Nash Implementation via Potential Game Form

Ville Korpela *

Public Choice Research Centre (PCRC) and Turku School of Economics (TSE),
University of Turku

July 2012

Abstract

Secure implementation demand that, in addition to every dominant strategy equilibrium, also every Nash equilibrium leads to a socially desirable outcome. Experiments have shown that secure mechanisms outperform dominant strategy mechanisms [Cason, T., Saijo, T., Sjöström, T. and Yamato, T., 2006. Secure implementation experiments: do strategy-proof mechanisms really work? Games and Economic Behavior, 57, pp. 206-235]. This is highly surprising and deserves to be called a puzzle. One would expect that dominant strategies are played whenever they exist, and hence that there would be no statistically significant difference in the success ratio of these two implementation forms. In this paper we provide a theoretical explanation why dominant strategy implementation may not work and secure implementation is needed instead. We do this by deriving an unexpected connection between dominant strategy implementation and potential games.

JEL Classification: C72; C92; D71; D78

Keywords: Dominant strategy implementation; Experiment; Potential game; Robust implementation; Secure implementation

*E-mail address: ville.korpela@utu.fi, telephone number: +358 2 333 6428.
1. INTRODUCTION

Dominant strategy equilibrium is the single most important and widely used solution concept in mechanism design. It has the nice feature of not being dependent on the beliefs of the players, and therefore, it expel the need of a mechanism designer to know anything about these. Which, after all, is a rather demanding assumption. However, although one would expect a player to always choose a dominant strategy equilibrium when available, experimental evidence does not support this (e.g. Attiyeh et al., 2000, Kagel and Levin, 1993, and Kawagoe and Mori, 2001).

Recently, Cason et al. (2006) has shown that a certain sub-class of dominant strategy mechanisms, called secure mechanisms, perform a lot better in experimental settings (in the sense that equilibria are played more often). Secure mechanisms demand that, in addition to every dominant strategy equilibrium, also every Nash equilibrium leads to a socially acceptable outcome. Saijo et al. (2007) has provided several explanations for this empirical fact, while theoretically, it is still very much a puzzle why someone would play a Nash equilibrium, which is not a dominant strategy equilibrium, while there are also dominant strategy equilibria to choose form. Indeed, at first glance one would not expect there to be a statistically significant difference in the success of dominant strategy mechanisms and secure mechanisms. We provide a deeper explanation that suggest otherwise. To this end, we need to introduce the idea of a potential game.

A game in strategic form is called a potential game if there exists a function (common to all players) from the set of strategy profiles to the set of real numbers such that, for any player, changes in utility resulting from changes in own strategy (keeping others strategies fixed) are reflected in the corresponding changes of the potential function (see Monderer and Shapley, 1996, and Dubey et al., 2006).

\footnote{We shall later criticize this standard wisdom.}

\footnote{Rosenthal (1973) was the first to use the idea of a potential game, familiar from the physical sciences, in a game theoretic setting.}
function. Therefore, as the underlying potential function works to align the incentives of all players, one would expect to see Nash equilibria played in such games. Many different types of potential games have been defined in the literature, some weaker than the others.

In this paper we show that if a social choice rule is dominant strategy implementable, and therefore strategy-proof in particular, any mechanism that is used to implement it is necessarily a pseudo-potential game form (Lemma 3 and the discussion following it). This observation suggest that there may be a problem in the idea of dominant strategy implementation: In a implementing mechanisms players tend to coordinate into some of the Nash equilibria since there always exists an underlying pseudo-potential function that works to align the incentives of all players. Theoretically this problem can be fixed by using secure mechanisms (when possible), which may explains why they seem to outperform dominant strategy mechanisms in experimental settings. We also define a new implementation concept that we call Nash implementation via pseudo-potential game form. Moreover, we show that this is almost equivalent to secure implementation (Lemma 4 and 5).

The first paper that studied the connection between implementation and potentials is Jehiel, Meyer-ter-Vehn and Moldovanu (2008). In contrast to our paper, the solution concept is ex-post equilibrium. There is no obvious connection between our paper and Jehiel et al. (2008). This is due to the fact that potentials must be defined in a different way for ex-post equilibrium. However, the idea of a potential is much more natural when it is coupled with Nash equilibrium. After all, the whole point of a potential is to provide a unified representation of best-replies.

