Safe Implementation in Mixed Nash Equilibrium*

Anand Chopra† — University of Liverpool
Malachy James Gavan‡ — University of Liverpool

Antonio Penta§
ICREA-UPF and BSE

July 23, 2025

Abstract

Safe Implementation (Gavan and Penta, 2025) combines standard implementation with the requirement that the implementing mechanism is such that, if up to k agents deviate from the relevant solution concept, the outcomes that are induced are still 'acceptable' at every state of the world. In this paper, we study Safe Implementation of social choice correspondences in *mixed Nash Equilibrium*. We identify a condition, *Set-Comonotonicity*, which is both necessary and (under mild domain restrictions) almost sufficient for this implementation notion.

Keywords: Set-Comonotonicity, Implementation, Mechanism Design, Mixed Implementation, Robustness, Safe Implementation

JEL Codes: C72; D82.

*We thank Ritesh Jain and Michele Lombardi for their feedback.The BSE acknowledges financial support from the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2024-001476-S)

†*email:* Anand.Chopra@liverpool.ac.uk.
‡*email:* malachy.gavan@liverpool.ac.uk.
§*email:* antonio.penta@upf.edu.

1 Introduction

Implementation theory studies which social outcomes may ensue as the result of the strategic interaction of rational individuals. The central question is whether, for a given set of agents and states of the world, it is possible to construct a mechanism where, at each state, the set of equilibrium outcomes coincides with the designer's objectives, represented as a Social Choice Correspondence (SCC). Since Maskin's seminal contributions (1977; 1999), traditional models assume that the designer can freely assign outcomes both on and off the equilibrium path, including imposing arbitrary punishments if agents deviate. In many contexts, however, such flexibility may not be realistic. The designer's ability to punish deviations may be restricted, for instance, due to institutional, ethical, or informational considerations. Moreover, deviations from equilibrium may occur due to mistakes, bounded rationality, or misspecified uncertainties about the environment. In these cases, the designer want to pursue implementation with a mechanism that 'performs well' even if these events materialize.

These observations motivated Gavan and Penta (2025)'s notion of *Safe Implementation*, whereby the designer specifies not only a Social Choice Correspondence (SCC), to be induced by the equilibrium outcomes, but also an *Acceptability Correspondence*, which restricts the outcomes of the mechanism if a bounded number of agents deviates from the equilibrium. Gavan and Penta (2025) study Safe Implementation in *pure* Nash equilibrium, and identify necessary and sufficient conditions (*Comonotonicity* and *Safe No-Veto*) that generalize the analogous conditions for (non safe) Nash Implementation (Maskin (1977)).[1]

In this paper, we extend the analysis to general (i.e., pure or *mixed*) Nash equilibrium, following the ordinal approach of Mezzetti and Renou (2012), who require robustness across different cardinal representations of agents' preferences: rather than fixing a par-

[1] *Comonotonicity* restricts the joint behavior of the SCC and acceptability correspondence, and it coincides with Maskin monotonicity when the acceptability correspondence is vacuous, in the sense of allowing all allocations at all states. Formally, the closest condition in the earlier literature is that of *extended monotonicity* (Bochet and Maniquet, 2010), which characterizes virtual implementation where stochastic mechanisms are permitted, with support restrictions (Bochet and Maniquet, 2010), and which restricts the joint behaviour of the SCC and the (state dependent) support, in a similar fashion to the joint restriction that Comonotonicity imposes on the SCC and the acceptability correspondence.

ticular von-Neumann Morgenstern (vNM) utility function over deterministic outcomes, Mezzetti and Renou (2012) require that the implementing mechanisms achieve perform well *for all* cardinal representations that are consistent with the underlying ordinal preferences over certain outcomes.

