooser C looks up

The Chooser's utilities have the following values, depending on the S or F type and on whether they look up or not Assume that $1<\alpha y$ (otherwise both F and S types choose A_S)

F type

In the Social Image treatment (as in Baseline above), the F types always look up because they want to know if $1<\alpha y$ which triggers the choice of A_F. In this treatment there is an additional motive which is utility from social image (or guilt-avoidance).

$$U_F(A,\lambda)=1+\gamma[0-E_F[E_O[\pi_O(A)|\lambda=1]]_0 \qquad \lambda=1, \text{ choose } A_S$$

$$0+\alpha y+\gamma[y-E_F[E_O[\pi_O(A)|\lambda=1]]_0 \qquad \lambda=1, \text{ choose } A_F$$

Since $1<\alpha y$, and the expectation terms above are larger for the choice of A_F than for the choice of A_F, the F type chooses $\lambda=1$ and A_F. Note that in this case, the F type behaves in the same way as in the Baseline when $1<\alpha y$ (choosing A_F) but the F type's utilities are different. Psychobiological measurements of emotion, attention, or neural activity could therefore find differences in underlying sources of utility even though the observed A_F choice is the same.

S type

$U_S(A,\lambda)=1+\gamma[0-E_S[E_O[\pi_O(A)|\lambda=1]]_0$ $\lambda=1$, choose A_S

$0+\gamma[y-E_S[E_O[\pi_O(A)|\lambda=1]]_0$ $\lambda=1$, choose A_F

Let's unpack the expression $U_S(A,\lambda=1)=1+\gamma\,[0-E_S[E_O[\pi_O(A)|\lambda=1]]_0$ just to clearly understand the nature of these decision utilities. This expression describes the utility to a selfish S type chooser who looks up payoffs ($\lambda=1$). The lookup creates an expectation $E_O[\pi_O(A)|\lambda=1]$ in the observer about their earning $\pi_O(A)$. The S type maintains their own iterated expectation $E_S[E_O[\pi_O(A)|\lambda=1]]_0$ because in equilibrium type S's expectation about O's expectation is correct (that is an imposed assumption). S's second-order expectation creates "guilt" disutility $\gamma[0-E_S[E_O[\pi_O(A)|\lambda=1]_0$ if they choose A_S so that the observer O gets nothing and is disappointed. The guilt comes from S having built up an expectation, based on lookups, that the observer O might get y but only get 0.

The S type chooses A_F over A_S iff $\gamma y>1$. Intuitively, if avoiding the γ-weighted guilt of "leading the observer on" through lookups is greater than the gain of 1 from the A_S choice, then the S type will choose $\lambda=1$ and A_F.

Willful ignorance (or plausible deniability)

An interesting and plausible equilibrium is one in which the S type never looks up the observer's payoffs at all. Aspects of this kind of "willful ignorance" (or plausible deniability) has been shown experimentally by Dana et al. (2006, 2007), Lazear et al. (2012) and extensively in social psychology (Vu et al., 2023).

Remember that all F types look at all observer payoffs ($\lambda=1$) because they will choose A_F iff $y>1/\alpha$. They look up (if costless) because they would like to know if this threshold condition is met. To compute when there is willful ignorance of the observer payoffs among the choosers, we need to know the overall chance, for the experimentally-designed distribution of y values, that an observed y will lead the F type to choose A_F. Denote this value by $y^*=E(y|y\geq\tau)P(y\geq\tau)$ where $\tau=1/\alpha$. The second term $P(y\geq\tau)$ is the probability that the fair F type looks up a y value that makes her choose A_F. The term $E(y|y\geq\tau)$ is the conditional expectation of the y values that, when looked up, lead the fair type to choose A_F.

Now we're back to the S types. If they don't look up at all, they think there is a p chance (the prior probability P(F)) that the observers expect y*. The observers' unconditional expectation is therefore $E_O[\pi_O(A)]=py^*$. Now if the S player never looks up, and always chooses selfishly, the conditional expectations of the observer is $E_O[\pi_O(A)|\lambda=0]=0$ and the iterated expectation is $E_S[E_O[\pi_O(A)|\lambda=0]]=0$.

In this willful ignorance scenario, S gets $1+\gamma 0=1$ from not looking ($\lambda=0$) and choosing A_S. If S does look up (but will always choose A_S), then $E_S[E_O[\pi_O(A)|\lambda=1]]_0=py^*$. Then S gets a guilt-adjusted utility of $1+\gamma[0-py^*]$ which is less than1. If she plans to choose A_S no matter what, she prefers not to look up at all; going so just creates an expectation of possible gain in O, which creates a little guilt in the chooser (the term $-\gamma py^*$).

As is common in signaling games, there are other equilibria. In the next section, we contrast two of them which make different predictions about lookups and choices.

Pooling equilibrium on $\lambda=1$ and A_F: When $1<\alpha y$, F type choosers prefer to choose A_F because the altruism weight α is sufficiently large. When $\gamma y>1$, there is an equilibrium in which S types choose to look, then choose A_F to avoid guilt. This combination creates a "pooling equilibrium" in which both types take the same actions—looking $\lambda=1$ and choosing A_F. Remember that in the baseline condition the S types always choose A_S. Thus, because of this change, the prediction is that if the pooling equilibrium occurs there will be more overall looking and more A_F choice in the Social Image treatment.

In this equilibrium, both F and S chooser types look at the observer payoffs and choose A_F. When $1<\alpha y$ and $1<\gamma y$ the pooling equilibrium exists. This means that the lower of the two values α (fair players' altruism) and γ (selfish players' guilt) determines whether this is an equilibrium.

Remark: In the willful ignorance never-look separating equilibrium, S earns the simple payoff of 1 and does not suffer from guilt. In the pooling equilibrium, S earns less, 0, because otherwise she would earn A_S from choosing $1-\gamma y<0$, and $\gamma>1/y$ is a sufficiently high guilt parameter to lead to an A_F choice and sustain that pooling equilibrium. What keeps the pooling equilibrium together is the off-equilibrium belief (among observers, filtered through to choosers given assumed rational expectations) that if a player deviates and does not look ($\lambda=0$), that player is probably a fair F type and the Observer then expects to get a payoff. This perception creates guilt if the deviating S type chooses $\lambda=0$, which prevents deviation.

Separating equilibrium: There is also a separating equilibrium in which the S and F types *do not* make the same λ or A choices—that is, the types can be *ex post* "separated" or Bayesian-distinguished based on observable lookup behavior. This occurs when $1<\alpha y$ (F types prefer A_F) but $\gamma y<1$, so that S types choose $\lambda=0$ and A_S.

With observation (Social Image): O *does* know $\lambda(A)$. The observer updates the belief about the chooser's type if she observes the chooser look up her observer payoffs, and depending on what is actually chosen after those lookups.

The fair-minded F type chooser always looks up $\lambda(A)=1$ and chooses A_F. The selfish S type chooser chooses $\lambda(A)=1$ but chooses A_S. Anticipating these patterns of lookup and choice, the observer's expected payoff (given the prior P(F)=p) is $E_O[\pi_O(A)|\lambda=1]=py$. Rational expectations leads to $E_S[E_O[\pi_O(A)|\lambda=1]=py$. The S type chooser compares earning $1-\gamma E_S[E_O[\pi_O(A)|\lambda=1]=1-\gamma py$ from looking up and choosing A_S (feeling *py* units of guilt), and $0-\gamma E_S[E_O[\pi_O(A)|\lambda=1]]_0=0$. She separates by choosing A_S iff $1-\gamma py>0$, or equivalently, $\gamma<1/py$.

