BANK OF ENGLAND

Staff Working Paper No. 601

Robustness of subgame perfect implementation

Peter Eccles and Nora Wegner

May 2016

BANK OF ENGLAND

Staff Working Paper No. 601
Robustness of subgame perfect implementation

Peter Eccles[1] and Nora Wegner[2]

Abstract

In this paper we consider the robustness of subgame perfect implementation in situations when the preferences of players are almost perfectly known. More precisely, we consider a class of information perturbations where in each state of the world players know their own preferences with certainty and receive almost perfectly informative signals about the preferences of other players. We show that implementations using two-stage sequential move mechanisms are always robust under this class of restricted perturbations, while those using more stages are often not.

Key words: Implementation, subgame perfect equilibrium, robustness.

JEL classification: D04, D82.

[1] Universidad Carlos III de Madrid. Email: eccles.peter@gmail.com
[2] Bank of England. email: nora.wegner@bankofengland.co.uk

The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England or its committees.

We thank Luis Corchón, Angel Hernando-Veciana, Natalia Fabra, Sebastian Panthöfer, Alessandro Pavan, Hamid Sabourian and participants at the AEA NASM 2014 for useful comments and suggestions.

Information on the Bank's working paper series can be found at
www.bankofengland.co.uk/research/Pages/workingpapers/default.aspx

Publications Team, Bank of England, Threadneedle Street, London, EC2R 8AH
Telephone +44 (0)20 7601 4030 Fax +44 (0)20 7601 3298 email publications@bankofengland.co.uk

ISSN 1749-9135 (on-line)

1 Introduction

This paper studies the robustness of implementation in subgame perfect equilibrium (SPE) in the fashion of Moore & Repullo (1988) and Aghion *et al.* (2012).

A social choice function (SCF) is said to be implemented fully, if there exists a mechanism such that the outcome prescribed by the SCF is the unique equilibrium of the mechanism in all states. Subgame perfect implementation is relevant when sequential mechanisms are used. Although the existing literature on implementation in SPE characterizes the set of SCFs which can be implemented under different informational assumptions, these papers do not provide a distinction between SCFs that are seen to be implemented in practice and those that are not. This distinction is an important aim of implementation, as in any situation it allows a social planner to fully understand the set of SCFs he can choose from.

In this paper we show that placing a very reasonable restriction on the information players have about their own preferences and on the information they have about the preferences of others, allows to distinguish between SCFs which we are seen to be implemented in practice and those that do not appear. More precisely we focus on environments where information is almost complete and introduce information perturbations where each player has more precise information about his own preferences than do other players. These perturbations are referred to as *restricted information perturbations.*

Moore & Repullo (1988) show that under complete information almost any SCF can be implemented in SPE. Taking a step away from implementation under complete information, Aghion *et al.* (2012) (henceforward AFHKT) show that any implementation of a non-Maskin monotonic SCF is not robust to a general class of information perturbations we refer to as *full perturbations.* Maskin monotonicity is a very restrictive requirement and is violated by many SCFs that are implemented in practice, for example firms paying a higher wage to workers with higher outside options. The result obtained by AFHKT therefore questions the usefulness of subgame perfect implementation.

In this paper we argue that typically each player is better informed about his own preferences than is any other player. We restrict attention to a class of perturbations by requiring that players know their own preferences with certainty. This is a reasonable restriction as there are many situations where each player knows his preferences, while others may be slightly uncertain.[1] One example of such a setting is that studied by Bester & Kraehmer (2012) who consider a seller making an offer to a buyer who has private information about how much he values the good.

We show that these restrictions provide a good distinction between SCFs seen to be implemented in practice and those that are not. In particular we demonstrate that under these restricted information perturbations, a wide range of SCFs can be robustly implemented, including many that are not Maskin monotonic. The class of SCFs that can be implemented robustly under the *restricted perturbations* considered here is therefore strictly larger than those that can be implemented robustly under a wider range of full perturbations.

Informally, the reason why the implementability of certain SCFs is robust to restricted perturbations but not full perturbations is the following: Under restricted perturbations, players know their preferences with certainty and do not gain information about their own preferences from the actions of another player. Meanwhile when using full perturbations players have some uncertainty about their own preferences, and hence may update their beliefs about their own preferences after the moves of other players. In particular, in a two-stage game the result of AFHKT relies on off-equilibrium beliefs which ensure that the second-mover gains a significant amount of information about his own preferences after observing an off-equilibrium move from the first-mover. The lack

[1] Our logic also applies to cases where a player is slightly uncertain about his own preferences, as long as he is more certain about them than is any other player.

of belief updating considered here leads to a much larger class of robust mechanisms under restricted perturbations.

Consider the example of a single firm and two types of workers, who differ in their outside option. A 'bad' sequential equilibrium is one where a high type worker accepts a wage that is below his outside option. These equilibria may arise under full information perturbations and rely on the fact that the worker is less informed about his ability and therefore his preferences than the firm. This may occasionally be the case for example when the firm has more information about the job description than the worker. However in most situations this is unlikely to hold, for instance when the worker is more informed of his preferences or outside options. Hence in many applications *restricted perturbations* are the more appropriate tool for assessing whether a certain mechanism is robust. Using this analysis, subgame perfect implementation is very robust in settings where players are confident about which allocations they value.

For most of the paper, we restrict attention to non-stochastic mechanisms where players move sequentially. This restriction is motivated by the fact that in many situations mechanisms where players move simultaneously are not feasible. For instance when bargaining a player must observe the offer made by his opponent before deciding whether to accept or reject the offer made: indeed in most bargaining models - for instance Rubinstein (1982) - players move sequentially. In contrast Baliga (1999) and Bergin & Sen (1998) study implementation in a similar setting with incomplete information and extensive form games, but where players choose their actions simultaneously. These papers show that allowing players to move simultaneously leads to much more permissive results than those presented here.

Meanwhile Corchón & Ortuno-Ortín (1995) and in a generalisation Yamato (1994) consider similar information structures where each player perfectly knows the preferences of other players in his own group but has imperfect information about players outside his group. Using Bayesian and dominant strategy implementation as equilibrium concepts they find that Nash implementation in complete information is a necessary and sufficient condition for robust implementation. In this paper we focus on a two player setting and study subgame perfect implementation which is particularly relevant in sequential move games.

Our main result relates to the concept of *exact implementation* studied by Moore & Repullo (1988) as well as Abreu & Matsushima (1994). The term exact implementation in a setting with information perturbations is used to mean that the desired allocation is always implemented whenever players observe correct signals about the state. The main result of our paper then proves a sufficient condition for a SCF to be exactly implementable with restricted information perturbations. In particular we show that any SCF which can be implemented in a two-stage sequential move game in complete information can be implemented exactly with restricted information perturbations. Moreover requiring two stage implementation is more permissive than requiring Maskin monotonicity, but more restrictive than requiring only three stage implementation.

Since the necessary and sufficient conditions for two stage implementation do not provide great insight, the relevance of two stage implementation is illustrated using a number of examples. Many standard settings of principal agent interaction proceed in two stages, where the principal offers a contract. The agent can reject the contract, accept it - or in some cases - choose an action. Indeed, the examples given in this paper can be interpreted as classic principal agent settings. More precisely, the analysis can be be interpreted as studying the robustness of the outcome of principal agent interactions to small levels of asymmetric information.

Finally, we consider the weaker concept of virtual implementation studied by Abreu & Sen (1991). Virtual implementation with information perturbations requires that the desired allocation is implemented almost always, but does not exclude the possibility for the wrong allocation to be occasionally implemented even when players observe the correct signals. In a deviation from most literature we do not consider virtual implementation using a stochastic element in the mechanism.[2] Instead we follow an approach introduced by Serrano & Vohra (2010) and allow players to choose mixed strategies. We say that an SCF is virtually implementable when in the only equilibrium of the game with information perturbations, players choose mixed strategies, such that the outcome prescribed by the SCF is reached almost always and the probability with which any type chooses a different path becomes arbitrarily small when the information perturbations tend to zero.

Using an example, we show that requiring only virtual implementation some SCFs are robust to restricted information perturbations, although they are not robust when exact implementation is required. This argument shows that the set of SCFs that can be considered robust to information perturbations become larger when considering weaker concepts of implementation. The decision of which concept is appropriate may depend on the situation one has in mind.

