Process Manipulation in Unique Implementation

Hitoshi Matsushima
University of Tokyo

November 2012

Process Manipulation in Unique Implementation[+]

Hitoshi Matsushima[*]

Faculty of Economics, University of Tokyo

First Version: December 8, 2008

This Version: July 29, 2012

Abstract

We incorporate social influence into implementation theory, and highlight the manner in which an informed agent feels guilty with regard to disobeying an uninformed principal's wishes. The degree of this feeling depends on the agent's expectation of others' behavioral modes. We demonstrate a method of process manipulation, through which the principal employs psychological tactics for incentivizing agents to announce information in keeping with his/her wishes. We indicate that with a version of incentive compatibility, the principal can implement any alternative that he/she wishes as the unique Nash equilibrium without employing any contractual devices. Each agent's psychological cost would be negligible.

Keywords: Nash Implementation, Uniqueness, Expectation-Based Obedience, Process Manipulation, Tail-Chasing Competition, Panopticon.

Journal of Economic Literature Classification Numbers: C72, D78, D81, D86

---

[*]Faculty of Economics, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. E-mail: hitoshi at mark e.u-tokyo.ac.jp. Fax: +81-3-5841-5523

1. Introduction

This paper investigates a decision problem that involves a principal's attempt to select an alternative that is most compatible with his wishes; however, he/she is unaware of which alternative would be the most desirable. Besides the principal, numerous agents possess private signals with regard to such possible alternatives. Hence, the principal requires these agents to disclose the private signals by means of an announcement. In order to obtain this information from the agents, the principal is required to formulate various methods of incentivizing each agent so that they reveal information that would enable the principal to determine the desired alternative. However, in this case, the agent's obedient announcement for satisfying incentive compatibility is insufficient because they may make other self-enforcing announcements that are not in keeping with the principal's wishes, thereby preventing the principal from arriving at his/her desirable alternative. Therefore, in addition to incentive compatibility, the principal is required to employ additional incentive devices that *eliminate* unwanted equilibria. In other words, the principal is required to obtain agents' obedient announcements as the *unique* Nash equilibrium.

The issue of uniqueness has been studied intensively in the standard theory of implementation, in which it was generally assumed that agents are concerned merely with regard to acquiring material benefits and enjoy full autonomy in making their announcements. Following this assumption, the authors in literature pertaining to this field have generally focused on inventing material-based contractual devices, such as the modulo (Maskin (1977/1990)) and Abreu-Matsushima mechanisms (Abreu and Matsushima (1992)). These material-based contractual devices implement, at least in the virtual sense, any value of the fixed social choice function as the unique Nash equilibrium outcome in compensation for artificial tailoring.[1]

In contrast with this standard theory, any real person is more or less concerned not

---

[1] For surveys on implementation theory, see Moore (1992), Palfrey (1992), Osborne and Rubinstein (1994, Chapter 10), and Maskin and Sjöström (2002).

only with regard to material benefit but also with various psychological inclinations. Among them, inclinations toward *social influence* could be particularly relevant to the decision problem, according to which, *an individual experiences feelings of guilt for disobeying the authority's wishes and this feeling is intensified when he/she expects his/her reference group to obey this authority's wishes*. In this regard, several studies in social psychology such as Ash (1955), Milgram (1974), Zimbardo et al. (1977), and Hofling et al. (1966) have commonly reported that subjects in laboratories and fields tended to obey their authorities and sought conformity to their reference groups' behavioral modes. Several of these experiments reported that the subjects are obedient to authority even if the authority attempts to disturb social order.

In contrast, the present paper will focus on more moderate social influences. We do not take into account the possibility that the impact of social influences is extremely strong. It is implicit to assume that the principal mainly concerns about the wellness for the agents, and that it is preferable for any agent to participate in the society rather than be absent from the society. With these assumptions, this paper excludes the consideration of the case that the principal's wishes are unreasonable such as depressing a particular religious group.