Figures 1-2 illustrate some of these results graphically. (The willful ignorance result is not shown because it is graphically uninteresting.) In Figure 1, the x-axis is O's possible payoff $\pi_O(A_F)$ which is y. The y-axis is the altruism weight α which the F type gives to the O payoff. The hyperbola $\alpha y=1$ divides the parameter space into two: When values of y and α are low (to the lower left of the thick line), both types choose A_S. In this region, O does not expect any payoff so there is also no guilt for the choosers.

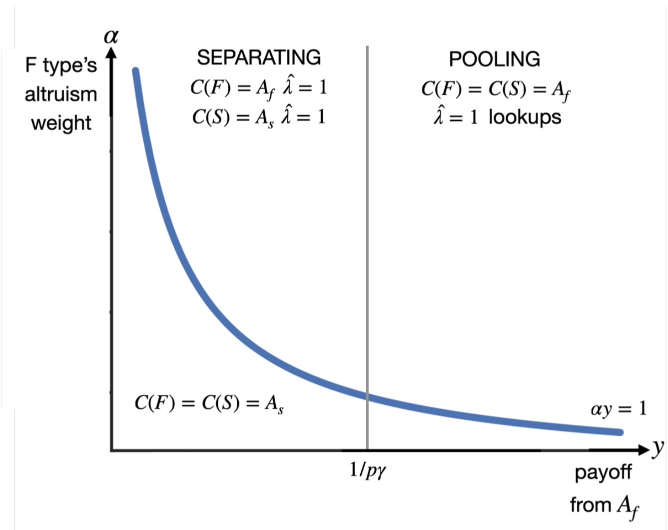


Figure 1: Graphical depiction of choices and lookups and selected equilibrium. Choices of fair and selfish types are C(F),C(S). Lookups are λ=1 (hats denote observed lookups of Observer payoffs). Note: A similar graph could be constructed to show a willful ignorance separating equilibrium in which F types choose λ=1 and A_F (αy>1) or A_S (αy<1), and S types always choose λ=0 and A_S.

In the upper right region, when γ<1/py the separating equilibrium can occur. Both types look up, F's choose A_F and S's choose A_S. When γ<1/py the pooling equilibrium can occur. Both F and S types look up payoffs and choose A_F. The S types are choosing to avoid guilt.

Figure 2 illustrates a simple comparative static result. If the prior P(F)=p and guilt weight γ fall in a combination so that if 1/pγ increases, the area of separating equilibrium is larger so there are fewer λ=1 and A_F choices from the S type. The reduction in the prior reduces the observer's expectation $E_O[\pi_O(A)|\lambda=1]=py$, together with the reduction in γ, reduce the S type's weighted guilt and create fewer A_F choices.

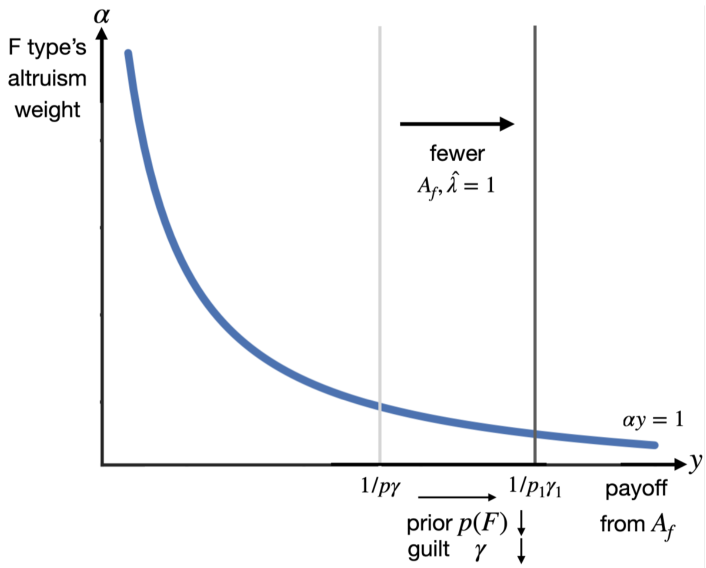


Figure 2: The effect of reducing the prior p(F)=p and guilt γ. Reducing those values to p_1, γ_1 shifts the (thicker) vertical line $y=1/p_1\gamma_1$ to the right. This leads to a larger parametric range in which is a larger region of the separating equilibrium, which implies fewer choices (by S types) of A_F.

The analysis leads to three hypotheses about the possible differences between observation and choice in the Baseline and Social Image treatments:

1. No effect: One possibility is that observability makes no difference overall. This is just a benchmark null hypothesis.

2. Willful ignorance: There is an equilibrium, as shown above, in which the S type chooser never looks up payoffs and always chooses A_S. In this equilibrium, there will be no difference between lookups and choices in the Baseline and Social Image case. The prediction here is the same as "no effect" but it is emerges from equilibrium reasoning.

	Prediction	
Hypotheses	**Prosocial choices**	**Visual looking durations**
No effect	**B = SI**	**B = SI**

Social image influence	**B < SI**	**B < SI**
Willful Ignorance	**B = SI**	**B=SI**
Lip service	**B = SI**	**B < SI**

Table 1. Summary of theory predictions about choices and visual durations of payoff looking. The experimental conditions are baseline (B) and social image (SI), in which the Observers can see which payoffs Choosers are looking at.

3. More observation and prosociality in the Social Image condition: The pooling equilibrium derived above is one mechanism by which SI will increase looking and prosociality. It depends on the parametric condition $\gamma y>1$. There are other intuitive possibilities which fall outside the analytic framework. Choosers may simply have a tastes for non-instrumental information (curiosity) about the Observer's payoff[6], even if it does not produce guilt or affect their choices in other ways.

4. Lip service: As noted in the introduction, it could be that choosers look at the observer payoffs because they think that merely looking generates some benefit to themselves, or to observers. In the guilt-driven model above, if payoffs are observed, but do not affect the chooser's choice, there is a negative utility from guilt. Lip service ignores such utility.

Remark: Notice that the no effect, willful ignorance, and lip service hypotheses cannot be distinguished from choices alone. However, they have different predictions about the chooser's attention to observer payoffs. Furthermore, in the lip service case it may be that social observers derive utility from knowing that their possible outcomes were considered (i.e., visually attended to by the chooser), and/or that choosers see the observers' utility as enhancing their social image. There would be no basis for this conclusion without the visual attention data from mousetracking.

IV. Experimental Design and Procedures

We focus on simple binary social allocation decisions where the decision-maker often faces a

[6] See Eliaz and Schotter (2007, 2010), Kang et al. (2009), Alan and Mumcu (2024).

tradeoff between her own payoff and the payoff for others. To study the hypotheses about the impact of an affected observer on the decision-maker's attention and choice, we run a Social Image treatment where an observer can see exactly what the decision-maker pays attention to before her choice, but not the actual choice. In the baseline condition, there are no observers. The Deniability hypothesis predicts less attention to the observer's payoffs while the Lip service and Social image influence hypotheses predict more attention. Not showing the observer the decision-maker's actual choice allows us to distinguish between the Lip service hypothesis which predicts that decision-maker will not make more pro-social choices (because they can signal through attention alone) and the Social image influence hypothesis which predicts that they will.

There were three types of allocation games, each offering a choice between two possible allocations. (The full set of allocations is shown in the Appendix). We use allocation games from Charness and Rabin (2002) and Engelmann and Strobel (2004) as well as some variants of those games.

