IMPLEMENTATION WITH INDIVIDUALS' LIMITED WILLPOWER

by
GIZEM KILIÇGEDIK

Submitted to the Graduate School of Social Sciences
in partial fulfilment of
the requirements for the degree of Master of Arts

Sabancı University
July 2022

IMPLEMENTATION WITH INDIVIDUALS' LIMITED WILLPOWER

Approved by:

Assoc. Prof. Mehmet Barlo ....

(Thesis Supervisor)

Asst. Prof. Esra Durceylan Kaygusuz ....

Assoc. Prof. Mehmet Yiğit Gürdal ....

Date of Approval: July 22, 2022

ABSTRACT

IMPLEMENTATION WITH INDIVIDUALS' LIMITED WILLPOWER

GIZEM KILIÇGEDIK

Economics M.A. Thesis, July 2022

Thesis Supervisor: Assoc. Prof. Mehmet Barlo

Keywords: Behavioral Implementation, Nash Implementation, Limited Willpower

This thesis studies the implementation problem under complete information in an environment where agents have limited willpower stocks to exert self-control when faced with temptation. We integrate the limited willpower representation model of Masatlioglu, Nakajima, and Ozdenoren (2020) into the behavioral implementation setting. We present a slight modification of the notion of Nash equilibrium, the concept of Nash equilibrium under willpower, and identify the associated consistency notion, consistency under willpower. We show that consistency under willpower is necessary for implementation under willpower and that it is also a sufficient condition when paired with the economic environment assumption. Moreover, we provide examples to illustrate that the implementation result is not monotonically dependent on the willpower stock variable.

ÖZET

SINIRLI İRADE İLE UYGULAMA

GIZEM KILIÇGEDIK

Ekonomi Yüksek Lisans Tezi, Temmuz 2022

Tez Danışmanı: Doç. Dr. Mehmet Barlo

Anahtar Kelimeler: Davranışsal Uygulama, Nash Uygulaması, Sınırlı Irade

Bu tezde, tam bilgi altında uygulamalar, toplumdaki bireylerin sınırlı irade stoklarına sahip oldukları durumlar için incelenmenktedir. Masatlioglu, Nakajima, and Ozdenoren (2020)'in sınırlı irade modelini davranışsal uygulama ortamına entegre ediyoruz. Nash dengesi kavramının bir modifikasyonu olarak irade gücü altında Nash dengesi kavramını sunuyoruz ve ilgili tutarlılık kavramını, irade gücü altında tutarlılığı tanımlıyoruz. İrade gücü altında tutarlılığın, irade gücü altında uygulama için gerekli olduğunu ve ekonomik ortam varsayımı ile birleştirildiğinde ise yeterli bir koşul olduğunu gösteriyoruz. Devamında, uygulama sonucunun irade stoğuna monoton olarak bağlı olmadığını gösteren örnekler sunuyoruz.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my thesis advisor, Mehmet Barlo, for his invaluable support and enormous help during my study. I could not have written this thesis without his patient guidance and encouragement.

Further, I am grateful to my thesis jury members Mehmet Yiğit Gürdal and Esra Durceylan Kaygusuz for their valuable comments and time.

I also owe my thanks to all of my friends, particularly my graduate cohort at Sabancı University for being unwavering in their interest and encouragement throughout this process.

Last but not least, I would like to thank my dear parents, my sisters, and my beloved husband for their love and support. Without their presence, I would not succeed.

In memory of Ismail Serhat Oğuz

TABLE OF CONTENTS

1. INTRODUCTION

Consider a society composed of individuals having preferences over a set of alternatives. In this context, a state of the world captures a particular preference profile of the society. Moreover, the collective goal, a pre-determined social choice rule, describes the so-called socially optimal alternative(s) for each state of the world, and it is common knowledge among the individuals. So, every individual in the society knows the realized state and the socially desirable outcome(s) at that state. We have a social planner who wishes to implement the pre-determined social choice rule. What makes implementation theory interesting is that the social planner does not know the preferences of the individuals (i.e., he does not know the realized state of the world), yet, he needs to elicit this information from the individuals. To this end, he designs a mechanism (or a game-form) that consists of a message space for the individuals and an outcome function that specifies an outcome (an alternative) for each message profile. In other words, the individuals of the society become players of the game induced by the mechanism designer, and they choose a strategy that fits their interests. A social choice rule is said to be implemented by a mechanism whenever the set of equilibrium outcomes overlaps with the set of socially optimal alternatives in every state of the world. The social planner needs to be careful while designing the mechanism because the players may have an incentive to misreport the realized state of the world in order to obtain higher utilities.

Eric Maskin, who won the Nobel prize in economic sciences in 2007, jointly with Leonid Hurwicz and Roger B. Myerson for establishing the principles of mechanism design theory, provides necessary and sufficient conditions for a social choice rule to be Nash implementable (Maskin 1999). He shows that monotonicity is a necessary condition for Nash implementability, meaning that every Nash implementable social choice rule must be monotonic. A social choice function is said to be monotonic whenever an alternative that is socially optimal in one state continues to be optimal in another state when the ranking of that alternative does not get worse for any individual in the latter state. He also demonstrates that when there are at least three individuals in the society, monotonicity together with the no-veto-power property implies Nash implementability. In other words, if a social choice rule is monotonic and satisfies the no-veto-power property, then it can be implemented by a mechanism. No-veto-power property requires that an alternative that is top-ranked by all individuals except one in a state of the world should be socially optimal in that state, and the agent who does not top-rank it should not have any right to veto it.

Following Maskin (1999)'s seminal work, Moore and Repullo (1990) proposes a property labeled condition μ and argues that it closes the gap between the necessary and the sufficient conditions presented by Maskin. Condition μ is weaker but more complex than monotonicity and the no-veto-power property requirements. They also use a constructive proof similar to that of Maskin (1999). The contribution of Moore and Repullo (1990) is to show that when there are three or more agents, condition μ is both necessary and sufficient for Nash implementability. As in Maskin (1999)'s, their study is based on rational preference relations.

On the other hand, Hurwicz (1986) is one of the first studies that investigated the implementation problem without rationality assumptions (Korpela 2012). Hurwicz (1986) realizes that the conditions proposed by Maskin (1999) do not require individual preference relations to be transitive and complete. By defining a new equilibrium concept termed Generalized Nash Equilibrium and using choice functions based on binary comparisons, he shows that the implementation results due to Maskin (1999) are still valid when the preferences are allowed to be intransitive and incomplete. Please note that Hurwicz (1986) is the first to use choice functions in implementation theory. Subsequent work by Korpela (2012) extends the results of Hurwicz (1986) by providing a new solution concept termed Behavioral Nash Equilibrium. He also relaxes the restriction on the choice sets presented by Hurwicz (1986) by allowing individuals to choose from the sets with more than two alternatives. Korpela (2012) argues that compared to Hurwicz (1986)'s solution concept, Behavioral Nash Equilibrium is more compatible with the idea that human behavior can be described by a choice function. Moreover, he generalizes condition-μ due to Moore and Repullo (1990) to condition-λ without restricting any assumptions on choice behavior. However, obtaining necessity of implementation in Behavioral Nash Equilibrium demands Korpela (2012) to adopt one of the well-known rationality assumptions, Chernoff (1954)'s α (alternatively, Arrow (2012)'s independence of irrelevant alternatives). De Clippel (2014) generalizes Nash implementation results due to Maskin (1999) by investigating implementation when the choices of the individuals are allowed to be inconsistent with their rational preferences. He uses choice correspondences instead of preference relations. He introduces a condition termed consistency and shows that it is necessary for Nash implementability without any rationality assumptions. He demonstrates that consistency is also a sufficient condition when combined with unanimity and strong consistency when there are at least three individuals in the society. The implementation results he provides are valid for the cases where individuals might have cognitive biases such as status quo bias, choice overload, cognitive dissonance, endowment, and framing effects.

Integrating behavioral insights into the mechanism design theory is not confined to these studies. Research in cognitive psychology and behavioral economics establishes that human beings are prone to make systematic errors when making decisions. In mechanism design theory, a social planner would like to implement a social choice rule relying on the information coming from the choices of the individuals that are not necessarily rational. Therefore, designing mechanisms robust to human mistakes and behavioral aspects becomes a concern in contemporary research on implementation theory. For instance, Eliaz (2002) investigates the implementation problem when some agents are 'faulty' in a way that they behave sub-optimally. Amorós (2009) examines the case where the mechanism is designed for eliciting the socially optimal rankings of contestants from unfair jurors who might favor some contestants. Bierbrauer and Netzer (2016) introduces reciprocity into the mechanism design by incorporating intention-based preferences into the implementation theory.