On the basis of the arguments stated above, we demonstrate a new concept for the implementation of the principal's desirable alternative; based on the vast store of information pertaining to social psychology and daily life, it is evident that the principal pragmatically considers the manner in which social psychology tactics must be employed for influencing agents with respect to their choice of announcements.[2] To be more precise, with a discrete and infinite time horizon, and given a modified version of incentive compatibility named *sequential incentive compatibility*, the principal manipulates the decision process in the following manner:

(i)   At the beginning of the decision process, the principal explains his/her wishes to all the agents, and makes them accept that whether or not they were

---

[2] Attempts to incorporate social psychology into economics are not new but are now acquiring growing popularity. For example, see Akerlof and Dickens (1982), Geanakoplos, Pearce, and Stacchetti (1989), Bernheim (1994), Gneezy (2005), Charness and Dufwenberg (2006), and Bébabou (2007).

obedient to his/her wishes before this process is started is irrelevant to the decision;

(ii)   Subsequently, every agent is repeatedly and alternately required to announce messages to the principal until the decision process is terminated;

(iii)   During the decision process, every agent is prohibited from monitoring other agents' announcements or communicating with each other;

(iv)   The principal randomly terminates the decision process at a constant hazard rate;

(v)   According to the decision function, the principal selects the alternative that corresponds to the agents' announcements at the terminal round.

We further introduce a concept of social influence that we refer to as *expectation-based obedience*. Expectation-based obedience implies that the degree to which each agent experiences feelings of guilt with regard to disobeying the principal's wishes depends on his/her expectations with regard to other agents' behavioral modes; *an agent experiences greater feeling of guilt with regard to disobeying the principal's wishes if he expects that no other agent has been disobedient in the past.*

This paper assumes that the principal can successfully manipulate the decision process, drawing the psychological inclination toward expectation-based obedience from other psychological inclinations that each agent has potentially. With this assumption, the conclusion of this paper on the issue of *uniqueness* is quite permissive; with sequential incentive compatibility and expectation-based obedience, *the principal can implement any alternative that he wishes as the unique Nash equilibrium outcome by employing the process manipulation, where an agent's psychological cost of disobeying could be even negligible as compared to his/her whole material benefits.* We do not need any contractual device in this case.

Since the decision process is randomly terminated, any point-wise change in announcement rarely influences the selection of the alternative. Moreover, according to expectation-based obedience, each agent can slightly reduce his/her psychological cost by waiting for another agent to disobey. This psychological cost reduction along with random termination is sufficient to trigger a tail-chasing competition among the agents,

eliminating any possibility of them disobeying the principal's wishes in due order.

It is important to note that expectation-based obedience is conceptually different from the altruism such as Gneezy (2005) that an agent's propensity not to lie depends positively on the cost thereby put on the others; the agent does not feel guilty from the respect of altruism because a point-wise change in announcement rarely influences the others' welfare. In contrast, in the present paper, an agent has intrinsic preference for honesty that he/she dislikes the idea of lying that goes against the intention of the principal even if it does not influence the welfare, i.e., it is just a *white lie*. The basic concept of disliking a white lie in this manner was introduced by Matsushima (2008a, 2008b) in implementation literature[3], showing the permissive result by employing detail-free contractual devices. The present paper extends this concept to expectation-based obedience by making the propensity not to tell a white lie depend on whether he/she expects to be the first person to lie.

The concept of tail-chasing competition stems from the Abreu-Matsushima mechanisms (Abreu and Matsushima (1992)), which was explored in the standard theory of implementation. The standard theory was generally devoted to inventing material-based contractual devices. In contrast, this paper indicates that if there is a scope for the principal to infringe on the agents' autonomy, the principal can apply the same logic as that of the standard theory for inventing *mind-control* methods rather than contractual devices.

Since the process manipulation prohibits monitoring and communication, there is no scope for an agent to influence other agents by being disobedient him/herself, incite disobedience, or eliminate feelings of guilt experienced by members of his/her reference group. In other words, manipulating the decision process in the abovementioned manner is assumed to permit a principal to successfully defend him/herself against any possibility of civil disobedience from his/her agents, a real-life example of which is the Montgomery Bus Boycott that eventually led to the modern civil rights movement in the United States.