In the first type, S2O, the decision-maker (DM) decides between two allocations to herself and two other people. In the second type, S1O, the DM chooses between two allocations to herself and one other person. Finally, in the third type, 2O, the DM chooses an allocation for two others only; the choice does not affect DM's payoff.

In the first two baseline sessions, there were 80 and 84 trials respectively. In the first 40 trials, we alternated between blocks of 10 S2O and S1O trials. In the later 40 trials, the alternation was between blocks of 10 S2O and 2O trials.[7] In all subsequent sessions, the 20 S2O, 20 S1O, and 10 2O games that were the most diagnostic out of the 84 (i.e., generated closest to 50% choice) were used.

In the Social Image condition, half of the participants are decision-makers in the first 25-trial block (10 S2O, 10 S1O, and 5 2O) where in all games with two others involved, Other 1 is the

---

[7] In the second session, four more S1O games were added at the end that had the same payoff difference for self but different payoff differences for the other participant.

observer. In the next 25-trial block, the other half of participants are now the decision-makers and Other 2 is the observer. In the third block, the decision-makers are the same as in the first block and Other 2 continues to be the observer. Finally, in the last block of 25 trials, the second group of decision-makers have Other 1 as the observer. Participants are randomly rematched after each trial. With this ABAB design, all participants were decision-makers for the same 50 games over 100 trials. All participants were also observers as Other 1 for 15 games, Other 2 for 15 games, and the only possible observer for the other 20 games. Thus, we have 50 independent observations for each participant.

In the baseline condition, all payoffs are initially hidden. The decision-maker can move the mouse and click on any cell to reveal the payoff. Once the mouse is moved from that cell, the payoff becomes hidden again and the decision-maker must click on the cell again to see the payoff. Thus, we can track the exact sequence and duration of lookups until the choice between the two allocations is made in each trial. The entire payoff table was revealed to the decision-maker at the end of each trial.

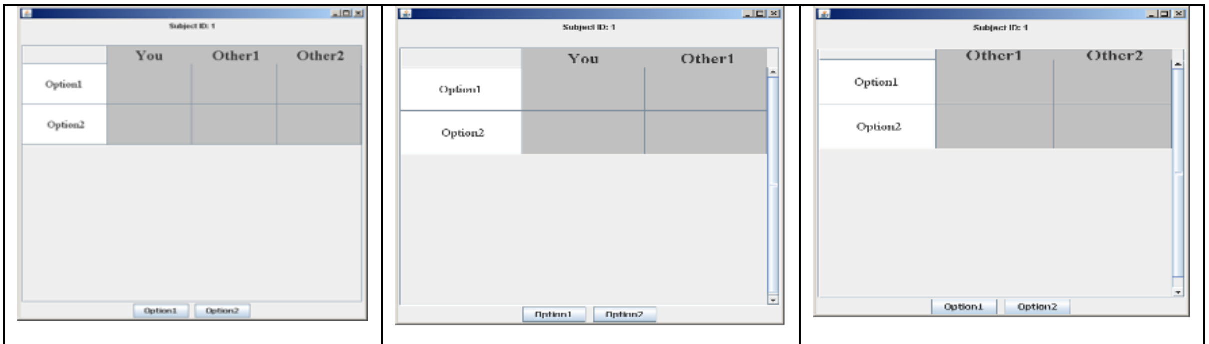
The only difference in the Social Image (SI) condition is that in addition to the decision-maker, there is an observer in each trial who can see what the decision-maker has clicked on in real-time through screen-sharing. After the trial, the full payoff matrix is also revealed to the observer but not the choices made by the decision-maker. In the last Social Image session, we did not reveal the payoffs to the decision-maker and the observer after each trial and verified that this change did not affect the looking patterns.

Participants were recruited via email at CASSEL at UCLA except for a small number of Caltech students (n=6) at the SSEL lab. There were four Baseline sessions with 66 participants in total and three Social Image sessions with 62 participants in total.[8] All participants were paid based on two randomly chosen trials, one in which they were the decision-maker and another in which they were the observer in the allocation group.[9] Figures 1-2 contain sample screenshots for both the baseline and SI conditions. Sample instructions can be found in the Appendix.

---

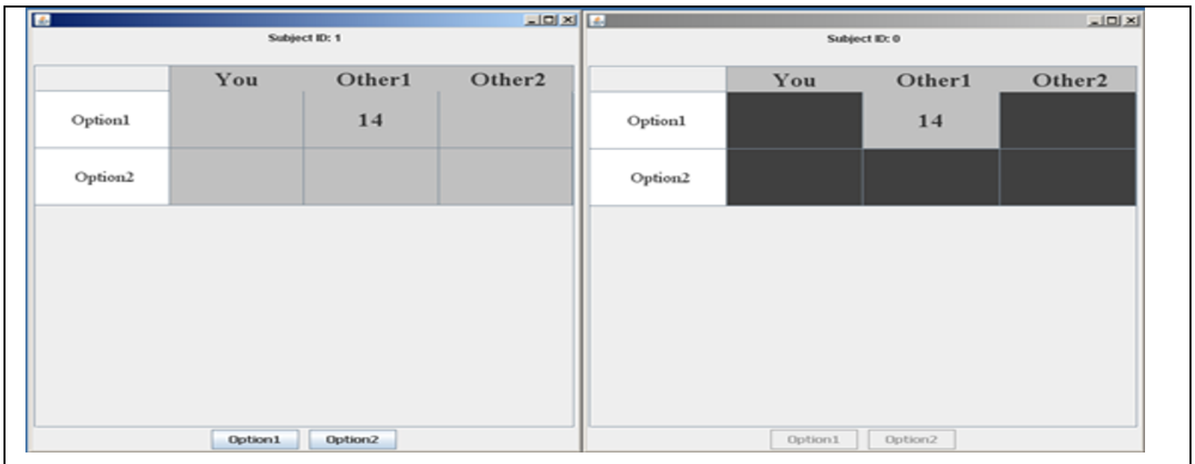
[8] See Table A1 in the Appendix for a more detailed summary.

[9] We used the Random Problem Selection mechanism for its incentive compatibility properties (Azrieli et al., 2018).



S2O	S1O	2O

Figure 1. Sample screenshots of the three types of games with payoffs hidden



Decision Maker Screen	Observer Screen

Figure 2. Sample screenshots of decision-maker's and observer's screens in real-time when the decision-maker clicks in a cell.

V. Results

V.1. Choices and Social Preference Theories

The first question is what social preferences are generally reflected by choices. Being observed clearly reduces the rate of purely selfish choice. Overall, 74.2% of choices are selfish in the

Baseline case (maximizing own payoff), compared to 62.04% in Social Image[10]. Thus, there is a simple and significant (Mann-Whitney: $p<0.01$) treatment effect.

Table 1 shows the results of a simple logit specification, at the trial level, which estimates the weights placed on differences in self payoff, total payoff for the group, and the minimum payoff in the group that best explain the observed choices (following Charness and Rabin, 2002). In the aggregate, self payoff and (to a lesser extent) minimum payoff are both significant predictors of choices. Perhaps surprisingly, total payoff is not a significant influence. All the coefficients are also jointly significantly different across the two conditions (Chow test: $p<0.01$).

	Baseline	Social Image
Diff in Self Payoff	13 (8.25)	5.2 (7.70)
Diff in Minimum Pay	2.0 (8.20)	0.63 (3.58)
Diff in Total Pay	0.090 (0.39)	-0.16 (-1.09)
Constant	-72 (1.92)	-27 (0.82)
Log-Likelihood	-1517	-1641
N	2553	2457

Table 1. Logit regression of choice in S2O and S1O trials on payoff differences (t-statistics in parentheses). Standard errors are clustered at the individual level; all coefficients multiplied by 1000).