The remainder of the paper proceeds as follows. In section two we provide an example to illustrate the differences between implementability under complete information, full perturbations and restricted perturbations respectively, as well as present the intuition behind these differences. Section three introduces the model and formal definitions. The sufficient condition for robust implementation under restricted perturbations is presented in section four. In section five we consider the case of virtual implementation. Section six concludes.

2 Example

Suppose a firm (P) is bargaining with a worker (A). There are two states of the world $\Theta = \{L, H\}$, which represent the fact that workers may either be high type (H) or

[2] This approach is often criticised, because implementation relies on the mechanism designer committing to occasionally implement an allocation that he knows is not Pareto efficient at the point of implementing it.

low type (L). The probability that the worker is high type is $\alpha_H \in (0,1)$, while the probability that the worker is low type is $\alpha_L = 1 - \alpha_H$. There are three outcomes $X = \{w_H, w_L, d\}$. First a high wage w_H may be agreed, secondly a low wage w_L may be agreed and thirdly a default option d may be reached. Both types of workers are equally productive when working for the firm and so the preferences of the firm do not depend on the type of the worker. The firm prefers to pay a low wage rather than a high wage, and prefers to pay a high wage rather than failing to make an agreement:

$$\text{Firm's preferences:} \quad u_P(w_L;\theta) > u_P(w_H;\theta) > u_P(d;\theta) \quad \text{for } \theta \in \{L,H\}$$

Meanwhile, all workers prefer the high wage to any other alternative. However, low type workers prefer to receive the low wage rather than the outside option, while the high type workers prefer the outside option to the low wage. Therefore the preferences of each type of worker are given as follows:

$$\begin{array}{rll} \text{Low type's preferences:} & u_A(w_H;\theta) > u_A(w_L;\theta) > u_A(d;\theta) & \text{for } \theta = L \\ \text{High type's preferences:} & u_A(w_H;\theta) > u_A(d;\theta) > u_A(w_L;\theta) & \text{for } \theta = H \end{array}$$

All of the above is commonly known. Players negotiate according to the following two-stage sequential move bargaining procedure. In the first stage the firm makes an offer $w \in \{w_L, w_H\}$, and then in the section stage the worker chooses to accept (Y) or decline (N) the offer. If the worker accepts the wage offer this agreement is made, and otherwise the default option is reached. The extensive-form version of this game is given in Figure 1.[3]

We analyse this game under three different information structures. In the first case we consider complete information, where both players know the worker's type. In the second and third case, we assume that one player knows the worker's type, while the

[3] Each node is an information sets and there are no moves by nature, as we assume that workers are born with their preferences.

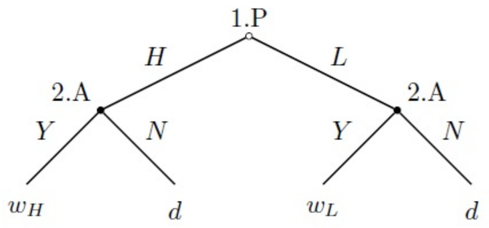

Figure 1: Two stage mechanism.

other receives a signal $s \in \{s_L, s_H\}$ which is highly correlated with the worker's type. More precisely after observing a signal s_L the probability of the worker being low type is equal to $(1-\epsilon)$, while after observing a signal s_H the probability of the worker being high type is equal to $(1-\epsilon)$. After receiving such a signal a player is highly confident - although not completely sure - about the worker's type: in this case we say the worker's type is ϵ-known. Throughout the example it is assumed that $\epsilon > 0$ and ϵ is sufficiently small. A more formal approach is taken in the next section.

Complete information

First consider the case of complete information, where the worker's type is commonly known. In this case there is a unique SPE, where on the equilibrium path the firm offers the low type worker the low wage, the firm offers the high type worker the high wage and all offers are accepted. Off the equilibrium path, low type workers accept a high wage and high type workers reject a low wage. Therefore in complete information this mechanism implements a SCF $f(\theta)$, where $f(L) = w_L$ and $f(H) = w_H$. Note that this SCF is not Maskin monotonic, since both types of workers prefer a high wage to a low wage and yet only the high type workers receive a high wage while the low type workers receive a low wage. Formally Maskin monotonicity is defined as follows:

Definition 1 (Maskin monotonicity) *An SCF ψ is Maskin monotonic, if for all $\theta, \theta' \in \Theta$:*

$\psi(\theta) = x$ *and* $\theta' \in L_i(x, \theta_i)$ *for* $i = A, B$ *imply* $\psi(\theta') = x$

where $L_i(x, \theta_i)$ is the lower contour set of player i with preferences θ_i at allocation x.

An information perturbation where workers know their own preferences

Secondly consider the case where the worker's type is known by the worker and ϵ-known by the firm. Since the worker knows his own type, high type workers always reject the low wage, while low type workers always accept it. Given ϵ is sufficiently small it follows that:

$$\alpha_L(1-\epsilon)\Big(u_P(w_L;\theta) - u_P(w_H;\theta)\Big) > \alpha_H\epsilon\Big(u_P(w_H;\theta) - u_P(d;\theta)\Big)$$

The left hand side represents the firm's gains when offering a low wage rather than a high wage to a low type player having received a signal s_L which was correct. Meanwhile the right hand side denotes the losses that the firm incurs when offering a low wage - which is rejected - rather than a high wage after an incorrect signal s_L. If ϵ is sufficiently small and the signal is sufficiently reliable, it is clear that the gains from offering a low wage outweigh the loss of occasionally reaching the default after an incorrect signal. It follows that there is a unique sequential equilibrium where the firm offers a low wage after observing a signal s_L and a high wage after observing a signal s_H. Note that the unique sequential equilibrium is very close to the complete information SPE. Hence this mechanism can be considered robust to those information perturbations where the worker knows his own preferences.

An information perturbation where workers do not know their own preferences

Finally, consider the case where the worker's type is known by the firm and ϵ-known by the worker. In this case there are two distinct sequential equilibria. First there is a separating equilibrium, which is almost outcome-equivalent to the complete information SPE. In the first stage the firm nearly always offers a high type worker the high wage and a low type worker the low wage. Then in the second stage the workers always accept if they receive a high wage or if they receive a low wage and have a low signal. They play mixed strategies when the firm offers a low wage and they receive a high signal. If ϵ is small this third case happens rarely, and the complete information outcome is nearly always reached. In this 'trusting' sequential equilibrium, workers believe the firm is very likely to have made the appropriate offer unless they have reason to believe otherwise.

However, there is also another pooling equilibrium which leads to a very different outcome. In the first stage the firm offers all workers the high wage, and in the second stage all workers accept. To ensure that this is indeed a sequential equilibrium, it is assumed that workers have the following off-equilibrium beliefs: if the firm makes a low offer (which does not happen in equilibrium), then the worker believes he is very likely to be a high type regardless of his initial signal. This means that the off-equilibrium beliefs are such that the firm's off-equilibrium move is *much more informative than the worker's original signal*. Therefore when a worker who has received a low signal s_L receives a low offer w_L, he believes there is a significant chance that he is high type and rejects the offer. In this 'suspicious' pooling equilibrium workers do not believe that the firm has made the appropriate offer when the firm makes an off-equilibrium move. These suspicious off-equilibrium beliefs sustain what AFHKT refer to as a 'bad' sequential equilibrium.

AFHKT prove that any mechanism implementing a non-Maskin monotonic SCF in complete information is not robust to certain information perturbations. This example suggests that this result relies on the fact that players learn about their own preferences from the actions of other players. The main result of this paper formalises this. We show that bad sequential equilibria arise precisely in the case where the second mover significantly updates his belief about his own preferences from observing the other player's move. We prove that any SPE implementation in complete information which uses a two stage sequential mechanism is robust to those information perturbations where players remain certain of their own preferences. This shows that many SPE implementations in complete information are robust to the class of perturbations which are most relevant for many situations.

