Procedural Fairness and the Cost of Control[*]

By Judd B. Kessler[†] and Stephen Leider[‡]

This Draft: August 7, 2014

ABSTRACT

A large and growing literature has demonstrated that imposing control on agents has the potential to backfire, leading agents to withhold effort. Consistent with principles of procedural fairness, we find that the way in which control is imposed — in particular whether control is imposed symmetrically on both principals and agents and whether both parties have a say in whether control is imposed — affects how agents respond to control. In our setting, control leads agents to withhold effort only when control is imposed unilaterally with an asymmetric affect on the agent.

[†] The Wharton School, University of Pennsylvania, 3620 Locust Walk, 1454 Steinberg Hall-Dietrich Hall, Philadelphia, PA 19104, judd.kessler@wharton.upenn.edu
[‡] The Ross School of Business, University of Michigan, 701 Tappan Street R4486, Ann Arbor, MI 48109, leider@umich.edu.

I. Introduction

In many principal-agent, supply chain, and partnership relationships, parties can take opportunistic actions that harm their counterparts. Agents can shirk, principals can refuse to pay earned discretionary bonuses; suppliers can fail to deliver high quality, buyers can fail to pay on time; and partners can withhold effort or resources from one another. To prevent these opportunistic actions, parties often consider the use of control mechanisms such as establishing contractual restrictions that require perfunctory performance. Standard economic theory suggests that these contractual tools can more effectively align the interests of parties and avoid inefficient outcomes, and control is regularly used in principal-agent settings.[1]

Recent literature has demonstrated, however, that such control strategies may come at a cost. In principal-agent settings, control that restricts an agent's action has been shown to backfire, leading to more opportunistic behavior rather than less. The intuition behind these results is that imposing control in a contract demonstrates distrust and may lead controlled parties to withhold effort (e.g. Frey 1993, Barkema 1995, Falk and Kosfeld 2006).[2] These findings highlight the need for researchers to better understand when agents will withhold effort in response to control and when control will work in the way standard economic theory predicts.

One dimension that may affect whether control backfires is how control is imposed. Our research in this area is motivated by the observation that, unlike the

---

[1] Between 1960 and 1995 average supervisor-employee ratios in the non-farm economy increased for many developed countries (Vernon 2003). Similarly, in light of the potential for crowd out of intrinsic motivation, 37% of individuals in the U.S. have some form of pay-for-performance incentives (Lemieux et al. 2009).

[2] These findings are related to work showing that extrinsic incentives put in place by a principal to motivate an agent might undermine an agent's intrinsic motivation and lead to lower effort (see e.g. Titmuss 1970, Frey 1994, Gneezy and Rustichini 2000a) or might undermine a norm and make misbehavior more transactional (Gneezy and Rustichini 2000b). See also a rich literature in Psychology, which has shown that extrinsic incentives can undermine intrinsic motivations (see Lepper and Greene 1978; Deci 1975; Deci 1971; Kruglansky, Freedman, and Zeevi 1971), a notion which has been more recently studied in the economics literature (see for example Frey, Oberholzer and Eichenberger 1996 and Frey and Oberholzer-Gee 1997). A survey by Bowles and Polania-Reyes (2012) identifies four major mechanisms for a crowding out effect of incentives: (1) incentives providing "bad news" about the principal, (2) framing effects that lead to "moral disengagement", (3) aversion to a loss of autonomy, and (4) influence on the formation and updating of preferences.

asymmetric settings in which control has been shown to backfire, control is often applied consistently across individuals and affected parties often have a voice in establishing the control mechanisms imposed on them.[3] A robust literature in procedural fairness (Thibault and Walker 1975, Lind and Tyler 1988) finds that fairer processes lead to greater trust and better performance and suggests that control mechanisms are likely to be more effective when applied symmetrically (i.e. consistently across individuals) and bilaterally (i.e. when affected parties have a voice in imposing control).[4]

We explore the implications of procedural fairness on the efficacy of control in a laboratory experiment.[5] In our experiment, pairs of subjects interact anonymously in a one-shot game. In the game, one party has the opportunity to take a costly action that benefits the other. If control is in place[6], the most opportunistic actions are eliminated from the action space so that the actor is required to provide perfunctory performance. Our design allows us to vary whether control is imposed asymmetrically on one party or has a symmetric effect on both parties and whether control is imposed unilaterally by one party or bilaterally by an agreement of both parties.

In order to vary how control is imposed, our design randomly assigns subjects to play in the role of a principal or an agent and control that affects the agent is sometimes

---

[3] For example, in employment relationships, pay and work rules are often set for job categories rather than for individuals, and unions and joint consultative committees frequently help determine pay and work practices. Even without these formal structures, managers often seek input from employees. In the 2011 UK Workplace Employment Relations Study, 52% of employees rated management as "very good" at seeking input in decisions, and 46% rated it as very good at responding to suggestions (van Wanroy et al. 2013). In particular, 95% of managers consulted employees if laying off two or more workers, and 40% or more consulted on decisions such as changing work techniques, changing work organization, introducing performance pay, or addressing health and safety concerns.

[4] In a labor relationship, a requirement to arrive to work by 9am or face punishment could apply only to front-line employees (asymmetric) or to managers as well (symmetric). The rule could have been imposed by the managers (unilaterally) or negotiated with a worker union (bilaterally). Similarly, in a supply chain relationship, enforceable deadlines might only control the delivery of goods (asymmetric) or also apply to accounts receivables (symmetric). These contractual requirements could be part of a non-negotiable contract to transact with the supplier (unilateral) or negotiated between the parties (bilateral).

[5] Schnedler and Vadovic (2011) consider a similar notion of "legitimacy of control," however they focus on the issue of who control is targeting — for example, varying whether there is a chance control will be imposed on a computer player rather than a real agent — while we address the process by which control is imposed.

[6] In our setting "higher" actions are more prosocial. Controlling the agent is therefore imposing a minimum actions. This was described to subjects as "restricting" the other player.

imposed before and sometimes imposed after subjects learn who is which role. To observe behavior when control is imposed unilaterally and asymmetrically, in our baseline treatment we reveal the identities of the principal and agent first and then allow the principal to impose control on the agent.[7] To observe behavior in a more symmetric environment where control is applied consistently across individuals, we run a treatment in which one subject is randomly selected to impose control before the identities of the principal and agent are revealed, so that control affects whichever of the two subjects is revealed to be the agent. Finally, to model an environment in which control is imposed bilaterally and both participants have a voice in imposing control, we run a treatment in which players must agree on control for it to be imposed on whichever subject is later revealed to be the agent.

Our results are consistent with theories of procedural fairness. We find that control only backfires when it is imposed unilaterally and asymmetric on the agent. Control becomes somewhat more effective when we allow it to be imposed symmetrically, and it is most effective when imposed bilaterally by an agreement of both agents.[8]

In addition to demonstrating the importance of procedural fairness on the efficacy of control, our experimental results speak to an ongoing experimental debate about the robustness of results in this literature. In previous experimental work, principals were shown to be worse off on average when they imposed control on agents rather than leaving them unconstrained (Falk and Kosfeld 2006). This result is not always observed, even in somewhat similar experimental settings (Hagemann 2007, Schnedler and Vadovic 2011, Ploner et al. 2011). We find that for control to make principals worse off, control must not only be imposed unilaterally and asymmetrically on the agent, but it must also be sufficiently weak — so that it cannot induce significant effort from otherwise low-contributing agents — and average effort in the absence of control must be

---

[7] This treatment allows us to embed a design quite similar to the experiment run in Falk and Kosfeld (2006).

[8] As will be discussed further in Section IV, our results also speak to a growing theoretical literature that has arisen to explain why control may lead agents to withhold effort. While many models do not map onto our setting in a straightforward way, we find that Ellingsen and Johannesson's (2008) model of esteem can rationalize our pattern of results.

sufficiently high.[9] If these conditions are not met, principals are no worse off from imposing control and may be better off doing so.

The totality of our results provides insight about when agents will respond negatively to control and when the contractual tools available can be expected to work as standard theory predicts.

The paper proceeds as follows. Section II highlights related literature. Section III describes the experimental design. Section IV motivates our main hypotheses. Section V presents our main experimental results. Section VI describes extensions to our experiment and their associated results. Section VII discusses the implications of our results for economic theory and firm behavior and concludes.

II.   Related Literature

Firms regularly use control, monitoring, and incentives to manage agency problems in principal-agent settings and in markets with supply chains. Research has addressed the role of contracts in implementing these incentive strategies and revealed a striking fact that many contracts are much simpler and less complete than standard theory would predict. Traditional explanations of this contractual incompleteness appeal to transaction costs (e.g. Coase 1937, Williamson 1975, 1985) or bounded rationality (e.g. Simon 1981) to argue that more complete contracts are impractical. Another line of research has suggested that leaving contracts incomplete may be suboptimal but necessary given that agents are asked to multitask (Holmstrom and Milgrom 1991). Additionally, some authors have provided theoretical justifications for why incomplete contracts may be optimal, such as complete contracts signaling negative information about the contract proposer (Allen and Gale 1992, Spier 1992), complete contracts leading the agent to infer that a less prosocial norm prevails (Sliwka 2007), or that incompleteness creates strategic ambiguity that helps enforce implicit agreements (Bernheim and Whinston 1998).

---

[9] As will be described in detail in the experimental design section, we increase average actions absent control by giving the subjects the opportunity to mutually agree to play a high effort level in advance of the revelation of roles and the decision to control, which has been shown to work in related settings (Kessler and Leider 2012).

In the spirit of these results is an experimental literature demonstrating that control, monitoring, and incentives can demoralize agents.[10] Falk and Kosfeld (2006) suggest that contractual incompleteness in control mechanisms could also arise to signal trust. They demonstrate that imposing control on agents — by eliminating their most opportunistic actions and forcing them to provide at least minimum perfunctory effort — can lead to worse outcomes for the principal. Falk and Kosfeld (2006) find both a "hidden cost of control" — in which agents who would provide high effort when unconstrained display a behavioral response and provide less effort when controlled — and an "average cost of control" — in which this behavioral response is so large that it swamps the beneficial effect of control of raising low actions and makes principals worse off from imposing control.[11]

Several theoretical models have been proposed to explain such a cost of control. Sliwka (2007) describes a model in which some agents are uncertain whether the prevailing norm is selfish or prosocial. In this case, control imposed by the better-informed principal is a negative signal about the norm, which induces lower effort. Ellingsen and Johannesson (2008) lays out model where individuals are (heterogeneously) prosocial, care about others' belief about their type, care more about the opinions of high types, and types and beliefs are positively correlated. In this model, controlling provides a negative signal about the principal's type, leading agents to care less about signaling their own prosociality to such a principal. Von Siemens (2013) uses intentions-based reciprocity to explain the cost of control. In this model, control is perceived as harmful to selfish agents, and reciprocal agents reward principals who do not control.

Despite the numerous theories motivated by the cost of control findings, a number of recent papers have attempted and failed to replicate the average cost of control result

---

[10] In addition to the control mechanisms described in detail below, extrinsic incentives have been shown to crowd out intrinsic motivation (see Deci et al. 1999 and Gneezy, Meier and Rey-Beil 2011 for surveys).

[11] Notice that we say that an "average cost of control" arises when a principal receives on average less effort from controlling an agent than from giving the agent a larger action space. We say that we have observed a "behavioral response" or a "hidden cost of control" when a subset of subjects respond negatively to the imposition of control by providing less effort when control is imposed than when it is not. Falk and Kosfeld (2006) use the term "hidden cost of control" to title their paper.

found in Falk and Kosfeld (2006) and have consequently argued that costs of control are unlikely to be the cause of contractual incompleteness (Hagemann 2007;[12] Schnedler and Vadovic 2011 and Ploner et al. 2012[13]). While these papers fail to replicate the average cost of control results, they generally do replicate the behavioral response or "hidden cost of control" results in which a subset of agents contribute less when they are controlled than when they are not controlled.

Research in other settings, however, has observed the beneficial effect of control mechanisms without the offsetting behavioral response (Kessler and Leider 2012).[14] One striking difference between the settings with and without the behavioral response is the manner in which control is implemented, namely who is affected by control and who imposes it.[15]

Such findings are consistent with an extensive literature in psychology and law considering the role of "procedural justice" and "procedural fairness," that is the fairness of the process by which decisions are made and outcomes are determined (see Thibault and Walker 1975, Lind and Tyler 1988 for seminal works). Leventhal (1980) identifies six principles of procedural justice, two of which highlight the importance of symmetry in the imposition of control. The "consistency rule" argues that procedures to determine payments and outcomes should be consistent across persons, and the "bias suppression rule" states that in a process the decision-makers should seek to be impartial, rather than advancing their own personal self-interest. In addition, "voice" — the ability of those affected by a decision process to participate and state their interests and desires — has been identified as an important contributor to perceptions of procedural justice,

---

[12] However, with only 30 agents in each treatment, Hagermann's experiment may be underpowered to identify a treatment effect in the baseline case. Hagermann (2007) finds a difference in average effort of 5.3, which is very similar to the difference of 5.5 (23 without control and 17.5 with control) in the equivalent treatment in Falk and Koslfeld (2006), which has 72 agents and identifies the effect as significant.