The architecture of the process manipulation may resemble the basic concept of

---

[3] For studies in the field of implementation theory that are relevant to psychological aspects, see Eliaz (2002), Dutta and Sen (2011), Kartik and Tercieux (2012), and Lombardi and Yoshihara (2012), for instance.

Panopticon, a prison building designed by Jeremy Bentham, in a metaphorical sense. Panopticon is a circular building that is divided into cells and has a central tower. An observer in the tower can scrutinize the occupants of the cells individually and collectively, whereas the occupants of the cells are isolated from one another by walls.[4]

This paper is organized in the following manner. Section 2 illustrates the model. Section 3 introduces sequential incentive compatibility and expectation-based obedience, and presents the main theorem. Section 4 gives discussions. Section 5 concludes.

---

[4] See Foucault (1995).

2. The Model

Let $N \equiv \{1, 2, ..., n\}$ denote the set of all *agents*, where $n \geq 2$, and A denotes the set of all *alternatives*. Let us consider the following decision problem with a discrete and infinite time horizon. Let M_i and $M \equiv \underset{i \in N}{\times} M_i$ denote the set of all *messages* for each agent $i \in N$ and the set of all message profiles, respectively. At the beginning of the decision process, a principal explains his/her wishes to all agents by using expressions like "I wish to aid the poorest person." Subsequently, he/she requires each agent to provide any relevant information that he/she is not aware of, by putting forth questions like "Where does the poorest person live?" Each agent is repeatedly required to announce messages. During the decision process, no agent can monitor the other agents' announcements. A *strategy* for each agent i is defined as a function $s_i : \{1, 2, ...\} \to M_i$; at each round $t \in \{1, 2, ...\}$, he/she announces a message $m_i = s_i(t) \in M_i$, which is independent of the other agents' past announcements. Let S_i and $S \equiv \underset{i \in N}{\times} S_i$ denote the set of all strategies for agent i and the set of all strategy profiles, respectively. Let $S_{-i} \equiv \underset{j \neq i}{\times} S_j$ for each $i \in N$. We assume that during each round t, the agents *alternately* announce their messages; agent i announces message $s_i(t)$ before agent $j > i$ announces message $s_j(t)$.[5]

The principal randomly terminates the decision process at the end of each round according to a constant *hazard rate* $\delta \in (0,1)$; the probability of termination at round t is given by

$$\delta(1-\delta)^{t-1}.$$

When the decision process is terminated at the end of round t, the principal selects an alternative based on the message profile $s(t) = (s_i(t))_{i \in N} \in M$ at round t. Let us denote by $g : M \to A$ the *decision function*; the principal selects the alternative

---

[5] They alternately announce messages but cannot observe the others' messages.

$g(s(t)) \in A$. In this case, he/she regards $g(s(t)) \in A$ as the desirable alternative, in keeping with his/her wishes.

Based on the above observations, the decision process could be denoted by $\Gamma \equiv (M, g, \delta)$. For each $i \in N$, let us establish a message $m_i^* \in M_i$ as the *truthful* message for agent i, i.e., the obediently announced message by agent i; $m_i^* \in M_i$ is the most suitable answer announced by agent i, and is in keeping with the principal's wishes. Let $m^* = (m_i^*)_{i \in N} \in M$ denote the truthful message profile. We define the truthful strategy $s_i^* \in S_i$ for agent i by

$$s_i^*(t) = m_i^* \quad \text{for all} \quad t \in \{1, 2, ...\}.$$

According to s_i^*, agent i continues to obediently announce information. Let $s^* = (s_i^*)_{i \in N} \in S$ denote the truthful strategy profile.