Another prominent model of social preference is inequality aversion where people weigh and add their own payoffs and measures of "envious" and "guilty" inequality (measured by differences in their own payoffs and payoffs of others).

$$U_i(x) = x_i - \alpha_i \max\{x_j - x_i, 0\} - \beta_i \max\{x_i - x_j, 0\}$$

In the three-parameter (ρ, σ, and γ) specification in Charness and Rabin (2002), $\sigma<0<\rho<1$ would be consistent with this model of inequality aversion. The probability of choosing option 1 for player i is:

[10] For the Social Image treatment, we compare the percentage of selfish choices for participants who encountered the initial block of 25 allocation decisions first (60.28%) and the percentage of self choices for participants who encountered the second block of 25 allocation decisions first (63.83%) and find no significant order effects (Mann-Whitney: $p=0.12$).

$$P_{i1} = \frac{e^{[(r+s)x_{1j}+(1-r-s)x_{1i}]}}{e^{[(r+s)x_{1j}+(1-r-s)x_{1i}} + e^{[(r+s)x_{2j}+(1-r-s)x_{2i}}}$$

where r=1 if $x_i>x_j$, and r=0 otherwise and s=1 if $x_i<x_j$, and s=0 otherwise.

The two specifications can be easily extended to accommodate the three player case. Table 2 shows that participants in the baseline condition display "envious" but not "guilty" inequality aversion. For the Social Image treatment, there is no significant evidence of either type of inequality aversion.

	Baseline	Social Image	Charness-Rabin
ρ	0.13 (4.26)	-0.10 (1.28)	0.42 (25.5)
σ	-0.0044 (0.24)	-0.011 (0.30)	-0.014 (0.73)
γ	0.013 (7.70)	0.0039 (5.08)	0.014 (11.6)
Log-Likelihood	-1543	-1646	-528
N	2553	2457	903

Table 2. Three-parameter logit specification in Charness-Rabin (t-statistics in parentheses; SEs clustered at individual level)

The ρ and γ coefficients are significantly different between the Baseline and the SI condition and all the coefficients are also jointly significantly different across the two conditions (Chow test: $p<0.01$)

Result 1: Participants make more pro-social choices, consistent with the Social Image Influence hypothesis, and inconsistent with the No Effect, Deniability, and Lip Service hypotheses which predict more or equally selfish choices.

V.2. Overall Duration and Click Statistics

This section reports basic statistics about attention to payoffs. Table A2 in the Appendix summarizes overall statistics on how long participants look at different types of payoffs (denoted “duration”) and how frequently they click different payoff boxes in the three types of games and the two information conditions.
On average, participants look up each box 3-4 times per trial. The lookup patterns and durations per lookup are very regular across game types. Durations are 250-300 msec per click. In most trials the participants clicked on all payoff cells at least once.

It does appear that the number of clicks is about 10-15% lower in the Social Image condition than in the Baseline condition games. However, this difference can be largely attributed to learning effects, since the Social Image condition has more trials, and later trials have faster lookup durations.[11] Furthermore, most of the analyses below focus on relative differences in looking at payoffs for oneself and for others, which are robust to any learning-based decrease in looking.

The persistence and duration of attention to self and other payoffs across information treatments is evidence against the deniability hypothesis. Participants do not appear to purposely avoid looking at the payoffs of others, in order to make a selfish choice without creating guilt or harming social image. In fact, there is substantial evidence that they look more at payoffs of others in the Social Image condition. For example, in S1O games, the ratio of average lookup times in other and own payoff cells is higher in Social Image than in Baseline (1.25 in SI vs. .97 in baseline; t-test: $p<0.01$).

Keep in mind that nothing in the instructions, design, or technology forces participants to pay attention to payoffs of others at all. A participant could just look at her own payoffs and choose selfishly. Indeed, in the first baseline session there was one (Caltech) participant who made

[11] Recall that in the baseline sessions 50 trials are used for analysis, and can therefore be compared with the first and second blocks of 50 trials in the longer 100-trial SI sessions. The average total durations in the first block of 50 trials for the baseline and SI conditions are 2847 msec and 2684 msec (two-tailed t-test: $p=0.07$). The average total duration in the second 50-trial SI block is significantly lower at 2114 msec (compared to 2684), showing learning curve effects. Thus, while there is a small difference in total duration (around 200 msec) between the two conditions even in the early block of trials, most of the difference in durations is due to the different number of trials. Furthermore, a regression of total lookup durations on trial number shows an estimated reduction in duration per trial of 22 msec and 18 msec for Baseline and SI, respectively (a drop that is highly significantly positive).

100% selfish choices, and did not look at the potential payoffs of others at all for 56% of the S2O and S1O trials. Thus, the textbook economic agent who is selfish and highly inattentive to how his choices affect other people, actually does exist— but there was only one such participant.[12]

Table A3 in the Appendix shows the average number of transitions between payoff boxes of different types. There are generally more within-choice row transitions than within-column transitions, but the difference is small and is not present in all type-condition combinations. Given that there is no clear guidance from theory about which transitions are more likely, the data suggest there are multiple lookup routes to similar judgments.

However, inequality-aversion models do make a specific algorithmic prediction: If working memory is constrained, there should be a lot of Self-O1 and Self-O2 payoff box transitions (since each payoff difference enters the utility function separately, and cannot be computed from an O1-O2 comparison alone). There should be essentially no O1-O2 transitions (since this comparison plays absolutely no role in inequality-aversion computations). Contrary to this tentative hypothesis, there are very few Self-O2 transitions. Instead, most of the transitions are between adjacent payoff cells and there are many O1-O2 transitions. A strong bias toward adjacent-payoff transitions is, of course, not surprising as this kind of visual search is common in reading, online shopping, and many other activities.

Result 2: Participants look more at the payoffs of others in the Social Image treatment compared to the Baseline, again consistent with the Lip Service and Social Image Influence hypotheses and contradicting the Deniability hypothesis which predicted avoidance of these payoffs.

V.3. Influence of Social Observation on Visual Attention to Payoffs

In the S2O and 2O games, participants know which one of the two other players is observing their lookups. Table 3 shows differences in total durations of attention to observer payoffs and

---

[12] Another participant in a Social Image session only looked at the others' payoffs in three of the 40 S2O and S1O trials (although that participant did not always choose selfishly).

non-observer payoffs.

Condition	Type	Observer Payoff Duration		Non-Observer Payoff Duration	p-value
Baseline	S2O	1199		1205	0.37
Baseline	2O	1214		1162	0.72
Social Image	S2O	1132	>	1014	<0.001
Social Image	2O	953	>	864	0.03

Table 3. Lookup Duration in Observer vs. Non-Observer Payoff Boxes (p-value based on Wilcoxon sign-rank test)

In both S2O and 2O trials, the decision-makers do look significantly more at the observer's payoffs than at the non-observer's payoffs. Keep in mind that in S2O and 2O trials, the identity of the observer was switched from O1 to O2 regularly within a session, so powerful within-subject comparisons can be made. The preferential looking longer at the payoffs of the observer rather than those of the non-observer is significant in these within-subjects tests in both S2O (t-test: p=0.03) and 2O (t-test: p<0.001) games.[13]

In S2O games, where being observed could conceivably increase the total duration to payoff cells of both of the other players combined, there is no difference in that total duration. However, there are also differences in observer payoff duration percentages across the Baseline and Social Image conditions that are small, but statistically significant for all three types of games (Wilcoxon sign-rank at the individual level: p<0.05). Furthermore, the percentage of total attention allotted to non-observer payoffs declines a little in the Social Image conditions from Baseline (from 29 to 27% of total duration for S2O, and from 50% to 47% in 2O). Thus, the observer influence on visual attention seems to be restricted entirely to looking a little longer at the payoff of the person who is observing you.