Behavioral implementation results are also valid for the environments featuring boundedly-rational individuals who suffer self-control problems when faced with temptation. Intuitively, people get tempted whenever there exists an alternative that gives immediate satisfaction rather than long-lasting benefits. To illustrate this behavior, consider an individual who needs to submit a term paper to graduate from college, and he has limited time to do so. The other options available to him, such as watching TV or scrolling through his phone that give more instantaneous pleasure but less long-term benefits than writing his term paper, are called tempting alternatives. Given his limited time, if he chooses to watch TV instead of writing the term paper, then he succumbs to temptation. However, when he does not procrastinate and choose writing over watching TV, he uses some cognitive resource to exert self-control. Often this resource is referred to as willpower, and research in experimental psychology shows that willpower is a limited resource that is depletable (e.g., Heatherton, Tice et al. (1994), Bratslavsky et al. (1998), Baumeister and Vohs (2003)). For instance, Bratslavsky et al. (1998) conducts a two-phase experiment in which they show that people who exerted self-control over eating chocolates (by eating radishes) in the first period tend to spend less time with a complex and frustrating puzzle than people who succumbed to temptation in the first period. This result suggests that the self has a limited capacity for willpower. When willpower is used too much for a task, it remains less for a subsequent demanding one that requires the use of willpower.

Research in economic theory focuses on modeling temptation, self-control, and willpower to understand their effect on decision-making. There are different perspectives on the study of temptation and willpower in the literature. While one group of studies addresses time-inconsistent preferences (e.g., (Laibson 1997)), another part of the research focuses on self-control costs (e.g., (Gul and Pesendorfer 2001)). In addition to these, there exists research on the models of dual-self representations (e.g., Thaler and Shefrin (1981), Fudenberg and Levine (2006)). For a more detailed discussion regarding the literature on temptation, we refer the interested reader to Lipman, Pesendorfer et al. (2013). These studies focus on the characterization of choices over menus that capture *preferences for commitment*. People commit themselves by restricting alternatives they can reach for future consumption. To illustrate this, consider the student who needed to write his term paper. Suppose he decides to go to a study hall where the possible distractive factors are minimized. The only factor that can refrain him from focusing on his study is his mobile phone. To increase his productivity, he can choose not to bring his phone to the library; in this case, he limits his options for study time. When this is not possible, while he is studying, he may exert self-control not to get distracted by his phone.

In this thesis, we integrate the *limited willpower model* of Masatlioglu, Nakajima, and Ozdenoren (2020) into the behavioral Nash implementation setting. This model differs from the standard temptation models in the sense that the preferences of the agents are not menu-dependent. In other words, the individuals do not choose menus for future consumption. They have *ex-ante preferences* over a set of alternatives, and their *ex-post choice* from the very same set of alternatives depends on the *commitment utilities* and *temptation values* of the alternatives and on their *willpower stocks*. Masatlioglu, Nakajima, and Ozdenoren (2020) argues that compared to revealing ex-ante preferences across all menus of options, this is more natural and straightforward. According to this model, every individual has a *willpower stock* representing his ability to overcome temptations. The more an agent has willpower stock, the more he can exert self-control in cases of temptation. Furthermore, every alternative has a *temptation value*, and also a *commitment utility* that represents the individual's commitment preferences. An agent can consider an alternative from a choice set only if he has enough willpower to choose it, and he chooses the alternative that maximizes his commitment utility among the ones that pass the willpower threshold.

This thesis aims to apply the limited willpower representation model to the implementation theory. We revisit the necessary and sufficient conditions for Nash implementation in an environment with individuals having limited willpower stocks to exert self-control when faced with temptation. We first describe the limited willpower representation of Masatlioglu, Nakajima, and Ozdenoren (2020), and by using this model, we redefine Nash equilibrium and Nash implementation notions. Then, we introduce a slightly different version of De Clippel (2014)'s consistency in limited willpower structure as consistency under willpower. This condition requires that if an alternative is socially optimal, then it should be chosen by every individual in the society at that state, meaning that it should be the alternative that gives the highest utility among the ones that satisfy the willpower constraint for every agent. It also requires that if an alternative is socially optimal in one state but not in another state, then there must be a person for whom one of the two conditions is true: either he does not have enough willpower to choose this alternative in the second state, or he has enough willpower, but there is another alternative that gives a higher commitment utility at the second state.

We prove that consistency under willpower is necessary for Nash implementation under willpower.

For the sufficiency part, we define the economic environment assumption that implies no-veto-power property and unanimity condition presented in the literature. The economic environment assumption in our setting demands that for each alternative, there must be at least two individuals who do not choose it from the set of all alternatives that satisfy their willpower constraint. We show that the economic environment assumption, together with consistency under willpower, is a sufficient condition for Nash implementability under willpower.

For the proof we use the canonical mechanism that is employed by Maskin (1999), Saijo (1988), Moore and Repullo (1990), Bergemann and Morris (2008), Barlo and Dalkiran (2009), Kartik and Tercieux (2012), Korpela (2012), De Clippel (2014), Barlo and Dalkıran (2022), among others. In this mechanism, every individual is asked to announce the realized state, a socially optimal alternative at the state that the individual announced as the realized state, a reward alternative, and an integer. The outcome function is defined by three rules which ensure that the social choice correspondence can be implemented. The first rule realizes if all the individuals announce the same state and the same alternative which is socially optimal at that state; in such cases, the first rule requires the outcome to be equal to this announced alternative. Meanwhile, the second rule realizes if an individual (the odd-man-out) makes an announcement involving a state and an alternative that is different than those of all the others; in such situations, the social planner believes the odd-man-out (who claims that the realized state and socially optimal alternative pair is not the one announced by all the others) only if the odd-man-out chooses a reward alternative that would have made him worse off if the realized state were to be the one that the others announce. That is why the mechanism is robust to a collective lie, thanks to this rule. The last rule, on the other hand, is for any other situation not included in the first two rules. According to this rule, the outcome would be the alternative stated by the winner of an integer game.

To show the impact of the willpower stock variable on implementation results, we provide three simple examples with two individuals, three alternatives, and two different states of the world. In the first example, we provide a mechanism that implements a social choice correspondence, with given values of willpower stocks for two individuals for two states of the world. In the other examples, we change the levels of individuals' willpower stocks, and as a result, we obtain 'bad' equilibria. Therefore, we show that the implementation result is critically dependent upon levels of willpower stocks. We also demonstrate there is no monotone relationship between the strength of willpower and Nash implementability. In other words, an increase or decrease in any individual's willpower stock may cause an SCC to cease to be implementable. Thus, it is important for a social planner to identify the willpower stock levels of the society in every possible state of the world.

The rest of this thesis is organized as follows: Chapter 2 provides notations and definitions. Chapter 3 presents our necessity and sufficiency results. Chapter 4 provides three examples to illustrate the non-monotonicity of the implementation result with respect to willpower. Chapter 5 concludes.

2. NOTATIONS AND DEFINITIONS

Let $N = \{1,2,...,n\}$ be a finite set of individuals constituting a society, X be a non-empty set of alternatives, and $\mathcal{X} := 2^X \setminus \{\emptyset\}$ be all non-empty subsets of X. We denote the set of all possible (payoff relevant) states of the world by $\Theta = \Theta_1 \times \Theta_2 \times ... \times \Theta_n$, where $\theta \in \Theta$ is a generic element. In our setting, a state determines individuals' preference profiles.

Next, we define the *limited willpower representation* of Masatlioglu, Nakajima, and Ozdenoren (2020) which we will employ in our construction and formalities. Let $u_i : X \times \Theta \to \mathbb{R}$ be the *commitment utility function* which represents individual i's rational preferences on the set of alternatives. This can be thought of as the utility an agent gets from an alternative $x \in X$ when he is rational and considers his long-term goals. Let $v_i : X \times \Theta \to \mathbb{R}$ be the *temptation value function* where $v_i^\theta(x)$ indicates how tempting the alternative x is for i at state θ. Further, $w_i : \Theta \to \mathbb{R}_+$ is the *willpower stock function* which plays a crucial role in the decision-making process of individual i: for any given $\theta \in \Theta$ and $i \in N$, $w_{i,\theta} \geq 0$ determines how limited the willpower of individual i at state θ is. The particular details are explained in what follows.