3 The model

There are two players $i = \{A, B\}$ and the payoff type of each player is denoted by $\theta_i \in \Theta_i$. The state is given by the pair of payoff types $\theta = (\theta_A, \theta_B) \in \Theta_A \times \Theta_B = \Theta$. We let X denote the set of allocations, while players' Bernoulli utilities are denoted by $u_i(x; \theta_i)$. These utilities depend only on the eventual allocation $x \in X$ and the player's type θ_i. It is assumed that the state space Θ and the set of outcomes X are finite. A complete information SCF f is a one to one mapping from a state to an outcome, $f: \ \Theta \mapsto X$.

Before any move is made, player A observes a signal $s^A = (s^A_A, s^A_B) \in S^A$ and player B observes a signal $s^B = (s^B_A, s^B_B) \in S^B$ where s^i_j is a signal about player $j's$ preferences. We identify the signal sets with the state space so that $S^A = S^B = \Theta$. Signals are drawn from a common prior described by $\nu \in V$, where $\nu : \Theta \times S^A \times S^B \mapsto [0, 1]$ and $\sum \nu = 1$.

We restrict our focus to extensive form mechanisms Γ with a finite number of stages where players move sequentially and every move is immediately and perfectly observed by the other player. Without loss of generality it is assumed that player A moves first, players move alternately and the number of stages is $2N$ for some $N \in \mathbb{N}$.

In any stage n, if n is odd then player A chooses a strategy $\sigma_{A,n} \in \Sigma_{A,n}$, while if n is even then player B chooses a strategy $\sigma_{B,n} \in \Sigma_{B,n}$. Therefore in the first stage player A makes a move, in the second stage player B moves and so on. Let $\sigma_A = (\sigma_{A,1}, \sigma_{A,3} .... \sigma_{A,2N-1})$ and $\sigma_B = (\sigma_{B,2}, \sigma_{B,4} .... \sigma_{B,2N})$ denote a possible set of strategies for player A and player B respectively. Furthermore let $\sigma = (\sigma_A, \sigma_B)$, and write $\Gamma(\sigma) \in X$ to mean the allocation implemented when players choose strategies σ. It is assumed that all strategy sets $\Sigma_{A,n}$ and $\Sigma_{B,n}$ are finite.

Players may condition their strategies on their signal and previously observed moves. Hence a strategy profile $h_{i,n}$ at stage n for player i maps a vector $(s^i, \sigma_{i,1}, \sigma_{i,2}, ..., \sigma_{i,n-1})$ to a strategy σ_n. A complete strategy profile h_i for player i denotes a set of strategy profiles for each stage where that player moves. Hence the strategy profile $h = (h_A, h_B)$ is a subgame perfect equilibrium (SPE) of the complete information game Γ if players have no incentive to deviate from this strategy profile.

Players initially form their beliefs based on their signal and the initial common prior. As the game progresses, players may update their beliefs after the move of an opponent. A belief profile $\phi_{i,n}$ for player i at stage n maps a vector $(s^i, \sigma_{i,1}, \sigma_{i,2}, ..., \sigma_{i,n-1})$ to a prior ν. A complete belief profile ϕ_i denotes a set of belief profiles for every stage, and $\phi = (\phi_A, \phi_B)$ denotes a pair of such belief profiles. The pair (h, ϕ) is a sequential equilibrium (SE) induced by the game (Γ, v) if ϕ represents a set of consistent beliefs given that (i) players are playing according to the strategy profile h and (ii) given their beliefs ν players have no incentive to deviate from the strategy profile h in any information set.[4]

3.1 Three informational environments

We now outline three possible restrictions on the prior ν which capture three different informational environments. First consider a complete information environment where players are certain of each others preferences. This is only the case when players always receive the correct signal about their own preferences and the preferences of their opponent. Hence we say that ν^0 is a complete information prior if ν puts probability 1 on $s^A = s^B = \theta$.

[4] This definition follows Aghion *et al.* (2012) who provide a formal definition of a sequential equilibrium in these multistage games in their online appendix.

Definition 2 (Complete information) *The prior ν^0 is a complete information prior, if and only if*

$$\sum_{\theta\in\Theta}\nu^0(\theta,\theta,\theta)=1$$

Secondly consider the environment where both players observe a highly reliable signal about the preferences of both players as studied by AFHKT. In particular suppose that the reliability of the signal is such that a player is misinformed about either the preferences of his opponent or his own preferences with a probability lower than ϵ. Therefore $s^A=\theta$ and $s^B=\theta$ with probability greater than $1-2\epsilon$, and hence we define a full (ϵ)-perturbation as follows:

Definition 3 (Full (ϵ)-perturbations) *The prior ν^ϵ is a full (ϵ)-perturbation if and only if*

$$\sum_{\theta\in\Theta}\nu^\epsilon(\theta,\theta,\theta)>1-2\epsilon$$

Finally consider an environment where players are certain of their own preferences and observe a highly reliable signal about the preferences of the other player. Suppose that players are misinformed about the preferences of his opponent with a probability lower than ϵ. As before, since players are almost always correctly informed about both their preferences and their opponent's preferences $s^A=\theta$ and $s^B=\theta$ with probability $1-2\epsilon$. However since players are certain of their own preferences there is an additional requirement, since both $s^A_A=\theta_A$ and $s^B_B=\theta_B$ with probability 1. Hence a prior ν^ϵ with restricted (ϵ)-perturbations is defined as follows:

Definition 4 (Restricted-(ϵ) perturbations) *The prior ν^ϵ is a restricted (ϵ)-perturbation if and only if*

1. *ν^ϵ is a full (ϵ)-perturbation*

2. *If $s^A_A\neq\theta_A$, then $\nu^\epsilon(\theta,s^A,s^B)=0$*

3. If $s^B_B \neq \theta_B$, then $\nu^\epsilon(\theta, s^A, s^B) = 0$

Finally define V_C to be the set of complete information priors, V^ϵ_F to be the set of full (ϵ)-perturbations and V^ϵ_R to be the set of restricted (ϵ)-perturbations. Note that $V_C \subset V^\epsilon_R \subset V^\epsilon_F$. The next two sections investigate under what conditions exact implementation and virtual implementation are robust to restricted (ϵ)-perturbations.

4 Exact implementation

We now give a definition of exact implementation in an environment with information perturbations. We say that a SCF f is robustly implementable with information perturbations if - when perturbations are sufficiently small - the desired outcome is implemented with probability one whenever players receive the correct signals.[5] Under information perturbations, the definition of exact implementation can be extended as follows:

Definition 5 *A mechanism Γ exactly implements a SCF $f : X \mapsto \Theta$ with restricted (full) perturbations if and only if given any complete information prior $\nu^0 \in V_C$ and any sequence of priors $\{\nu^\epsilon\}_{\epsilon>0}$ whenever*

1. $\nu^\epsilon \in V^\epsilon_R$ ($\nu^\epsilon \in V^\epsilon_F$)

2. The sequential equilibrium $(\sigma^\epsilon, \phi^\epsilon)$ is induced by the game (Γ, ν^ϵ)

then there exists some $\bar{\epsilon}$ such that $\Gamma(\sigma^\epsilon) = f(\theta)$ whenever i) $\epsilon < \bar{\epsilon}$ and ii) $s^A = s^B = \theta$

Using this definition, the main result of AFHKT applies in our setting:

Theorem 4.1 (AFHKT) *An SCF f can be robustly implemented with full perturbations if and only if*

[5] Note that the standard definition of exact implementation requires the desired allocation to be implemented with probability one in all cases. Under information perturbations this definition leads to trivial results, since clearly the wrong allocation will arise when players receive the wrong signals. For the analysis to be sensible, the definition is adapted to allow for other outcomes in the rare case, where players receive wrong signals.

1. *f is Maskin-monotonic*

2. *f is implementable in a complete information setting*

This result holds in a very general setting with $n \geq 2$ players, where moves may be either sequential or simultaneous. It relies on the fact that in extensive form games with several stages, additional equilibria can be formed by choosing off-equilibrium beliefs judiciously. We discussed an example of an additional bad equilibrium that arises when full perturbations are considered in the previous section. It follows that using additional stages does not increase the number of SCFs that can be implemented. As shown by AFHKT, certain small information perturbations can reduce the power of sub-game perfect implementation significantly.