The remainder of the paper is organized as follows: In Sect. 2 we fix notation and give all the central definitions. In particular, we introduce the new implementation concept that we dub Nash implementation via potential game form. Then, in Sect. 3, we present and prove our main results. Finally, Sect. 4 is devoted to a concluding discussion.
2. NOTATION AND DEFINITIONS

Our notation follows essentially that of Monderer and Shapley (1996) and Saijo et al. (2007). As usual, for any profile \( v = (v_1, \ldots, v_k) \), let \( v_{-i} \) denote the profile \( (v_1, \ldots, v_i-1, v_{i+1}, \ldots, v_k) \) and \( (v'_i, v_{-i}) \) the profile \( (v_1, \ldots, v_i-1, v'_i, v_{i+1}, \ldots, v_k) \).

2.1 Environments

Let \( A \) be a finite set of alternatives with no particular mathematical structure, and \( I = \{1, \ldots, n\} \) a set of players. We denote a typical element of \( I \) by \( i \) or \( j \), and assume that \( n \geq 2 \). Each player \( i \in I \) has a preference relation over \( A \) that admits a numerical representation \( u_i : A \to \mathbb{R} \) called the utility function. The designer of a social choice mechanism does not know the utility function of player \( i \) exactly, rather she only knows a set of possible utility functions \( U_i \). Let \( u = (u_1, \ldots, u_n) \in U \equiv \times_{i \in I} U_i \). An environment is any triplet \((I, A, U)\).

2.2 Social Choice Rules

A Social Choice Rule (SCR) is any function \( f : U \to A \) that associates a unique alternative \( f(u) \in A \) with every \( u \in U \). The idea is that this mapping connects all profiles of utility functions to a socially acceptable (desirable, optimal) alternative. Some properties of SCRs will be needed.

**Definition 1 (Strategy-Proofness).** The SCR \( f \) is strategy-proof if for all \( i \in I \), all \( u_i, u'_i \in U_i \), and all \( u'_{-i} \in U_{-i} \), we have \( u_i(f(u_i, u'_{-i})) \geq u_i(f(u'_i, u'_{-i})) \).

Strategy-proofness is a strong incentive compatibility condition. It says, roughly speaking, that no player can gain by lying regardless of what everyone else is doing.

**Definition 2 (Rectangular Property).** The SCR \( f \) satisfies the rectangular property if for all \( u, u' \in U \), if \( u_i(f(u_i, u'_{-i})) = u_i(f(u'_i, u'_{-i})) \) for all \( i \in I \), then \( f(u') = f(u) \).

Rectangular property is a technical condition that is necessary for secure
Definition 3 (Weak Non-Bossiness). The SCR $f$ satisfies weak non-bossiness if for all $u, u' \in U$ and all $i \in I$, if $f(u_i, u_{-i}) \neq f(u'_i, u_{-i})$, then there exists some $u''_{-i}$ such that $u_i(f(u_i, u''_{-i})) \neq u_i(f(u'_i, u''_{-i}))$. □

Weak non-bossiness requires that if some individual can change the outcome by giving a different utility function, this must affect her own utility at least in some cases.

2.3 Pseudo-Potential Games

Consider a normal form game $\Gamma = (I; S_1, \ldots, S_n; \pi_1, \ldots, \pi_n)$, where $I$ is the set of players, $S_i$ is the strategy set of player $i$, and $\pi_i$ is the payoff function of player $i$. Denote $S = \times_{i \in I} S_i$ and $S_{-i} = \times_{j \neq i} S_j$. We say that $\Gamma$ is a pseudo-potential game if there exists a real-valued function $P : S \to \mathbb{R}$ (called the pseudo-potential)$^3$ such that for all $i \in I$ and all $s_{-i} \in S_{-i}$, the inclusion

$$\emptyset \neq \arg \max_{s'_i \in S_i} P(s'_i, s_{-i}) \subseteq \arg \max_{s'_i \in S_i} \pi_i(s'_i, s_{-i}).$$

holds. The underlying idea is that the pseudo-potential $P$ works as a common proxy for the payoff functions $\pi_1, \pi_2, \ldots, \pi_n$ when solving for Nash equilibria of the game $\Gamma$. At least when one is trying to find some Nash equilibrium. However, it is important to keep in mind that even if a pseudo-potential game $\Gamma$ has many Nash equilibria, the game defined by the pseudo-potential $\Gamma_P = (I; S_1, \ldots, S_n; P, \ldots, P)$ does not necessarily have any.