[13] Ploner et al. (2012) find both directionally negative and directionally positive effects for the principal of imposing control, depending on the subject pool.

[14] In Kessler and Leider (2012), subjects play a number of two-person public good games, and in three of the four games, adding control to a pre-game contract does not generate a behavioral response.

[15] Kessler and Leider (2012) investigates control imposed both bilaterally and unilaterally, but when a minimum restriction is imposed it is always imposed on both agents simultaneously.

highlighting the importance of giving parties that might be affected by control a role in determining whether it is imposed (Bies and Shapiro 1988, Lind et al. 1990).

The literature has also shown that perceptions of procedural justice (or injustice) affect behavior and performance. Alexander and Ruderman (1987) show that procedural justice, specifically giving workers voice, increases trust in management. Lind et al. (1993) finds that procedurally fair dispute resolution mechanisms (i.e. those with impartial authorities and participant voice) were more successful in leading participants to accept the outcomes of the arbitration. De Cremer and van Knippenberg (2002, 2003) show that voice for participants increased cooperation in social dilemmas, particularly when decision-makers do not privately benefit from their power. Finally, a meta-analysis by Cohen-Charash and Spector (2001) finds that across both lab and field studies, procedural justice was positively associated with job performance (including both effort and outcome metrics) and negatively associated with counterproductive work behavior (such as improper work, theft, and damaging equipment).

An economics literature on procedural fairness, surveyed by Frey et al. (2004), finds similar results. Random allocation procedures such as lotteries have received particular attention, with lotteries being perceived of as quite fair, particularly if they are "symmetric" in the sense of choosing outcomes favorable to each party with equal probability (Kahneman et al. 1986, Bolton et al. 2005). Research on how procedural fairness affects performance mirror the psychology literature. Frohlich and Oppenheimer (1990) found that workers who had the ability to vote on a tax system were more productive over time than those who had a tax system imposed. Greenberg (1990) finds that fair processes led to less negative reactions to pay decreases, while Benz and Stutzer (2003) finds that voice in pay determination led to increased job satisfaction. Benz and Frey (2004) compares self-employed workers to workers in firm hierarchies (holding fixed pay, hours, etc.) and finds that the more self-directed employees had higher job satisfaction.

The results from these literatures suggest that more procedurally fair impositions of control will lead control to backfire less often and be more effective. The experimental design in this paper, which is presented in the next section, investigates this hypothesis by

analyzing whether the procedure by which control is imposed impacts the efficacy of control.

III.   Experimental Design

In the experiment, subjects played an anonymous transfer game a total of 20 times. Subjects were randomly matched with another subject in the laboratory in each round of the game.[16]

In each round of the game, the agent (called "Player A" in the instructions) started with 120 experimental units (EUs) worth $0.05 each. The agent could transfer these units to the principal (called "Player B" in the instructions) and any units transferred to the principal were doubled. Consequently, the payoffs for the principal agent game were:

Agent ("Player A"): $\pi_A = 120 - x$

Principal ("Player B"): $\pi_P = 2x$

where x represents the number of units transferred by the agent to the principal.

If control (called "a restriction on Player A's transfer" in the instructions) was not imposed, agents could choose to transfer any amount x from 0 to 120. If control was imposed, agents were restricted to transfer at least 4 EUs, so agents could transfer any amount x from 4 to 120.

The experiment has three main treatments, which vary whether control asymmetrically affects one subject or symmetrically affects both subjects and whether control is imposed unilaterally or bilaterally. Table 1 displays the three treatments.

Table 1: Experimental Treatments

		Symmetry of Control	
		Asymmetric	**Symmetric**
Control Imposed	**Unilaterally**	Baseline Treatments	Mutual Minimum Treatment
	Bilaterally		Consent Treatment

Table 1 shows the main experimental treatments in the experiment.

---

[16] The experiment was run on z-Tree (Fischbacher 2007).

In the *Baseline Treatment*, the roles of principal and agent were assigned before control was imposed. After the principal and the agent were assigned their roles, the principal was given the option of whether to impose control (the principal decided between: "No restriction" and "A restriction that Player A must transfer at least 4 EUs"). This choice was revealed to the agent who decided how many experimental units to transfer, with the transfer restricted to be at least 4 EUs when control was imposed. Notice that for the *Baseline Treatment*, the minimum is imposed asymmetrically and unilaterally.

We add symmetry to control in the other two treatments by having control affect both players; we implemented this by giving subjects the opportunity to impose control before they learned who was the agent.[17] In the *Mutual Minimum Treatment*, we randomly gave one of the players the option to impose control on *whichever player* became the agent. After the subject decided whether to impose control, we assigned the roles of principal and agent. If the subject had imposed control, whichever of the two players was randomly selected to be the agent was restricted to transfer between 4 and 120 EUs. If control was not imposed, the agent could choose any transfer between 0 and 120 EUs. Notice that for the *Mutual Minimum Treatment*, the minimum is imposed symmetrically and unilaterally.

In both of the previous treatments, one subject has the opportunity to impose control unilaterally. In our third treatment, both subjects must bilaterally agree on control for it to be imposed. In the *Consent Treatment*, before we assigned the roles of principal and agent, we allowed both players to suggest whether or not control should be imposed on *whichever player* became the agent. In particular, each player could either suggest that the restriction be in place or not suggest it. Only if both players suggested the restriction be in place was control imposed. After each subject made a decision, the players were told who suggested the restriction and whether the restriction was imposed. We then

---

[17] This mechanism is similar to Rawls' (1971) "veil of ignorance," where individuals must establish the rules of a society before knowing their roles. The "veil of ignorance" has been used experimentally to examine issues such as taxation and redistribution (Frohlich and Oppenheimer 1990, Sutter and Weck-Hannemann 2003, Krawczyk 2010, Schildberg-Hörisch 2010). However, we are not interested here in analyzing the impact of the veil of ignorance, nor do we expect this veil to exist in practice. Rather, we use it as a design tool to create initial symmetry between the two paired subjects so that we can estimate the effect of changes in procedural fairness.

assigned the roles of principal and agent. If both players suggested the restriction, the agent was restricted to transfer between 4 and 120 EUs. If at least one of the players did not suggest the restriction, then the agent could choose to transfer any amount between 0 and 120 EUs. Notice that for the *Consent Treatment*, the minimum is imposed symmetrically on both players and is imposed bilaterally.

The cell in Figure 1 that is not associated with a treatment would require both subjects to agree bilaterally to impose control on one specific subject. For control to be imposed in this setting, a subject would need to choose to control himself, knowing that he alone would be affected. We do not consider this setting to be particularly relevant to our endeavor and so we did not run a treatment associated with this cell.

It is worth noting that two things change as we move from the *Baseline Treatment* to the *Mutual Minimum Treatment*. Control is being imposed symmetrically on both agents but we have also imposed control before we assign the roles of principal and agent. To ensure that any difference between treatments was driven by the symmetry of control and not the difference in timing, we also ran the *Unknown Agent Treatment* in which control was imposed asymmetrically on one player but was imposed before the role of agent as been assigned. That is, before we assigned the roles of principal and agent, we randomly gave one of the players the option to impose control on *the other player* if that other player became the agent. If the player who decided about control became the agent, he was unconstrained. As we will show in the following section, results from the *Unknown Agent Treatment* are never significantly different from results in the *Baseline Treatment*. We collapse these treatments together in our main analysis and call them jointly the *Baseline Treatments* (or just *Baseline*); we also show the results of the two treatments separately in Appendix B.

Subjects in our experiment always played 10 rounds in the *Baseline Treatment* and 10 rounds in one of the other three treatments. Whether they played the *Baseline Treatment* first or second was randomly assigned by session. Control can have two main effects on agent behavior. Control may raise transfers that would have been in the range 0 to 3 EUs to be at least 4 EUs when control is imposed. Additionally, control might lead some subjects who would have made a large transfer to transfer less, what we call a "behavioral response" or a "hidden cost of control". The net effect of these two forces in

a given treatment will determine whether we observe an "average cost of control" (i.e. whether principals receive less effort on average from agents when control is imposed). As we analyze the results starting in the next section, we will look both for a behavioral response as well as identify the net effect of both forces on average transfers.

Since the experiment aimed to investigate the behavioral response and the average cost of control, we made two additional experimental design choices. First, we implemented a pre-stage to the game, i.e. before the control decision, with the intent of raising the average action when the agent was not controlled, so that we would have a better chance of observing a behavioral response. In particular, before subjects were assigned to the role of principal or agent for the round, and before they knew whether control would be imposed, they had the opportunity to make a non-binding agreement to transfer 40 EUs (i.e. $x=40$) if they ended up being the agent.[18] Results from our earlier work (Kessler and Leider 2012), as well as the work of others,[19] suggests that allowing subjects to make such a non-binding agreement will lead to higher actions in the population and so would make behavioral responses to control easier to observe and measure. We choose to make the agreement amount 40 EUs since that is the payoff-equalizing transfer, leading both the principal and the agent to receive 80 EUs.

As will be documented in Section V when we introduce additional experimental treatments, including a treatment in which no agreements are allowed, giving subjects the opportunity to make such agreements raised average actions. Consequently, we were able to observe the behavioral response and an average cost of control more clearly due to this innovation.

---

[18] Agreements were made in the same way control was implemented in the Consent Treatment. In particular, each of the players independently decided whether or not to suggest: "An agreement that says 'We agree that if we are Player A, we will transfer 40 EUs to Player B.'" If both players suggested the agreement, then the agreement was made. If one or both of the players did not suggest the agreement, then no agreement was made. After both players had decided whether or not to suggest the agreement, the players were told what the other had chosen and whether they had made an agreement.

[19] Other studies have found benefits of unilateral promises in holdup games (Ellingsen and Johannesson 2004), trust games (Charness and Dufwenberg 2006), and dictator games (Vanberg 2008). In related work, Dufwenberg et al. (2011) provide a theoretical model that identifies what agreements should form as binding contracts or as non-binding informal agreements, and test their model with a lost wallet game. Binding contracts are predominantly 50-50 splits, while non-binding informal agreements lead to higher payoffs for the second mover, which one can think of as the agent.

Second, we chose a minimum to be 10% of the value of payoff-equalizing transfer, which set the minimum at 4 EUs, slightly below the benchmark minimums of 5, 10, and 20 EUs in Falk and Kosfeld (2006). We again made this design choice with the hope of being able to identify a behavioral response and an average cost of control in the baseline case. We investigate the effect of raising the minimum in Section V.

Our design also allows us to speak to an additional question about how agents respond to the imposition of control. Since we observed subjects play this one-shot game a number of times, and since we randomly assigned the roles of principal and agent in each round, we observe the same subject playing as both a principal and an agent. Consequently, in addition to identifying how agent behavior responds to the symmetry of control and whether control is imposed unilaterally or bilaterally, the experiment addresses whether individuals' use control as a principal is correlated with how they respond to control as an agent. We address this question towards the end of the results section.

IV.   Behavioral Hypotheses

We base our hypotheses on intuitions from the procedural fairness literature. Control imposed in the *Baseline Treatments* exhibits little procedural fairness. The principal imposes control only on the agent (violating Leventhal's consistency rule); the principal directly and uniquely benefits from controlling the agent (contrary to the bias suppression rule); and the agent has no voice in the decision process. Hence we would expect controlled agents to feel untrusted and to choose low effort. This would manifest as a behavioral response and potentially an average cost of control. Note that whether control is imposed unilaterally and asymmetrically on the agent after the role of agent is assigned to a subject (as in the *Baseline Treatment*) or before the role is assigned to a subject (as in the *Unknown Agent Treatment*) does not change the process with respect to procedural fairness, hence we would expect no difference between these two *Baseline Treatments*.

Control in the *Mutual Minimum Treatment* displays more procedural fairness. Since both players face the same constraints when making an effort choice, the consistency rule is now satisfied. Additionally, the subject who can choose to impose

13

control cannot privately benefit from this decision-making right, since the other subject would receive the same benefits of control if placed in the role of principal. Hence we should expect agents to feel less distrusted by control. Effort should therefore be higher, and the behavioral costs of control should be lower, in *Mutual Minimum* as compared to the *Baseline Treatments*.

The *Consent Treatment* further increases procedural fairness, since both subjects now have a voice in the decision to impose control, which is imposed symmetrically as in the *Mutual Minimum Treatment*. Therefore, we should expect control to be most effective in the *Consent Treatment*.