However if we rule out the possibility that players are mistaken about their own preferences and only consider this smaller class of restricted perturbations, the situation is not nearly so bleak. Our example has already shown that the implementability of some SCFs are robust to restricted perturbations and not full perturbations. We now generalise this result and give a sufficient condition for exact implementation under restricted perturbations.

4.1 Sufficient condition

In this section we introduce a sufficient condition for exact implementation with restricted information perturbations which is significantly weaker than Maskin-monotonicity. This shows that restricting the set of information perturbations in an intuitive way significantly increases the set of SCFs that are robustly implementable. We first make the following definition:

Definition 6 (F_2) *An SCF $f \in F_2$ if it can be implemented under complete information by a two stage mechanism with sequential moves.*

We now state our sufficient condition for robust implementation with restricted information perturbations:

Theorem 4.2 (Sufficiency) *If an SCF $f \in F_2$, then it can be robustly implemented with restricted information perturbations.*

In order to prove this result we first characterize the SPEs under full information in two stage sequential move games. We have to show that the strategy profile used in any sequential equilibrium with sufficiently small restricted information perturbations coincides with a SPE in complete information. It is easy to show that the second mover - assuming he receives the correct signal - chooses his strategy in the same way as he does under complete information, because when making his decision, the second mover has not updated his preferences and simply chooses the allocation he likes most. Given that the second-mover behaves as he does under complete information, it is then possible to show that the first-mover also behaves as he does under complete information as long as his signal is correct and perturbations are sufficiently small. The complete proof can be found in the appendix.

4.2 Comparison with complete information

In order to illustrate that restricted perturbations provide an appropriate criterion for distinguishing between SCFs which are seen to be implemented in practice and those that are not, we now provide a comparison with the case of complete information. We show that robustness to restricted perturbations is more restrictive than implementation under complete information.

We consider the canonical mechanism introduced by Moore & Repullo (1988). Although this mechanism can be used to implement a wide-range of SCFs under complete information, it is not robust to restricted perturbations. More precisely there exist SCFs which can be exactly implemented using this mechanism under complete information, but cannot exactly be implemented under restricted perturbations. Hence

Preferences	
Firm: $\theta \in \{L, H\}$	$u_P(w_L;\theta) > u_P(x_H;\theta) > u_P(w_H;\theta) > u_P(x_L;\theta)$
Low type: $\theta = L$	$u_A(w_H;\theta) > u_A(w_L;\theta) > u_A(x_L;\theta) > u_A(x_H;\theta)$
High type: $\theta = H$	$u_A(w_H;\theta) > u_A(w_L;\theta) > u_A(x_H;\theta) > u_A(x_L;\theta)$

Table 1: Example: Simple three stage mechanism: Implementable under complete information, not implementable under restricted perturbations

exact implementation under restricted perturbations is a more restrictive criterion than exact implementation under complete information. In particular many SCFs that require complex mechanisms to be implemented under complete information can not be implemented when allowing for restricted information perturbations.

This is illustrated using the following example. Again consider a setting where a firm denoted by P wants to hire a a worker denoted by A. The worker may be a high type or a low type. In this example there are two outside options denoted x_H and x_L respectively. The players' preferences are given in Table 1. Now consider the mechanism represented in Figure 2.

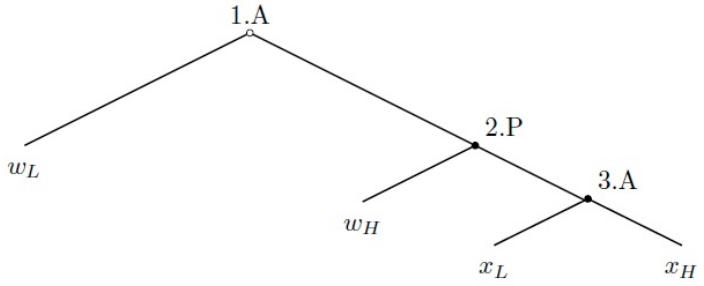

Figure 2: Moore and Repullo mechanism

Under complete information this Moore and Repullo mechanism implements the SCF where the high type worker receives w_L and the low type worker receives w_H. However note that the separating equilibrium implemented under complete information is not robust to restricted information perturbations. If the firm believes that it faces a high type whenever a worker starts by choosing the branch on the right, the firm will react by offering the worker the choice between the two outside options. This creates a 'bad' pooling sequential equilibrium in which all workers receive w_L. The SCF where the low type worker receives w_H and the high type worker receives w_L, is therefore not robust to restricted information perturbations. Hence this shows that the canonical Moore-Repullo mechanism is not robust to restricted perturbations.[6].

Other examples of two stage sequential move mechanisms seen in practice include a decision on a public good, where one agent announces how much he is willing to contribute, before a second agent decides to raise the amount to the critical threshold or to not contribute. Alternatively one can think of a principal agent setting, where the principal offers a menu of contracts and the agent chooses his preferred contract.

One should note that implementability in two stages under complete information is sufficient for exact implementation with restricted perturbations, but is not necessary. In the appendix we present an example of an SCF that can be exactly implemented in three stages with restricted perturbations but not in two stages.[7]

5 Virtual Implementation

In this section we show that the range of SCFs that are robust to restricted information perturbations becomes even larger, when considering the weaker concept of virtual implementation. Formally virtual implementation requires that for each $\epsilon > 0$ there

[6] Note that this does not prove that the SCF cannot be implemented robustly. But it cannot be implemented robustly using the mechanism suggested by Moore & Repullo (1988)

[7] However, these examples are rare and difficult to construct. In particular the example we present is such that by allowing for simultaneous move in the first stage and then allowing one of the players to move again in the second stage, the SCF can be implemented in two stages. Hence by weakening condition $F2$ to implementability in two stages where the first stage allows for simultaneous moves, while only one player moves in the second stage.

exists a nearby game Γ^ϵ such that in any sequential equilibrium of this game the desired outcome is obtained with probability greater than $1-\epsilon$. This is weaker than the concept of exact implementation considered previously, since we now allow for the possibility that the desired outcome is occasionally not implemented even in cases when both players receive the correct signals. More precisely:

Definition 7 *A mechanism Γ virtually implements an SCF $f : X \mapsto \Theta$ with restricted (full) perturbations if and only if given any $\delta > 0$, any complete information prior $\nu^0 \in V_C$ and any sequence of priors $\{\nu^\epsilon\}_{\epsilon>0}$ whenever*

1. *$\nu^\epsilon \in V_R^\epsilon$ ($\nu^\epsilon \in V_F^\epsilon$)*
2. *The sequential equilibrium $(\sigma^\epsilon, \phi^\epsilon)$ is induced by the game (Γ, ν^ϵ)*

then there exists some $\bar{\epsilon}$ such that $P\Big(\Gamma(\sigma^\epsilon) = f(\theta)\Big) > 1 - \delta$ whenever $\epsilon < \bar{\epsilon}$

Most previous work on virtual implementation - see Serrano & Vohra (2010) for an exception - considers stochastic mechanisms where in equilibrium players play according to pure strategies. In these cases the slight uncertainty over the eventual outcome is caused by the stochasticity of the mechanism. In contrast, in the examples considered below slight uncertainty over the eventual outcome is caused by the fact that players do not play pure strategies, but rather play *almost pure* strategies, allowing them to deviate from the main strategy prescribed for their type occasionally.

Virtual implementation under restricted perturbations is less permissive than virtual implementation under complete information, while being more permissive than exact implementation under restricted perturbations. To show the first part of this claim it is sufficient to consider the canonical Moore-Repullo mechanism analysed above. It can immediately be seen that this mechanism - and hence the canonical mechanism - is not robust to restricted perturbations even when considering the weaker criterion of virtual implementation. This follows from the fact that this mechanism has a pooling equilibrium, as explained in the previous section. Whenever 'bad' sequential equilibria arise from pooling, both virtual implementation and exact implementation fail.

To show the second part of this claim we provide an example of an SCF which cannot be robustly implemented under restricted perturbations if exact implementation is required, but is robust when requiring only virtual implementation. This difference follows from the fact that exact implementation requires the complete information allocation to be implemented whenever both players receive the correct signal. Virtual implementation allows rare occasions where players deviate from their complete information strategy in which case a different allocation is implemented despite both players receiving the correct signal. Robust virtual implementation requires these 'differences' to become increasingly rare as signal precision increases. An example of such a setting is discussed below.