A *payoff* function for agent i is denoted by $U_i : S \to R$, where $U_i(s_i, s_{-i})$ implies the payoff for agent i when he/she follows $s_i \in S_i$ and he/she *expects* the other agents to follow $s_{-i} \in S_{-i}$. We define a *game* as a combination of the decision process and the profile of payoff functions $(\Gamma, (U_i)_{i \in N})$. A strategy profile $s \in S$ is said to be a *Nash equilibrium* in $(\Gamma, (U_i)_{i \in N})$ if for every $i \in N$,

$$U_i(s) \geq U_i(s_i', s_{-i}) \quad \text{for all} \quad s_i' \in S_i.$$

3. The Theorem

By focusing on the issue of uniqueness, we introduce a modified version of incentive compatibility, named *sequential incentive compatibility*, as follows. For every $s_i \in S_i \setminus \{s_i^*\}$, we define $t_i(s_i) \in \{1, 2, ...\}$ by

$$s_i(t_i(s_i)) \neq m_i^*, \text{ and}$$

$$s_i(t) = m_i^* \text{ for all } t \in \{1, ..., t_i(s_i) - 1\}.$$

Hence, $t_i(s_i)$ implies the first round at which agent i makes a disobedient (dishonest) announcement. Denote $t_i(s_i^*) = \infty$. For every $i \in N$, every $s_i \in S_i$, and every $t \geq t_i(s_i)$, we define a strategy $s_{i,t} \in S_i$ for agent i by

$$s_{i,t}(\tau) = m_i^* \text{ for all } \tau \in \{1, ..., t\}, \text{ and}$$

$$s_{i,t}(\tau) = s_i(\tau) \text{ for all } \tau \in \{t + 1, ...\}.$$

According to $s_{i,t}$, agent i continues to obediently announce information until and at round t, and follows s_i afterwards.

Sequential Incentive Compatibility: For every $s \in S$ and every $i \in N$, if

$$t_i(s_i) < t_j(s_j) \text{ for all } j \neq i,$$

then

$$U_i(s_{i,t_i(s_i)}, s_{-i}) > U_i(s).^6$$

Sequential incentive compatibility implies that whenever an agent expects the other agents to announce obediently until and at a round t, then he/she is willing to announce obediently until and at this round t. From sequential incentive compatibility, it follows that the truthful strategy profile s^* is a Nash equilibrium in $(\Gamma, (U_i)_{i \in N})$,

---

[6] Note that $s_i(t_i(s_i)) \neq m_i^* = s_{i,t_i(s_i)}(t_i(s_i))$.

inducing the desired alternative. However, s^* is not necessarily the unique Nash equilibrium in this case.

This paper assumes that $U_i(s)$ is separated into two parts:

$$U_i(s) = V_i(s) - W_i(s).$$

The first part $V_i(s)$ and the second part $W_i(s)$ are termed the *material benefit* and the *psychological cost,* respectively. The material benefit $V_i(s)$ implies the expected value of the intrinsic utility $v_i(a)$, which is directly derived from the selection of an alternative, and therefore, expressed in the following manner:

$$V_i(s) = \sum_{t=1}^{\infty} v_i(g(s(t)))\delta(1-\delta)^{t-1}.$$

Let us introduce an assumption for psychological cost $W_i(s)$, which we refer to as *expectation-based obedience*, as follows. Let us define the upper limit of differences in intrinsic utility for agent i by

$$L_i \equiv \max_{(a,a')\in A^2} |v_i(a) - v_i(a')|.$$

Expectation-Based Obedience: For every $i \in N$, every $j \in N \setminus \{i\}$, every $t \in \{1,2,...\}$, and every $s \in S$, if

$$j > i,$$

$$t = t_i(s_i) = t_j(s_j) \leq t_h(s_h) \quad \text{for all} \quad h \in N \setminus \{i, j\},$$

and for every $h \in N \setminus \{i, j\}$,

$$t < t_h(s_h) \quad \text{whenever} \quad h < i,$$

then,

(1) $$W_i(s) - W_i(s_{i,t}, s_{-i}) > L_i\delta(1-\delta)^{t-1}.$$

Expectation-based obedience implies that each agent feels guilty with regard to disobeying the principal's wishes, and the degree of this feeling is crucially dependent on his/her expectation with regard to the other agents' behavioral modes; each agent's

feeling of guilt with regard to being disobedient would increase if he expects that no other agent has been disobedient in the past, rather than otherwise.