This effect is similar to what one might expect (based on the behavioral literature) if a pair of eyes were placed over the observer column only. However, the manipulation here is more clearly

[13] Moreover, a counterfactual placebo comparison was done for the baseline condition with dummy observer and non-observer designations that are the same as the ones in the SI condition. Naturally, there are no significant differences in duration in this placebo test (t-test: p=0.37 for S2O games and p=0.72 for 2O games).

understood by participants, and is more lifelike because it represents actual observation of payoff search by the human observer participant).

Result 3: When there are two other players in the group, the decision maker spends significantly more time looking at the observer's payoffs than that of the player who is not observing the lookups.

V.4. Durations and Choices

In this section we explore further the central question: how does being observed, in the SI case, influence choice? The first next step is to look at the relation between attention to payoffs to others vs. self to selfish choice, at the individual level, and whether SI changes that relation[14]. Table 4 shows a simple linear OLS regression of individual-level rates of selfish choice on looking (relative other/self duration), an SI dummy variable, and the interaction. There are strong negative main effects of looking at others' payoffs and of the SI treatment on selfish choice. However, the strongest effect is a main effect of SI reducing the percentage of selfish choices. Being observed makes people choose less selfishly. The SI condition dummy variable does not appear to strongly modulate the relation between looking and choice.

	%Selfish	% Selfish	% Selfish
Other/Self Duration	-0.14 (3.76)	-0.23 (4.01)	-0.21 (3.75)
Social Image Dummy	-0.10 (2.87)		-0.10 (2.88)
(Other/Self)*Social Image		0.12 (1.54)	0.12 (1.57)
Constant	0.89 (19.33)	0.92 (14.90)	0.95 (15.60)
R^2	0.16	0.13	0.17

Table 4: Relationship between other/self payoff looking durations and percentage of selfish

[14] There is a vast literature on the effect of attention in decision-making, e.g. Orquin and Loose (2013) and Li and Camerer (2022) and in behavioral economics (Loewenstein and Wojtowiczl, 2023).

choice (dependent variable) in S2O and S1O games (Note: Other/self duration ratio in interaction with SI dummy variable is normalized by subtracting sample mean.)

We now focus on S1O games at the trial level. Figure 1 shows controls for the fraction of lookup time directed to self payoffs (x-axis) and reports the percentage of selfish choices for each level of self-payoff lookup time fraction (y-axis). Choices are more selfish in the Baseline condition for all lookup bins except one. Thus, the SI condition generally reduces selfish choice.

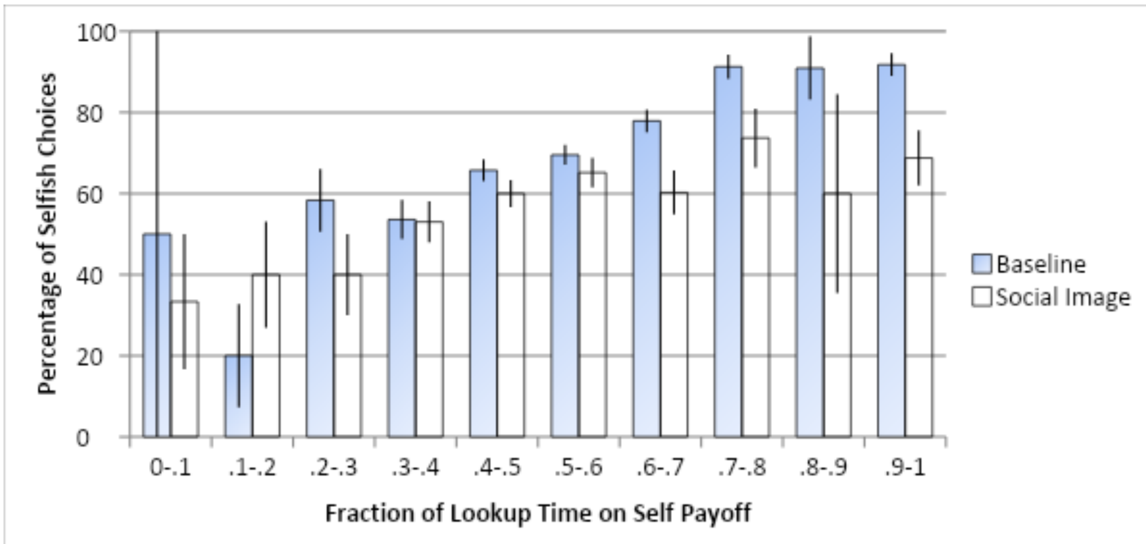


Figure 1. Percentage of selfish choices as a function of self payoff duration percentage in S1O trials. (Note: The overall percentage of selfish choices is 72% for the baseline condition and 59% for the Social Image condition.)

A more fine-grained analysis, shown in Table 5, uses all trials in a trial-level logit of selfish choice on the differences in percentage of lookup time[15] spent in other and self payoff cells, a SI condition dummy, and an interaction term between the SI dummy and the payoff lookup difference.

Variable	**Selfish Choice**
(Other - Self Duration) %	-1.48 (5.08)
Social Image	-0.52 (2.88)
(Other - Self Duration) % *Social Image	1.24 (3.34)
Constant	0.79 (4.92)
R^2	0.04

[15] Note that the difference in percentage duration is used in the trial-based analysis because the other/self ratios can vary dramatically from trial to trial (while percentage differences do not).

Table 5. S1O choices: Trial-level logit regression of selfish choice (0-1) on difference in other duration and self duration %, SI condition dummy variable, and interaction. (Note: Other duration % - Self duration % in interaction with SI dummy variable is normalized by subtracting sample mean. SEs are clustered at individual level.)

In this analysis, the main effect of the SI condition emerges strongly again (it lowers selfish choice), but there is also a strong positive interaction effect, indicating that SI also flattens out the negative relationship between other-self looking and choice. This pattern is consistent with the idea that in Baseline, the participants' attention and choices are closely aligned (selfish people look at other payoffs much less). In SI, that clear relation is disrupted, since some participants will choose selfishly but look quite a lot at other players' payoffs (consistent with the lip service hypothesis). The behavior of these participants reduces the attention-choice association.

There is no evidence in the S2O games that participants choose in a way which differentially benefits the other player who is observing them, compared to the non-observer. We suspect the allocation games we chose from the literature, and then extended, are not ideal for detecting such an effect, but better allocations could be created easily.[16]

Finally, we directly test whether visual attention is a statistical "mediator" of the influence of SI on choice. In general, a mediator carries the influence of an independent variable, partly or fully, on a dependent variable. In this case, the independent variable is the binary dummy variable for the Social Image condition and the dependent variable is either the percentage of selfish choices in S2O and S1O trials at the individual level, or the binary selfish choice dummy at the S1O trial level. The mediating variable is the ratio of average lookup time spent on payoffs of others (divided by two for S2O trials) to lookup time spent on self payoff at the trial level, or the mean of this ratio across trials at the individual level.