5.1 Comparison with exact implementation

We now give an example of an SCF which can be virtually implemented robustly, but cannot be exactly implemented robustly. Note also that the example is constructed such that the SCF can be virtually implemented robustly using a three stage mechanism, even though it cannot be virtually implemented using a two-stage mechanism.

Let $\Theta = \{L, H\}$, $X = \{w_L, w_H, x_H, x_L, y_H, y_L\}$ and consider the preference profile given in Table 2.

Now consider the SCF $f : \Theta \mapsto X$, where $f(H) = w_L$ and $f(L) = w_H$. This SCF is implementable using restricted perturbations but it is not implementable in a two stage sequential move mechanism in complete information.

To show that this SCF can be virtually implemented using restricted perturbations, consider the mechanism represented in Figure 3. This mechanism virtually implements

Preferences	
Firm: $\theta \in \{L, H\}$	$u_P(y_L;\theta) > u_P(w_L;\theta) > u_P(x_H;\theta) > u_P(w_H;\theta) > u_P(x_L;\theta) > u_P(y_H;\theta)$
Low type: $\theta = L$	$u_A(w_H;\theta) > u_A(w_L;\theta) > u_A(x_L;\theta) > u_A(x_H;\theta) > u_A(y_L;\theta) > u_A(y_H;\theta)$
High type: $\theta = H$	$u_A(w_H;\theta) > u_A(w_L;\theta) > u_A(x_H;\theta) > u_A(x_L;\theta) > u_A(y_H;\theta) > u_A(y_L;\theta)$

Table 2: Complex three stages: Virtually implementable under restricted perturbations not exactly implementable under restricted perturbations

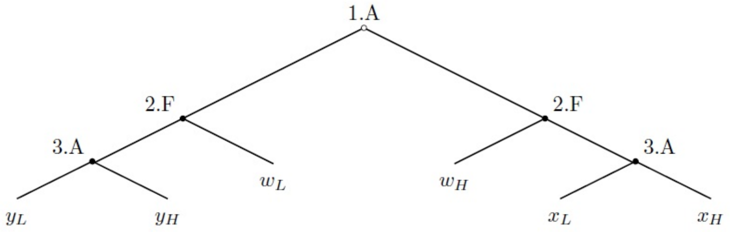

Figure 3: Complex three stage mechanism

the SCF described above both under complete information and with restricted perturbations. The extra off equilibrium outcomes y_L and y_H ensure that the bad sequential equilibrium that arises in the three stage example described in the previous section does not arise here. Note that in complete information this mechanism implements the allocation w_H if the worker is type L and w_L if the worker is type H.

When restricted information perturbations are introduced, the mechanism fails to implement this SCF exactly. To see this, consider the following equilibrium. Define m_L to be the proportion of low types and m_H to be the proportion of high types. Suppose perturbations happen with probability at most ϵ and that ϵ is sufficiently small. Finally choose mixing probabilities α and β such that the following equations are satisfied:

$$u_P(w_H) = (1-\alpha)m_H\nu(s^H|\theta_H)u_P(x_H) + m_L\nu(s^H|\theta_L)u_P(x_L)$$

$$u_H(w_L) = \Big[\nu(s^L|\theta_H) + \beta\nu(s^H|\theta_H)\Big]u_H(w_H) + (1-\beta)\nu(s^H|\theta^H)u_H(x_H)$$

In the first stage all low types choose the right branch. Meanwhile high types mix, with a proportion α choosing the left branch and a proportion $(1-\alpha)$ choosing the right branch. In the second stage if the worker chose the left branch the firm always chooses w_L. Meanwhile if the worker chose the right branch and the firm observes a signal s^L the firm always chooses w_H. Finally if the worker chose the right branch and the firm observes a signal s^H the firm mixes: with probability β the firm chooses w_H while with probability $(1-\beta)$ the firm proceeds to the third stage. In the third stage a high type worker chooses x_H or y_H while a low type worker chooses x_L or y_L.

It can be easily checked that the strategy profile above outlines a SPE whenever $\epsilon > 0$. In the appendix it is proved that this is indeed the unique SPE. Note that in the first round high types mix between choosing the left branch and the right branch, and so this mechanism does not exactly implement the desired SCF under restricted perturbations. However as $\epsilon \to 0$, then $\alpha \to 1$ where α denotes the fraction of high types who choose the left branch in the first round. This - together with the fact that the SPE outlined above is unique - shows that this mechanism does virtually implement the desired SCF under restricted perturbations. In particular if perturbations are sufficiently small, then the proportion of high types imitating low types can be made to be arbitrarily small. Hence the desired allocation is reached in almost all cases.

This example shows that when exact implementation is prevented by the behaviour of a small proportion of types, allowing players to mix with small probabilities, virtual implementation (as defined above) may still be possible. Note that as the precision of the signal increases, the proportion of players deviating from the complete information equilibrium becomes small. On the one hand - as shown in the previous section - exact and virtual implementation under restricted perturbations coincide when implementation is prevented by the creation of fully pooling 'bad' sequential equilibria. On

	Exact Implementation	Virtual Implementation
Full Perturbations	Maskin Monotonic	Maskin Monotonic
Restricted Perturbations	Two-stage mechanisms and Example 3	Also Example 2
Complete Information	Condition C	Condition C

Table 3: Summary (Example 3 can be found in the appendix)

the other hand, there exist other cases - particularly when perturbations only slightly change equilibrium outcomes - where virtual implementation is more permissive than exact implementation.

6 Discussion

The central message of this paper is that the power of SPE implementation depends on the relevant set of information perturbations and the strength of implementation required. At one extreme, if information perturbations are irrelevant and there is complete information, a wide range of SCFs can be implemented using Moore-Repullo mechanisms. Meanwhile, at the other extreme, if full perturbations are relevant, then AFHKT show that only Maskin-monotonic SCFs can be implemented. In this paper we have considered the intermediate case of restricted perturbations and provide results which lie somewhere between these two extremes. These results are summarised in Table 3.

The exact power of implementation under restricted perturbations depends on whether virtual implementation or exact implementation is required. One argument for considering virtual implementation is that the definition of exact implementation already allows for mistakes in the rare case when players receive the wrong signal. Hence the concept of exact implementation given here is already - in some sense - a restricted type of virtual implementation, and so it seems natural to instead consider the full version of virtual implementation instead. Meanwhile, an argument for considering exact implementation is that it requires players to follow pure strategies, which are more intuitive than the *almost* pure strategies players follow when considering virtual implementation.

There are two ways in which the results presented here could be easily extended. First the sufficiency result stated here can be extended to an n-player framework where each player moves exactly once. One extra restriction would be necessary: players who move earlier must not be able to communicate information about the preferences of any player who moves later. The proof would be very similar to the two-player case, albeit with extra notation.

The second extension involves considering a class of perturbations wider than those considered in this paper, but still more restricted than than full information restrictions. Note that the formation of 'bad' sequential equilibria relies on players changing their beliefs about their own type to a significant extent. Therefore the results above are also robust to a more general class of restricted perturbations. In particular consider the case where the second-mover receive a signal about their own preferences which is highly (but not perfectly) reliable, while the first-mover receives a significantly less reliable signal. In these cases the second-mover is much more informed than the first-mover, and hence only updates his beliefs about his own preferences by a small amount. This ensures 'bad' sequential equilibria cannot be formed, and that two-stage implementations continue to be robust.

7 Appendix

7.1 Proof of Proposition 4.2

Before proving this theorem we introduce some additional notation and definitions. We use $h_B(s^B, \sigma_A) \in \Sigma_B$ to denote the strategy chosen by player B when he observes signal s^B and player A has chosen strategy σ_A. Hence, $h_B \in H_B$ is a strategy profile of player B, where H_B is the set of all such profiles.