Additionally, if there is heterogeneity between subjects in how they view the procedural fairness of imposing control, this should affect both decisions as the principal and the agent. A subject's willingness to impose control as a principal in the *Baseline Treatment* may indicate that they do not see control as unfair. This would then lead those subjects to have a smaller decrease in their effort (i.e. a smaller behavioral response) when being controlled as an agent.

V.   Results

A total of 464 student subjects participated in 25 sessions in the Wharton Behavioral Lab at the University of Pennsylvania. As noted in Section III, all subjects participated in the *Baseline Treatment* and one other treatment. Of the 464 subjects, 158 subjects also participated in the *Mutual Minimum Treatment*, 158 subjects in the *Consent Treatment*, and the remaining 148 subjects in the *Unknown Agent Treatment*. Sessions lasted approximately one hour. Average subject pay was $17.28, including a $10 show-up fee.

Before we delve into analysis, we make two simplifications that allow us to more clearly communicate our data. First, in our main results we combine data from the two treatments where control is imposed asymmetrically and unilaterally, the *Baseline Treatment* and the *Unknown Agent Treatment*. As described in the Section III, the latter treatment was specifically designed to ensure the difference between *Baseline Treatment* and *Mutual Minimum* was not due to the timing of when principal and agent roles were revealed. Results demonstrate that there is no effect of the timing of role revelation and

14

results from the *Baseline Treatment* and the *Unknown Agent Treatment* are nearly identical (see Appendix A for a comparison of summary statistics and Appendix B to see the main regression tables with these two treatments presented separately). Throughout this section we call the combined data the *Baseline Treatments* (or just *Baseline*).

Second, as we expected, allowing subjects to make a non-binding agreement that whichever subject ends up as an agent will transfer 40 EUs raised the average transfer and made it easier to observe an average cost of control (see complete analysis in Section VI). In particular, most pairs decide to have the agreement, which significantly raises average actions in the absence of control and thus make it much more likely for us to observe a behavioral response when control is imposed.

To show the effect of control on different types of groups, the graphs in this section condition on whether a pair had an agreement. However, it is worth emphasizing that whether the subjects in the pair choose to have an agreement is endogenous. Consequently, we present all our regression specifications twice, once pooling across *all pairs* in each treatment (i.e. combining together both pairs that had an agreement and those that did not) and again analyzing only the pairs that endogenously choose to have an agreement where we are more likely to see a cost of control.

5.1 Effect of Control Across Treatments

In this subsection we analyze how control affects transfers and show that, as hypothesized, it varies by whether control is imposed asymmetrically or symmetrically and unilaterally or bilaterally.

For control to be implemented in the *Baseline Treatments* and *Mutual Minimum Treatment*, *one subject* needs to have imposed control. In the *Consent Treatment*, on the other hand, *both subjects* need to want the minimum for control to be imposed. To avoid selection issues arising from the fact that in the *Consent Treatment* control is only imposed on subjects who suggest the restriction themselves, throughout this section we analyze agents' behavior as a function of whether the *other subject* wants the minimum (i.e. wants to impose control by restricting the agent's action). Notice that whether the

other subject wants the minimum is exogenous to the agent who makes the transfer. This makes the *Consent Treatment* comparable with the other treatments.[20]

Figure 1 shows the average amount transferred in each treatment as a function of whether the other subject wanted to control the agent by having the minimum in place. Panel A shows the effect of the other subject wanting the minimum when the two subjects made an agreement to transfer 40 EUs. We can begin to see differences between the treatments. In the *Baseline Treatments*, the average transfer when the other subject wants the minimum is significantly lower than when the other does not want the minimum, reflecting an average cost of control when an agreement is in place (29.28 when other does not want minimum, 23.62 when other wants the minimum; subject level non-parametric permutation test: $p = 0.03$; session level permutation test: $p = 0.01$). This difference is smaller in the *Mutual Minimum Treatment* (27.27 when other does not want minimum, 24.95 when other wants the minimum; subject-level: $p = 0.58$; session-level: $p = 0.30$) and flips sign in the *Control Treatment* where asking for the minimum leads to an increase in average transfer (19.78 when other does not want minimum, 22.22 when other wants the minimum; subject-level: $p = 0.02$; session-level: $p = 0.08$).

Panel B shows average transfers when subjects did not make an agreement to transfer 40 EUs. When no agreement is in place, asking for the minimum always directionally increases transfers, with the increase being statistically significant in the *Baseline Treatments* ($p = 0.02$ for both) and marginally significant in the *Consent Treatment* (subject-level: $p = 0.09$; session-level: $p = 0.73$).

We see a similar picture when we turn to the CDFs of transfers in Figure 2. In each graph the red (dashed) line is the CDF when the other player wants the minimum and the blue (solid) line is the CDF when the other player does not ask for control. The left column displays CDFs when subjects have an agreement. Looking in the top row, we see that in the *Baseline Treatment* there is a behavioral response when control is imposed in that many fewer agents transfer 40 and many more make a transfer at the minimum of 4. This effect of asking for control is mitigated in the *Mutual Minimum Treatment*; while we still see directionally fewer subjects transfer 40 and directionally more transfer at the

---

[20] In the analysis, we exclude agents who were also the player who decided whether or not there should be a restriction in the *Unknown Agent* and *Mutual Minimum* treatments so the control in those treatments is always being imposed by the subject who ends up in the role of the principal.

minimum when control is imposed, the CDFs appear closer together than in the *Baseline Treatments*. Finally, in the *Consent Treatment* asking for the minimum increases transfers by shifting the distribution up: it raises transfers from 0 up to 4 and there is no accompanying behavioral response.

We can use nonparametric tests to see if the distributions are significantly different. When subjects have an agreement in the *Baseline Treatments*, imposing control shifts the whole distribution above the minimum to the left (Kolmogorov-Smirnov test: $p < 0.01$ for both). In the *Mutual Minimum Treatment* there is no difference in the overall distribution of transfers (KS test: $p > 0.20$). By contrast, in the *Consent Treatment* there is a marginally significant shift of the overall distribution of transfers to the right (KS test: $p = 0.10$).

Meanwhile, when there is no agreement in place we see little-to-no difference between the CDFs for agents transferring to principals who want control and do not want control in any of three treatments. The only change is that having control increases transfers from 0 to 4 when control is imposed.

The results from Figures 1 and 2 are reflected in regression specifications shown in Table 2 and Table 3 (all specifications include subject fixed effects and cluster standard errors by session). Table 2 reports regressions including all the data from each treatment, pooling pairs that made the agreement and pairs that did not. The first two columns show the effect on average transfer for all rounds (column 1) and for only the first treatment played in a session (column 2). The coefficient on *Other Restricted in Baseline* is negative and significant demonstrating an average cost of control in the *Baseline Treatments*. *Other Restricted in Mutual Minimum* is directionally positive and not significantly different from 0 and *Other Restricted in Consent* is positive and statistically significant, demonstrating that control increases transfers on average in the *Consent Treatment*.

Figure 1: Effect of Control on Transfers by Treatment

Panel A: Average Transfer with an Agreement
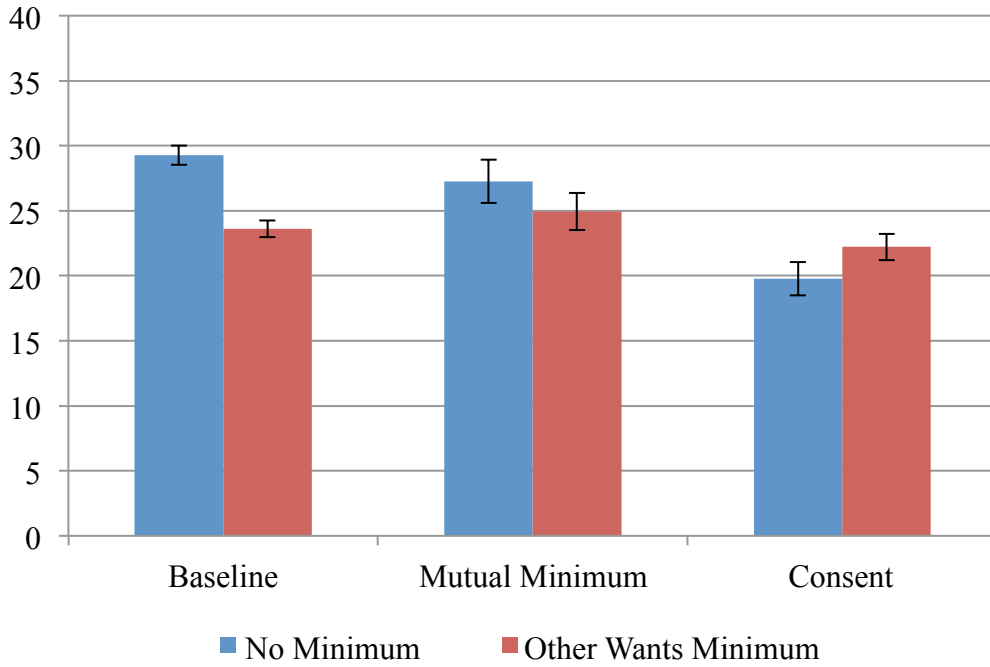
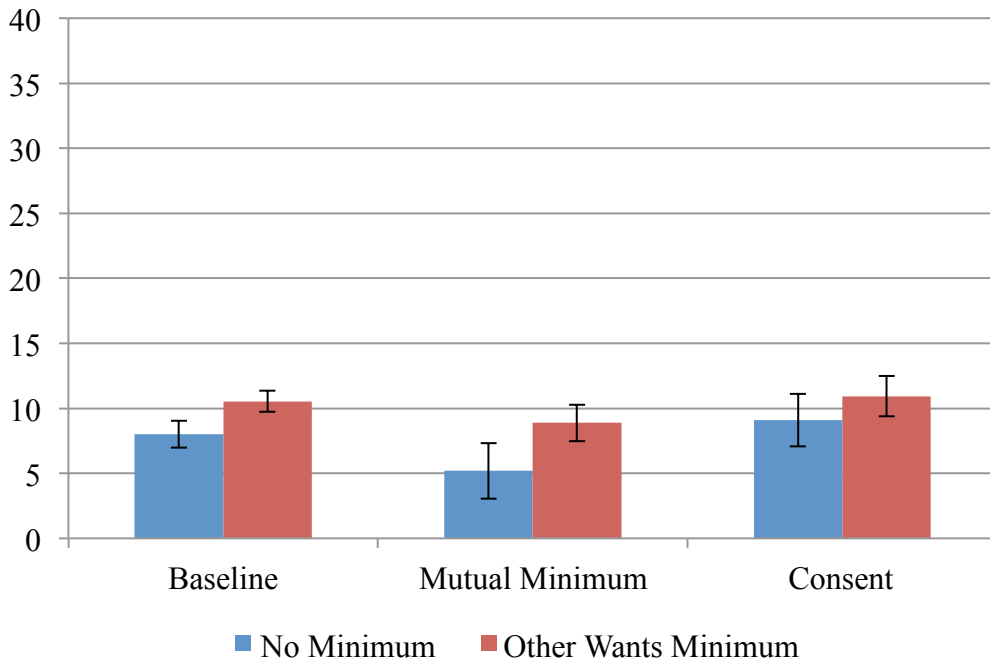Panel B: Average Transfer without an Agreement

Figure 1 shows the average transfer as a function of whether the other player wants the minimum (red bars) or does not want the minimum (blue bars). Panel A shows transfers when subjects have an agreement to play 40 EUs. Panel B shows transfers when no agreement is in place. Standard error bars are shown around each mean.