According to this model, an individual makes his decision based on these three factors. Individual i would be able to consider choosing an alternative x from a set $S \in \mathcal{X}$ only if his willpower stock is enough to include x in his choice set. More precisely, $x \in S$ would be in the *willpower compatible set of alternatives* (i.e., his choice set) only if

$$\max_{y \in S} v_i^\theta(y) - v_i^\theta(x) \leq w_{i,\theta}$$

In words, x would be in the choice set of i only if the difference between the temptation value of the most tempting alternative and the temptation value of x does not exceed his willpower stock. Then, the agent i chooses the alternative which maximizes his commitment utility u_i from his willpower-compatible set of alternatives. Formally, given $i \in N$, $\theta \in \Theta$ and $S \in \mathcal{X}$, we denote i's willpower compatible set of alternatives at state θ from the set of alternatives S by

$$W_i^\theta(S) := \left\{ y \in S \mid (\max_{\bar{y} \in S} v_i^\theta(\bar{y})) - v_i^\theta(y) \leq w_{i,\theta} \right\}$$

This is the set of feasible alternatives for i from the set $S \in \mathcal{X}$, at θ. Then the choice of i at θ from the set S will be given by

$$C_i^\theta(S) := \left\{ x \in W_i^\theta(S) \mid u_i^\theta(x) \geq u_i^\theta(y), \text{ for all } y \in W_i^\theta(S) \right\}$$

In words, at θ, individual i chooses the ones that give the highest payoff according to his commitment preferences among the alternatives which i has enough willpower to choose.

In the rational domain, individuals' preferences on the set of alternatives are continuous and complete preorders represented by a utility function. When it comes to the choice behavior of a rational agent, we have two properties of choice correspondences (α and β), formulated by Chernoff (1954) and Sen (1971), which together imply that the choice correspondences having these properties satisfy the weak axiom of revealed preference (WARP). In other words, if an individual choice correspondence $C : \mathcal{X} \to \mathcal{X}$ satisfies properties α and β, then we say that his choices satisfy the standard axioms of rationality. Property α is satisfied whenever an alternative x which is chosen from a set T continues to be chosen from a set $S \subset T$ whenever it is available. Formally, for any $S, T \in \mathcal{X}$ with $S \subseteq T$, and $x \in S$, if $x \in C(T)$, then it must be that $x \in C(S)$. Property β requires the following: If $x, y \in S \subset T$ for any $S, T \in \mathcal{X}$, and $x, y \in C(S)$, then $x \in C(T)$ if and only if $y \in C(T)$. In sum, we say that the choice behavior of an agent is rational, and it satisfies the WARP if and only if the choice correspondence of that individual satisfies properties α and β.

In the limited willpower model, individual choices may violate WARP. To see this consider the following example where $X = \{a, b, c\}$, $w_i = 5$, and commitment utilities and temptation values are given in Table 2.1. We wish to point out that a is the

Table 2.1 A specification with willpower implying the failure of Property α

X	u_i	v_i
a	10	1
b	2	5
c	1	10

best according to commitment utilities while it is the least tempting alternative, c is the worst in terms of commitment utilities, but it is the most tempting one, and b is the one in the middle in both terms. When we investigate the choice behavior of this agent given in Table 2.2, we see that the property α fails. This is because

Table 2.2 Choices with willpower violating Property α

$\mathcal{S} \subseteq \mathcal{X}$	$W_i(\mathcal{S})$	$C_i(\mathcal{S})$
$\{a,b,c\}$	$\{b,c\}$	$\{b\}$
$\{a,b\}$	$\{a,b\}$	$\{a\}$
$\{a,c\}$	$\{c\}$	$\{c\}$
$\{b,c\}$	$\{b,c\}$	$\{b\}$

i chooses b from the set $\{a,b,c\}$. With property α he would be expected to choose b from the set $\{a,b\}$ as well, since $b \in \{a,b\} \subset \{a,b,c\}$. However, when the choice set shrinks from $\{a,b,c\}$ to $\{a,b\}$, although b is still available in the smaller set, we observe that it is not chosen due to temptation. Although a is better than b and c in terms of commitment utilities, i does not have enough willpower stock to choose a when the most tempting alternative c is available in the choice set. In this case, i is tempted by the alternative c and the difference between the temptation values of c and a is greater than his willpower stock, i.e., $10 - 1 > 5$. Therefore a is not feasible for him when c is an option that he can go for. On the other hand, when c is not in the set of alternatives to choose from, although i is tempted by the alternative b, now he has enough willpower stock to choose a. These establish that these choices violate the WARP, meaning that we do not have rationality in this setting.

In what follows, we integrate the limited willpower representation into our implementation setting, where we go beyond the rationality domain. Before doing so, we need to formalize the collective goal (via social choice correspondences) and the mechanism (via a game-form).

Let $f : \Theta \to \mathcal{X}$ be a *social choice correspondence* (SCC) mapping states to alternatives. Given a state of the world $\theta \in \Theta$, it selects the socially desired alternatives for the society and $f(\theta)$ denotes *f-optimal alternatives.* In our setting, we have *complete information* meaning that every individual in the society knows both the state of the world and the socially optimal alternatives contingent upon that state. However, the social planner whose aim is to implement the social choice rule lacks information about the realized state of the world. Hence, all he has is the information he can get from the agents, and there is no guarantee that they will convey the truth. They may refrain from telling the truth in cases where doing so is more profitable for their benefit. Therefore, the social planner should design a mechanism in which the agents will not have any incentives to lie about the realized state.

The *mechanism* (alternatively, game-form) constructed by the social planner is denoted by $\mu = (M, g)$. This is a game-form which specifies a *message space* $M = M_1 \times M_2 \times ... \times M_n$ with $m_i \in M_i$ being a message profile for i, and an *outcome function* $g : M \to X$ mapping message profiles to alternatives. A mechanism is said to implement a social choice rule when the set of f-optimal alternatives coincides with the equilibrium outcomes of μ at $\theta \in \Theta$.

In this paper, we aim to investigate the implementation notion when individuals' choices are affected by temptation. This means that we need to go beyond the standard implementation models where rationality is assumed. To this end, we are going to borrow findings from articles studying *behavioral implementation* with settings where agents' choices are not necessarily rational (e.g., Hurwicz (1986), Korpela (2012), De Clippel (2014), Barlo and Dalkıran (2022)). According to the models in these studies, Nash equilibrium of a mechanism is defined via *opportunity sets* that the mechanism induces. An *opportunity set* of an agent i in mechanism μ consists of alternatives i can obtain through unilateral deviations. A *Nash equilibrium at a state θ with behavioral agents* is a message profile in which the associated alternative the message profile induces is chosen by all the agents at that state from their opportunity sets defined via the very same message profile. Formally, given $m_{-i} \in \times_{j \neq i} M_j =: M_{-i}$, i.e., the message profile for the players other than i, i's opportunity set is given by $O_i(m_{-i}) = \{g(m_i, m_{-i}) \mid m_i \in M_i\} \subseteq \mathcal{X}$. A message profile m^* is a Nash equilibrium of the μ at state $\theta \in \Theta$ if $g(m^*) \in C_i^{\theta}(O_i^{\mu}(m^*_{-i}))$, for all $i \in N$.

We proceed with the definitions of *behavioral Nash equilibrium under willpower* and *Nash implementation under willpower.*

Definition 2.1. *Given a mechanism $\mu = (M, g)$, m^* is a Nash equilibrium under willpower of μ at θ if for all $i \in N$*

$$g(m^*) \in C_i^{\theta}(O_i^{\mu}(m^*_{-i}))$$

where for any given $m_{-i} \in M_{-i}$

$$C_i^{\theta}(O_i^{\mu}(m_{-i})) := \Big\{x \in W_i^{\theta}(O_i^{\mu}(m_{-i})) \mid u_i^{\theta}(x) \geq u_i^{\theta}(y), \text{ for all } y \in W_i^{\theta}(O_i^{\mu}(m_{-i}))\Big\}$$

and

$$W_i^{\theta}(O_i^{\mu}(m_{-i})) := \Big\{y \in O_i^{\mu}(m_{-i}) \mid \max_{\bar{y} \in O_i^{\mu}(m_{-i})} v_i^{\theta}(\bar{y}) - v_i^{\theta}(y) \leq w_{i,\theta}\Big\}$$

and

$$O_i^{\mu}(m_{-i}) := g(M_i, m_{-i}) = \{x \in X \mid \text{ there is } m_i \in M_i \text{ such that } x = g(m_i, m_{-i})\}.$$

In words, a message profile m^* is a Nash equilibrium at θ if every individual in the society chooses the outcome of this message profile from the set of alternatives available to him via unilateral deviations, given that the others play m^*_{-i}. Note that individual i can consider an alternative only if it passes the willpower threshold.