Meanwhile $h_A(s^A, h_B) \in \Sigma_A$ denotes the strategy chosen by player A when he observes signal s^A and expects player B to play according to strategy profile h_B. Hence $h_A \in H_A$ denotes a strategy profile determining the choice of player A when he observes a certain signal and has a certain belief about the strategy profile of player B. H_A is the set of all such strategy profiles. We now define H_B^* and $H_A^*(h_B)$, which denote the possible SPE strategy profiles that occur in a complete information setting:

Definition 8 *$h_B \in H_B^*$ if and only if for all σ_A, for all θ and for all $\hat{\sigma}_B \in \Sigma_B$*

$$u_B(\Gamma(\sigma_A, h_B(\theta, \sigma_A)); \theta_B) \geq u_B(\Gamma(\sigma_A, \hat{\sigma}_B); \theta_B)$$

Definition 9 *$h_A \in H_A^*(h_B)$ if and only if for all θ and for all $\hat{\sigma}_A \in \Sigma_A$*

$$u_A(\Gamma(\sigma_A, h_B(\theta, \sigma_A)), \theta_A) \geq u_A(\Gamma(\hat{\sigma}_A, h_B(\theta, \hat{\sigma}_A)), \theta_A)$$

In a complete information setting with the complete information prior ν, the following proposition is immediately implied by the definitions above:

Proposition 7.1 *(h_A, h_B) denote a SPE of Γ iff $h_B \in H_B^*$ and $h_A \in H_A^*(h_B)$*

This characterizes the SPEs under full information in two stage sequential move games. Note that any sequential move game with finite strategy sets has at least one equilibrium. Hence in order to prove proposition 4.2 it is sufficient to show that the strategy profile used in any sequential equilibrium with sufficiently small restricted information perturbations coincides with a SPE in complete information.

To do this consider a game with restricted information perturbations (Γ, ν^{ϵ}) and corresponding sequential equilibrium strategy profiles $(h^{\epsilon}_A, h^{\epsilon}_B)$. It is sufficient to prove that for some $\bar{\epsilon} > 0$, $h^{\epsilon}_B \in H^*_B$ and $h^{\epsilon}_A \in H^*_A(h_B)$ whenever $\epsilon \leq \bar{\epsilon}$. The proof is now split into two parts.

First we prove that $h^{\epsilon}_B \in H^*_B$. This follows from the fact that player B knows his own preferences with certainty and hence in response to player A's move chooses his preferred alternative. [8]

Secondly we prove $h^{\epsilon}_A \in H^*_A(h^{\epsilon}_B)$. The proof relies on the fact that player A knows his own type with certainty and estimates the type of player B correctly with probability $(1-\epsilon)$. Hence as $\epsilon \to 0$ the incentives of player A are very similar to the incentives he has in complete information. In particular the probability ϵ event where he estimates the type of player B incorrectly becomes relatively unimportant.

We slightly abuse notation by defining $u_A(\sigma_A, \sigma_B; \theta_A) := u_A(\Gamma(\sigma_A, \sigma_B); \theta_A)$. Moreover throughout the proof we use the fact that perturbations are restricted: that is to say $s^A_A = \theta_A$ and $s^B_B = \theta_A$.

7.1.1 Proof of 4.2 Part (i): $h^{\epsilon}_B \in H^*_B$

Proof.

Suppose this does not hold. Then for some signal $\tilde{s}^B$, and strategy $\tilde{\sigma}_A$ there exists a deviating strategy $\hat{\sigma}_B$ such that:

$$u_B(\tilde{\sigma}_A, h^{\epsilon}_B(\tilde{s}^B, \tilde{\sigma}_A); \tilde{s}^B_B) < u_B(\tilde{\sigma}_A, \hat{\sigma}_B; \tilde{s}^B_B)$$

[8] Note that this is the part of the proof that does not hold in the setting AFHKT consider. In their setting player B may infer something about his own preferences from the move of player A. In particular, $u_B(\Gamma(\sigma_A, \sigma_B); \theta_B) \neq u_B(\Gamma(\sigma_A, \sigma_B); s^B_B)$.

Since information perturbations are restricted $s^B_B = \theta_B$, and it follows that:

$$u_B(\tilde{\sigma}_A, h^\epsilon_B(\tilde{s}^B, \tilde{\sigma}_A); \theta_B) < u_B(\tilde{\sigma}_A, \hat{\sigma}_B; \theta_B)$$

Consider the following strategy profile:

$$\hat{h}^\epsilon_B(s^B, \sigma_A) = \begin{cases} \hat{\sigma}_B \quad \text{if} \quad (s^B, \sigma_A) = (\tilde{s}^B, \tilde{\sigma}_A) \\ h^\epsilon_B(s^B, \sigma_A) \quad \text{otherwise} \end{cases}$$

Playing according to strategy profile $\hat{h}^\epsilon_B$ rather than strategy profile h^ϵ_B leads to a higher payoff in the subgame when $(s^B, \sigma_A) = (\tilde{s}^B, \tilde{\sigma}_A)$ and the same payoff otherwise. Hence h^ϵ_B cannot be a sequential equilibrium profile of the game (Γ, ν^ϵ). This is a contradiction, and completes the proof.

□

7.1.2 Proof of 4.2 Part (ii): $h^\epsilon_A \in H^*_A(h^\epsilon_B)$

Proof.

First define the following:

$$\begin{aligned} \underline{u} &:= \min_{x, s^A}\{u_A(x; s^A_A)\} \\ \overline{u} &:= \max_{x, s^A}\{u_A(x; s^A_A)\} \\ \sigma_A(s^A) &= h^\epsilon_A(s^A, h^\epsilon_B) \\ u(s^A) &:= u_A(\sigma_A(s^A), h^\epsilon_B(s^A, \sigma_A(s^A)); s^A_A) \\ \hat{u}(s^A) &:= \max_{\tilde{\sigma}_A}\{u_A(\tilde{\sigma}_A, h^\epsilon_B(s^A, \tilde{\sigma}_A); s^A_A)\} \end{aligned}$$

We use $\overline{u}$ and $\underline{u}$ to refer to the maximum and minimum payoffs player A could receive, while $u(s^A)$ is the utility player A obtains when he plays according to strategy $\sigma_A(s^A) = h_A(s^A, h^\epsilon_B)$, player B has the same signal as him ($s^B = s^A$) and plays according to a strategy profile h^ϵ_B. Meanwhile $\hat{u}(s^A)$ is the maximum utility player A could obtain in this situation by choosing some arbitrary strategy. Let $\hat{\sigma}_A(s^A)$ be one of these maximizing strategies, so that $\hat{u}(\theta) = u_A(\hat{\sigma}_A(\theta), h^\epsilon_B(\theta, \hat{\sigma}_A(\theta)); \theta)$.

Now suppose $h^\epsilon_A \notin H_A(h^\epsilon_B)$. Remembering that h_A is a strategy profile of a sequential equilibrium, we aim for a contradiction. Since $h^\epsilon_A \notin H_A(h^\epsilon_B)$, it follows that for some signal $\tilde{s}^A$ there exists a profitable deviation $\tilde{\sigma}_A$. That is to say:

$$u_A(\sigma_A(\tilde{s}^A), h^\epsilon_B(\tilde{s}^A, \sigma_A(\tilde{s}^A)); \tilde{s}^A_A) < u_A(\tilde{\sigma}_A, h^\epsilon_B(\tilde{s}^A, \tilde{\sigma}_A); \tilde{s}^A_A) \tag{1}$$

Using the definition of $\hat{\sigma}_A$, note that the strategy $\hat{\sigma}_A(s^A)$ maximizes the payoff of player A given his signal is s^A. Therefore:

$$u_A(\tilde{\sigma}_A, h^\epsilon_B(\tilde{s}^A, \tilde{\sigma}_A); \tilde{s}^A_A) \leq u_A(\hat{\sigma}_A(\tilde{s}^A), h^\epsilon_B(\tilde{s}^A, \hat{\sigma}_A(\tilde{s}^A)); \tilde{s}^A_A) \tag{2}$$

Putting these equations 1 and 2 together and using the definition of $u(s^A)$ and $\hat{u}(s^A)$ leads to the following:

$$\begin{aligned} u_A(\sigma_A(\tilde{s}^A), h^\epsilon_B(\tilde{s}^A, \sigma_A(\tilde{s}^A)); \tilde{s}^A_A) &< u_A(\hat{\sigma}_A(\tilde{s}^A), h^\epsilon_B(\tilde{s}^A, \hat{\sigma}_A(\tilde{s}^A)); \tilde{s}^A_A) \\ u(\tilde{s}^A) &< \hat{u}(\tilde{s}^A) \end{aligned}$$