Incorporating this requirement into the Safe Implementation framework presents additional challenges, as safety must be maintained uniformly across all such cardinal representations. To pin things down, a mechanism is said to (A, k)-Safely Implement a SCC if, at every state and for every cardinal representation of the underlying preferences over certain outcomes, the set of (pure or mixed) Nash equilibrium outcomes coincides with those admitted by the SCC, and any deviation by up to k players leads to an outcome within the Acceptability Correspondence. This definition generalises both standard mixed implementation and the Safe Implementation framework of Gavan and Penta (2025) to environments where agents may randomise.

Our main results identify a condition which is both necessary and 'almost sufficient' for Safe Implementation in mixed Nash equilibrium. This condition, which we call *Set-Comonotonicity*, restricts the joint behavior of the the SCC and the Acceptability Correspondence. Informally, *Set-Comonotonicity* requires that, when moving from one state to another, both the SCC and acceptability correspondence at the second state must include all of alternatives included at the first state, whenever one of the following two conditions hold: (i) the alternatives selected by the SCC at the first state are also top-ranked at the second state, for all agents, *among all the alternatives that are acceptable at the first state*; or (ii) for each alternative included in the SCC at the first state, all allocations that are weakly or strictly worse at the frist state, among those that are acceptable at that state, remain weakly and strictly worse within the same set, also according to the preferences at the second state, for all agents. This notion generalises Mezzetti and Renou (2012)'s Set-Monotonicity, and it is weaker than Gavan and Penta (2025)'s *Comonotonicity*, which in turn generalizes Maskin monotonicity.

In Theorem 1 we establish necessity of *Set-Comonotonicity*. Then, in Theorem 2, we show that when *Set-Comonotonicity* is combined with a *Safe No-Veto condition*, under mild domain restrictions, then (A, k)-Safe Mixed Nash Implementation is possible for any

k lower than half the total number of agents.

To further examine some key implications of these results, we turn to some applications, and obtain both possibility and impossibility results. In particular, first we show that while the strong Pareto correspondence is implementable on the domain of single-top preferences in mixed Nash equilibrium (Mezzetti and Renou, 2012), it is *not* Safely Implementable for *any* non-vacuous acceptability correspondence. Similarly, on the domain of strict preferences, the top-cycle correspondence is mixed Nash implementable (Mezzetti and Renou, 2012), but it is *not* Safely Implementable for *any* non-vacuous acceptability correspondence.

On the other hand, we construct interesting and economically relevant rules that are not implementable in Nash equilibrium (and therefore not Safe Nash Implementable for any acceptability correspondence), and yet are implementable under Safe mixed Nash equilibrium. Specifically, we consider a social choice rule defined over a domain of single-peaked preferences that extends the top-cycle correspondence to address fairness concerns. On such a domain, the top-cycle correspondence selects the most preferred alternatives of the median-peaked individual, which can be seen as making no concessions to those who favour "lower" or "higher" allocations. In contrast, our proposed rule introduces a fairness adjustment: In order to curb the 'tyranny of the median' entailed by the baseline Condorcet rule, we allow some concessions to the groups to the left and right-neighborhood of the median voter to increase the fairness of the resulting allocation. We show that, while this modified rule is not implementable in Nash equilibrium, it is Safe (mixed) Nash Implementable for an Acceptability Correspondence that excludes alternatives lower than the lowest peak and higher than the highest peak, when the median peak remains unchanged.

The rest of the paper is organised as follows. In section 2 we introduce the model and the key definitions. In section 3.1 we provide a number of necessary conditions related to *Set-Comonotonicity*. We then turn to sufficiency in section 3.2. Applications are considered in section 4, before turning to the related literature in section 5. Section 6 concludes.

2 Model

An environment is given by $\langle N, X, \Theta \rangle$ where $N = \{1, 2, ..., n\}$ is a finite set of agents, X a finite set of alternatives, and Θ a finite set of states of the world. Each state is associated with a preference profile $\succsim^\theta = (\succsim_1^\theta, ..., \succsim_n^\theta)$ where $\succsim_i^\theta$ is player i's preference relation over X at state θ. As usual, we let $x \succ_i^\theta y$ if and only if $x \succsim_i^\theta y$ but not $y \succsim_i^\theta x$, interpreted as the *strict* preference. An environments has *strict preferences* if there is no state $\theta \in \Theta$ where some agent is indifferent between any two alternatives.