Now suppose that a social planner wishes to implement a social choice rule f at θ. Then $f(\theta)$ identifies the socially desired outcomes at θ. To that end, the social planner needs to make sure that the Nash equilibrium outcomes under willpower of the game-form he designs coincide with the socially desired alternatives at θ.

Definition 2.2. *An SCC $f : \theta \to \mathcal{X}$ is implementable by $\mu = (M, g)$ in Nash Equilibrium under willpower if*

1.1 For all $\theta \in \Theta$ and for all $x \in f(\theta)$, there exists $m^{(x,\theta)} \in M$ such that $g(m^{(x,\theta)}) = x$ and $x \in C_i^{\theta}(O_i^{\mu}(m_{-i}^{(x,\theta)}))$ for all $i \in N$, and

1.2 For all $\theta \in \Theta$, if $m^* \in M$ is such that $g(m^*) \in C_i^{\theta}(O_i^{\mu}(m_{-i}^*))$ for all $i \in N$, then $g(m^*) \in f(\theta)$.

In words, an SSC f is Nash implementable if the set of f-optimal alternatives at state θ coincides with the Nash equilibrium outcomes at θ. In particular, the *first* item tells that for all f-optimal alternatives x at a state θ, there is a Nash equilibrium message profile $m^{(x,\theta)}$ sustaining the alternative x in Nash equilibrium under willpower. On the other hand, the *second* item demands that if a message profile m^* is a Nash equilibrium of the mechanism at a state θ, then the outcome of this message profile, $g(m^*)$, must be an f-optimal alternative at θ.

When the first item of implementation holds, it is said that f is partially implementable in Nash equilibrium under willpower. Notwithstanding, if the second item does not hold, it means that there is a "bad" Nash equilibrium, a Nash equilibrium message profile that results in an outcome not f-optimal at θ. As we concentrate on full implementation rather than partial implementation, we focus on designing mechanisms that partially implement the given SCC while admitting no bad Nash equilibria.

3. NECESSITY AND SUFFICIENCY

In this section, we provide the necessary and sufficient conditions for Nash implementability in a setting where individuals' choices may be affected by temptation. In that regard, we are going to use the *consistency* result of De Clippel (2014) that can be thought of as an extension of Maskin monotonicity to the behavioral domain.

Definition 3.1. *We say that a profile of sets* $\mathcal{S} = (S_i(x,\theta))_{i\in N,\ \theta\in\Theta,\ x\in f(\theta)}$ *is consistent under willpower with the given SCC* $f : \Theta \to \mathcal{X}$ *if*

2.1 For all $i \in N$*, for all* $\theta \in \Theta$*, and for all* $x \in f(\theta)$*,*

$$u_i^\theta(x) \geq u_i^\theta(y) \text{ for all } y \in S_i(x,\theta) \text{ such that} \max_{\tilde{y}\in S_i(x,\theta)} v_i^\theta(\tilde{y}) - v_i^\theta(y) \leq w_{i,\theta}$$

and $\max_{\tilde{y}\in S_i(x,\theta)} v_i^\theta(\tilde{y}) - v_i^\theta(x) \leq w_{i,\theta}$*; and*

2.2 $x \in f(\theta)$ *and* $x \notin f(\tilde{\theta})$ *implies there exists* $j^* \in N$ *such that*

$$\begin{aligned} &\text{either} \max_{\tilde{y}\in S_{j^*}(x,\theta)} v_{j^*}^{\tilde{\theta}}(\tilde{y}) - v_{j^*}^{\tilde{\theta}}(x) > w_{j^*,\tilde{\theta}}, \\ &\text{or} \max_{\tilde{y}\in S_{j^*}(x,\theta)} v_{j^*}^{\tilde{\theta}}(\tilde{y}) - v_{j^*}^{\tilde{\theta}}(x) \leq w_{j^*,\tilde{\theta}}, \text{ and there is } \bar{y} \text{ such that} \\ &\qquad \max_{\tilde{y}\in S_{j^*}(x,\theta)} v_{j^*}^{\tilde{\theta}}(\tilde{y}) - v_{j^*}^{\tilde{\theta}}(\bar{y}) \leq w_{j^*,\tilde{\theta}} \text{ and } u_{j^*}^{\tilde{\theta}}(\bar{y}) > u_{j^*}^{\tilde{\theta}}(x). \end{aligned}$$

Equivalently, $\mathcal{S} = (S_i(x,\theta))_{i\in N,\ \theta\in\Theta,\ x\in f(\theta)}$ *is consistent under willpower with* f *if*

3.1 For all $i \in N$*, for all* $\theta \in \Theta$*, and for all* $x \in f(\theta)$

$$u_i^\theta(x) \geq u_i^\theta(y) \text{ for all } y \in W_i^\theta(S_i(x,\theta)), \text{ and } x \in W_i^\theta(S_i(x,\theta));$$

and

3.2 $x \in f(\theta) \setminus f(\tilde{\theta})$ *implies there exits* $j^* \in N$ *such that*

either $x \notin W^{\tilde{\theta}}_{j^*}(S_{j^*}(x,\theta))$,
or $x \in W^{\tilde{\theta}}_{j^*}(S_{j^*}(x,\theta))$ *and there is* $\bar{y} \in W^{\tilde{\theta}}_{j^*}(S_{j^*}(x,\theta))$ *with* $u^{\tilde{\theta}}_{j^*}(\bar{y}) > u^{\tilde{\theta}}_{j^*}(x)$.

In words, consistency under willpower of a profile of sets $\mathcal{S} = (S_i(x,\theta))_{i\in N,\theta\in\Theta,x\in f(\theta)}$ demands the following requirements: The *first* is that for every player i and for all states θ and for every f-optimal alternative x in that state, x must be in i's willpower compatible set of alternatives at θ from $S_i(x,\theta)$ (which trivially implies that $x \in S_i(x,\theta)$) as well as x should provide maximal commitment utilities among all the alternatives in i's willpower compatible set of alternatives at θ from $S_i(x,\theta)$. On the other hand, the second demands that if an alternative x is f-optimal at state θ but not at state $\tilde{\theta}$, then there must be an individual, $j^* \in N$, (often referred to as the "whistleblower" in the implementation literature (Bergemann and Morris 2008)) for whom the following holds: either x is not in j^*'s willpower compatible set of alternatives at $\tilde{\theta}$ from $S_{j^*}(x,\theta)$, or x is in j^*'s willpower compatible set of alternatives at $\tilde{\theta}$ from $S_{j^*}(x,\theta)$ but there is an alternative $\bar{y}$ that is in j^*'s willpower compatible set of alternatives at $\tilde{\theta}$ from $S_{j^*}(x,\theta)$ and provides strictly higher commitment utilities than those obtained from x at state $\tilde{\theta}$.

The following result establishes that regardless of whether one accepts consistency under willpower "as natural or has qualms about its restrictiveness, it is an inescapable requirement for implementability in Nash equilibrium" (Maskin 1999) under willpower. Indeed, the existence of a profile of sets of alternatives consistent under willpower with the given SCC is a necessary condition for Nash implementation under willpower of that SCC:

Theorem 3.1 (Necessity). *If* $f : \Theta \to \mathcal{X}$ *is Nash implementable under willpower, then there exists a profile of sets consistent under willpower with* f.

Proof. Let $\mu = (M,g)$ be a mechanism such that $f(\theta) = NE^{\mu}(\theta)$, for all $\theta \in \Theta$. Then for any $x \in f(\theta)$, there is $m^{(x,\theta)} \in M$ such that $g(m^{(x,\theta)}) = x$ and $g(m^{(x,\theta)}) \in C^{\theta}_i(O^{\mu}_i(m^{(x,\theta)}_{-i}))$, for all $i \in N$.