Now let $\delta = \hat{u}(\tilde{s}^A) - u(\tilde{s}^A)$ and note that $\delta > 0$. Define an alternative strategy profile $\hat{h}^\epsilon_A$ as follows:

$$\hat{h}^\epsilon_A(s^A) = \begin{cases} \hat{\sigma}_A(s^A) & \text{when} \quad s^A = \tilde{s}^A \\ \sigma_A(s^A) & \text{when} \quad s^A \neq \tilde{s}^A \end{cases}$$

We now show that $\hat{h}^\epsilon_A$ is a profitable deviation. When $s^A \neq \tilde{s}^A$, payoffs under both strategy profiles are equal under both strategy profiles z, so we focus on the case where $s^A = \tilde{s}^A$. Note that in this case $\hat{h}^\epsilon_A(s^A) = \hat{\sigma}_A(\tilde{s}^A)$ and $h^\epsilon_A(s^A, h^\epsilon_B) = \sigma_A(\tilde{s}^A)$. Since information perturbations are restricted, $\theta_A = \tilde{s}^A_A$. Hence it is enough to show that:

$$S = E_{s^B \in \Theta_B}[u_A(\hat{\sigma}_A(\tilde{s}^A), h^\epsilon_B(s^B, \hat{\sigma}_A(\tilde{s}^A)); \tilde{s}^A_A)] - E_{s^B \in \Theta_B}[u_A(\sigma_A(\tilde{s}^A), h^\epsilon_B(s^B, \sigma_A(\tilde{s}^A)); \tilde{s}^A_A)] > 0$$

First note that with probability $p > (1-\epsilon)$, $s^B = \tilde{s}^A$. In this case the left hand side is equal to $\hat{u}(\tilde{s}^A)$, while the right-hand side is equal to $u(\tilde{s}^A)$. Moreover with probability ϵ any payoff between $u_A \in [\underline{u}, \overline{u}]$ may be obtained. These observations lead to the following bounds:

$$\begin{aligned} E_{s^B \in \Theta_B}[u_A(\hat{\sigma}_A(\tilde{s}^A), h^\epsilon_B(s^B, \hat{\sigma}_A(\tilde{s}^A)); \tilde{s}^A_A)] &\geq (1-\epsilon)\hat{u}(\tilde{s}^A) + \epsilon\underline{u} \\ E_{s^B \in \Theta_B}[u_A(\sigma_A(\tilde{s}^A), h^\epsilon_B(s^B, \sigma_A(\tilde{s}^A)); \tilde{s}^A_A)] &\leq (1-\epsilon)u(\tilde{s}^A) + \epsilon\overline{u} \end{aligned}$$

Using these bounds, the fact that $\delta = \hat{u}(\tilde{s}^A) - u(\tilde{s}^A) > 0$ and assuming $\epsilon < \frac{1}{2}$ gives:

$$\begin{aligned} S &\geq (1-\epsilon)\hat{u}(\tilde{s}^A) + \epsilon\underline{u} - (1-\epsilon)u(\tilde{s}^A) - \epsilon\overline{u} \\ &> \delta - 2\epsilon(\overline{u} - \underline{u}) \end{aligned}$$

$\delta > 0$ and both δ and $(\overline{u} - \underline{u})$ are fixed parameters. Therefore there exists some $\overline{\epsilon}$ such that $S > 0$ whenever $\epsilon \in (0, \overline{\epsilon})$. This shows that $\hat{h}^\epsilon_A$ is a profitable deviation and hence h^ϵ_A cannot be the strategy profile of a sequential equilibrium. This proves the result. □

7.2 Example: $F2$ is sufficient but not necessary

Consider again the initial example of the firm and the worker. Now however there is a third type of worker, ($\theta_B = S$). This worker has an outside option that he prefers

	Preferences	
Norm fi	$u_P(w_L;\theta) > u_P(w_H;\theta) > u_P(d;\theta) > u_P(S;\theta)$	for $\theta \in \{(N,L),(N,H),(Y,L),(Y,H)\}$

Spec fi	$u_P(w_L;\theta) > u_P(w_H;\theta) > u_P(S;\theta) > u_P(d;\theta)$
	$u_P(w_L;\theta) > u_P(w_H;\theta) > u_P(S;\theta) > u_P(d;\theta)$
		for $\theta \in \{Y,S\}$
		for $\theta \in \{(N,L),(N,H)\}$

Low t	$u_A(S,\theta) > u_A(w_H;\theta) > u_A(w_L;\theta) > u_A(d;\theta)$	for $\theta_B = L$
High t	$u_A(S,\theta) > u_A(w_H;\theta) > u_A(d;\theta) > u_A(w_L;\theta)$	for $\theta_B = H$
Spec t	$u_A(S,\theta) > u_A(w_H;\theta) > u_A(d;\theta) > u_A(w_L;\theta)$	for $\theta_B = S$

Table 4: Example: F2 is not necessary

to w_H, but otherwise has the same preferences as the high type worker. This outside option can be thought of as another job offer with a high salary. In case he does not reach an agreement with the firm he takes the outside offer. Also suppose that there are two types of firms ($\theta_A \in \{Y,N\}$). One firm would like to hire this special worker by offering him a wage that is even higher than the outside option. The other type of the firm does not want to pay such a high wage.

The references are given in Table 4.

The social choice function where $f(N,L) = f(Y,L) = w_L$, $f(N,H) = f(Y,H) = w_H$, $f(N,S) = d$ and $f(Y,S) = S$ can be implemented in three stages in complete information, where the worker first chooses between the special branch and the normal branch. In case the worker has chosen the special branch, the firm decides to pay a very high wage S if the worker is indeed the special type and the firm is special, too. It chooses outside option d otherwise. On the other hand, if the worker chooses the normal branch, the game continues as in the basic example.

If the proportion of special firms is sufficiently small and normal workers dislike allocation O sufficiently, then this mechanism is robust to restricted perturbations and the SCF can be implemented robustly, despite requiring three stages. Note however, that this mechanism can be reduced to two stages, when allowing players to move simultaneously in the first stage. Workers report that they are *normal* or *special* and the firm chooses one of S and the default d and one of w_H and w_L. If the worker chooses the special branch the game ends and S or d as chosen by the firm is implemented. If the worker chooses the normal branch then if the firm chose w_H this is implemented. In the final case, where the worker has chosen the normal branch and the firm chose w_L, the worker gets to make a final choice between accepting w_L and rejecting the offer to implement the default d.

7.3 Simultaneous moves

We now provide an example to show that the credible threat condition is not necessary for robust implementation under restricted information perturbations when allowing players to move simultaneously.

Consider the case where there are two players A and B. For simplicity assume that the preferences of player B are fixed, while player A's preferences are given by θ or $\hat{\theta}$. We assume that player A knows his preferences with certainty while the signal player B receives is equal to player A's preferences with probability $1-\epsilon$ and equal to the other preference with the remaining probability ϵ. Now consider the mechanism described by Figures 4 and 5.

In the first stage of the game both players simultaneously choose between reporting θ and reporting $\hat{\theta}$. This is described in Figure 4. If both players report $\hat{\theta}$ then the game ends and players receive the payoffs given in brackets. The first number corresponds to the payoff of player A if he is type θ, the second number is the payoff of player A if he is type $\hat{\theta}$ and the third number is the payoff of player B. Similarly if player B reports $\hat{\theta}$ and player A reports θ, the payoffs are $(0,0,0)$ and the game ends.