To be more precise, let us suppose that agent i is the first person to be disobedient, and agent $j > i$ is the first person except agent i to be disobedient. Let us suppose also that agent j announces disobediently at the same round as that at which agent i firstly announces disobediently, i.e., $t = t_i(s_i) = t_j(s_j)$. In this case, by postponing his/her first act of disobedience from round t to round $t+1$, agent i can avoid being the first person to be disobedient. This enables him/her to save considerable psychological cost given by $W_i(s) - W_i(s_{i,t}, s_{-i})$. The inequality of (1) implies that the reduction in costs that ensue in this manner is greater than $L_i \delta (1-\delta)^{t-1}$. Since the principal can select the hazard rate δ as close to zero as possible, expectation-based obedience bears an implicit assumption that the degree to which each agent i can save his psychological cost is negligible as compared to his/her material benefit. In this case, in order to prevent the principal's decision from delaying too long, it is implicit to assume that the time interval between neighboring periods shrinks as the hazard rate approaches zero.

The plausibility of sequential incentive compatibility and expectation-based obedience, compared with the standard definition of incentive compatibility that is based only on the material benefits crucially depends on the type of the principal's wishes. This paper assumes that the principal mainly concerns about the wellness for the agents, and that it is preferable for each agent to participate in the society rather than be absent from the society. Without these assumptions, there could be the case that the principal's wishes are too unreasonable to be made compatible with sequential incentive compatibility and expectation-based obedience. For example, by making these assumptions, this paper can exclude the consideration of the principal's unreasonable wishes such as suppressing a particular religious group; if the principal wishes to suppress the group, any agent is likely to prefer to hide the information about who belongs to this group rather than obey his/her wishes, irrespective of whether he/she him/herself belongs to this group or not. In this case, it might be plausible to assume that each

agent has non-negligible psychological cost for obeying such evil wishes irrespective of whether he/she expects the other agents to obey them. This contradicts sequential incentive compatibility and expectation-based obedience. We also need to exclude from this paper's consideration the case of the principal's unreasonable wishes such as killing members who belong to a particular group in order to make the upper limits L_i of differences in intrinsic utility compatible with the inequality of (1).

We demonstrate the main theorem that with sequential incentive compatibility and expectation-based obedience, the principal can implement any alternative that he/she wishes as the *unique* Nash equilibrium outcome.

Theorem: *With the assumptions of sequential incentive compatibility and expectation-based obedience, the truthful strategy profile s^* is the unique Nash equilibrium in $(\Gamma,(U_i)_{i\in N})$.*

Proof: Consider an arbitrary strategy profile $s \in S \setminus \{s^*\}$. Suppose that there exist $i \in N$, $j \in N \setminus \{i\}$, and $t \geq 1$ such that

$$j > i,$$

$$t = t_i(s_i) = t_j(s_j) \leq t_h(s_h) \quad \text{for all} \quad h \in N \setminus \{i, j\},$$

and for every $h \in N \setminus \{i, j\}$,

$$t_i(s_i) < t_h(s_h) \quad \text{whenever} \quad h < i.$$

From the definition of L_i, it follows that

$$V_i(s) - V_i(s_{i,t}, s_{-i}) = \{v_i(g(s(t))) - v_i(g(m_i^*, s_{-i}(t)))\}\delta(1-\delta)^{t-1}$$

$$\leq L_i\delta(1-\delta)^{t-1},$$

which along with the inequality of (1) implies that

$$U_i(s) - U_i(s_{i,t}, s_{-i}) = V_i(s) - V_i(s_{i,t}, s_{-i}) - \{W_i(s) - W_i(s_{i,t}, s_{-i})\} < 0.$$

Hence, s is not a Nash equilibrium.