Figure 2: Distribution of Transfers by Treatment
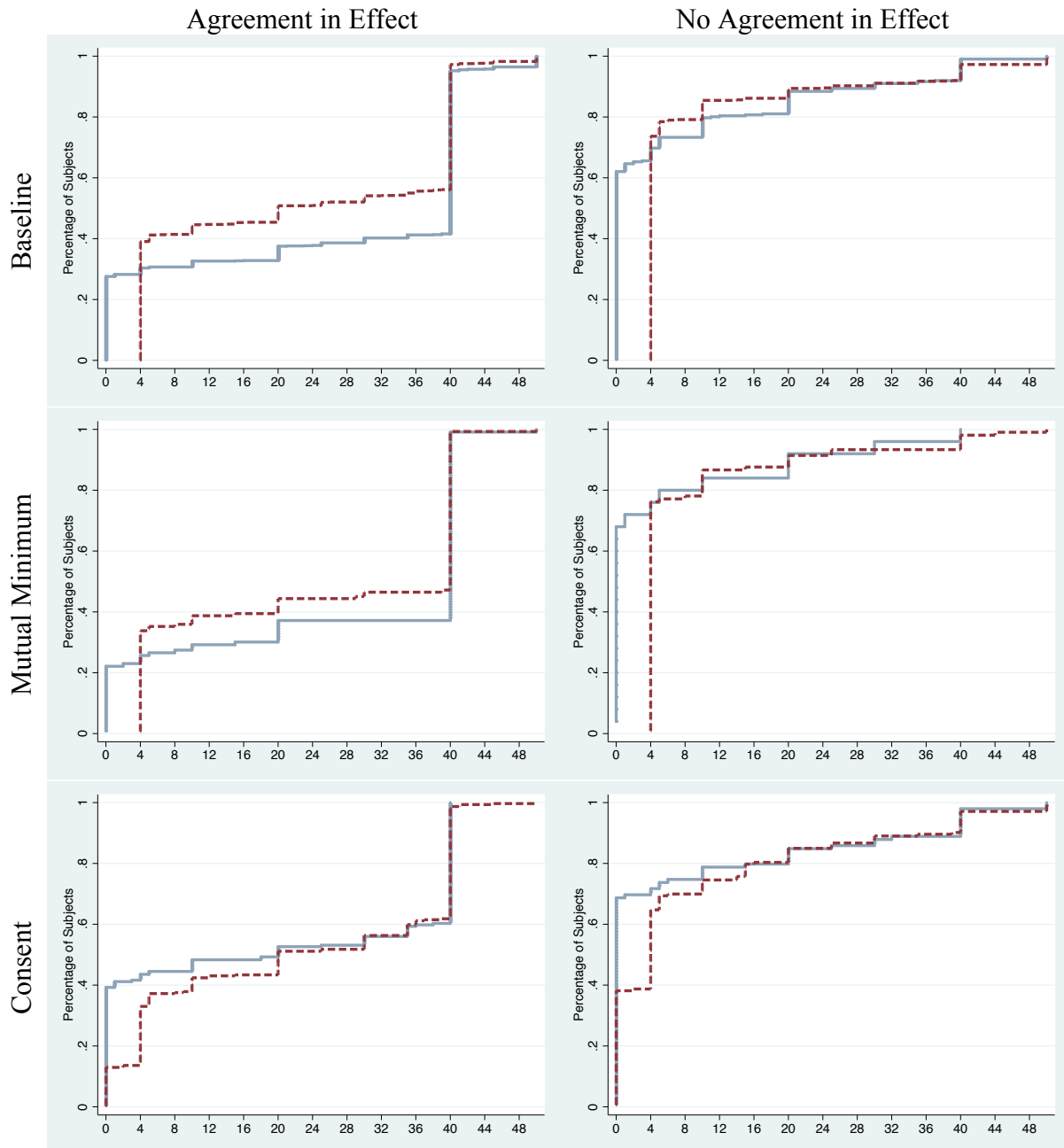
Notes: Figure 2 shows the CDF of transfers (with transferred censored at 50). The horizontal axis reports transfer size and the vertical axis reports the percentage of subjects. The blue (solid) lines are the CDFs when the other subject does not control, the red (dashed) lines are the CDFs when the other subject wants to control. In the *Consent Treatment* the other subject wanting control only leads to control if the agent also wanted control and so some transfers can be below 4 when the agent asks for control in that treatment. In the *Baseline Treatments* and *Mutual Minimum Treatment* control is always imposed when the other subject wants it.

Focusing on differences in average transfers can mask a behavioral response to control because control generates two opposing effects (as was seen in the CDFs in Figure 2). Control increases transfers that would have been below the minimum up to the minimum transfer allowed. In addition, control may also induce a behavioral response in which subjects lower their transfers in response to being controlled. In Tables 2 and 3 we investigate two ways of identifying a behavioral response. The first is to look at the fraction of subjects who transfer 4 EUs or less. If the restriction only affects those subjects who otherwise would have transferred less than the minimum, then the fraction of subjects transferring at or below the minimum of 4 should stay the same when control is imposed. Alternatively, if subjects who would otherwise transfer more than the minimum react negatively to the imposition of control by transferring only the imposed minimum, than this fraction transferring 4 units or less should increase when control is imposed. We analyze the probability the subjects take an action of 4 or less in Columns (3) and (4) of Table 2. The second way of identifying a behavioral response is to look at the share of subjects who transfer 40 EUs or more. A transfer of 40 leads to equal earnings for both subjects and is the amount subjects promise to transfer when an agreement is in place. Subjects who display a behavioral response may be inclined to decrease their transfer to be below 40. We analyze the probability the subjects take an action of 40 or more in Columns (5) and (6) of Table 2.

Analyzing these additional regression specifications, we see that the cost of control identified in the *Baseline Treatments* is associated with a large behavioral response as agents are more likely to transfer 4 units or less and are less likely to transfer 40 units or more when control is imposed than when it is not. Meanwhile, the coefficient for wanting the restriction in the *Mutual Minimum Treatment* is close to zero in most specifications and leads to a marginally significant positive increase in the likelihood of transferring 40 or more in the first half of the study. For the *Consent Treatment*, we find that the restriction *decreases* the frequency of transferring 4 units or less.

Table 3 replicates the results in Table 2 but only includes subjects who end up as agents in pairs who made an agreement to transfer 40. This allows us to look at the behavior of agents who we think are likely to transfer a large amount in the absence of control. It is worth noting that this means we are looking at subjects who endogenously

20

chose to ask for the agreement (about 80% of subjects in each round in all treatments) and whose randomly chosen partner also asked for the agreement in that round.

We find essentially the same overall pattern of results from the agreement subsample as in the full sample. The *Baseline Treatment* has a strong cost of control across all measures: average transfers are significantly lower with control, the frequency of transferring 4 or less increases by 7 percentage points, and the frequency of transferring 40 or more decreases by 9 percentage points. Control has the opposite effect in the *Consent Treatment*: average transfers increase and the frequency of transferring 4 or less decreases. The results for the *Mutual Minimum* treatment are much more mixed, with both significance and direction of effect varying across specifications.

Restricting the data to subjects who demand the agreement with high frequency in both treatments (columns (3), (6) and (9)) does not change our results, suggesting the difference in the impact of the restriction between treatments is not driven by a selection effect.[21] Overall, we find that imposing control is detrimental to the Principal in the Baseline and Unknown Agent treatments, has no effect in the Mutual Minimum treatment, and is beneficial in the Consent treatment.

---

[21] We also run a specification in the Consent treatment where we separately control for the agent wanting the restriction, the principal wanting the restriction and both wanting the restriction (full regression results are available from the authors on request). We find a marginally significant negative effect of the agent asking for the restriction if the other subject did not ($\beta = -5.57$, $p = 0.053$), suggesting that the agent may be punishing the principal, possibly interpreting the principal's failure to ask for a restriction as a signal that the principal was intending to make a low transfer if the principal had instead ended up as the agent. We find no significant effect of only the principal requesting the restriction ($\beta = -2.61$, $p = 0.279$), but a strong positive effect if the principal joined the agent in requesting the restriction ($\beta = 10.80$, $p = 0.006$). This strengthens our result that control is beneficial, as there is both a positive effect of controlling and a negative effect of failing to control an agent who wants the restriction. Additionally, this "punishment" effect by agents helps explain the low average transfer shown in the "No Minimum" bar in Panel A of Figure 1. If neither subject requests the restriction, the average transfer is 26.97, which is comparable to the other treatments.

Table 2: Transfers by Treatment (all pairs)

VARIABLES	Transfer		Transfer <= 4		Transfer >= 40	
	(1)	(2)	(3)	(4)	(5)	(6)
Other Restricted in Baseline	-2.237***	-2.056***	0.0727***	0.0827***	-0.0947***	-0.0969***
	(0.728)	(0.694)	(0.0168)	(0.0199)	(0.0211)	(0.0241)
Other Restricted in Mutual Minimum	0.248	2.744**	0.00327	-0.0325	-0.0339	0.0564*
	(1.462)	(1.276)	(0.0421)	(0.0537)	(0.0411)	(0.0280)
Other Restricted in Consent	4.141***	6.509***	-0.128***	-0.162***	0.0370	0.0761*
	(1.045)	(2.094)	(0.0260)	(0.0386)	(0.0290)	(0.0427)
Had Agreement	14.57***	16.72***	-0.360***	-0.378***	0.397***	0.471***
	(0.955)	(1.447)	(0.0246)	(0.0344)	(0.0251)	(0.0367)
Mutual Minimum	-1.221		0.0128		-0.0120	
	(1.917)		(0.0550)		(0.0483)	
Consent	-6.547***		0.185***		-0.126***	
	(1.652)		(0.0298)		(0.0357)	
First Treatment	5.631***		-0.169***		0.145***	
	(0.815)		(0.0204)		(0.0187)	
Constant	9.492***	11.63***	0.731***	0.616***	0.0877***	0.149***
	(1.010)	(1.067)	(0.0277)	(0.0265)	(0.0244)	(0.0274)
Observations	3,837	1,908	3,837	1,908	3,837	1,908
Number of Subjects	464	458	464	458	464	458
R-squared	0.187	0.197	0.218	0.203	0.247	0.266

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Standard errors clustered at the session level reported in parentheses. All specifications include subject fixed effects. The sample is restricted for the Unknown Agent and Mutual Minimum treatments to include only observations where the principal had the opportunity to restrict the agent. In columns (2), (4) and (6) the sample is further restricted to only the first treatment of a session, for these specifications treatment controls are dropped as they are collinear with the fixed effects. The dependent variable in columns (1) and (2) is the transfer of the agent, in columns (3) and (4) it is an dummy variable that equals one if the transfer was less than or equal to 4, in columns (5) and (6) it is a dummy variable that equals one if the transfer was greater than or equal to 40.

Table 3: Transfers by Treatment (pairs with an Agreement)

VARIABLES	Transfer			Transfer <= 4			Transfer >= 40		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Other Restricted in Baseline	-3.142***	-2.306**	-3.343**	0.0732***	0.0719***	0.0814***	-0.110***	-0.0910**	-0.111***
	(0.994)	(1.019)	(1.255)	(0.0220)	(0.0245)	(0.0270)	(0.0258)	(0.0347)	(0.0315)
Other Restricted in Mutual Minimum	-2.328*	1.541**	-2.318	0.0472	-0.0349	0.0466	-0.0940**	0.0174	-0.0900
	(1.299)	(0.689)	(1.577)	(0.0379)	(0.0715)	(0.0356)	(0.0435)	(0.0221)	(0.0563)
Other Restricted in Consent	3.868***	2.008	4.322***	-0.125***	-0.0493	-0.144***	0.0401	0.0290	0.0628
	(0.983)	(2.503)	(1.304)	(0.0331)	(0.0513)	(0.0307)	(0.0338)	(0.0909)	(0.0423)
Mutual Minimum	0.0190		-1.068	-0.00423		0.0211	0.0324		0.00538
	(1.843)		(1.889)	(0.0441)		(0.0441)	(0.0525)		(0.0581)
Consent	-8.772***		-8.992***	0.192***		0.174***	-0.175***		-0.196***
	(1.651)		(1.769)	(0.0360)		(0.0401)	(0.0410)		(0.0447)
First Treatment	5.718***		5.289***	-0.162***		-0.146***	0.170***		0.166***
	(1.005)		(1.035)	(0.0234)		(0.0229)	(0.0252)		(0.0255)
Constant	24.98***	29.11***	25.34***	0.364***	0.229***	0.353***	0.486***	0.632***	0.490***
	(0.978)	(0.377)	(0.993)	(0.0241)	(0.00933)	(0.0218)	(0.0242)	(0.0132)	(0.0259)
Observations	2,653	1,333	2,056	2,653	1,333	2,056	2,653	1,333	2,056
Number of Subjects	443	410	306	443	410	306	443	410	306
R-squared	0.065	0.007	0.065	0.081	0.013	0.074	0.088	0.014	0.091

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Standard errors clustered at the session level reported in parentheses. All specifications include subject fixed effects. The sample is restricted to observations where there was an agreement, and for the Unknown Agent and Mutual Minimum treatments only observations where the principal had the opportunity to restrict the agent are included. In columns (2), (5) and (8) the sample is further restricted to only the first treatment of a session. In columns (3), (6) and (9) only subjects who requested the agreement in at least 80% of periods for both treatments are included. The dependent variable in columns (1) to (3) is the transfer of the agent, in columns (4) to (6) it is an dummy variable that equals one if the transfer was less than or equal to 4, in columns (7) to (9) it is a dummy variable that equals one if the transfer was greater than or equal to 40.

5.2 Asking for Control Across Treatments

When principals in the *Baseline Treatments* have an agreement, they earn significantly less when they ask for the minimum than when they do not impose control on the agent. Do subjects in the *Baseline Treatment*s learn that imposing control is not optimal?