Define $S_i(x,\theta) := O^{\mu}_i(m^{(x,\theta)}_{-i})$, for all $i \in N$, for all $\theta \in \Theta$, and for all $x \in f(\theta)$. Since $m^{(x,\theta)}$ is a Nash equilibrium at θ, $g(m^{(x,\theta)}) = x \in C^{\theta}_i(S_i(x,\theta))$, which implies by the definition of choice under willpower that $x \in W^{\theta}_i(S_i(x,\theta))$ as well as $u^{\theta}_i(x) \geq u^{\theta}_i(y)$ for all $y \in W^{\theta}_i(S_i(x,\theta))$. Therefore, (1) of consistency under willpower is obtained.

Next, suppose for contradiction that $x \in f(\theta)$ but $x \notin f(\tilde{\theta})$, and there is no $j \in N$ such that either $x \notin W^{\tilde{\theta}}_j(S_j(x,\theta))$, or $x \in W^{\tilde{\theta}}_j(S_j(x,\theta))$ and there exists $\bar{y} \in$

$W_j^{\tilde{\theta}}(S_j(x,\theta))$ with $u_j^{\tilde{\theta}}(\bar{y}) > u_j^{\tilde{\theta}}(x)$. In this case $x \in C_i^{\tilde{\theta}}(S_i(x,\theta)) = C_i^{\tilde{\theta}}(O_i^{\mu}(m_{-i}^{(x,\theta)}))$, for all $i \in N$, implying that $m^{(x,\theta)}$ is a Nash equilibrium with outcome x at $\tilde{\theta}$. However this implies by (2) of Nash implementability that $x \in f(\tilde{\theta})$, contradicting to $x \in f(\theta) \setminus f(\tilde{\theta})$. □

In words, if a mechanism implements a social choice rule f in Nash equilibrium under willpower, then for any f-optimal alternative x at any state $\theta \in \Theta$, there exists a message profile the outcome which is x and this message profile constitutes a Nash equilibrium under willpower at that state. This means that every individual in the society chooses x out of their opportunity sets when other individuals play according to their equilibrium strategies. This establishes (1) of consistency under willpower when we take the opportunity sets as the choice sets constituting the profile of sets $\mathcal{S}$. Furthermore, (2) of consistency under willpower requires that if an alternative x is f-optimal at state θ but not at $\tilde{\theta}$, then there should be an individual $j \in N$ who does not choose x out of his opportunity set $S_j(x,\theta)$ at $\tilde{\theta}$. To demonstrate this, we suppose for contradiction that there is no $j \in N$ who does not choose x from $S_j(x,\theta)$ at state $\tilde{\theta}$. This means that every individual i in the society chooses x from $S_i(x,\theta)$ at $\tilde{\theta}$. In other words, every individual should have enough willpower to choose x from $S_i(x,\theta)$ at $\tilde{\theta}$, and x should be the alternative with highest commitment utility in i's willpower compatible set of alternatives at state $\tilde{\theta}$ from the set of alternatives $S_i(x,\theta)$. This further means that x is an outcome sustained via a Nash equilibrium under willpower at $\tilde{\theta}$, and (2) of Nash implementability under willpower implies that x should be f-optimal at $\tilde{\theta}$ as well. However, this contradicts the statement that $x \notin f(\tilde{\theta})$. Therefore, we establish (2) of consistency under willpower and show that consistency under willpower is a necessary condition for Nash implementability under willpower.

This result is important in the sense that it is also an *almost sufficient* result for Nash implementability under willpower. That is, in what follows, we establish that the social planner can make use of the existence of a profile of sets of alternatives consistent under willpower with the SCC in the construction of a mechanism that implements this SCC when we adopt a condition (implying standard restrictions in the literature; no-veto-power property and unanimity (e.g., Jackson (1991), Bergemann and Morris (2008), Kartik and Tercieux (2012), Barlo and Dalkıran (2022)). This condition, *the economic environment assumption*, coupled with consistency under willpower, delivers the sufficient condition for Nash implementability under willpower.

Definition 3.2 (**The Economic Environment Assumption**). *For all* $\theta \in \Theta$*, for all* $i \in N \setminus \{j\}$ *for some* $j \in N$*, there is no* $x \in X$ *such that*

$$x \in W_i^\theta(X) \text{ and } u_i^\theta(x) \geq u_i^\theta(y), \text{ for all } y \in W_i^\theta(X).$$

Equivalently, for all $\theta \in \Theta$*, for all* $i \in N \setminus \{j\}$ *for some* $j \in N$*, there is no* $x \in X$ *such that*

$$(\max_{\tilde{y} \in X} v_i^\theta(\tilde{y})) - v_i^\theta(x) \leq w_{i,\theta} \text{ and } u_i^\theta(x) \geq u_i^\theta(y),$$

$$\text{for all } y \text{ satisfying } (\max_{\tilde{y} \in X} v_i^\theta(\tilde{y})) - v_i^\theta(y) \leq w_{i,\theta}.$$

The economic environment assumption requires that an alternative $x \in X$ should not be chosen from the set of all alternatives X by at least two individuals in the society in each state of the world. In other words, at most $n-2$ individuals can choose x from X. For example, there might be an alternative that gives the highest commitment utility to every individual in the society. However, to satisfy the economic environment assumption, it is necessary that at least two individuals do not have enough willpower stock to choose this top-ranked alternative. Or, consider an alternative y that is middle-ranked for most of the agents. In this case, y might be the best alternative in every individual's willpower-compatible set of alternatives from X at θ. However, this would violate the economic environment assumption because every individual i would choose y from X. In short, the economic environment assumption says that there must be at least two individuals who do not choose an alternative $x \in X$ from their willpower-compatible set of alternatives from X at any one of the states $\theta \in \Theta$.

At this point, it is appropriate to discuss the sufficiency conditions used in the literature. The first is the no-veto-power property of Maskin (1999).

Definition 3.3 (**The No-Veto-Power Property**). *A SCC* $f : \Theta \to \mathcal{X}$ *satisfies the no-veto-power property if for all* $\theta \in \Theta$*,* $x \in \cap_{i \in N \setminus \{j\}} C_i^\theta(X)$ *for some* $j \in N$ *implies* $x \in f(\theta)$.

The second is the following unanimity condition:

Definition 3.4 (**The Unanimity of a SCC**). *A SCC* $f : \Theta \to \mathcal{X}$ *is unanimous if for all* $\theta \in \Theta$*,* $x \in \cap_{i \in N} C_i^\theta(X)$ *implies* $x \in f(\theta)$.

Trivially, every SCC satisfying the no-veto-power property is unanimous, while the reverse of that relationship does not hold. On the other hand, it is also fairly straightforward to notice that if the economic environment assumption is satisfied, then at no state θ we can have either $x \in \cap_{i\in N\setminus\{j\}} C_i^{\theta}(X)$ for some $j \in N$ or $x \in \cap_{i\in N} C_i^{\theta}(X)$. Thus, when the economic environment assumption holds, then every SCC satisfies both the no-veto-power property and the unanimity condition.

The following is our sufficiency theorem under willpower.

Theorem 3.2 (**Sufficiency**). *Let $n \geq 3$. If there is a collection of sets consistent under willpower with the SCC $f : \Theta \to X$ that satisfies economic environment assumption, then f is Nash implementable.*

Proof. Suppose that the economic environment assumption holds and there exists profile of sets $\mathcal{S} = (S_i(x,\theta))_{i\in N,\ \theta\in\Theta,\ x\in f(\theta)}$ that is consistent under willpower with the given SCC $f : \Theta \to \mathcal{X}$.

For the proof, we will use the following canonical mechanism as employed by Maskin (1999), Saijo (1988), Moore and Repullo (1990), Bergemann and Morris (2008), Barlo and Dalkiran (2009), Kartik and Tercieux (2012), Korpela (2012), De Clippel (2014), Barlo and Dalkıran (2022):

Let $M_i = \Theta \times X \times X \times \mathbb{N}$ where for each $i \in N$, $m_i = (\theta^{(i)}, x^{(i)}, y^{(i)}, k^{(i)}) \in M_i$ is such that $\theta^{(i)} \in \Theta$, $x^{(i)} \in f(\theta)$, $y^{(i)} \in X$, and $k^{(i)} \in \mathbb{N}$. That is, each player is asked the realized state, an alternative which is f-optimal at the announced state, a reward alternative to be used in terms of single agent deviations (the reward alternative), and an integer. The outcome function is defined via the following rules of the mechanism as follows:

Rule 1: If $m_i = (\theta, x, \cdot, \cdot)$ with $x \in f(\theta)$ for all $i \in N$, then $g(m) = x$.