We denote the lower contour set of an alternative x at state θ for player i by $L_i(x, \theta) := \{y \in X | x \succsim_i^\theta y\}$ and the strict lower contour set as $SL_i(x, \theta) := \{y \in X | x \succ_i^\theta y\}$. For $B \subseteq X$, let $\max_i^\theta B := \{x \in B | x \succsim_i^\theta y, \forall y \in B\}$.

It is assumed that any preference relation $\succsim_i^\theta$ can be represented by a cardinal utility function $u_i(\cdot, \theta) : X \to \mathbb{R}$. $\mathcal{U}_i^\theta$ is the set of all possible cardinal representations at state θ for agent i. The set of all possible cardinal representations for all players at θ is $\mathcal{U}^\theta = \times_{i \in N} \mathcal{U}_i^\theta$.

A social choice correspondence $F : \Theta \to 2^X \backslash \{\emptyset\}$ selects a non-empty set of alternatives for each state of the world. For any subset $Y \subseteq X$, let $\Delta(Y)$ denote the set of all probability measures over Y.

A (possibly stochastic) mechanism is a pair $\langle (M_i)_{i \in N}, g \rangle$ where M_i is the set of messages available to agent i and $g : \times_{i \in N} M_i \to \Delta(X)$ is the (possibly random) outcome function. As standard, we let $M = \times_{i \in N} M_i$ and $M_{-i} = \times_{i \neq j} M_j$. Mixed strategies are denoted by $\sigma_i \in \Delta(M_i)$, and their profiles as $\sigma = (\sigma_i)_{i \in N}$ and $\sigma = (\sigma_{-j})_{j \in N \backslash \{i\}}$. The probability that $m \in M$ is realised, given mixed strategy profile σ, is denoted by $\sigma(m)$.

Given a mechanism $\mathcal{M} = \langle (M_i)_{i \in N}, g \rangle$, a state θ, and a cardinal representation $(u_i(\cdot, \theta))_{i \in N}$ of $(\succsim_i^\theta)_{i \in N}$, the expected utility of an agent when playing $m_i \in M_i$, while her opponents play $m_{-i} \in M_{-i}$ is given by:

$$U_i(g(m_i, m_{-i}), \theta) = \sum_{x \in X} g(m_i, m_{-i})[x] \cdot u_i(x, \theta)$$

where $g(m_i, m_{-i})[x]$ denotes the probability that x is chosen by the mechanism when the profile of messages is (m_i, m_{-i}). This induces a strategic form game of $G(\theta, u) = \langle N, (M_i, U_i(g(\cdot), \cdot), \theta)_{i \in N} \rangle$. Given a mixed strategy profile σ, we let $P(\sigma, g)$ denote the probability distribution over alternatives induced by the outcome function g and by σ. That is, $P(\sigma, g)[x] = \sum_{m \in M} \sigma(m) \cdot g(m)[x]$ when M is countable and similarly expressed via integrals when uncountable. Let $\mathcal{C}^{\mathcal{M}}(\theta, u) \subseteq \times_{i \in N} \Delta(M_i)$ denote the set of (mixed) Nash equilibria of $\mathcal{M}$ when the cardinal representation of the preference profile at each state θ is given by $u(\cdot, \theta)$.

Next we introduce the primitives required for *Safe Implementation*. As in Gavan and Penta (2025), to account for Safety concerns, we introduce an acceptability correspondence that dictates which outcomes can be used by the designer when up to k players do not play as expected at each state.