Figure 3 shows the percentage of subjects who ask for the minimum in the first 5 rounds and the last 5 rounds of each treatment conditional on having an agreement. We see that subjects are not learning that control decreases principal payoff in the *Baseline Treatments*. In fact, the rate of asking for a restriction increases from the first five rounds to the second five rounds in the *Baseline Treatments* (non-parametric signed-rank tests: subject-level, $p = 0.01$; session-level: $p = 0.05$).

Figure 3: Percent of Subjects who want the Minimum by Treatment (Agreement)
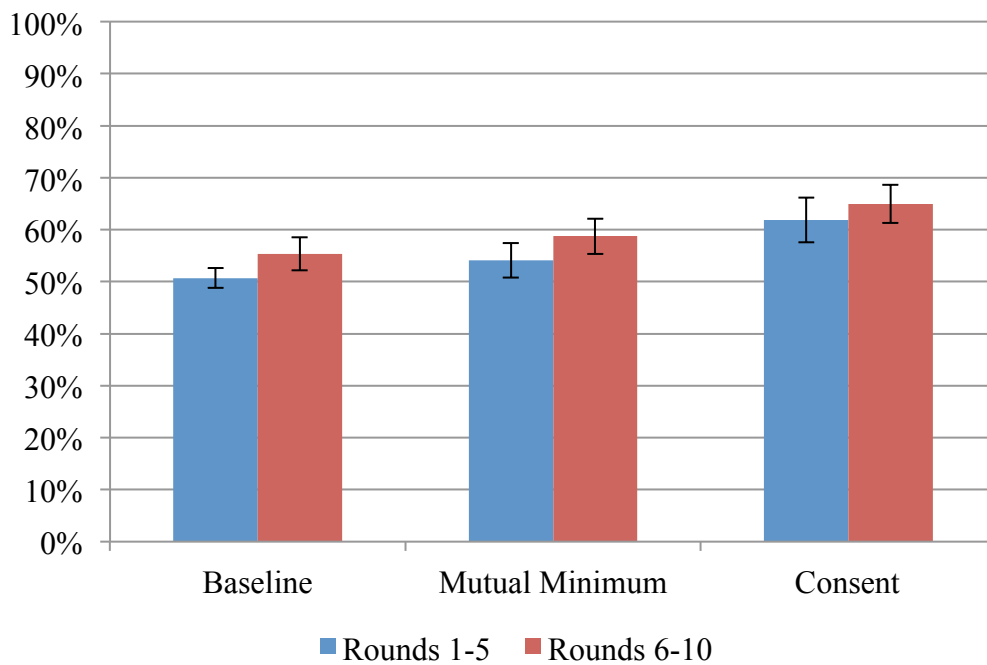

Rounds 1-5   Rounds 6-10

Figure 3 shows the percentage of subjects who want the minimum in each treatment for the first 5 rounds (blue bars) and the latter 5 rounds (red bars) of each treatment. Even though asking for the minimum leads to an average cost of control in the *Baseline Treatments* (i.e. principals are worse off when control is imposed) subjects become more likely to ask for it in the latter 5 rounds. Standard error bars (clustered by session) are shown around each mean.

5.3 Who Responds Negatively To Control

Because we observe all subjects playing the role of the principal in the *Baseline Treatment*,[23] we can use a subject's frequency of imposing control when a principal in the *Baseline Treatment* as a measure of their attitude towards control. This attitude towards control may affect how subjects respond to having control imposed upon them. For example, subjects who see control as a signal of distrust may be reluctant to impose control others and may react more negatively with a larger behavioral response when they are controlled. Conversely, subjects who see control as a reasonable precaution may prefer to restrict others and may not respond negatively to being controlled.

In the *Baseline Treatment*, the median subject imposed control in 2/3 of periods as a principal. To identify whether there is a different response for subjects with high and low usage of control, we estimate separate coefficients for *Other Subject Restricted* in each treatment for subjects above and below the median usage. The results are reported in Table 4.

We find results that are quite reasonable across the treatments. In the *Baseline Treatment*, we find a behavioral response only among agents who used control infrequently as principals. For this group, being restricted as an agent led to an estimated transfer decrease of 6 units, a 14 percentage point increase in the likelihood of making a transfer of 4 units or less, and a 20 percentage point decrease in the likelihood of transferring 40 units or more. By contrast, subjects in the *Baseline Treatment* who used control frequently as a principal had essentially zero response to the restriction as an agent. In the Mutual Minimum treatment, we find somewhat insignificant results for all subjects, although subjects who used control frequently have directionally more positive reactions to being controlled. In the *Consent Treatment*, the positive effect of the other subject asking for control was only observed among subjects who asked for the restriction frequently — for these subjects transfers increased by an estimated 7 units and the frequency of transfers of 4 or less decreased by 17 percentage points. Subjects who used the restriction infrequently have essentially a zero response to the restriction in the *Consent Treatment*. Overall, the pattern of results suggests that there is important

---

[23] Here we exclude data from the Unknown Agent Treatment and only look at the data from the Baseline Treatment, which everyone played either first or second in the session.

heterogeneity in how subjects perceived the restriction, with usage of the restriction as a principal being correlated with more positive reactions to the restriction as an agent.

One concern with interpreting the results in the baseline condition is that subjects switch between playing as a principal and as an agent over the course of the 10 rounds in the *Baseline Treatment*. We are tempted to interpret these results as supportive of a story in which subjects who are inclined to impose control as a principal respond less negatively (or more positively) to control as an agent. This interpretation would allow for a prescriptive suggestion that principals should feel comfortable controlling agents who themselves use control in settings where they are a principal (e.g. a CEO could feel comfortable controlling middle managers who are observed to control their front-line employees). However, an alternative explanation for this pattern of results is that subjects who respond negatively to control eventually learn to avoid using it. To show that the former interpretation is still valid, we conduct a similar analysis but divide subjects by whether they chose to impose control the *first time* they were a principal in the treatment being analyzed (rather than whether they used control more than 2/3 of the time in the *Baseline Treatment*) and then look only at behavior as an agent in all subsequent rounds of that treatment. We replicate the results above and so can assert that subjects who are observed to use control as a principal respond more favorably toward control when they are subsequently an agent.

Table 4: Effect of Subject Behavior as Principal in Baseline Treatment

VARIABLES	Average Transfer		
	Baseline	Mutual Minimum	Consent
	(1)	(2)	(3)
Other Restricted & Used	-5.917***	-2.697	0.796
Restriction < 2/3 in Baseline	(1.397)	(3.097)	(1.880)
Other Restricted & Used	0.787	-2.537	6.967**
Restriction >= 2/3 in Baseline	(1.280)	(2.109)	(2.113)
Constant	27.57***	27.43***	19.01***
	(0.524)	(1.310)	(0.908)
Observations	1,880	255	518
Number of Subjects	0.022	127	140
R-squared	0.022	0.013	0.026

VARIABLES	Transfer <=4		
	(4)	(5)	(6)
Other Restricted & Used	0.141***	0.0611	-0.0547
Restriction < 2/3 in Baseline	(0.0283)	(0.0760)	(0.0456)
Other Restricted & Used	-0.0120	0.0741	-0.172**
Restriction >= 2/3 in Baseline	(0.0302)	(0.0910)	(0.0531)
Constant	0.297***	0.264***	0.438***
	(0.0107)	(0.0419)	(0.0227)
Observations	1,880	255	518
Number of Subjects	436	127	140
R-squared	0.028	0.010	0.025

VARIABLES	Transfer >=40		
	(7)	(8)	(9)
Other Restricted & Used	-0.203***	-0.124	-0.00531
Restriction < 2/3 in Baseline	(0.0387)	(0.0993)	(0.0549)
Other Restricted & Used	-0.00581	-0.117	0.101
Restriction >= 2/3 in Baseline	(0.0310)	(0.0768)	(0.0574)
Constant	0.572***	0.640***	0.361***
	(0.0124)	(0.0408)	(0.0281)
Observations	1,880	255	518
Number of Subjects	436	127	140
R-squared	0.047	0.032	0.008

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Standard errors clustered at the session level reported in parentheses. All specifications include subject fixed effects. The sample is restricted to observations where there was an agreement, and for the Unknown Agent and Mutual Minimum treatments only observations where the principal had the opportunity to restrict the agent are included. The dependent variable in panel A is the transfer of the agent, in panel B it is an dummy variable that equals one if the transfer was less than or equal to 4, in panel C it is a dummy variable that equals one if the transfer was greater than or equal to 40.

VI.   Additional Experiments

6.1 Control when there is no opportunity for an Agreement

In the introduction and experimental design sections, we explained that we introduced the agreement in order to raise average actions in the absence of control so that we would have a better opportunity to observe a behavioral response and an average cost of control. We now show results demonstrating that giving subjects the opportunity to make the agreement had the intended effect. To do this test, we ran additional sessions with a *No Agreement Allowed Treatment* in which subjects did not have the opportunity to make an agreement. The No Agreement Allowed treatment is the same as the Baseline treatment, except that subjects were not given the opportunity to make an agreement. We conducted an additional 5 sessions, with a total of 94 subjects, in which we ran the *No Agreement Allowed Treatment* followed by the *Baseline Treatment*. We had subjects always play the *Baseline Treatment* second so that subjects would not have been previously exposed to the agreement when playing in the *No Agreement Allowed Treatment*.

Figure 4 shows the average transfer with and without a restriction in the *No Agreement Allowed Treatment* and compares it to behavior in the *Baseline Treatment* (including *Baseline* data from the main experiment and these new sessions).[24] There is a significant increase in transfers associated with giving subjects the opportunity to make the agreement. In addition, the opportunity to make an agreement allows us to more cleanly identify the average cost of control that arises when the principal imposes control in the *Baseline*.

In the *No Agreement Allowed Treatment*, transfers decrease slightly from 16.41 when control is not imposed to 14.91 when control is imposed, and the difference is not significant ($p > 0.20$ for both subject-level and session-level permutation tests). Similarly, in the same data, the fraction of subjects transferring 4 or less increases from 30% to 36% in response to control, but the difference is only marginally significant (subject-level: $p =$

---

[24] We obtain essentially the same results whether or not we include *Unknown Agent Treatment* data in with *Baseline Treatment* data and whether or not we exclude any observations that took place in the second half of the experiment (i.e. comparing the *No Agreement Allowed Treatment*, which was always played first in a session, to data from the *Baseline Treatment* when it was played first in a session).

0.06, session-level: $p > 0.20$). It is worth pointing at that these small and insignificant differences contrast with the results in Falk and Kosfeld (2006), which has a design very similar to our *No Agreement Allowed* except that in Falk and Kosfeld (2006), subjects play a one-shot game via the strategy method and control imposes a minimum transfer of 5 rather than 4. They find that imposing a minimum transfer of 5 leads to a decrease in average transfer from 25.1 to 12.2, and an increase in the fraction of subjects transferring 5 or less from approximately 20% to approximately 50%.

In the absence of control, transfers are much higher in Falk and Kosfeld's data than in ours, suggesting that there may be a difference in the willingness to transfer in the absence of control for their subject pool as compared to our subject pool. We are only able to get subjects to transfer an average of 25.1 units in the absence of control when subjects establish an agreement in our data (and only there do we see an average cost of control). Taken together, we only expect an average cost of control in settings where there is a strong willingness to transfer units in the absence of control, either by default as in Falk and Kosfeld (2006) data, or due to a specific agreement as in our data.

Figure 4: Transfers and Effect of Control by whether Agreements are Allowed
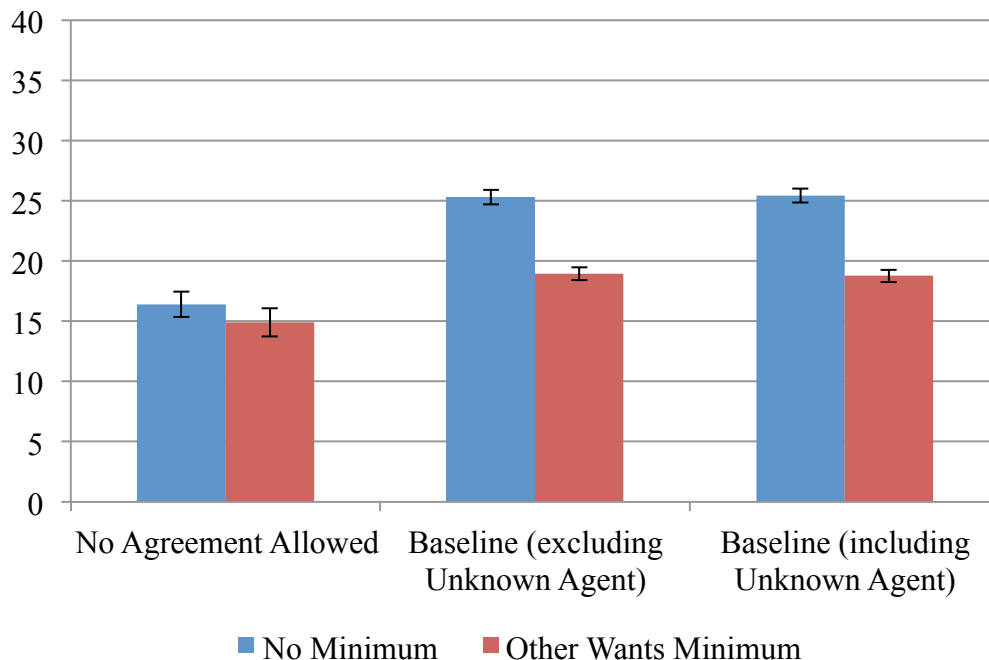

Figure 4 shows the average transfer as a function of whether the other player wants the minimum (red bars) or does not want the minimum (blue bars) when agreements are not allowed (left pair of bars) and in the Baseline [and Unknown Agent treatments] from the original experiment and the additional sessions. Standard error bars (clustered by session) are shown around each mean.