Suppose that there exist no such $i \in N$, $j \in N \setminus \{i\}$, and $t \geq 1$. In this case, there

exists $i \in N$ such that

$$t_i(s_i) < t_j(s_j) \quad \text{for all} \quad j \neq i.$$

From sequential incentive compatibility, it follows that

$$U_i(s_{i,t_i(s_i)}, s_{-i}) > U_i(s).$$

Hence, s is not a Nash equilibrium.

Since s^* is a Nash equilibrium because of sequential incentive compatibility, we have proved this theorem.

Q.E.D.

4. Discussions

In order to reduce the psychological cost, each agent may prefer postponing his/her first act of disobedience until after such time that another agent is disobedient. However, it is difficult for an agent to postpone his/her disobedient announcement because he/she is caught between the reduction in psychological cost and the loss in material benefits. Expectation-based obedience can overcome this difficulty. Expectation-based obedience implies that the hazard rate for the process termination is kept low enough for rendering the expected value of loss in intrinsic utility lower than the reduction in psychological cost on white lies. This low hazard rate can trigger a tail-chasing competition among the agents, which perpetually urges them to undertake their first acts of disobedience later on, thereby eliminating unwanted equilibria.

With respect to the functioning of this tail-chasing competition, our model is associated with the basic concept of the Abreu-Matsushima mechanism (Abreu and Matsushima (1992)). In this mechanism, each agent announces multiple messages and is motivated to avoid becoming the first person who makes an announcement that is inconsistent with the initial messages, triggering a tail-chasing competition among the agents. There are substantive points of difference between our model and the Abreu-Matsushima mechanism. In order to trigger the tail-chasing competition, the Abreu-Matsushima mechanism employs any contractual device of side payments (or similar), stipulating that any agent will be fined a small amount of money provided he/she is the first person to make an announcement that is inconsistent with the initial messages. It employs complicated contractual devices that incentivize the agents to announce truthful messages in the first instance. On the contrary, our model does not employ any such contractual device.

We have assumed an infinite time horizon; therefore, the principal can infinitely postpone the decision. Based on this assumption, we can constantly maintain a low hazard rate, thereby permitting the tail-chasing competition to continue until the terminal period. However, with a finite time horizon, it is impossible to constantly

maintain a low hazard rate. Hence, the agents are motivated to announce a disobedient message over the last several periods, provided the decision process does not terminate on the way. In spite of this, by permitting a sufficiently long finite time horizon, the principal *can* consistently maintain a low hazard rate except the few last periods, thereby implementing the desirable alternative in the virtual sense.

During the principal's process manipulation, it is crucial to assume that the agents *alternately* announce their messages. In the absence of this assumption, the agents are willing to prepare for announcing a disobedient message simultaneously, in order to reduce their feelings of guilt. This makes it difficult to eliminate unwanted equilibria, i.e., the agents are not obedient since the beginning.

We have focused on pure strategy Nash equilibrium. It is evident that the basic logic in the proof of the theorem implied the uniqueness of iteratively undominated strategy profile, thereby eliminating any mixed strategy Nash equilibria.

It was implicit to assume that the width in time between adjacent periods was set close to zero and the principal could ignore the direct loss arising from postponing his/her decision. Without this assumption, there could be the trade-off between the timely implementation and elimination of unwanted strategies.

It was implicit to assume that the principal could make the agents accept that whether or not they have been obedient to his/her wishes before the decision process is started is irrelevant to the decision. With this assumption, at the first round, agent 1, who is required to announce a message first, could experience feelings of guilt for disobeying the principal's wishes, even if he/she or any other agent has been disobedient before.

It was assumed that each agent cannot monitor other agents' announcements. If we permit each agent to monitor other agents' announcements, we will be required to investigate a version of the repeated games and struggle with the multiplicity of equilibria implied by the folk theorem or some similar principle. By permitting agents to monitor other agents' announcements, any agent may be motivated to initiate a disobedient announcement at an early stage in order to enable the other agents to prevent the principal from arriving at his/her desired alternative. This point would be

substantial with respect to the issue of implementation with mind control; however, this is beyond the scope of this paper.