Formally, let $A : \Theta \to 2^X \setminus \{\emptyset\}$, where $A(\theta)$ denotes the set of outcomes that the social planner deems *acceptable* at state θ. We maintain a natural requirement throughout the paper that $F(\theta) \subseteq A(\theta)$ for all $\theta \in \Theta$.[2]

Let $k \in \{1, \ldots, n\}$ represent the *safety threshold* the designer intends to enforce. This threshold specifies the maximum number of agents who may deviate from any given equilibrium $\sigma^* \in \mathcal{C}^M(\theta, u)$ while still requiring the mechanism to produce outcomes within $A(\theta)$, for every θ. To this end, for each k, define N_k as the collection of all subsets of N with exactly k members (i.e., $N_k := \{C \subseteq 2^N : |C| = k\}$). Take the distance metric $d_N(\sigma, \sigma') := |\{i \in N : \sigma_i \neq \sigma'_i\}|$ and a neighbourhood

$$B_k(\sigma) := \{\sigma' \in \times_{i \in N} \Delta(M_i) : d_N(\sigma, \sigma') \leq k\},$$

which consists of the set of mixed strategy profiles σ' that differ from σ for at most k agents. $A^* : \Theta \to 2^X \setminus \{\emptyset\}$ is a *sub-correspondence* of $A : \Theta \to 2^X \setminus \{\emptyset\}$ if it is such that $A^*(\theta) \subseteq A(\theta)$ for all $\theta \in \Theta$.

With this, (A, k)-*Safe Mixed Implementation* adds Safety concerns to the mixed Nash implementation of Mezzetti and Renou (2012), and is defined as follows:

[2]In fact, it will be immediately implied as a necessary condition for the following definition of implementation.

Definition 1 ((A, k)-Safe Mixed Implementation)**.** *The mechanism* $\langle (M_i)_{i \in N}, g \rangle$ (A, k)-*Safe Mixed Implements the social choice correspondence* F *if for all* $\theta \in \Theta$*, for all cardinal representations* $u(\cdot, \theta) \in \mathcal{U}^\theta$ *of* $\succsim_i^\theta$*, the following conditions hold:*

1. *For each* $x \in F(\theta)$ *there exists a mixed Nash equilibrium* $\sigma^* \in \mathcal{C}^{\mathcal{M}}(\theta, u)$ *such that* x *is in the support of* $P(\sigma^*, g)$*,*

2. *For any mixed Nash equilibrium* $\sigma^* \in \mathcal{C}^{\mathcal{M}}(\theta, u)$*, it holds that:*

   (a) the support of $P(\sigma^*, g)$ *is included in* $F(\theta)$*, and*

   (b) for all $\sigma \in B_k(\sigma^*)$ *the support of* $P(\sigma, g)$ *is included in* $A(\theta)$*.*

Furthermore, if $A : \Theta \to 2^\Theta \setminus \{\emptyset\}$ *admits no sub-correspondence* A' *for which* (A', k)*-Safe Mixed Implementation of* F *is possible, then we say that* A *is* Maximally Safe.

The notion of *Safe Implementation* allows for a rich set of interpretations and can encompass, amongst others: limited commitment of the designer, robustness to mistakes in play of the agents, state-dependent feasibility restrictions, direct safety concerns of the designer. To illustrate some of these and the expressive power of the model, we next provide some examples of natural acceptability correspondences (see Gavan and Penta (2025) for further details).

Some Examples of Acceptability Correspondences

1. ***Minimal Planner Welfare Guarantee:*** *If the SCC can be seen as the maximisation of the planner's objective function, say* $W : X \times \Theta \to \mathbb{R}$*, it is natural to think of the acceptability correspondence selecting outcomes that are above some minimal level of welfare the planner wants to ensure even in case of deviations. That is,* $A(\theta) = \{x \in X | W(x, \theta) \geq \bar{W}(\theta)\}$ *for some* $\bar{W}(\theta)$*.*

2. ***Pareto Interval:*** *When a space of outcomes can be ordered and preferences are single-peaked, a natural acceptability correspondence would be to select the out-*

comes between the two most extreme peaks. Therefore, even deviations cannot lead to inefficient outcomes.[3]