Rule 2: If $m_i = (\theta, x, \cdot, \cdot)$ with $x \in f(\theta)$ for all $i \in N \setminus \{j\}$ and $m_j = (\theta', x', y', \cdot)$ with $(\theta', x') \neq (\theta, x)$ and $x' \in f(\theta')$, then

$$g(m) = \begin{cases} y' & \text{if } y' \in S_j(x,\theta), \\ x & \text{if } y' \notin S_j(x,\theta). \end{cases}$$

Rule 3: For any other situation, $g(m) = x^{(i)}$ where $i^* = \arg\max_{i\in N} \max_{k^{(i)}} k^{(i)}$.

In words, if all the players have the same strategy $m_i = (\theta, x, \cdot, \cdot)$, and the alternative they have proposed is f-optimal in their proposed state of the world, then the outcome of the mechanism is that alternative. On the other hand, if there exists a player who has a different message in the first two components when compared with the rest of the players, then the outcome will be his reward alternative, i.e., y', whenever y' is in $S_j(x,\theta)$. Otherwise, the outcome would be the alternative proposed by the other players, i.e., x. For any other situation, the outcome of the mechanism is the alternative proposed by the player whose index is the highest among the players proposed the highest integer.

The following claim establishes that (1) of Nash implementability under willpower of f is sustained.

Claim 3.1. *For all $\theta \in \Theta$, for all $x \in f(\theta)$, letting $m_i^{(x,\theta)} = (\theta, x, x, 1)$ for all $i \in N$ ensures that $m^{(x,\theta)}$ is a Nash Equilibrium under willpower of μ such that $g(m^{(x,\theta)}) = x$.*

Proof. Let $\theta \in \Theta$, $x \in f(\theta)$ and $m_i^{(x.\theta)} = (\theta, x, x, 1)$ for all $i \in N$. Then, since each player has the same strategy, Rule 1 applies. Therefore, the outcome of the mechanism would be $g(m^{(x,\theta)}) = x$. To show that this message profile constitutes a Nash equilibrium under willpower, we need to show that none of the players has a profitable deviation. Suppose player $j \in N$ deviates and plays $m_j = (\theta', x', y', \cdot)$ with $(\theta', x') \neq (\theta, x)$. Then, Rule 2 applies. In this case, the outcome of the mechanism would be what j proposed as the reward alternative, i.e., y', only if $y' \in S_j(x, \theta)$. That is, the planner believes j and rewards him with the alternative y' only if choosing m_j would either be incompatible for j at θ from $S_j(x, \theta)$ or would have hurt player j if the other players $i \neq j$ were to tell the truth. Consequently, we have $O_j^{\mu}(m_{-j}^{(x,\theta)}) = S_j(x, \theta)$. By (1) of consistency under willpower, we know that $x \in \cap_{i \in N} C_i^{\theta}(S_i(x, \theta))$; alternatively, $x \in W_i^{\theta}(S_i(x, \theta))$ and $u_i^{\theta}(x) \geq u_i^{\theta}(y)$ for all $y \in W_i^{\theta}(S_i(x, \theta))$. Thus, for all $i \in N$, we have that $x \in W_i^{\theta}(O_i^{\mu}(m^{(x,\theta)})$ and $u_i^{\theta}(x) \geq u_i^{\theta}(y)$ for all $y \in W_i^{\theta}(O_i^{\mu}(m^{(x,\theta)})$. Therefore, x must be among j's chosen alternatives out of $S_j(x, \theta) = O_j^{\mu}(m^{(x,\theta)})$. Thus, $x = g(m^{(x,\theta)}) \in C_i^{\theta}(O_i^{\mu}(m_{-i}^{(x,\theta)}))$ for all $i \in N$, meaning that $m^{(x,\theta)}$ is a Nash equilibrium under willpower of μ at θ. □

The following claim establishes (2) of Nash implementability under willpower.

Claim 3.2. *If $m^* \in M$ is a Nash Equilibrium under willpower of μ, then $g(m^*) \in f(\theta)$.*

Proof. Let $\theta \in \Theta$. Then if m^* were to be a Nash Equilibrium under willpower at θ under Rule 3, we would have $O_i^{\mu}(m_{-i}^*) = X$, for all $i \in N$, because any player can obtain any outcome if he deviates from this message profile by proposing the highest integer. As $g(m^*)$ is an outcome sustained via a Nash Equilibrium under willpower at θ, it must be that $g(m^*) \in C_i^{\theta}(X)$ for all $i \in N$. However, this contradicts the economic environment assumption because it says that there is no $x \in X$ such that $x \in C_i^{\theta}(X)$ for all $i \in N \setminus \{j\}$ for some $j \in N$.

Similarly if m^* were to be a Nash Equilibrium under willpower at θ under Rule 2, then we would have $O_i^{\mu}(m^*_{-i}) = X$, for all $i \in N \setminus \{j\}$ for some $j \in N$. In words, the opportunity set for all individuals except j would be the set of all alternatives because if i deviates from his equilibrium strategy, then Rule 3 will apply, and he can obtain any alternative by proposing the highest integer. For $j \in N$, $O_j^{\mu}(m^*_{-j}) = S_j(x,\theta)$, because if he deviates, he can obtain any outcome he wants as long as that outcome is in $S_j(x,\theta)$. Since m^* is a Nash Equilibrium under willpower, we have $g(m^*) \in C_i^{\theta}(X)$, for all $i \in N \setminus \{j\}$, a contradiction to the economic environment assumption.

So, suppose that m^* is a Nash Equilibrium under willpower at θ under Rule 1. Then, for all $i \in N$ let $m_i^* = (\theta', x', \cdot, \cdot)$ with $x' \in f(\theta')$. Therefore, by Rule 1, $g(m^*) = x'$. Any agent $i \in N$ deviating from m^* would induce Rule 2 and his opportunity set would be $O_i^{\mu}(m^*_{-i}) = S_i(x', \theta')$. Suppose, for contradiction, that $g(m^*) = x' \notin f(\theta)$. Then by (2) of consistency under willpower there exists $j \in N$ such that $x' \notin C_j^{\theta}(S_j(x', \theta'))$. But this is a contradiction to m^* being a Nash Equilibrium at θ under Rule 1. □

Claims 1 and 2 establish our sufficiency result. □

4. NON-MONOTONICITIES DUE TO WILLPOWER

In this section, we are going to provide three examples with different willpower stocks for two players in order to illustrate that the strength of willpower is important while designing mechanisms.

Let the set of alternatives be $X = \{salad, pizza, hamburger\}$ and we let $N = \{Ann, Bob\}$. Suppose there are two different states of the world: θ_1 where the agents are *very hungry*, and θ_2 where they are *not so hungry*. Therefore, when the state of the world is θ_1, both of the agents feel very hungry, and this affects their self-control in the sense that it becomes more difficult to resist the temptation of choosing a high temptation and low commitment utility alternative. Thus, they have lower willpower stock in this state. On the other hand, at θ_2, since they are not so hungry, they find it easier to resist temptation. In this case, they have higher levels of willpower.

The commitment utilities and the temptation values of alternatives for each individual are given in Table 4.1. For Ann, *pizza* is the most tempting alternative with

Table 4.1 Example 1: Commitment Utilities of Ann and Bob

	u_A	v_A	u_B	v_B
salad	8	2	10	2
pizza	4	10	3	6
hamburger	6	7	7	9

a temptation value of 10 while *salad* is the one with the highest commitment utility. Similarly, *hamburger* is the most tempting food for Bob, and commitment utility of *salad* is the highest. Willpower stocks of Ann and Bob at θ_1 and at θ_2 are given in Table 4.2. In θ_2, they both have stronger willpower. As they are not so hungry in this state, they can exercise relatively more stringent self-control.