		B	
		θ	$\hat{\theta}$
A	θ	Γ	$(0, 0, 0)$
	$\hat{\theta}$	$\theta : (1, 0, 10)$ $\hat{\theta} : (0, 1, 0)$	$(7, 7, 3)$

Figure 4: Simultaneous moves

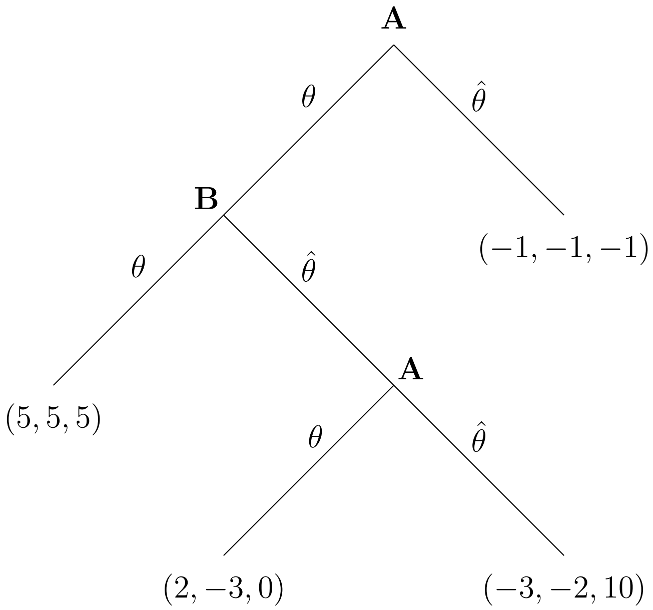

Figure 5: Mechanism Γ

Now consider the case where player A reports $\hat{\theta}$ and player B reports θ. In this case player A has got a second move and chooses again between the reports θ and $\hat{\theta}$ which correspond to payoff vectors of $(1, 0, 10)$ and $(0, 1, 0)$ respectively.

In the case where both players report θ, they start playing the mechanism Γ given by the game tree in Figure 5 in the second stage.

It can easily be checked that the underlying preferences do not satisfy the credible threat condition.

We now show that despite this fact, the simultaneous move mechanism described above robustly implements the social choice function with payoffs $(5, 5)$ in state θ and $(7, 3)$ in state $\hat{\theta}$ under restricted information perturbations. For simplicity we assume that the states θ and $\hat{\theta}$ are ex-ante equally likely.

First note that the unique equilibria under complete information are given by the reports $(\theta, \theta, \theta, \theta)$ in state θ and $(\hat{\theta}, \hat{\theta})$ in state $\hat{\theta}$. Hence the desired SCF is implemented under complete information.

Now consider the case where player A's realised preferences are θ. If the mechanism Γ is reached, player A has got a dominant strategy to re-report his preferences as θ. Moreover whenever player A's preferences are θ his initial report is θ. This ensures him a payoff of 2 which is greater than any payoff he can hope to achieve by reporting $\hat{\theta}$, since the reports $(\hat{\theta}, \hat{\theta})$ are not an equilibrium. Knowing this, player B assigns a high probability to player A's preferences being θ whenever he observes A re-reporting himself as θ and mechanism Γ is played. As a consequence B also reports θ and the desired allocation is implemented. There cannot be a case, where player A re-reports his preferences as θ and player B then assigns a higher probability to A's preferences being $\hat{\theta}$ than before the first stage.

Secondly consider the case where player A's preferences are given by $\hat{\theta}$. Then the reports $(\hat{\theta}, \hat{\theta})$ are an equilibrium. Player A cannot gain by deviating as there does not exist an allocation which gives him a higher payoff. Player B cannot gain by deviating to another report either: If he reports θ player A has got another move where he has a dominant strategy to re-report $\hat{\theta}$, leaving player B with a payoff $0 < 3$. Hence the report $(\hat{\theta}, \hat{\theta})$ is an equilibrium if the state is $\hat{\theta}$.

Note also that it is the only equilibrium in this state. In particular the mechanism Γ played when the reports are (θ, θ) cannot be an equilibrium, as it would implement an allocation $(-1, -1)$, which neither of the players likes.

7.4 Virtual Implementation

We now prove that the SPE equilibrium in mixed strategies stated in section 5.1 is indeed the unique equilibrium of the mechanism described and hence virtually implements the desired SCF. **Proof.** Let $\delta = \sqrt{\epsilon}$ and suppose ϵ is sufficiently small. In this case, if more than fraction δ of low types choose the left branch, the principal - on observing signal s^L will challenge the report. This is because the report is sufficiently likely to originate from a low type and hence:

$$u_P(w_L) < \frac{\delta(1-\epsilon)m_L}{\delta(1-\epsilon)m_L + \epsilon m_H} u_P(y_L) + \frac{\epsilon m_H}{\delta(1-\epsilon)m_L + \epsilon m_H} u_P(y_H)$$

Secondly note that if more than fraction δ of low types choose the right branch, the principal - on observing signal s^L will accept the report. This is because the report is sufficiently likely to originate from a low type and hence:

$$u_P(w_H) > \frac{\delta(1-\epsilon)m_L}{\delta(1-\epsilon)m_L + \epsilon m_H} u_P(x_L) + \frac{\epsilon m_H}{\delta(1-\epsilon)m_L + \epsilon m_H} u_P(x_H)$$

Suppose there is a SPE where more than δ low types choose the left branch in the first round. Then these low types with probability greater than $(1-\epsilon)$ would receive payoff $u_L(y_L)$. If ϵ is sufficiently low, it is optimal for these low types to deviate and choose the right branch in the first round guaranteeing a payoff higher than $u_L(y_L)$. It follows that in any SPE a fraction at least $(1 - \delta)$ low types chooses the right branch in the first stage.

Since a high fraction of low types report L in the first round, it follows from above that the principal - on observing a report of L and signal s^L - will always accept the report and implement w_H. Since this is the highest payoff a low type can receive it follows that all low types will report L in the first round.

Since only high types choose the left branch, it follows that the firm will accept to pay w_L, whenever a worker chooses the left branch in the first stage. Therefore high type workers have a choice between (i) choosing the left branch and receiving a guaranteed payoff of $u_H(w_L)$ and (ii) choosing the right branch. Suppose all high types choose the right branch. Then the firm - on observing a worker has chosen the right branch and a signal s^H - will challenge the worker by moving to the third stage - and x_H will be implemented. In this case high types - preferring w_H to x_H - would have an incentive to deviate and choose the left branch initially. Suppose now on the other hand that all high types choose the left branch. Then the firm - on observing that the right branch has been chosen and a signal s^H - will not challenge and w_H will be implemented. In this case high types - preferring w_H to w_L - would have an incentive to deviate.

It follows from the two observations above that high types must mix in the first stage. Moreover for high types to be indifferent over their mixing, it follows that the principal must mix in the second stage after observing a report L and a signal s^H. The mixing parameters α and β are calculated above, and hence this is the unique SPE.

□

References

Abreu, D., & Matsushima, H. 1994. Exact Implementation. *Journal of Economic Theory*, **64**, 1–19.

Abreu, D., & Sen, A. 1991. Virtual Implementation in Nash Equilibrium. *Econometrica*, **59**, 997–1021.

Aghion, P., Fudenberg, D., Holden, R., Kunimoto, T., & Tercieux, O. 2012. Subgame Perfect Implementation under Information Perturbations. *Quarterly Journal of Economics*, 1843–1881.

Baliga, S. 1999. Implementation in Economic Environments with Incomplete Information: The Use of Multi-Stage Games. *Games and Economic Behaviour*, **27**, 173–183.

Bergin, S., & Sen, A. 1998. Extensive Form Implementation in Incomplete Information Environments. *Journal of Economic Theory*, **80**, 222–256.

Bester, H., & Kraehmer, D. 2012. Exit options in incomplete contracts with asymmetric information. *Journal of Economic Theory*, **147**, 1947–1968.

Corchón, L., & Ortuno-Ortín. 1995. Robust implementation under altemative information structures. *Review of Economic Design*, **1**, 159–171.

Moore, J., & Repullo, R. 1988. Subgame Perfect Implementation. *Econometrica*, **56**, 1191–1220.

Rubinstein, A. 1982. Perfect Equilibrium in a Bargaining Model. *Econometrica*, **50**(1), 97–109.

Serrano, R., & Vohra, R. 2010. Multiplicity of Equilibria in Mechanisms: a Unified approach to Exact and Approximate Implementation. *Journal of Mathematical Economics*, **46**, 775–785.

Yamato, T. 1994. Equivalence of Nash implementatbility and Robust implementatbility with incomplete information. *Social Choice and Welfare*, **11**, 289–303.