3. ***Perfect Safety:*** *Another appealing acceptability correspondence is to say that* only *the outcomes selected by the SCC are acceptable. Specifically, $A(\theta) = F(\theta)$ for all $\theta \in \Theta$. This is the most demanding form of Safety. In the case of social choice functions, i.e. $|F(\theta)| = 1$ for all $\theta \in \Theta$, Gavan and Penta (2025) show that perfect safety implies a constant rule.*

4. ***State Dependent Feasibility:*** *It is natural that in some cases the set of alternatives that is available is itself state dependent. In this sense, it may not be possible to use certain outcomes to incentivise behaviour at certain states, but not in others. By allowing the acceptable outcomes at a state to only be those outcomes that are feasible at that state, this framework can accommodate this.*[4]

Also note that (A, k)-*Safe Mixed Implementation* generalizes the notion of Mezzetti and Renou (2012), which obtains for the special case where the acceptability restriction is vacuous in the sense that $A(\theta) = X$ for all $\theta \in \Theta$. Outside of this case, the Safety requirement makes the notion of implementation more demanding. Hence, the necessary condition identified by Mezzetti and Renou (2012) are also necessary for our notion, whereas the sufficient conditions that we will provide in Section 3.2 will also be sufficient for implementation *à la* Mezzetti and Renou (2012).

Finally, this notion is monotonic in two ways: first, for any k, if $k' < k$ and F is (A, k)-Safely Mixed Implementable then it is also (A, k')-Safely Mixed Implementable. Second, if a SCC is (A, k)-Safe Mixed Implementable, then it is $(\hat{A}, k)$-Safe Mixed Implementable for any 'less stringent' correspondence, $\hat{A} : \Theta \to 2^X \setminus \{\emptyset\}$, such that $A(\theta) \subseteq \hat{A}(\theta)$ for all $\theta \in \Theta$. This is the reason for introducing the notion of *Maximally Safe* acceptability correspondence: if A is maximally safe in the sense of Definition 1, then A describes the

[3]Section 4 will consider a version of this acceptability correspondence where the interval is the most extreme peaks for *any* state where the median peak leads is the same.

[4]Postlewaite and Wettstein (1989) have previously considered state dependent feasibility for implementation in a Walrasian economy.

most stringent safety requirements that can be attained, since safe implementation would be impossible for any sub-correspondence of A.

2.1 Related Definitions and Discussion of Model

In this subsection we discuss the main results and definitions from both the implementation in mixed Nash equilibrium (without safety concerns) from Mezzetti and Renou (2012), and the conditions for Safe (pure) Nash Implementation in Gavan and Penta (2025). These will provide natural benchmarks for the results presented in the remainder of the paper.

The next condition, *Set-Monotonicity*, is necessary for mixed Nash implementation, absent safety concerns (Mezzetti and Renou, 2012).

Definition 2 (Set-Monotonicity (Mezzetti and Renou, 2012))**.** *F satisfies Set-Monotonicity if, for any θ, θ', $F(\theta) \subseteq F(\theta')$ whenever for all $i \in N$ one of the following holds:*

1. *$F(\theta) \subseteq \max_i^{\theta'} X$, or*
2. *for all $x \in F(\theta)$:*
   - *(a) $L_i(x, \theta) \subseteq L_i(x, \theta')$ and*
   - *(b) $SL_i(x, \theta) \subseteq SL_i(x, \theta')$.*

Set-Monotonicity is a relaxation of *Maskin monotonicity*, which is necessary for (non-mixed) Nash Implementation. Intuitively, as the state transitions from θ to θ', Set-Monotonicity requires that the set $F(\theta')$ includes $F(\theta)$ only if, for all players, *all* alternatives in $F(\theta)$ either are all top-ranked at θ' or do not move down in the weak or strict rankings. In contrast, *Maskin monotonicity* demands that $F(\theta')$ include *each individual* alternative $x \in F(\theta)$ that does not move down in any player's weak ranking when moving from θ to θ'.