The hypothesis is that the SI condition is at least partially affecting the amount of selfish choices

[16] A better design would allow a choice among two allocations A=(100, 100+x, 100) and B=(100, 100, 120) for various x. If A is chosen for x<20 that would indicate a strict preference to make the observer happy (at the expense of the non-observer).

by shifting the attention to the payoff cells of others. (Note that this hypothesis could not be tested with a mediation analysis if there were no lookup data.) The Sobel test (1982) is used to quantify the significance of the mediator with bootstrapping (5000 replications; Preacher and Hayes, 2008) to correct bias and avoid making any distributional assumptions. The Sobel test reveals a significant mediator with an indirect effect of -0.026 compared to the total effect of -0.12 for the SI dummy. Thus, about 18% of the effect of the SI condition is mediated by its association with lookup patterns. The same analysis at the trial level on S1O trials uses the binary dummy for selfish choice as the dependent variable and other/self lookup time ratio as the mediating variable. The result is that the lookup variable is a significant mediator with an indirect effect of -0.011 out of a total effect of -0.13. Approximately 8.5% of the total effect of the SI condition is mediated through visual attention. Taken together, these results show that adding social observation in the SI condition partly influences choices (around 15% on average) by directing attention to payoffs of others.

Result 4: Attention paid to the other players' payoffs as measured by the ratio of time spent looking at others' payoffs to the time spent looking at own payoffs mediates the Social Image treatment's effect on selfish choice.

VI. Discussion and Conclusion

This paper is the first to directly implement observability of visual attention, by an interested observer, as another participant makes choices that affect the observer's payoff.
The key experimental condition is the Social Image (SI) treatment where screen-sharing is used to show the observer what payoffs the decision maker is looking at in real time.

In this SI condition, the decision makers are less selfish and also look more often at the payoffs of those who are watching them. (and whose payoffs depend on what the choosing decision makes do). Other studies have found such effects (e.g., Dana et al., 2007) without the stronger treatment of knowing that observers are seeing what the decision makers look at. The results go against a "lip service" hypothesis, because participants in the Social Image treatment do look more often at the payoff of the affected observers, but also choose less selfishly in general.

In discussing this design informally, the typical guess by experimental economists was the willful ignorance equilibrium prediction (a/k/a plausible deniability): Choosers will avoid looking at the observer payoffs, when they know observers are watching, in order to not feel guilty if they look at those payoffs then choose selfishly. This hypothesis is consistent with guilt-aversion in a psychological game theory signaling model. However, the main features of the data clearly go against willful ignorance.

The hypothesis that is most consistent with the main results is social image influence which states that exogenous social observation positively affects both visual attention and choice. The mediation analysis suggests that much of the effect just comes from being observed, but part of the effect (15% or so) comes from observation changing visual attention, which changes choice.

Why should economists care about these types of paradigms and results? The answer is that in social image models, what people perceive about what an agent attends to, and weighs, is an input to the observed agent's perceptions of what other people think of her. Procedural details therefore can matter greatly in these models.

Exploring those models, and eventually testing them carefully, will benefit from new choice designs, and also from measuring attentional data when possible. One can also conceive of field experiments and naturally-occurring data sets which would use similar observability treatments like these.[17]

Furthermore, the mediation pattern suggests the possibility that changing visual attention directly, toward payoffs or consequences of others, might increase prosociality. Economists are usually wary of causality claims of this type (reasonably so). If one thinks of visual attention as consciously and endogenously chosen to achieve a goal, then it seems unlikely that involuntarily

[17] Two examples: (1) Have confederates pose as panhandlers and solicit money from people passing by. The hypothesis is that looking at the panhandlers (probably their eyes, which play a powerful role in facial emotion recognition) is associated with giving. (2) Send email solicitations which include or do not include an "e-ceipt" so that an experimenter can observe when the email is "read" (or at least opened). Suppose the emails can be constructed so that the likelihood they will be opened is uncorrelated with perceptions of deservingness. Then we predict that requiring an e-ceipt will lead to more opened emails and more giving.

manipulating attention could change behavior.

However, there is ample evidence that low-level visual attention (and other types of attention, such as auditory) is not completely endogenously chosen, and hence can be easily manipulated, at least briefly and temporarily (e.g. Li and Camerer 2022). Accumulator models in which evidence is gathered or processed by a sensory system (Forstmann et al., 2016) also predict that hijacking attention will influence choice, and there is some evidence that exogenous manipulation of attention changes choice (Shimojo et al, 2003; Pachur, 2018). Thus, it is at least conceivable that exogenously manipulating visual attention to the consequences an action could have for a particular person could influence prosociality toward that person.

References

Alan, S. and I. Mumcu (2024). "Nurturing Childhood Curiosity to Enhance Learning: Evidence from a Randomized Pedagogical Intervention." American Economic Review, 114 (4), 1173-1210.

Andreoni, J. and B. D. Bernheim (2009). "Social Image and The 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects," Econometrica, 77(5), 1607-1636.

Andreoni, J., Rao, J. M., and H. Trachtman (2017). "Avoiding the Ask: A Field Experiment on Altruism, Empathy, and Charitable Giving," Journal of Political Economy, 125(3), 625-652.

Azrieli, Y., Chambers, C. P., and P. J. Healy (2018). "Incentives in Experiments: A Theoretical Analysis," Journal of Political Economy, 126(4), 1472-1503.

Baker, N. (1995). The Fermata. Vintage Books.

Bateson M., Nettle D., and G. Roberts (2006). "Cues of Being Watched Enhance Co-operation in a Real-World Setting," Biology Letters, 2, 412–414.

Battigalli, P., and M. Dufwenberg (2007). "Guilt in Games," American Economic Review, 97(2), 170-176.

Benabou, R. and J. Tirole (2006). "Incentives and Prosocial Behavior," American Economic Review, 96(5), 1652-1678.

Bertoni, M., Brunello, G., and L. Rocco (2013). "When The Cat Is Near, The Mice Won't Play: The Effect of External Examiners in Italian Schools," Journal of Public Economics, 104, 65-77.

Bolton, G. E. and A. Ockenfels (2000). "ERC: A Theory of Equity, Reciprocity, and Competition," American Economic Review, 90(1), 166-193.

Bourrat, P., Baumand, N., and R. McKay (2011). "Surveillance Cues Enhance Moral

Condemnation," Evolutionary Psychology, 2(9), 193-199

Bruhin, A., Fehr, E., and D. Schunk (2019). "The many Faces of Human Sociality: Uncovering the Distribution and Stability of Social Preferences," Journal of the European Economic Association, 17(4), 1025-1069.

Burnham, T. C. and B. Hare (2007). "Engineering Human Cooperation: Does Involuntary Neural Activation Increase Public Goods Contributions?" Human Nature, 18, 88-108.

Camerer, C. (2003). Behavioral Game Theory. Princeton University Press.

Camerer, C. and R. H. Thaler (1995). "Anomalies: Ultimatums, Dictators and Manners," Journal of Economic Perspective, 9(2), 209-219.

Charness, G. and M. Rabin (2002). "Understanding Social Preferences with Simple Tests," Quarterly Journal of Economics, 117(3), 817-869.

Cooper, D. and J. Kagel (2016). "Other Regarding Preferences: A Survey," in The Handbook of Experimental Economics, vol. 2, eds. Kagel and Roth. Princeton University Press.

Coricelli, G., Joffily, M., Montmarquette, C., and M. C. Villeval (2010). "Cheating, Emotions, and Rationality: An Experiment on Tax Evasion," Experimental Economics, 13, 226-247.

Dana, J., Cain, Daylian M., and R. M Dawes (2006). "What You Don't Know Won't Hurt Me: Costly (but quiet) Exit in Dictator Games. Organizational Behavior and Human Decision Processes, 100(2), 193-201.