Most potential games satisfy some kind of finite improvement property. The subsequent one, which is an adaptation of Lemma 2.3. in Monderer and Shapley (1996) into the case of pseudo-potential games, will be sufficient for our needs. A path in $S$ is any sequence $\omega = (s^1, s^2, s^3 \ldots)$, such that for every $k \geq 1$ there exists a unique player $i$ and a unique strategy $s_i \in S_i$, such that $s^{k+1} = (s^k_{-i}, s_i)$. A path $\omega$ is called a best-reply improvement path if we have

$$s^{k+1}_i \in \arg \max_{s_i \in S_i} \pi_i(s_i, s^k_{-i}) \quad \text{and} \quad \pi_i(s^{k+1}) > \pi_i(s^k)$$

$^3$Pseudo-potential games are studied in Schipper (2004).
for all \( k \geq 1 \), where \( i \) is the unique deviator at the step \( k \). In other words, the deviator always moves to a best-reply that increase her utility strictly.

**Lemma 1** Suppose that \( \Gamma \) is a finite pseudo-potential game, that is \(|S| < \infty\). Starting from any strategy profile \( s \in S \), there must exist at least one best-reply improvement path that leads to a Nash equilibrium.

**Proof.** Take any \( s \in S \). Assume, for the sake of contradiction, that every maximally long best-reply improvement path is infinite. Let \( P \) be the pseudo-potential of \( \Gamma \). By (2) at least one of these best-reply improvement paths, say \( \omega = (s, s^1, s^2, \ldots) \), must be such that (where \( i \) is the unique deviator at step \( k \))

\[
s_{i}^{k+1} \in \arg \max_{s_i \in S_i} P(s_i, s_{-i}^{k})
\]

for all \( k \geq 1 \). The fact that \( \omega \) is a best-reply improvement path then gives us

\[
P(s) < P(s^1) < P(s^2) < P(s^3) < \ldots ,
\]

which is impossible since \( S \) is finite. Hence, there must exist at least one best-reply improvement path that leads from \( s \) to a Nash equilibrium. \( \blacksquare \)

This lemma will play a central role in the proof of our main result in Lemma 4.

### 2.4 Secure Implementation

A *Game form*, or a *mechanism*, is a tuple \( G = (S; g) \) where \( S = \times_{i \in I} S_i \) is the *strategy space* and \( g : S \rightarrow A \) is the *outcome function*. For a fixed profile of utility functions \( u \in U \), this game form defines a game \( \Gamma(u) = (I, S_1, \ldots, S_n, u_1(g(\cdot)), \ldots, u_n(g(\cdot))) \) in normal form. A strategy profile \( s \in S \) is a *Nash equilibrium* of the game \( \Gamma(u) \) if \( u_i(g(s)) \geq u_i(g(s'_i, s_{-i})) \) for all \( i \in I \) and all \( s'_i \in S_i \). The set of all Nash equilibria of the game \( \Gamma(u) \) is denoted by \( NE[\Gamma(u)] \). We say that the game form \( G \) *implements the SCR* \( f \) in *Nash equilibrium* if

\[
g\left(NE[\Gamma(u)]\right) = f(u) \text{ for all } u \in U. \tag{3}
\]

Berger (2007, Lemma 7) derives an analogous result for ordinal potential games using a similar technique.
Likewise, a strategy profile \( s \in S \) is a dominant strategy equilibrium of the game \( \Gamma(u) \) if \( u_i(g(s, s_{-i})) \geq u_i(g(s')) \) for all \( i \in I \) and all \( s' \in S \).

The set of all dominant strategy equilibria of the game \( \Gamma(u) \) is denoted by \( DSE[\Gamma(u)] \), and we say that the game form \( G \) implements the SCR \( f \) in dominant strategies if Equation (3) above holds for dominant strategy equilibrium. Following Saijo et al. (2007), if the game form \( G \) implements \( f \) both in Nash equilibrium and dominant strategy equilibrium at the same time, then we call it securely implementable.

**Definition 4 (Secure Implementation).** The game form \( G \) securely implements the SCR \( f \) if for each \( u \in U \), (i) there exists \( s \in DSE[\Gamma(u)] \) such that \( g(s) = f(u) \) and (ii) for any \( s \in NE[\Gamma(u)] \), \( g(s) = f(u) \). The SCR \( f \) is securely implementable if there exists a game form that securely implements it. □

The class of all securely implementable SCRs is fully characterized by strategy-proofness and the rectangular property.