6.2 Restrictions with a Higher Minimum Transfer

In the *Baseline Treatment* of our main experiment, we observe a behavioral response among agents who are controlled. We find that agents are much less likely to transfer 40 units and much more likely to transfer the minimum of 4 units when they are controlled. This leads to an average cost of control, since the behavioral response is large relative to the benefit from raising transfers that were below 4 units to the minimum of 4. Principals who know that control will lead to a large behavioral response may decide to avoid using control. Alternatively, however, such principals may decide to invest in more powerful control, for example by finding a way to require a greater perfunctory performance (e.g. a better monitoring technology).

To test the impact of more powerful control, we ran 5 additional sessions with 94 subjects of the *Baseline* and *Consent* treatments in which control required a minimum transfer of 10 units rather than 4 units. Figure 5 shows the average transfer in each treatment as a function of whether an agreement was in place.

Figure 5: Effect of Control on Transfers by Treatment when Minimum is 10

Panel A: Average Transfer with an Agreement
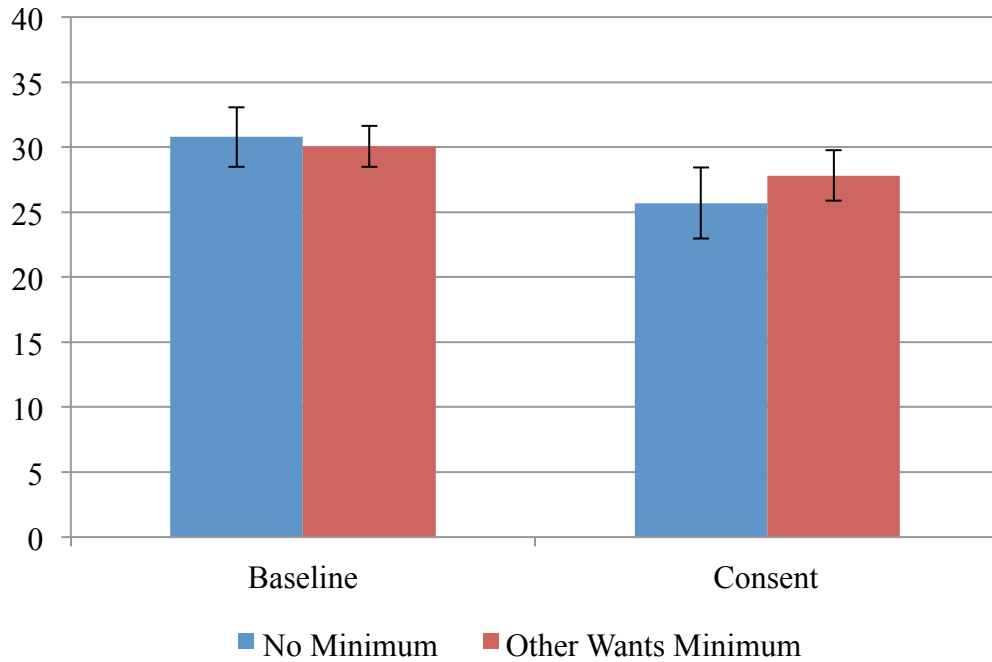
Panel B: Average Transfer without an Agreement
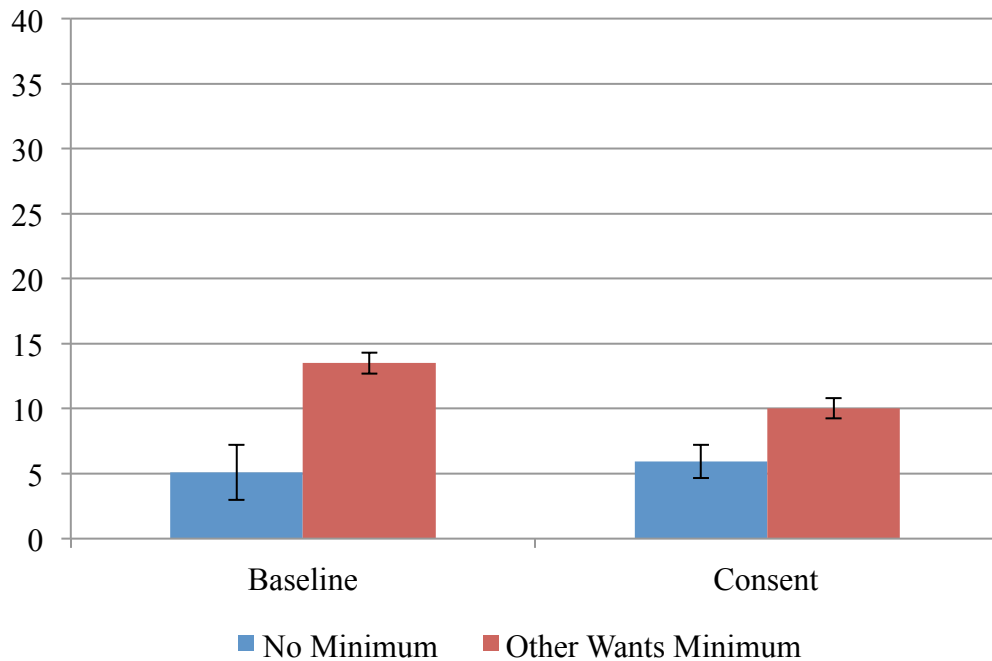

Figure 5 shows the average transfer as a function of whether the other player wants the minimum (red bars) or does not want the minimum (blue bars) when the minimum is 10 EUs. Panel A shows transfers when subjects have an agreement to play 40 EUs. Panel B shows transfers when no agreement is in place. Standard error bars (clustered by session) are shown around each mean.

When there is an agreement, we no longer find an average cost of control in the *Baseline Treatment*. When control forces a minimum transfer of 10, the average transfer decreases slightly from 29.5 without control to 28.2 with control, but the difference is not significant ($p > 0.20$ for both non-parametric permutation tests). As expected, there is still evidence of a behavioral response: the fraction of transfers of 10 or less increases from 29% to 40% in the presence of control (subject-level: $p = 0.05$; session-level: $p > 0.20$) while the fraction of transfers greater than or equal to 40 decreases from 57% to 46% (subject-level: $p = 0.02$; session-level: $p = 0.06$). In this case, however, the benefit of the increase due to the binding minimum counteracts the behavioral response. In the *Consent Treatment*, we find that the minimum is somewhat beneficial for the Principal, directionally increasing average transfers from 25.0 to 26.0 (subject-level: $p = 0.10$; session-level: $p > 0.20$). There is a directional decrease in the fraction of subjects transferring 10 or less with a restriction from 42% to 37% ($p > 0.20$ for both), and a directional increase in the number of subjects transferring 40 or more from 47% to 52% (subject-level: $p = 0.06$; session-level: $p > 0.20$).

As in our main experiment, we find that when there is no agreement, imposing the powerful restriction (or asking for the minimum in the Consent treatment) leads to higher average transfers from the agent.

Overall, these results suggest that the cost of control should only be a primary concern when the principal's ability to monitor and control her agent is relatively limited. As expected, we fail to find an average cost of control when control is powerful enough to compensate for any behavioral response.

VII.   Discussion

In this paper, we investigate the conditions under which an agent responds to control by withholding effort, a behavioral response that can lead a principal to be made worse off by imposing control. As hypothesized, the procedure by which control is imposed affects how subjects respond. We find a large behavioral response when control is imposed unilaterally and has an asymmetric effect on the agent as in the *Baseline Treatments*. In this *Baseline*, the behavioral response is so large that we observe an average cost of control in which principals are worse off when they impose control than

when they leave agents unconstrained. However, when control is imposed symmetrically, the behavioral response is mitigated; and when control is imposed bilaterally, there is no behavioral response and principals are strictly better off from asking for control. Our results are consistent with theories of procedural fairness that suggest control will be less offensive to the agent if it is applied symmetrically and if the agent had a voice in establishing the control.

Beyond demonstrating the importance of procedural fairness and voice in the efficiency of control, the results in this paper offer three additional insights. First, our results can help guide parties deciding whether or not to impose control. Second, our results speak to a class of models that aim to explain why a cost of control might arise. Third, our study can help reconcile results from other experimental papers that sometimes fail to replicate the average cost of control result from Falk and Kosfeld (2006). We address each of these three in turn.

First, from the perspective of parties deciding whether or not to impose control, our results highlight a number of factors that mitigate the risk associated with control. First, we do not observe a statistically significant behavioral response (and so do not observe an average cost of control) when control has symmetric impact (i.e. it affects both parties rather than just one) or when multiple parties have a voice in control. Second, we do not observe a behavioral response, nor an average cost of control, among agents who previously imposed control as principals. Third, we do not observe an average cost of control when the average action in the absence of control is too low (i.e. we only observe a cost of control when we allow parties to make an agreement that raises average actions). Fourth, we do not observe an average cost of control if control is strong enough (i.e. an average cost of control only arose with a minimum of 4 not with a minimum of 10).

To summarize, agents in our experiment only display the behavioral response when both: (1) control is imposed unilaterally and has an asymmetric effect on the agent and (2) the agent is does not use control himself when acting as a principal. Principals are only worse off from imposing control only when those two conditions are met as well as: (3) average transfers in the absence of control are high and (4) control is weak in that it cannot induce significant effort from agents. Our results suggest that principals and firms

should be most concerned about an average cost of control when they have otherwise high performing agents, when their monitoring and control technology is weak, and when their relationship with the agent is highly asymmetric (e.g. in an employment context or a supply chain setting with a dominant party). Control may be less problematic when both parties are on a more even footing (e.g. a joint venture). Firms may be able to diminish the cost if they can also credibly restrict their own bad actions or if they can allow agents to have a voice in the imposition of control.

Second, while our experiment was not designed to test any of these models, our results speak to models aiming to explain the cost of control. Some of the proposed models for the cost of control cannot explain our results given our setting. The model in Sliwka (2007), in which the principal's use of control signals a low norm of behavior, is not consistent with our setup. First, the model depends on the principal having greater information about the norm than the agent, which is not true in our setting. Additionally, the model cannot explain why an agent consenting to control would affect behavior, sine it adds no information about the norm. Von Siemens' (2013) model of intentions-based reciprocity also does not seem to apply to our setting. We do not generally model an individual's actions towards himself as representing kindness or unkindness, and hence the *Mutual Minimum* treatment should not differ from the *Baseline*.

However, the Ellingsen and Johannesson (2008) model of esteem is consistent with our data. In this model individuals care about others' beliefs of their prosociality ("esteem") and care more about esteem from prosocial individuals. Hence a prosocial agent may choose high effort to signal his type to a prosocial principal. A key assumption of that model is that an individual's beliefs are correlated with their type, and hence controlling is a signal of selfishness. In the *Mututal Minimum* and *Consent Treatments*, however, a controlling subject who is selfish imposes a cost on himself if he is chosen as the agent, while a prosocial subject is not harmed by controlling himself, since the minimum would not bind on him. This difference can disrupt the signaling equilibrium (depending on parameter values), which would explain the reduced cost of control. We show these results in Appendix C.

Third, our results help to reconcile a disagreement in the literature on the robustness of the cost of control results originally presented in Falk and Kosfeld (2006).

In particular, we show two factors that may make it hard to replicate their "average cost of control" results. First, we find that if the average transfer in the absence of control is not large enough, it is much more difficult to identify a behavioral response, let alone an average cost of control. To ensure that average actions are high enough, we implement the agreement protocol of Kessler and Leider (2012), which helps raise average effort high enough to observe a response. When we do not include the agreement phase of the game, we are not able to observe an average cost of control in the *Baseline Treatment*. Second, we find that if the minimum is too high, then the beneficial effect of the minimum raising low actions may swamp any behavioral response so that there is no longer an average cost of control.