In light of the information regarding commitment utilities and temptation values of the alternatives, and the willpower stocks, we obtain Table 4.3 showing the choice

Table 4.2 Example 1: Willpower Stock Levels of Ann and Bob

	θ_1	θ_2
w_A	3	6
w_B	4	8

behavior of Ann and Bob at θ_1 and θ_2 The only parameter that changes from θ_1 to

Table 4.3 Example 1: Choices of Ann and Bob under Willpower

S	$W_A^{\theta_1}(S)$	$C_A^{\theta_1}(S)$	$W_B^{\theta_1}(S)$	$C_B^{\theta_1}(S)$	$W_A^{\theta_2}(S)$	$C_A^{\theta_2}(S)$	$W_B^{\theta_2}(S)$	$C_B^{\theta_2}(S)$
$\{s,p,h\}$	$\{p,h\}$	$\{h\}$	$\{p,h\}$	$\{h\}$	$\{p,h\}$	$\{h\}$	$\{s,p,h\}$	$\{s\}$
$\{p,h\}$	$\{p,h\}$	$\{h\}$	$\{p,h\}$	$\{h\}$	$\{p,h\}$	$\{h\}$	$\{p,h\}$	$\{h\}$
$\{s,p\}$	$\{p\}$	$\{p\}$	$\{s,p\}$	$\{s\}$	$\{p\}$	$\{p\}$	$\{s,p\}$	$\{s\}$
$\{s,h\}$	$\{h\}$	$\{h\}$	$\{h\}$	$\{h\}$	$\{s,h\}$	$\{s\}$	$\{s,h\}$	$\{s\}$

θ_2 is the willpower stock for both players, and we observe that it plays a crucial role in the decision of the chosen alternatives.

The SCC $f : \Theta \to \mathcal{X}$ is given such that the f-optimal alternative is *hamburger* at θ_1, while it is *salad* at θ_2.

Next, we provide a simple mechanism which implements $f : \Theta \to \mathcal{X}$ in Nash equilibrium under willpower.

Table 4.4 Example 1: The Mechanism Nash Implementing f under Willpower

State is θ_1:

Ann \ Bob	U	M	D
L	p_A	s	h_A^B
C	s	h_A^B	p
R	p_A	s^B	s^B

State is θ_2:

Ann \ Bob	U	M	D
L	p_A	s_A^B	h_A
C	s^B	h	p
R	p_A	s_A^B	s

This mechanism implements $f : \Theta \to \mathcal{X}$ in Nash equilibrium under willpower. Observe that (C,M) and (L,D) are two Nash equilibria under willpower of this game at θ_1, both giving the same outcome, *hamburger*. So, $g(NE(\theta_1)) = \{h\} = f(\theta_1)$. Similarly, at θ_2, the Nash equilibria under willpower of the game are the message profiles (L,M) and (R,M), both giving *salad* as the outcome. Therefore, we have $g(NE(\theta_2)) = \{s\} = f(\theta_2)$.

From Theorem 3.1, we know that if a social choice rule $f : \Theta \to X$ is Nash implementable under willpower, then there exists a profile of sets consistent under willpower with f. So, we would expect to have a consistent under willpower profile of sets for the mechanism we constructed above.

Let $\mathcal{S} = \{S_A(h,\theta_1),\ S_A(s,\theta_2),\ S_B(h,\theta_1),\ S_B(s,\theta_2)\}$ be a profile of sets where a generic element $S_i(x,\theta)$ denotes the opportunity set of $i \in \{Ann, Bob\}$ at $\theta \in \{\theta_1,\theta_2\}$ when the other player plays according to his or her equilibrium strategy. We are going to show that $\mathcal{S}$ is consistent under willpower with f.

First condition of consistency under willpower requires that if an alternative x is f-optimal at a state θ, then it should be chosen by each individual at that state from the choice set $S_i(x,\theta)$. This condition is satisfied at Nash equilibrium under willpower $m^{(h,\theta_1)} = (L,D)$, because at θ_1, *hamburger* is the f-optimal alternative and $h \in C_A^{\theta_1}(S_A(h,\theta_1))$ where $S_A(h,\theta_1) = \{s,p,h\}$, and $h \in C_B^{\theta_1}(S_B(h,\theta_1))$ where $S_B(h,\theta_1) = \{s,p,h\}$. Similarly, at θ_2, *salad* is the f-optimal alternative consistency (1) condition is satisfied at Nash equilibrium under willpower $m^{(s,\theta_2)} = (L,M)$ as $s \in C_A^{\theta_2}(S_A(s,\theta_2))$ where $S_A(s,\theta_2) = \{s,h\}$, and $s \in C_B^{\theta_2}(S_B(s,\theta_2))$ where $S_B(s,\theta_2) = \{s,p,h\}$.

Second condition of consistency under willpower says that if $x \in f(\theta)$ but $x \notin f(\theta')$, then there must be an individual $j \in N$ such that $x \notin C_j^{\theta'}(S_j(x,\theta))$. Observe that *hamburger* is f-optimal at state θ_1, but not in state θ_2. Therefore, there must be $j \in \{Ann, Bob\}$ who does not choose h at θ_2 from $S_j(h,\theta_1)$. Bob does not choose h at θ_2 from $S_B(h,\theta_1) = \{s,p,h\}$ since $C_B^{\theta_2}(S_B(h,\theta_1)) = s$. Also, *salad* is f-optimal at θ_2 while it is not at θ_1. Observe that $s \notin C_A^{\theta_1}(S_A(s,\theta_2))$. In words, Ann does not choose *salad* at θ_1 from $S_A(s,\theta_2) = \{s,h\}$. Thus, we have shown that (2) of consistency under willpower is satisfied as well.

We proceed by decreasing the willpower stock of Bob from 4 to 3 at state θ_1, while all of the other variables stay the same. In this case, Example 2, the choices of Ann and Bob at θ_1 and θ_2 are as given in Table 4.5.

Table 4.5 Example 2: Choices of Ann and Bob under Willpower

S	$W_A^{\theta_1}(S)$	$C_A^{\theta_1}(S)$	$W_B^{\theta_1}(S)$	$C_B^{\theta_1}(S)$	$W_A^{\theta_2}(S)$	$C_A^{\theta_2}(S)$	$W_B^{\theta_2}(S)$	$C_B^{\theta_2}(S)$
$\{s,p,h\}$	$\{p,h\}$	$\{h\}$	$\{p,h\}$	$\{h\}$	$\{p,h\}$	$\{h\}$	$\{s,p,h\}$	$\{s\}$
$\{p,h\}$	$\{p,h\}$	$\{h\}$	$\{p,h\}$	$\{h\}$	$\{p,h\}$	$\{h\}$	$\{p,h\}$	$\{h\}$
$\{s,p\}$	$\{p\}$	$\{p\}$	$\{p\}$	$\{p\}$	$\{p\}$	$\{p\}$	$\{s,p\}$	$\{s\}$
$\{s,h\}$	$\{h\}$	$\{h\}$	$\{h\}$	$\{h\}$	$\{s,h\}$	$\{s\}$	$\{s,h\}$	$\{s\}$

Now, let us investigate whether the above mechanism implements f or not. The outcomes sustained via Nash equilibrium under willpower are circled and squared in Table 4.6. The mechanism that worked before the change in Bob's stock willpower level (by just one unit) does not implement f anymore. This is because at θ_1,

Table 4.6 Example 2: The Mechanism Failing to Nash Implement f under Willpower

State is θ_1:

Ann \ Bob	U	M	D
L	p_A	s	h_A^B
C	s	h_A^B	p
R	$\boxed{p_A^B}$	s	s

State is θ_2:

Ann \ Bob	U	M	D
L	p_A	s_A^B	h_A
C	s^B	h	p
R	p_A	s_A^B	s

the message profile (R,U) is one of the Nash equilibria of the mechanism with the outcome p, however p is not f-optimal at state, i.e., $g((R,U)) = p \notin f(\theta_1)$. In other words, (R,U) (squared in the mechanism) is a bad Nash equilibrium at θ_1. Hence, we have shown that when the willpower stock of an individual decreases at one state, the mechanism which implemented f before does not implement it anymore.