**Proposition 1 (Saijo et al., 2007).** An SCR is securely implementable if and only if it satisfies strategy-proofness and the rectangular property. ■

In addition, securely implementable SCRs have the nice extra feature that they satisfy a very strong form of the revelation principle.

**Lemma 2 (Saijo et al., 2007).** If the SCR \( f : U \rightarrow A \) satisfies strategy-proofness and the rectangular property, then the associated direct revelation mechanism \( G = (U; f) \) implements it securely. ■

It will be useful to know when an SCR is dominant strategy implemented by its associated direct revelation mechanism. This is due to the fact that the gap between these SCRs, and those that are securely implementable, is not that large.

The next proposition gives a complete characterization.

---

5We write \( u_i(g(s)) \) simply as \( u_i(s) \) when there is no real danger of confusion.

6Notice that in this paper the word implementation always refers to full implementation.

7An exact characterization of those SCRs that are dominant strategy implemented by their associated direct revelation mechanism, but not securely implementable, can be
**Proposition 2** (Saijo et al., 2007, Mizukami and Wakayama, 2007). An SCR \( f : U \to A \) is dominant strategy implemented by its associated direct revelation mechanism \( G = (U; f) \) if and only if it satisfies strategy-proofness and the weak non-bossiness. □

### 2.5 Our Model of Implementation

We introduce the following new concept of implementation. Notice that our definition encompass a multitude of different implementation forms depending on what kind of potential game is used.

**Definition 5.** The game form \( G = (S; g) \) implements the SCR \( f \) in Nash equilibrium via potential game form if, (i) \( G \) implements \( f \) in Nash equilibrium and (ii) the game \( \Gamma(u) \) is a potential game for all \( u \in U \). □

The game form \( G \) in this definition is called a potential game form. One way to interpreted this kind of implementation is to think it as a form of robust implementation: Nash equilibrium is a better prediction in games that admit a potential function than in those which do not.

### 3. UNEXPECTED LINKS BETWEEN DIFFERENT IMPLEMENTATION CONCEPTS

After introducing all this machinery, we are finally ready to prove our main results. In particular, we show that there is a close connection between Nash implementation via pseudo-potential game form, secure implementation and dominant strategy implementation. Basically, our results are motivated by the following remarkably general observation that is a little bit awkward for dominant strategy implementation.

**Lemma 3.** Suppose that the SCR \( f : U \to A \) is strategy-proof. The associated direct revelation mechanism \( G = (U; f) \) is a pseudo-potential

---

*See Dubey et al. (2006), Monderer and Shapley (1996), Voorneveld (2000) and Schipper (2004) for more on different types of potential games.

*Many concepts of robust implementation appear in the literature. See, for example, Bergemann and Morris (2005) on detail-free implementation or Eliaz (2002) on fault tolerant implementation.
Proof. Choose any \( u \in U \). We have to find a pseudo-potential \( P_u : U \to \mathbb{R} \) for the game \( \Gamma(u) = \left( I, U_1, \ldots, U_n, u_1(f(\cdot)), \ldots, u_n(f(\cdot)) \right) \). To this end, let us define \( P_u \) as follows (note that \( u \) is now fixed):

\[
\chi^i(\tilde{u}_i) = \min_{u'_{-i} \in U_{-i}} \left( u_i \left[ f(\tilde{u}_i, u'_{-i}) \right] - u_i \left[ f(u_i, u'_{-i}) \right] \right),
\]

\[
P_u(\tilde{u}) = \sum_{i \in I} \chi^i(\tilde{u}_i).
\]

Observe that

\[
\arg \max_{\tilde{u}_i \in U_i} P_u(\tilde{u}_i, \tilde{u}_{-i}) = \arg \max_{\tilde{u}_i \in U_i} \chi^i(\tilde{u}_i).
\]