By focusing on procedural fairness and investigating how control is imposed, we have shown one way in which the behavioral response to control can be mitigated and eliminated. There are certainly other factors that will influence the behavioral response that arises due to control, which future work can and should address. More generally, demonstrations that control *can* undermine effort (or that incentives can undermine intrinsic motivation) are important first steps in improving our models by including behavioral phenomenon. The next step in developing these models is to understand when such perverse effects of control and incentives will arise. Identifying the boundaries of the cost of control, as well as other behavioral phenomenon,[25] will help us to write better models of behavior and provide better guidance to practitioners deciding how to motivate workers and contract with counterparties.

---

[25] We see identifying boundaries on behavioral phenomena as a generally useful activity that pushes the field toward richer theories that incorporate these phenomena. Recent papers such as Ariely, Bracha and Meier (2009) pursue a similar approach.

VIII. References

Alexander, Sheldon, and Marian Ruderman, (1987). "The role of procedural and distributive justice in organizational behavior," *Social Justice Research,*1(2), 177-198.

Allen, Franklin and Douglas Gale, (1992). "Measurement Distortions and Missing Contingencies in Optimal Contracts." *Economic Theory*. 2(1), 1-26.

Ariely, Dan, Anat Bracha, and Stephan Meier, (2009). "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially." *American Economic Review*. 99(1), 544–555.

Barkema, Harry G., (1995). "Do Top Managers Work Harder when They Are Monitored?" *Kyklos*, 48(1), 19–42.

Benz, Matthias, and Bruno S. Frey, (2004). "Being independent raises happiness at work." *Swedish Economic Policy Review*, 11(2), 95-134.

Benz, Matthias, and Alois Stutzer (2003). "Do workers enjoy procedural utility?." *Applied Economics Quarterly*, 49(2), 149-172.

Bernheim, B. Douglas and Michael Whinston, (1998). "Incomplete Contracts and Strategic Ambiguity." *American Economic Review*. 88 (4), 902-932.

Bies, Robert J., and Debra L. Shapiro, (1988). "Voice and justification: Their influence on procedural fairness judgments." *Academy of Management Journal*, 31(3), 676-685.

Bolton, Gary E., Jordi Brandts, and Axel Ockenfels, (2005). "Fair procedures: Evidence from games involving lotteries." *The Economic Journal*, 115(506), 1054-1076.

Bowles, Samuel and Sandra Polania-Reyes (2012). "Economic Incentives and Social Preferences: Substitutes and Complements." *Journal of Economic Literature*. 50(2), 368-425.

Cohen-Charash, Yochi and Paul Spector, (2001). "The Role of Justice in Organizations: A Meta-Analysis," *Organizational Behavior and Human Decision Processess*, 86(2), 278-321.

Charness, Gary and Martin Dufwenberg, (2006). "Promises and Partnerships." *Econometrica*, 74(6), 1579-1601.

Coase, Ronald (1937). "The Nature of the Firm." *Economica*. 4, 386-405.

Deci, Edward, (1971) "Effects of externally mediated rewards on intrinsic motivation," *Journal of Personality and Social Psychology*, 18, 105-115.

Deci, Edward, (1975) *Intrinsic Motivation*. Plenum Press, New York and London.

Deci, Edward, Richard Koestner, and Richard Ryan, (1999). "A Meta-analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation." *Psychological Bulletin*, 125 (6), 627-668.

De Cremer, David, and Daan van Knippenberg, (2002). " How Do Leaders Promote Cooperation? The Effects of Charisma and Procedural Fairness." *Journal of Applied Psychology,* 87(5), 858-866.

De Cremer, David, and Daan van Knippenberg, (2003). "Cooperation with leaders in social dilemmas: On the effects of procedural fairness and outcome favorability in structural cooperation." *Organizational Behavior and Human Decision Processes,* 91(1), 1-11.

Dufwenberg, Martin, Simon Gächter, and Heike Hennig-Schmidt, (2011). "The framing of games and the psychology of play." *Games and Economic Behavior*, 73 (2), 459–478.

Ellingsen, Tore and Magnus Johannesson. (2004). "Promises, Threats and Fairness." *The Economic Journal*, 114(495), 397-420.

Ellingsen, Tore and Magnus Johannesson, (2008). "Price and Prejudice: The Human Side of Incentive Theory." *The American Economic Review*, 98(3), 990-1008.

Falk, Armin and Michael Kosfeld, (2006). "The Hidden Cost of Control." *American Economic Review*, 96(5), 1611-1630.

Fischbacher, Urs, (2007). "z-Tree: Zurich Toolbox for Ready-made Economic experiments." *Experimental Economics*, 10(2), 171-178.

Frey, Bruno S. (1993). "Does Monitoring Increase Work Effort? The Rivalry with Trust and Loyalty." *Economic Inquiry*, 31(4): 663–70.

Frey, Bruno S. (1994) "How Intrinsic Motivation is Crowded in and Out," *Rationality and Society*, 6(3), 334-352.

Frey, Bruno S., Matthias Benz, and Alois Stutzer, (2004). "Introducing procedural utility: Not only what, but also how matters." *Journal of Institutional and Theoretical Economics*, 160(3), 377-401.

Frey, Bruno, S., Felix Oberholzer-Gee and Reiner Eichenberger, (1996) "The old lady visits your backyard: a tale of morals and markets," *Journal of Political Economy*, 104(6), 1297-1313.

Frey Bruno S. and Felix Oberholzer-Gee, (1997). "The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out," *American Economic Review*, 87(4), 746-755.

Frohlich, Norman and Joe A. Oppenheimer, (1990). "Choosing Justice in Experimental Democracies with Production," *American Political Science Review*, 84(2), 461-477.

Gneezy, Uri and Aldo Rusticini, (2000a). "Pay Enough or Don't Pay at All." *Quarterly Journal of Economics*, 115(2), 791-810.

Gneezy, Uri and Aldo Rusticini, (2000b). "A Fine is a Price." *Journal of Legal Studies*, 29(1), 1-18.

Gneezy, Uri, Stephan Meier and Pedro Rey-Biel, (2011). "When and Why Incentives (Don't) Work to Modify Behavior." *The Journal of Economic Perspectives*, 25(4), 191-209.

Greenberg, Jerald, (1990). "Employee theft as a reaction to underpayment inequity: The hidden cost of pay cuts." *Journal of Applied Psychology,* 75(5), 561-570.

Hagemann, Petra, (2007). "What's in a Frame? Comment on: The Hidden Costs of Control," *Unpublished manuscript*, University of Cologne.

Holmstrom, Bengt and Paul Milgrom, (1991). "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, & Organization*, 7, 24-52.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler, (1986). "Fairness and the assumptions of economics." *Journal of Business*, 59(4), S285-S300.

Kessler, Judd B. and Stephen Leider (2012). "Norms and Contracting," *Management Science*, 58(1), 62–77.

Krawczyk, Michał, (2010). "A glimpse through the veil of ignorance: Equality of opportunity and support for redistribution." *Journal of Public Economics*, 94(1), 131-141.

Kruglansky, Arie, Sarah Alon, and Tirtzah Lewis, (1972) "Retrospective misattribution and task enjoyment," *Journal of Experimental Social Psychology*, 8, 493-501.

Lemieux, W. Thomas, Bentley MacLeod, and Daniel Parent, (2009), "Performance Pay and Wage Inequality" *The Quarterly Journal of Economics*, 124(1), 1-49.

Lepper, Mark R. and Greene, David, (1978), *The Hidden Costs of Reward: New Perspectives in the Psychology of Human Motivation*. Lawrence Elbaum Associates, Publishers; John Wiley and Sons.

Leventhal, Gerald S., (1980). "What Should be Done with Equity Theory? New Approaches to the Study of Fairness is Social Relationships." In K.J. Gergen, M.S. Greenberg and R. H. Willis (eds), *Social Exchange: Advances in Theory and Research* (pp. 27-55). New York: Plenum.

Lind, E. Allan, and Tom R. Tyler, (1988). *The social psychology of procedural justice*. Springer.

Lind, E. Allan, Ruth Kanfer, and P. Christopher Earley, (1990). "Voice, control, and procedural justice: Instrumental and noninstrumental concerns in fairness judgments." *Journal of Personality and Social Psychology,* 59(5), 952-959.

Lind, E. Allan, Carol T. Kulik, Maureen Ambrose and Maria V. de Vera Park, (1993). "Individual and Corporate Dispute Resolution: Using Procedural Fairness as a Decision Heuristic," *Administrative Science Quarterly*, 38, 224-251.

Ploner, Matteo, Katrin Schmelz, and Anthony Ziegelmeyer, (2012). "Hidden Costs of Control: Four Repetitions and an Extension," *Experimental Economics,* 15(2), 323-340.

Rawls, John, (1971). *A Theory of Justice*. Cambridge, MA: Harvard Press.

Schildberg-Hörisch, Hannah, (2010). "Is the veil of ignorance only a concept about risk? An experiment." Journal of Public Economics, 94(11), 1062-1066.

Schnedler, Wendelin and Radovan Vadovic, (2011). "Legitimacy of Control," *Journal of Economics and Management Strategy*, 20(4), 985-1009.

Simon, Herbert (1981). *The Sciences of the Artificial*. MIT Press, Cambridge, MA

Sliwka, Dirk, (2007). "Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes." *American Economic Review*, 97 (3), 999-1012.

Spier, Kathryn E. (1992) "Incomplete Contracts in a Model with Adverse Selection and Exogenous Costs of Enforcement." *RAND Journal of Economics*. 23, 432-443.

Sutter, Matthias, and Hannelore Weck-Hannemann, (2003). "Taxation and the Veil of Ignorance–A real effort experiment on the Laffer curve." *Public Choice,*115(1), 217-240.

Thibaut, John, and Laurens Walker (1975). *Procedural justice: A Psychological Analysis*. Hillsdale, New Jersey.Thibaut, John, and Laurens Walkers (1978). "A Theory of Procedure." *California Law Review*, 66(3), 541-566.

Titmuss, Richard M., (1970) *The Gift Relationship*. Allen and Unwin, London.

Vanberg, Cristoph, (2008). "Why do people keep their promises? An experimental test of two explanations." *Econometrica*, 76(6), 1467-1480.

van Wanrooy, Brigid, Helen Bewley, Alex Bryson, John Forth, Stephanie Freeth, Lucy Stokes and Stephen Wood (2013). "The 2011 Workplace Employment Relations Study First Finding" *Workplace Employment Relations Study Report*.

Vernon, Guy (2003). "Comparative work organization, managerial hierarchies and occupational classification", *Employee Relations*, 25(4), 389 – 404.

von Siemens, Ferdinand A., (2013). "Intention-based Reciprocity and the Hidden Costs of Control," *Journal of Economic Behavior and Organizations*, 92, 55 – 65.

Williamson, Oliver (1975). *Markets and Hierarchies: Analysis and Antitrust Implications*. Free Press, New York.

Williamson, Oliver (1985). *The Economic Institutions of Capitalism*. Free Press: New York.

Appendix A: Comparing Baseline Treatment with Unknown Agent

			No Restriction			Restriction		
	Want Agreement	Want Restriction	Transfer	% At Min	% 40+	Transfer	% At Min	% 40+
Baseline	0.840	0.592	25.27	0.323	0.514	18.93	0.511	0.306
Unknown Agent	0.845	0.553	26.62	0.386	0.503	17.24	0.587	0.291
p-value different	p=0.744	p=0.291	p=0.586	p=0.123	p=0.817	p=0.409	p=0.179	p=0.727
Pooled	0.841	0.583	25.46	0.332	0.512	18.73	0.520	0.305

Notes: p-value comes from a regression clustered by the 25 sessions run in the main experiment

Appendix B: TABLES 2 & 3 WITH UNKNOWN AGENT SPLIT OUT

VARIABLES	Transfer		Transfer <= 4		Transfer >= 40	
	(1)	(2)	(3)	(4)	(5)	(6)
Other Restricted in Baseline	-1.780**	-1.910**	0.0602***	0.0814***	-0.0849***	-0.100***
	(0.800)	(0.711)	(0.0175)	(0.0217)	(0.0228)	(0.0258)
Other Restricted in Unknown Agent	-5.032***	-3.580	0.150***	0.0969***	-0.155***	-0.0636
	(1.211)	(2.184)	(0.0345)	(0.0294)	(0.0325)	(0.0598)
Other Restricted in Mutual Minimum	0.253	2.742**	0.00315	-0.0325	-0.0338	0.0564*
	(1.469)	(1.276)	(0.0422)	(0.0537)	(0.0412)	(0.0280)
Other Restricted in Consent	4.146***	6.507***	-0.128***	-0.162***	0.0371	0.0761*
	(1.044)	(2.095)	(0.0260)	(0.0387)	(0.0289)	(0.0427)
Had Agreement	14.57***	16.71***	-0.360***	-0.378***	0.397***	0.472***
	(0.959)	(1.451)	(0.0245)	(0.0345)	(0.0252)	(0.0368)
Unknown Agent	1.493		-0.0327		0.0235	
	(1.688)		(0.0368)		(0.0434)	
Mutual Minimum	-0.929		0.00485		-0.00573	
	(1.961)		(0.0556)		(0.0490)	
Consent	-6.285***		0.178***		-0.121***	
	(1.644)		(0.0294)		(0.0355)	
First Treatment	5.764***		-0.173***		0.148***	
	(0.818)		(0.0205)		(0.0192)	
Constant	9.177***	11.69***	0.738***	0.616***	0.0817***	0.147***
	(1.092)	(1.086)	(0.0285)	(0.0265)	(0.0269)	(0.0275)
Observations	3,837	1,908	3,837	1,908	3,837	1,908
Number of Subjects	464	458	464	458	464	458
R-squared	0.188	0.198	0.219	0.203	0.247	0.266

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Standard errors clustered at the session level reported in parentheses. All specifications include subject fixed effects. The sample is restricted for the Unknown Agent and Mutual Minimum treatments to include only observations where the principal had the opportunity to restrict the agent. In columns (2), (4) and (6) the sample is further restricted to only the first treatment of a session, for these specifications treatment controls are dropped as they are collinear with the fixed effects. The dependent variable in columns (1) and (2) is the transfer of the agent, in columns (3) and (4) it is an dummy variable that equals one if the transfer was less than or equal to 4, in columns (5) and (6) it is a dummy variable that equals one if the transfer was greater than or equal to 40.