Dana, J., Weber, R. A., and J. X. Kuang (2007). "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness," Economic Theory, 33(1), 67-80.

DellaVigna, S., List, J. A., and U. Malmendier (2012). “Testing for Altruism and Social Pressure in Charitable Giving,” Quarterly Journal of Economics, 127(1), 1-56.

Dillenberger, D. and P. Sadowski (2012). “Ashamed to Be Selfish,” Theoretical Economics, 7, 99-124.

Dufwenberg, M. and G. Kirchsteiger (2004). “A Theory of Sequential Reciprocity,” Games and Economics Behavior, 47, 268-298.

Eliaz, Kfir, and Andrew Schotter. 2007. "Experimental Testing of Intrinsic Preferences for NonInstrumental Information," American Economic Review, 97(2): 166-169.

Eliaz, K., & Schotter, A. (2010). “Paying for confidence: An experimental study of the demand for non-instrumental information,” Games and Economic Behavior, 70(2), 304–324.

Engelmann, D. and M. Strobel (2004). “Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments,” American Economic Review, 94(4), 857-869.

Falk, A. and U. Fischbacher (2006). “A Theory of Reciprocity,” Games and Economics Behavior, 54, 293-315.

Fehr, E. and K. M. Schmidt (1999). “A Theory of Fairness, Competition, and Cooperation,” Quarterly Journal of Economics, 114(3), 817-868.

Fehr, E., and F. Schneider (2010). “Eyes Are on Us, but Nobody Cares: Are Eye Cues Relevant for Strong Reciprocity?” Proceedings of the Royal Society B: Biological Sciences, 277, 1315–1323.

Forstmann, B. U., Ratcliff, R., and E.-J. Wagenmakers (2016). “Sequential Sampling Models in Cognitive Neuroscience: Advatnages, Applications, and Extensions,” Annual Review of

Psychology, 67, 641-666.

Guerin, B. (1986). "Mere Presence Effects in Humans: A Review." Journal of Experimental Social Psychology, 22(1), 38-77.

Haley, K. J. and D. M. T. Fessler (2005). “Nobody’s Watching? Subtle Cues Affect Generosity in an Anonymous Economic Game,” Evolution and Human Behavior, 26, 245-256.

Ho, B. and E. Liu (2011a). “Does Sorry Work? The Impact of Apology Laws on Medical Malpractice,” Journal of Risk and Uncertainty, 43(2), 141-167.

Ho, B. and E. Liu (2011b). “What’s an Apology Worth? Difference-in-Differences Analysis of State Apology Laws on Medical Malpractice Payouts,” Journal of Empirical Legal Studies, 8(1), 179-199.

Izuma, K., Saito, D. N., and N. Sadato (2010). “Processing of the Incentive for Social Approval in the Ventral Striatum during Charitable Donation,” Journal of Cognitive Neuroscience, 22(4), 621-631.

Izuma, L, Matsumoto, K., Camerer, C. F., and R. Adolphs (2011). “Insensitivity to Social Reputation in Autism,” Proceedings of the National Academy of Sciences, 108(42), 17302-17307.

Kang MJ, Hsu M, Krajbich IM, Loewenstein G, McClure SM, Wang JT, Camerer CF. “The wick in the candle of learning: epistemic curiosity activates reward circuitry and enhances memory,” Psychological Science, 20(8), 963-73.

Krátký J., McGraw, J.J., Xygalatas, D., Mitkidis, P., and P. Reddish (2016). “It Depends Who Is Watching You: 3-D Agent Cues Increase Fairness,” PLOS ONE, 11(2).

Lazear, E. P., Malmendier, U., and R. A. Weber (2012). “Sorting in Experiments with Application to Social Preferences,” AEJ: Applied Economics, 4(1), 136-163.

Li, X. and C. F. Camerer (2022). “Predictable Effects of Visual Salience in Experimental Decisions and Games,” Quarterly Journal of Economics, 137(3), 1849-1900.

Lind, F. A. and T. R. Tyler (1988). The Social Psychology of Procedural Justice. Springer Science.

Loewenstein, G. and Wojtowicz, Z. (2023). The Economics of Attention. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4368304

Manesi, Z., Van Lange, P., and T. Pollett (2016). “Eyes Wide Open: Only Eyes That Pay Attention Promote Prosocial Behavior,”Evolutionary Psychology, 14(2), 1–15.

Munger, K. and S. Harris (1989). “Effects of an Observer on Hand Washing in a Public Restroom,” Perceptual and Motor Skills, 69, 733-734.

Orquin, J. L. and S. M. Loose (2013). “Attention and Choice: A Review on Eye Movements in Decision-Making,” Acta Psychologica, 144(1), 190-206.

Pachur, T., Schulte-Mecklenbeck, M., Murphy, R. O., and R. Hertwig (2018). “Prospect Theory Reflects Selective Allocation of Attention,” Journal of Experimental Psychology: General, 147(2), 147-169.

Panagopoulos, C. and S. van der Linden (2016). “Conformity to Implicit Social Pressure: The Role of Political Identity,” Social Influence, 11(3), 177-184,

Plato (trans. B. Jowett) (1941). Plato's The Republic. New York: The Modern Library.

Preacher, K. J. and A. F. Hayes (2008). "Asymptotic and Resampling Strategies for Assessing and Comparing Indirect Effects in Multiple Mediator Models,” Behaviour Research Methods, 40(3), 879–891.

Rabin, M. (1993). "Incorporating Fairness Into Game Theory and Economics," American Economic Review, 83, 1281-1302.

Rigdon, M., Ishii, K., Watabe, M., and S. Kitayama (2009). “Minimal Social Cues in the Dictator Game,” Journal of Economic Psychology, 30(3), 358-367.

Rotemberg, J. J. (2008). “Minimaly Acceptable Altruism and the Ultimatum Game,” Journal of Economic Behavior & Organization, 66, 457-476.

Rotemberg, J. J. (2014). “Models of Caring, or Acting as if One Cared, about the Welfare of Others.” Annual Review of Economics, 6, 129–154.

Shimojo, S., Simeon, C., Shimojo, E., and Y. Scheier (2003). “Gaze Bias Both Reflects and Influences Preference,” Nature Neuroscience, 6, 1317–1322.

Sobel, M. E. (1982). "Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models,” Sociological Methodology, 13, 290–312.

Sparks, A. and P. Barclay (2013). “Eye Images Increase Generosity, But Not for Long: The Limited Effect of a False Cue,” Evolution and Human Behavior, 34, 317-322.

Taylor, M. S., Tracy, K. B., Renard, M. K., Harrison, J. K., and S. J. Carroll (1995). “Due Process in Performance Appraisal: A Quasi-experiment in Procedural Justice,” Administrative Science Quarterly, 40, 495-523.

Trachtman, H., Steinkruger, A., Wood, M., Wooster, A., Andreoni, J., Murphy, J. J., and J. M. Rao (2015). “Fair Weather Avoidance: Unpacking the Costs and Benefits of “Avoiding the Ask,” Journal of the Economic Science Association, 1, 8-14.

Vu, Linh, Ivan Soraperra, Margarita Leib, Joël van der Weele, and Shaul Shalvi (2023).

“Ignorance by Choice: A Meta-Analytic Review of the Underlying Motives of Willful Ignorance and Its Consequences” Psychological Bulletin, 149(9-10), 611–635.