The fact that \( f \) is strategy-proof implies two things: \( \chi^i(u_i) = 0 \) and \( \chi^i(\tilde{u}_i) \leq 0 \) for all \( \tilde{u}_i \in U_i \). Hence \( u_i \in \arg \max_{\tilde{u}_i \in U_i} \chi^i(\tilde{u}_i) \), and if also \( u'_i \in \arg \max_{\tilde{u}_i \in U_i} \chi^i(\tilde{u}_i) \), then the equality

\[
u_i[f(u_i, \tilde{u}_{-i})] = u_i[f(u'_i, \tilde{u}_{-i})]
\]

must hold in particular. This is so because \( u_i[f(u_i, \tilde{u}_{-i})] = u_i[f(u'_i, \tilde{u}_{-i})] \) must then hold for all \( \tilde{u}_{-i} \in U_{-i} \). These observations together give us the inclusion required by Equation (1),

\[
\arg \max_{\tilde{u}_i \in U_i} u_i[f(\tilde{u}_i, \tilde{u}_{-i})] \supseteq \arg \max_{\tilde{u}_i \in U_i} P_u(\tilde{u}_i, \tilde{u}_{-i}) \text{ for all } \tilde{u}_{-i} \in U_{-i},
\]

which proves our claim. ■

Notice that this result holds also for any indirect mechanism. Suppose that \( G = (S; g) \) implements \( f \) in dominant strategies. Fix \( u \in U \). We have to find a pseudo-potential \( P_u : S \to \mathbb{R} \) for the game \( \Gamma(u) \). Let \( s^d_i(u_i) \) be some

\[\text{Schipper (2004) and Voorneveld (2000) presents some connections between different types of potential games. In particular, any exact-, weighted-, ordinal-, generalized ordinal- or best-response potential game is also a pseudo-potential game. Although pseudo-potential game is one of the weakest notions of a potential game, it still has several of the good properties associated with potential games in general.}\]
(any) dominant strategy of player $i$ when the profile of utility functions is $u$. Define

$$
\chi^i(\tilde{s}_i) = \min_{s'_{-i} \in S_{-i}} \left( u_i \left[ g(\tilde{s}_i, s'_{-i}) \right] - u_i \left[ g(s^d_i(u_i), s'_{-i}) \right] \right),
$$

and set

$$
P_u(\tilde{s}) = \sum_{i \in I} \chi^i(\tilde{s}_i).
$$

It is straightforward to show that $P_u$ is a pseudo-potential of $\Gamma(u)$. Rather than using the strategy-proofness of $f$, this time we have to use the fact that $s^d_i(u_i)$ is a dominant strategy of player $i$. Otherwise the proof is exactly as in Lemma 3.

So why, exactly, is this awkward for dominant strategy implementation? When a planner wish to implement in dominant strategies, the mechanism is necessarily a pseudo-potential game form. Therefore, one would expect that at least some of the Nash equilibria may be played. This is, in fact, precisely what the experiments seem to suggest — why else would secure implementation outperform dominant strategy implementation. However, it is reasonable to assume that this happens only when the information is nearly complete, so that Nash equilibrium is a good prediction in the first place. Still, and this is important, dominant strategy implementation may not be as robust to informational assumptions as has previously been assumed (see Bergemann and Morris, 2005).

The question is, then, what type of mechanisms work well for a given solution concept. If the solution concept is Nash equilibrium, then a natural candidate as the class of appropriate mechanisms would obviously be the set of all potential game forms. Furthermore, since pseudo-potential game is the weakest notion of a potential game, it is only natural to start by studying Nash implementation via pseudo-potential game form. Surprisingly, as will soon be clear, this is a very demanding form of implementation. The following result is a direct corollary to Lemma 2 and Lemma 3.

**Corollary.** If the SCR $f : U \rightarrow A$ is securely implementable, then it is Nash implementable via pseudo-potential game form.
Proof. By Lemma 2 the direct revelation mechanism $G = (U; f)$ implements $f$ securely, and hence in Nash equilibrium in particular. Since $f$ must be strategy-proof by Proposition 1, this mechanism is a pseudo-potential game form by Lemma 3. ■

Unfortunately, this result does not hold the other way around. Therefore, we do not have an equivalence between secure implementation and Nash implementation via pseudo-potential game form. This is confirmed by the following example.

Example. Let $I = \{1, 2, 3\}$ and $A = \{0, 1, 2, \ldots, 9\}$. A utility function $u_i : A \to \mathbb{R}$ is called quadratic single-peaked if $u_i(a) = -(a - p(u_i))^2$ for all $a \in A$, where $p(u_i) \in A$ is the peak of the utility function. Let $U_i$ be the set of all quadratic single-peaked utility functions and define the SCR $f : U \to \mathbb{R}$ by the condition:

$$f^m(u) = \text{median of } \{p(u_1), p(u_2), p(u_3), 3, 6\}.$$

This kind of SCRs are known as generalized median voter rules. It is easy to show that $f^m$ is strategy-proof, but does not satisfy the rectangular property. Hence, $f^m$ is not securely implementable. However, it is Nash implementable via pseudo-potential game form. Let the game form $G = (S; g)$ be such that $S_1 = S_2 = S_3 = \{1, 2, 3\}$ and $g(s) = s_1 + s_2 + s_3$. Even though it is a bit tedious to show, and therefore omitted here, this game form Nash implements $f^m$ and has a pseudo-potential

$$P_u(s) = -\left(s_1 + s_2 + s_3\right)^2 + 2\left(s_1 p(u_1) + s_2 p(u_2) + s_3 p(u_3)\right).$$