VARIABLES	Transfer			Transfer <= 4			Transfer >= 40		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Other Restricted in Baseline	-2.901**	-2.245**	-3.152**	0.0641***	0.0761***	0.0721**	-0.101***	-0.0934**	-0.101***
	(1.056)	(1.037)	(1.345)	(0.0221)	(0.0260)	(0.0273)	(0.0270)	(0.0375)	(0.0323)
Other Restricted in Unknown Agent	-4.763*	-2.979	-4.628*	0.132**	0.0257	0.139**	-0.174***	-0.0643	-0.179***
	(2.588)	(4.372)	(2.469)	(0.0518)	(0.0609)	(0.0571)	(0.0467)	(0.0619)	(0.0585)
Other Restricted in Mutual Minimum	-2.334*	1.541**	-2.321	0.0474	-0.0349	0.0467	-0.0943**	0.0174	-0.0902
	(1.303)	(0.690)	(1.580)	(0.0380)	(0.0715)	(0.0358)	(0.0436)	(0.0221)	(0.0565)
Other Restricted in Consent	3.874***	2.008	4.330***	-0.125***	-0.0493	-0.145***	0.0404	0.0290	0.0631
	(0.984)	(2.504)	(1.305)	(0.0331)	(0.0513)	(0.0307)	(0.0337)	(0.0909)	(0.0423)
Unknown Agent	2.724*		3.001*	-0.0425		-0.0532	0.0644		0.0703
	(1.532)		(1.662)	(0.0298)		(0.0328)	(0.0531)		(0.0518)
Mutual Minimum	0.160		-0.957	-0.00953		0.0158	0.0380		0.0116
	(1.897)		(1.947)	(0.0447)		(0.0444)	(0.0536)		(0.0590)
Consent	-8.639***		-8.892***	0.187***		0.169***	-0.169***		-0.190***
	(1.657)		(1.753)	(0.0362)		(0.0398)	(0.0415)		(0.0448)
First Treatment	5.779***		5.353***	-0.164***		-0.148***	0.173***		0.169***
	(1.003)		(1.024)	(0.0237)		(0.0232)	(0.0252)		(0.0250)
Constant	24.66***	29.14***	24.99***	0.371***	0.231***	0.361***	0.477***	0.631***	0.480***
	(1.035)	(0.417)	(1.037)	(0.0252)	(0.00925)	(0.0227)	(0.0262)	(0.0127)	(0.0266)
	-2.901**	-2.245**	-3.152**	0.0641***	0.0761***	0.0721**	-0.101***	-0.0934**	-0.101***
	(1.056)	(1.037)	(1.345)	(0.0221)	(0.0260)	(0.0273)	(0.0270)	(0.0375)	(0.0323)
Observations									
Number of Subjects	-4.763*	-2.979	-4.628*	0.132**	0.0257	0.139**	-0.174***	-0.0643	-0.179***
R-squared	(2.588)	(4.372)	(2.469)	(0.0518)	(0.0609)	(0.0571)	(0.0467)	(0.0619)	(0.0585)

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Standard errors clustered at the session level reported in parentheses. All specifications include subject fixed effects. The sample is restricted to observations where there was an agreement, and for the Unknown Agent and Mutual Minimum treatments only observations where the principal had the opportunity to restrict the agent are included. In columns (2), (5) and (8) the sample is further restricted to only the first treatment of a session. In columns (3), (6) and (9) only subjects who requested the agreement in at least 80% of periods for both treatments are included. The dependent variable in columns (1) to (3) is the transfer of the agent, in columns (4) to (6) it is an dummy variable that equals one if the transfer was less than or equal to 4, in columns (7) to (9) it is a dummy variable that equals one if the transfer was greater than or equal to 40.

Appendix C: Our Results and the Ellingsen and Johannesson (2008) Model of Esteem

Individuals have three components to their utility: material payoffs, inequity aversion (captured by parameter θ), and esteem (denoted by v). For simplicity, we assume there are two type: $\theta_L = 0$ and $\theta_H > 0$. The esteem that player i would feel from player j believing he was type θ_i is $v = g_i * s_j * \theta_i$, where g_i denotes player i's general concern for esteem, and s_j denotes the weight placed on the opinion of someone of player j's type. Individuals are assumed to care more about the opinions of high types than low types ($s_H > s_L$).

The principal's utility is then $U_P = 2x - |2x - (120 - x)|\theta_P + v_P$. The agent's utility is similarly $U_A = 120 - x - |(120 - x) - 2x|\theta_A + v_A$. Additionally, low types have belief p_L that their counterpart is a high type, while high types have believe $p_H > p_L$. Following Ellingsen and Johannesson (2008) we consider for the *Baseline Treatment* a simple separating equilibrium that follows the basic cost of control result,[26] namely low types control when possible and choose the minimum action as agents, while high types trust, choose 40 if trusted and choose 20 if controlled. If principals who control are systematically more selfish than principals who trust, then a high-type agent will have a stronger incentive to choose a high action in order to signal their type (since esteem from a high type is more valuable than esteem from a low type). In a pooling equilibrium esteem will be the same in the controlling and trusting cases, so we would expect the same action (and therefore no cost of control).

For the equilibrium to hold we need the following conditions:

a) Low type agent controlled does not pool with high types: $116 \geq 100 + g_L s_L \theta_H$ or $16 \geq g_L s_L \theta_H$

b) Low type agent trusted does not pool with high types: $120 \geq 80 + g_L s_H \theta_H$ or $40 \geq g_L s_H \theta_H$

c) High type agent controlled does 20 instead of x>20: $100 - 60\theta_H \geq 120 - x - |120-3x|\theta_H$ or $1/3 \geq \theta_H$

High type agent controlled does 20 instead of 4: $100 - 60\theta_H + g_H s_L \theta_H \geq 116 - 108\theta_H$ or $g_H s_L \theta_H \geq 16 - 48\theta_H$

---

[26] For simplicity we set aside here the role of agreements. Since almost everyone makes the agreement it should not have a signaling effect. We can also adapt the Elligsen and Johannesson (2008) model to include norm sensitivity instead of inequity aversion to explicitly allow for the effect of an agreement — results are similar.

d) High type agent trusted chooses to play 40: $80 + g_H s_H \theta_H \geq 120 - 120\theta_H$ or $g_H s_H \theta_H \geq 40 - 120\theta_H$

e) Low type controls: $p_L(40) + (1-p_L)(8) \geq p_L(80 + g_L s_H \theta_H) + (1-p_L)(g_L s_L \theta_H)$ or $p_L \leq (8 - g_L s_L \theta_H)/(48 + g_L \theta_H[s_H - s_L])$

f) High type trusts: $p_H(80 + g_H s_H \theta_H) + (1-p_H)(-120\theta_H + g_H s_L \theta_H) \geq p_H(40-60\theta_H) + (1-p_H)(8-108\theta_H)$ or $p_H \geq (8 + 12\theta_H - g_H s_L \theta_H)/(48 + 72\theta_H + g_H \theta_H[s_H - s_L])$

Hence we need inequality aversion to be not too strong, low types to be sufficiently pessimistic and not too esteem concerned, and high types to be sufficiently optimistic and sufficiently esteem concerned.

For the *Mutual Minimum* treatment, conditions (a) to (d) remain the same. However condition (e) is now $\frac{1}{2}[p_L(40) + (1 - p_L)(8)] + \frac{1}{2}[116] \geq \frac{1}{2}[p_L(80 + g_L s_H \theta_H) + (1-p_L)(g_L s_L \theta_H)] + \frac{1}{2}[120]$ or $p_L \leq (4 - g_L s_L \theta_H)/(48 + g_L \theta_H[s_H - s_L])$, which is a stricter criterion. Similarly, condition (f) is now $\frac{1}{2}[p_H(80 + g_H s_H \theta_H) + (1 - p_H)(-120\theta_H + g_H s_L \theta_H)] + \frac{1}{2}[80 + p_H(g_H s_H \theta_H) + (1 - p_H)(g_H s_L \theta_H)] \geq \frac{1}{2}[p_H(40-60\theta_H) + (1 - p_H)(8 - 108\theta_H)] + \frac{1}{2}[116 - 108\theta_H]$ or $p_H \geq (22 - 48\theta_H - g_H s_L \theta_H)/(24 + 36\theta_H + g_H \theta_H[s_H - s_L])$. For this to be a stricter condition than the Baseline we need $432(2 - 3\theta_H - 9\theta_H^2) + (14 - 60\theta_H)(g_H \theta_H[s_H - s_L]) - (24 + 36\theta_H)g_H s_L \theta_H \geq 0$. Having $\theta_H < 7/30$ is a sufficient condition for this to be true. Therefore, the separating equilibrium is less likely to be sustainable in the *Mutual Minimum Treatment* than in the *Baseline Treatment*. If instead a pooling equilibrium holds then we would expect similar agent choices between controlling and trusting (since esteem will be equal), and hence a smaller cost of control.

The *Consent Treatment* further changes the criteria for the decisions to control or trust. The condition for the low types to control is slightly relaxed compared to the *Mutual Minimum Treatment* (since in some cases when he will be the agent the other player trusted): $\frac{1}{2}[p_L(40) + (1 - p_L)(8)] + \frac{1}{2}[p_L(120) + (1 - p_L)(116)] \geq \frac{1}{2}[p_L(80 + g_L s_H \theta_H) + (1 - p_L)(g_L s_L \theta_H)] + \frac{1}{2}[120]$ or $p_L \leq (4 - g_L s_L \theta_H)/(44 + g_L \theta_H[s_H - s_L])$. However, this condition is still stricter than the *Baseline Treatment*. The condition for the high types to trust is now $\frac{1}{2}[p_H(80 + g_H s_H \theta_H) + (1 - p_H)(-120\theta_H + g_H s_L \theta_H)] + \frac{1}{2}[p_H(80 + g_H s_H \theta_H) + (1 - p_H)(100 - 60\theta_H + g_H s_L \theta_H)] \geq \frac{1}{2}[p_H(40 - 60\theta_H) + (1 - p_H)(8 - 108\theta_H)] + \frac{1}{2}[p_H(120 - 120\theta_H) + (1 - p_H)(116 - 108\theta_H)]$ or $p_H \geq (12 - 18\theta_H - g_H s_L \theta_H)/(12 + 72\theta_H + g_H \theta_H[s_H - s_L])$. For this to be a stricter condition than in the *Baseline Treatment* we need:

$480(2 + 3\theta_H)(1 - 3\theta_H) + (14 - 60\theta_H)(g_H\theta_H[s_H - s_L]) - 36g_Hs_L\theta_H \geq 0$. Having $\theta_H < 7/30$ is again a sufficient condition for this to be true. Additionally, this is always a stricter condition than the *Mutual Minimum* condition. Therefore, the separating equilibrium is less likely to be sustainable in the *Consent Treatment* than in the *Baseline Treatment*. Again, since the separating equilibrium (but not a pooling equilibrium) generates the cost of control this means we would be less likely to see a cost of control in the *Consent Treatment* than in the *Baseline Treatment*.

Hence, there is a wide range of parameters for the Ellingsen and Johannesson (2008) model that are consistent our primary results: that there is a cost of control in the *Baseline Treatments* and this cost is reduced or eliminated in the *Mutual Minimum Treatment* and *Consent Treatment*.