Appendix

Session	Condition	# of participants
1	Baseline	6 (Caltech)
2	Baseline	24
3	Baseline	12
4	Baseline	24
5	Social Image	20
6	Social Image	22
7	Social Image	20
Total	Both	128

Table A1. Summary of Experimental Sessions

Condition	Type	total duration	self duration	other* duration	total clicks	self clicks	other* clicks	% trials all clicked
Baseline	S2O	3635	1231	1202	13.09	4.62	4.24	92
Baseline	S1O	1939	1021	918	7.37	4.18	3.19	92
Baseline	2O	2376		1188	8.71		4.36	98
Social Image	S2O	3275	1129	1073	10.69	3.87	3.41	83
Social Image	S1O	1792	865	926	6.32	3.37	2.95	92
Social Image	2O	1817		908	6.20		3.10	93

Table A2. Summary of Overall Mean Durations (in msec) and Clicks
Note: * other=(other1+other2)/2 for S2O and 2O.

Condition	**Type**	**All**	**Excluding repeated lookups**	**Within Payoff column**	**Within Choice Row**	**S-O1**	**S-O2**	**O1-O2**
Baseline	S2O	8.27	7.92	2.61	3.80	2.70	0.44	2.07
Baseline	S1O	8.49	8.17	2.92	3.67			
Baseline	2O	10.91	10.50	3.79	4.94			
Social Image	S2O	6.74	6.44	2.40	2.61	1.92	0.12	0.58
Social Image	S1O	7.22	6.88	2.62	2.77			

Social Image	2O	9.00	8.58	3.39	3.53			

Table A3. Summary of Transition Patterns (repeated lookups are two consecutive clicks in the same payoff cell)

Allocation Decisions

Decision	Self	Other 1	Other 2
1	400	1400	500
	400	800	700
2	600	0	300
	500	0	425
3	500	100	100
	500	500	0
4	575	575	575
	675	800	200
5	750	500	1600
	800	1300	300
6	1200	200	1100
	1200	400	500
7	1000	1600	900
	1100	2100	100
8	1200	300	2100
	1300	400	1700
9	400	300	300
	300	475	475
10	0	500	500
	100	0	0

11	400	400	
	400	750	
12	375	750	
	400	400	
13	200	800	
	0	0	
14	600	300	
	500	700	
15	600	600	
	700	200	
16	500	600	
	500	400	
17	900	100	
	800	200	
18	600	400	
	800	200	
19	600	300	
	500	425	
20	400	300	
	300	475	
21		400	400
		375	750
22		500	700
		600	300
23		800	0
		400	400

24		800	200
		900	100
25		600	300
		500	425
26	800	1600	500
	1000	1000	100
27	600	300	1300
	800	100	1000
28	500	400	400
	500	750	375
29	500	1200	0
	500	400	400
30	575	575	575
	600	900	300
31	560	820	460
	600	880	360
32	560	820	460
	560	940	260
33	640	940	260
	640	260	940
34	650	940	260
	700	740	380
35	1000	900	1700
	1200	1800	500
36	500	500	
	450	1000	

37	350	500	
	300	300	
38	700	1050	
	800	800	
39	800	800	
	900	550	
40	750	1050	
	850	550	
41	900	1200	
	900	1050	
42	400	400	
	375	800	
43	375	1200	
	400	400	
44	375	1500	
	400	400	
45	400	400	
	375	2000	
46	400	400	
	750	375	
47	1200	0	
	400	400	
48	900	300	
	575	575	
49	820	460	
	940	260	

50	900	1700	
	100	1900	

Sample Instructions (Social Image Treatment)

Thank you for participating in this experiment. You will receive a show-up fee and in addition will be given to earn additional money as a function of your decisions during the experiment.

If you have any questions during the instruction period, raise your hand and your question will be addressed.

During the experiment you will make 50 independent decisions. The entire experiment will take place through computer terminals, and all interaction between participants will take place through the computers. It is very important that you not talk or in any way try to communicate with other participants during the experiment.

The experiment is divided in trials. In every trial you will be paired with two other participants. You will not be told the identity of the participants that are grouped with.

In each type of trial one person in the group will be asked to make a choice between two potential sets of transfers from the experimenter to the members of your group, including yourself. Each option is represented by a row in the table. Each column represents what either you or one of the other two people in the group will receive under each option.

[Screen 1-sample game matrix-you, other 1, other 2]

The diagram illustrates the way the options are described in the experiment. For example, if Option 1 is chosen, then you will receive 3, Other 1 will receive 14, and Other 2 will receives 15. On the other hand, if option two is chosen, then you will receive 9, Other 1 will receive 26, and Other 2 will receive 5.

You will have to make decisions in three types of situations. The following screens describe each type.

[Screen 2-hidden matrix-you, other 1, other 2]

TYPE 1. Display 2 shows the first type of situations. Payoffs in the table are now **hidden in boxes**. This is the type of screen you will see during the experiment. You can see the payoffs by moving your computer mouse into the boxes, ***left- or right-clicking***, and ***holding down*** the mouse button. If you do not hold down the mouse button the payoff will disappear. When you move the mouse away from the box, the payoff will also disappear. If you move your mouse back into the same box, and click and hold, the payoff will appear again. In the experiment, clicking does not affect your earnings and you can click as few or as many times as you wish.

Whenever you are ready to make a choice, you will have to press the Option 1 or Option 2 button to indicate your choice.

[Screen 3-hidden matrix-you, other 1]

TYPE 2. In the second type of situation, the options pertain only to you and one other person. The third person in the group is not affected by the choice thus the payoff table has only two columns for You and Other 1.

[Screen 4-hidden matrix- other 1, other 2]

TYPE 3. In the third type of situation, the options pertain only to the two other people in the group. You are not affected by the choice thus the payoff table has only two columns for Other 1 and Other 2.

The person you paired with will be observe all of your clicks and see the payoffs that you click on their screen. In type 1, the observer may be Other 1 or Other 2. In type 2, the observer is Other 1. In type 3, again the observer may be Other 1 or Other 2.
No option choices will be revealed at any point.

Your earnings during the experiment are denominated in **points**. At the end of the experiment you will be paid $1.00 for every 100 **points** you have earned. You will be paid based on the decisions that you made in two randomly chosen situations, one in which you made the option choice and one in which another person in a group to which you belong made the choice. Everyone will be paid in private and you are under no obligation to tell others how much you earned.

[AUTHENTICATE clients]

Please double click on the icon on your desktop that says mouseclicker. When the computer prompts you for your name, type your First and Last name. Then click SUBMIT and wait for further instructions.

[START match]

We will begin with two practice situations so you can become familiar with the interface. You will not be paid for your decision in this situation. If you are making the decision in the situation, please use your mouse button to reveal the payoffs in the different boxes. Familiarize yourself with the click-and-hold method. For the time being do not make any option choices. If you are observing, please look at the screen to see what the decider in your group is clicking on and what the payoffs are.

If your subject ID is odd, choose Option A. If your subject ID is even, choose Option B. Once everyone has made a choice, the payoff table is revealed and we continue to go on to the next situation.

In the second situation, those who were making the decision before are not observers and the observers are now making the decisions. Please follow the same instructions as in the first practice trial.

Are there any questions before we begin with the paid experiment? We will now begin with the real paid trials. Please pull out your dividers for privacy. If there are any problems or questions from this point on, raise your hand and an experimenter will come and assist you.

There are 25 trials in Part 1. You will be assigned a role, decision maker or observer, and keep the same role for all 25 trials. However, you will be re-matched into a different group after each trial.

[END OF MATCHES]

That was the last trial. We will now have you answer some survey questions. When you are finished, please line up to be paid one by one. Thank you for your participation.