This verifies that $f^m$ is indeed Nash implementable via pseudo-potential game form. □

In addition to the fact that our Corollary does not hold the other way around, this example reveals a couple of other things that are of interest. In the following remarks, a quadratic single-peaked utility function is identified with its peak.

---

This example was provided by an anonymous referee.
The revelation principle does not hold for Nash implementation via pseudo-potential game form. A strategy profile \( u \in U \), such that \( p(u_1) = p(u_2) = p(u_3) = 3 \) or \( p(u_1) = p(u_2) = p(u_3) = 6 \), is a Nash equilibrium of the direct revelation game associated with \( f^m \). Since no one is able to change the outcome by deviating unilaterally, this is true under all possible profiles of utility functions. This observation suggest that a full characterization may be hard to find.

Lemma 3 cannot be strengthened. The direct revelation mechanism associated with a strategy-proof SCR is not necessarily a best-reply potential game form, and this is so even if all preferences are strict.\(^{13}\) Suppose that \( u \in U \) is such that \( p(u_1) = 3 \), \( p(u_2) = 5 \) and \( p(u_3) = 6 \). The sequence\(^{14}\)

\[
(3, 5, 6) \rightarrow (5, 5, 6) \rightarrow (5, 5, 5) \rightarrow (5, 5, 3) \rightarrow (3, 5, 3) \rightarrow (3, 5, 6)
\]

of strategy profiles is a cycle in the direct revelation mechanism associated with \( f^m \). Furthermore, the utility of the deviator is never decreased in this sequence, and in addition, the utility of player 1 is strictly increased in the fourth step. It is easy to see that this could never happen in a best-reply potential game.

Despite of (i) and (ii), there is one way to characterize those SCRs that are Nash implementable via pseudo-potential game form indirectly. Namely, we can show that if an SCR is Nash implementable via pseudo-potential game form, then the associated direct revelation mechanism implements it in dominant strategies. It then follows that, with a minor reservation, a full characterization lies somewhere between two known implementation concepts – secure implementation and dominant strategy implementation. This will be confirmed by the next two lemmata.

**Lemma 4.** If the SCR \( f : U \rightarrow A \) is Nash implementable via pseudo-potential game form that is finite, then it must be strategy-proof.

\(^{13}\)To be exact, a normal form game is a best-reply potential game if and only if Equation 1 holds as equality.

\(^{14}\)The corresponding sequence of outcomes is 5 \( \rightarrow 5 \rightarrow 5 \rightarrow 5 \rightarrow 3 \rightarrow 5 \).
Proof. Take any \( i \in I \), \( u_i, u'_i \in U_i \) and \( u_{-i} \in U_{-i} \). We have to show that \( u_i[f(u_i, u_{-i})] \geq u_i[f(u'_i, u_{-i})] \). Let \( G = (S; g) \) be a finite pseudo-potential game form that implements \( f \) in Nash equilibrium. Furthermore, let \( P_u \) be a pseudo-potential of the game \( \Gamma(u) \).

Choose any strategy profile \( s^1 \in NE [\Gamma(u'_i, u_{-i})] \). We can assume w.l.o.g. that \( f(u_i, u_{-i}) \neq f(u'_i, u_{-i}) \). Otherwise there is nothing to prove and we are done. Since \( \Gamma(u) \) is a finite pseudo-potential game, there must exist (by Lemma 1) a sequence of strategy profiles \( s^1, s^2, \ldots, s^{k+1} \in S \) and a sequence of players \( j(1), \ldots, j(k) \in I \), such that \( s^h \) and \( s^{h+1} \) only differ in the component \( j(h) \) (one deviator at each step), \( s^{k+1} \in NE [\Gamma(u)] \) and

\[
    u_{j(h)}(s^{h+1}) > u_{j(h)}(s^h) \quad \text{for all} \quad h \in \{1, \ldots, k\}.
\]

Moreover, these two sequence can be chosen in such a way that

\[
    s^{h+1}_{j(h)} \in \arg \max_{s_{j(h)}} P_u (s_{j(h)}; s^h_{-j(h)}) \quad \text{for all} \quad h \in \{1, \ldots, k\}.
\]

We divide the rest of this proof into a three mutually exhaustive cases: Either

(i) \( u_{j(k)}(s^k) = u_{j(1)}(s^2) \),

(ii) \( u_{j(k)}(s^k) > u_{j(1)}(s^2) \) or

(iii) \( u_{j(k)}(s^k) < u_{j(1)}(s^2) \).