Now, consider the original example, Example 1, where the willpower stock of Ann increases from 3 to 5 at state θ_1, and from 6 to 8 at state θ_2. We refer to this example in which Ann becomes less prone to temptation as Example 3. Table 4.7 displays the choice behavior of Ann and Bob in this situation. Based on these choices, the

Table 4.7 Example 3: Choices of Ann and Bob under Willpower

S	$W_A^{\theta_1}(S)$	$C_A^{\theta_1}(S)$	$W_B^{\theta_1}(S)$	$C_B^{\theta_1}(S)$	$W_A^{\theta_2}(S)$	$C_A^{\theta_2}(S)$	$W_B^{\theta_2}(S)$	$C_B^{\theta_2}(S)$
$\{s,p,h\}$	$\{p,h\}$	$\{h\}$	$\{p,h\}$	$\{h\}$	$\{s,p,h\}$	$\{s\}$	$\{s,p,h\}$	$\{s\}$
$\{p,h\}$	$\{p,h\}$	$\{h\}$	$\{p,h\}$	$\{h\}$	$\{p,h\}$	$\{h\}$	$\{p,h\}$	$\{h\}$
$\{s,p\}$	$\{p\}$	$\{p\}$	$\{s,p\}$	$\{s\}$	$\{s,p\}$	$\{s\}$	$\{s,p\}$	$\{s\}$
$\{s,h\}$	$\{s,h\}$	$\{s\}$	$\{h\}$	$\{h\}$	$\{s,h\}$	$\{s\}$	$\{s,h\}$	$\{s\}$

Nash Equilibria under willpower of the mechanism at states θ_1 and θ_2 are displayed as circled and squared in Table 4.8. In this example, the willpower stock of Ann increases to 8 at θ_2, and now she has enough willpower to consider any alternative in the set of all alternatives $\{s,p,h\}$. This means that she will behave as a rational

Table 4.8 Example 3: The Mechanism Failing to Nash Implement f under Willpower

State is θ_1:

Ann \ Bob	U	M	D
L	p_A	s_A	h_A^B
C	s	h^B	p
R	p_A	$\boxed{s_A^B}$	s

State is θ_2:

Ann \ Bob	U	M	D
L	p	s_A^B	h
C	s_A^B	h	p
R	p_A	s_A^B	s_A^B

decision-maker since her willpower stock allows her to choose the alternative with the highest commitment utility from each of her choice sets. In this case, there are two Nash equilibria under willpower arising at θ_1: (R, M) with outcome s, and (L, D) with outcome h. Observe that although s is an outcome sustained via a Nash equilibrium at θ_1, it is not f-optimal at that state of the world. This is a violation of (2) of Nash implementability under willpower. Hence, we conclude that the mechanism we construct does not implement f when Ann's willpower stock increased due to a bad Nash equilibrium under willpower.

With these three examples, we have shown that there is no monotone relationship between Nash implementability under willpower with respect to changes in the *willpower stock* variable. In other words, an increase or a decrease in the magnitude of individuals' willpower does not have a predictable effect on the Nash implementability of a mechanism. Knowing the willpower stocks might help the social planner to construct a mechanism that Nash implements a given social choice rule f in the sense that he uses these values to construct the opportunity sets. However, the mechanism he constructs is critically dependent upon the willpower stocks of the individuals in a non-monotone way so that a slight change in any player's willpower may result in the mechanism not implementing the very same social choice rule.

5. CONCLUDING REMARKS

This thesis focuses on Nash implementation in environments where individuals have limited willpower to exert self-control when faced with temptation. We use the limited willpower representation of Masatlioglu, Nakajima, and Ozdenoren (2020) to model the choice behavior of the agents when they have tempting alternatives in their choice sets. In this setting, individual choices may violate WARP; therefore, we have no rationality assumptions in our setting.

Starting from Hurwicz (1986), researchers have been studying Nash implementation under complete information with behavioral aspects. In these studies, choice correspondences are used instead of individual preferences, and the Nash equilibrium concept is defined via opportunity sets. In this paper, we define Nash equilibrium under willpower with opportunity sets induced by the limited willpower stocks of the individuals. We also define Nash implementability under willpower. For the necessary and sufficient conditions for an SCC to be implemented under willpower, we integrate limited willpower representation into the consistency condition provided by De Clippel (2014). Then we show that consistency under willpower is a necessary and almost sufficient condition for Nash implementability. Consistency under willpower becomes a sufficient condition when coupled with the economic environment assumption. We used the canonical mechanism that is employed in mainstream literature.

Lastly, we provide three examples with different willpower stocks for two individuals to analyze the effect of willpower stock on the implementation result. When the willpower stocks of an individual increase, his choices will be similar to the choices of a rational agent since he will have enough willpower to consider any alternative from the choice set available to him. In light of this, one may argue that increased levels of willpower stock would be associated with a positive result on Nash implementability in the sense that a non-implementable SCC becomes implementable when the choices become rational. Or, a decrease in the willpower stocks would affect the implementability result negatively. With the examples we provide, we show that none of these arguments are plausible and that there is no monotone relationship between the implementation result and the levels of willpower.

BIBLIOGRAPHY

Amorós, Pablo. 2009. "Eliciting socially optimal rankings from unfair jurors." *Journal of Economic Theory* 144(3): 1211–1226.

Arrow, Kenneth J. 2012. *Social choice and individual values.* Vol. 12 Yale university press.

Barlo, Mehmet, and Nuh Aygun Dalkiran. 2009. "Epsilon-nash implementation." *Economics Letters* 102(1): 36–38.

Barlo, Mehmet, and Nuh Aygün Dalkıran. 2022. "Computational implementation." *Review of Economic Design* pp. 1–29.

Baumeister, Roy F, and Kathleen D Vohs. 2003. "Willpower, choice, and self-control.".

Bergemann, Dirk, and Stephen Morris. 2008. "Ex post implementation." *Games and Economic Behavior* 63(2): 527–566.

Bierbrauer, Felix, and Nick Netzer. 2016. "Mechanism design and intentions." *Journal of Economic Theory* 163: 557–603.

Bratslavsky, Ellen, Mark Muraven, Dianne M Tice et al. 1998. "Ego depletion: Is the active self a limited resource?" *Journal of Personality and Social Psychology* 74(5): 1252–1265.

Chernoff, Herman. 1954. "Rational selection of decision functions." *Econometrica: journal of the Econometric Society* pp. 422–443.

De Clippel, Geoffroy. 2014. "Behavioral implementation." *American Economic Review* 104(10): 2975–3002.

Eliaz, Kfir. 2002. "Fault tolerant implementation." *The Review of Economic Studies* 69(3): 589–610.

Fudenberg, Drew, and David K Levine. 2006. "A dual-self model of impulse control." *American economic review* 96(5): 1449–1476.

Gul, Faruk, and Wolfgang Pesendorfer. 2001. "Temptation and self-control." *Econometrica* 69(6): 1403–1435.

Heatherton, T, Dianne M Tice et al. 1994. "Losing control: How and why people fail at self-regulation." *San Diego: Academic Press* .

Hurwicz, Leonid. 1986. "On the implementation of social choice rules in irrational societies." *Social Choice and Public Decision Making Essays.* .

Jackson, Mathew O. 1991. "Bayesian implementation." *Econometrica: Journal of the Econometric Society* pp. 461–477.

Kartik, Navin, and Olivier Tercieux. 2012. "Implementation with evidence." *Theoretical Economics* 7(2): 323–355.

Korpela, Ville. 2012. "Implementation without rationality assumptions." *Theory and decision* 72(2): 189–203.

Laibson, David. 1997. "Golden eggs and hyperbolic discounting." *The Quarterly Journal of Economics* 112(2): 443–478.

Lipman, Barton L, Wolfgang Pesendorfer et al. 2013. Temptation. In *Advances in economics and econometrics: Tenth World Congress*. Vol. 1 Citeseer pp. 243–288.

Masatlioglu, Yusufcan, Daisuke Nakajima, and Emre Ozdenoren. 2020. "Willpower and compromise effect." *Theoretical Economics* 15(1): 279–317.

Maskin, Eric. 1999. "Nash equilibrium and welfare optimality." *The Review of Economic Studies* 66(1): 23–38.

Moore, John, and Rafael Repullo. 1990. "Nash implementation: a full characterization." *Econometrica: Journal of the Econometric Society* pp. 1083–1099.

Saijo, Tatsuyoshi. 1988. "Strategy space reduction in Maskin's theorem: sufficient conditions for Nash implementation." *Econometrica: Journal of the Econometric Society* pp. 693–700.

Sen, Amartya K. 1971. "Choice functions and revealed preference." *The Review of Economic Studies* 38(3): 307–317.

Thaler, Richard H, and Hersh M Shefrin. 1981. "An economic theory of self-control." *Journal of political Economy* 89(2): 392–406.