5. Concluding Remark

This paper should be regarded as the first step toward the study of process manipulation in implementation. It would be important future researches to elucidate how general it would be for the principal to successfully manipulate the decision process for making the agents announce obediently. Models for the impact of process manipulation on agents' psychological aspects should be built in a more complete manner than the present paper. In this case, the purposes of researches that have priority over all others should be, not merely to demonstrate general analytical methods for effectively manipulating the decision process, but to elucidate whether, when, what is the extent to which, and what is the manner according to which, the authority infringes the autonomy of the individuals in case contractual arrangements such as contingent monetary transfers are restricted for some reasons.

References

Abreu, D. and H. Matsushima (1992): "Virtual Implementation in Iteratively Undominated Strategies: Complete Information," *Econometrica* 60, 993–1008.

Akerlof, G. and W. Dickens (1982): "The Economic Consequences of Cognitive Dissonance," *American Economic Review* 72, 307–319.

Ash, S. (1955): "Opinion and Social Pressure," *Scientific American* 193, 31–35.

Bébabou, R. (2007): "Groupthink: Collective Delusions in Organizations and Markets," Mimeo.

Bernheim, D. (1994): "A Theory of Conformity," *Journal of Political Economy* 102, 841–877.

Charness, G. and M. Dufwenberg (2006): "Promises and Partnerships," *Econometrica* 74, 1579–1601.

Dutta, B. and A. Sen (2011): "Nash Implementation with Partially Honest Individuals," forthcoming in *Games and Economic Behavior*.

Eliaz, K. (2002): "Fault Tolerant Implementation," *Review of Economic Studies* 69, 589–610.

Hofling, C., E. Brontzman, S. Dalrymple, N. Graves, and C. Piece (1966): "An Experimental Study of Nurse-Physician Relationships," *Journal of Nervous and Mental Disease* 143, 171–180.

Foucault, M. (1995): *Discipline and Punish: The Birth of the Prison*. Vintage.

Geanakoplos, J., D. Pearce, and E. Stacchetti (1989): "Psychological Games and Sequential Rationality," *Games and Economic Behavior* 1, 60–79.

Gneezy, U. (2005): "Deception: The Role of Consequences," *American Economic Review* 95, 384–394.

Kartik, N. and O. Tercieux (2012): "Implementation with Evidence," Theoretical Economics 7, 323-356.

Lombardi, M. and N. Yoshihara (2012): "Natural Implementation with Partially Honest Agents," mimeo.

Maskin, E. (1999): "Nash equilibrium and Welfare Optimality," *Review of Economic Studies* 66, 23–38.

Maskin, E. and T. Sjöström (2002): "Implementation Theory," in *Handbook of Social Choice and Welfare 1*, ed. by K. Arrow, A. Sen, and K. Suzumura. Amsterdam: Elsevier.

Matsushima, H. (2008a): "Role of Honesty in Full Implementation," *Journal of Economic Theory* 127, 353–359.

Matsushima, H. (2008b): "Behavioral Aspects of Implementation Theory," *Economics Letters* 100, 161–164.

Milgram, S. (1974): *Obedience to Authority: An Experimental View*. New York: Harper and Row.

Moore, J. (1992): "Implementation in Environments with Complete Information," in *Advances in Economic Theory: Sixth World Congress,* ed. by J. J. Laffont. Cambridge: Cambridge University Press.

Osborne, M. and A. Rubinstein (1994): *A Course in Game Theory*. Cambridge, MA: MIT Press.

Palfrey, T. (1992): "Implementation in Bayesian Equilibrium: The Multiple Equilibrium Problem in Mechanism Design," in *Advances in Economic Theory: Sixth World Congress*, ed. by J. J. Laffont. Cambridge: Cambridge University Press.

Zimbardo, P., C. Haney, W. Banks, and D. Jaffe (1977): "The Psychology of Imprisonment: Privation, Power, and Pathology," in *Contemporary Issues in Social Psychology*, ed. by J. Brigham and L. Wrightman. Cole Publishing Co.