REMARK. Notice two things. First, we must have \( j(1) = i \). This is due to the fact that only player \( i \) can deviate profitably from \( s^1 \in NE [\Gamma(u'_i, u_{-i})] \) when the truly prevailing preference profile is \( u = (u_i, u_{-i}) \). Second, we can assume w.l.o.g. that \( u_{j(k)}(s^{k+1}) = u_i(s^{k+1}) \). If this does not hold, choose \( d \in \mathbb{R} \) in such way that \( u_{j(k)}(s^{k+1}) + d = u_i(s^{k+1}) \) and define \( \tilde{u}_{j(k)}(\cdot) \equiv u_{j(k)}(\cdot) + d \). This is a virtual utility function and does not necessarily belong to \( U_{j(k)} \). Nevertheless, all dominat strategy equilibria, Nash equilibria and better-reply sequences in the two games, \( \Gamma(u_1, \ldots, u_{j(k)}, \ldots, u_n) \) and \( \Gamma(u_1, \ldots, \tilde{u}_{j(k)}, \ldots, u_n) \), are the same. Furthermore, these games also have the same pseudo-potentials. This is all that we need and make use of.

With these observations in mind, the rest of the proof is now fairly straightforward. Note that since \( s^1 \in NE [\Gamma(u'_i, u_{-i})] \) and \( s^{k+1} \in NE [\Gamma(u)] \), we must have \( u_{j(1)}(s^1) = u_{j(1)}(f(u'_i, u_{-i})) \) and \( u_{j(k)}(s^{k+1}) = u_{j(k)}(f(u_i, u_{-i})) \).
Case (i): By Equation (4) we have
\[ u_{j(k)}(s^{k+1}) - u_{j(k)}(s^k) + [u_{j(1)}(s^2) - u_{j(1)}(s^1)] > 0. \]
Using the assumption \( u_{j(k)}(s^k) = u_{j(1)}(s^2) \) gives us \( u_{j(k)}(s^{k+1}) > u_{j(1)}(s^1) \), which by the above Remark can be re-written as \( u_i[f(u_i, u_{-i})] > u_i[f(u_i', u_{-i})] \).\(^{15}\)

Case (ii): Again by Equation (4) we have
\[ u_{j(k)}(s^{k+1}) - u_{j(k)}(s^k) + [u_{j(1)}(s^2) - u_{j(1)}(s^1)] > 0. \]
Since now \( u_{j(k)}(s^k) > u_{j(1)}(s^2) \), so that \(-u_{j(k)}(s^k) + u_{j(1)}(s^2) < 0\), we must have \( u_{j(k)}(s^{k+1}) > u_{j(1)}(s^1) \) just as in the previous case.

Case (iii): This last case is not as straightforward as the previous two. Now we have to use all the information in the best-reply improvement path \( s^1, \ldots, s^{k+1} \). Using Equation (4) gives us (since all terms in the sum are positive)
\[ \sum_{h=1}^{k} [u_{j(h)}(s^{h+1}) - u_{j(h)}(s^h)] > 0. \]
By Equation (5), together with the fact that \( P_u \) is a pseudo-potential of the game \( \Gamma(u) \), also
\[ [u_{j(k)}(s^{k+1}) - u_{j(k)}(s^k)] + \sum_{h=2}^{k-1} [P_u(s^{h+1}) - P_u(s^h)] \]
\[ + [u_{j(1)}(s^2) - u_{j(1)}(s^1)] > 0. \]
must hold. By assumption we have \( u_{j(k)}(s^k) < u_{j(1)}(s^2) \), and since \( P_u(s^k) - P_u(s^2) > 0 \)\(^{16}\) there must exist a (positive) scaling factor \( M \), such that
\[ -u_{j(k)}(s^k) + u_{j(1)}(s^2) = \frac{P_u(s^k) - P_u(s^2)}{M}. \]
But also \( \frac{P_u(\cdot)}{M} \) is a pseudo-potential of the game \( \Gamma(u) \) (since \( M \) is positive). This means that Equation (5) must hold when we replace \( P_u(\cdot) \) by
\(^{15}\)To clarify, \( u_{j(k)}(s^{k+1}) = u_i(s^{k+1}) \) by the above Remark, and since \( s^{k+1} \in NE(\Gamma(u)) \), we have \( u_i(s^{k+1}) = u_i(\cdot) \).
\(^{16}\)By Equation (4) and (5) we must have \( P_u(s^{k+1}) > P_u(s^k) > \cdots > P_u(s^2) > P_u(s^1) \).
This finally gives us $u_{j(k)}(s^{k+1}) - u_{j(1)}(s^1) > 0$, which is the same as $u_i[f(u_i, u_{-i})] > u_i[f(u'_i, u_{-i})]$ by the above Remark, as required. ■

**Lemma 5.** If the SCR $f : U \rightarrow A$ is Nash implementable via pseudo-potential game form that is finite, then it must satisfy the weak non-bossiness.

**Proof.** Suppose that $f(u_i, u_{-i}) \neq f(u'_i, u_{-i})$ for some $u, u' \in U$ and $i \in I$. We have to find $u''_{-i} \in U_{-i}$, such that $u_i(f(u_i, u''_{-i})) \neq u_i(f(u'_i, u''_{-i}))$. Choose any $s^1 \in NE[\Gamma(u'_i, u_{-i})]$. Since $s^1$ cannot be Nash equilibrium under the utility profile $u$, it must be that player $i$ can deviate profitably. But then there must exist a finite best-reply improvement path just like the one given in Equations 4 and 5 (Lemma 4). This implies that $u_i(f(u_i, u_{-i})) > u_i(f(u'_i, u_{-i}))$ (since this inequality holds in all possible cases i - iii), so that we can simply choose $u''_{-i} = u_{-i}$. This proves the claim. ■

Together, Lemma 4 and 5 imply that if an SCR is Nash implementable via finite pseudo-potential game form, then it is dominant strategy implemented by its associated direct revelation mechanism (Proposition 2). Although we have made the additional assumption of a finite game form, this result does suggest that Nash implementation via pseudo-potential game form is very demanding. Notice that the SCR in the Example satisfies weak non-bossiness, as it should since it is Nash implementable via finite pseudo-potential game form. This follows directly from the fact that if the domain of a SCR consists of strict preferences only, then weak non-bossiness is trivially satisfied.
4. CONCLUDING DISCUSSION

It is well-known that any SCR which is implementable in dominant strategies is also truthfully implemented by its associated direct revelation mechanism (e.g. Dasgupta, Hammond and Maskin, 1979). This powerful tool of mechanism design, called the revelation principle, has the obvious drawback that direct revelation mechanisms may have other dominant strategy equilibria besides the truth-telling one. However, experiments show (e.g. Cason et al., 2006) that subjects do not necessarily play any of their dominant strategies (truthful or not), rather they sometimes coordinate into a Nash equilibrium. This phenomenon deserves to be called a puzzle − why would anyone play Nash equilibrium, which is not a dominant strategy equilibrium, when there are also dominant strategy equilibria to choose from?

Our goal was to give a theoretical explanation for these experimental results: If an SCR is dominant strategy implementable, then the associated direct revelation mechanism is a pseudo-potential game form (Lemma 3). In fact, any mechanism that the planner can use is necessarily a pseudo-potential game form. Therefore, there exists an underlying pseudo-potential that works to align the incentives of all players, and hence, one would indeed expect to see at least some of the Nash equilibria played at least some of the times (those that are Nash equilibria also in the game defined by the pseudo-potential). How often this will happen depends on the simplicity of the underlying pseudo-potential, as well as on the information structure.\(^\text{17}\)

This analysis suggest that it is important to think what is the appropriate class of mechanisms for a given solution concept. We have shown that if the appropriate class of mechanisms for Nash equilibrium is the set of all potential game forms, then Nash implementation is even more demanding than dominant strategy implementation.

\(^{17}\text{Symmetric oligopoly Cournot competition, for example, admits a potential that does not have any clear intuitive meaning (see Slade, 1994, and Monderer and Shapley, 1996).}\)
References


MIZUKAMI, H. and WAKAYAMA, T. (2007): "Dominant Strategy Im-