Together with a classical *No-Veto condition*, which requires that if all but one agent agree that an outcome is top-ranked at a state then it must be implemented, Mezzetti and Renou (2012) show that Set-Monotonicity is also sufficient for mixed Nash implementation when there are three or more agents.

Turning next to the Safety considerations, consider the following definition:

Definition 3 (Weak Comonotonicity (Gavan and Penta, 2025))**.** *A SCC, $F : \Theta \to 2^X \setminus \{\emptyset\}$, and an acceptability correspondence, $A : \Theta \to 2^X \setminus \{\emptyset\}$, are* weakly Comonotonic *if both of the following requirements hold:*

1. *If $\theta, \theta' \in \Theta$ and $x \in F(\theta)$ are such that $L_i(x, \theta) \cap A(\theta) \subseteq L_i(x, \theta') \cap A(\theta)$ for all $i \in N$, then $x \in F(\theta')$.*
2. *If $\theta, \theta' \in \Theta$ are such that, $\forall x \in F(\theta)$, $L_i(x, \theta) \cap A(\theta) \subseteq L_i(x, \theta') \cap A(\theta)$ for all $i \in N$, then $A(\theta) \subseteq A(\theta')$.*

Gavan and Penta (2025) show that weak Comotonicity is a necessary condition for *Maximally* Safe Nash Implementation. Weak Comotonicity requires that, whenever an outcome, x, is selected by the SCC at θ, if *within* the set of acceptable outcomes at θ any outcome that was weakly worse than x at θ remains weakly worse at θ', then x must also be within the SCC at θ'. Further, if all outcomes in the SCC at θ only rise in the rankings between θ and θ', then all the outcomes that are acceptable at θ must also be acceptable at θ'.

Conversely, Gavan and Penta (2025) show that, along with a No-Veto condition adapted for Safety concerns, a stronger Comonotonicity condition is sufficient for Safe (non-mixed) Nash Implementation when $n \geq 3$.

Definition 4 (Strong Comonotonicity (Gavan and Penta, 2025))**.** *A SCC, $F : \Theta \to 2^X \setminus \{\emptyset\}$, and an acceptability correspondence, $A : \Theta \to 2^X \setminus \{\emptyset\}$, are* strongly Comonotonic *if both of the following requirements hold:*

1. *If $\theta, \theta' \in \Theta$ and $x \in F(\theta)$ are such that $L_i(x, \theta) \cap A(\theta) \subseteq L_i(x, \theta') \cap A(\theta)$ for all $i \in N$, then $x \in F(\theta')$.*
2. *If $\theta, \theta' \in \Theta$ are such that, $\exists x \in F(\theta)$, $L_i(x, \theta) \cap A(\theta) \subseteq L_i(x, \theta') \cap A(\theta)$ for all $i \in N$, then $A(\theta) \subseteq A(\theta')$.*

Note that the only difference between the *Strong* and *Weak Comonotonicity* lies in the requirement on the acceptability correspondence between states, in point 2 of the two definitions. In Strong Comonotonicity, if *even one* of the outcomes within the SCC at a state

θ has weakly 'risen' in the preference rankings of the acceptable outcomes at θ', then the outcomes that were acceptable at θ must also be acceptable at θ'. Weak Comonotonicity, in contrast, requires inclusion only if the previous comparison holds *for all* outcomes within the SCC at state θ.

3 Main Results

In this section we provide the main results of the paper, which uncover general necessary and sufficient conditions for Safe Mixed Implementation. To this end, consider the following definition, which is the central notion for our main results:

Definition 5 (Set-Comonotonicity)**.** *An SCC and Acceptability Correspondence,* (F, A)*, satisfy Set-Comonotonicity if, for any* θ, θ' *we have that (i)* $F(\theta) \subseteq F(\theta')$ *and (ii)* $A(\theta) \subseteq A(\theta')$ *whenever for all* $i \in N$ *one of the following holds:*

1. $F(\theta) \subseteq \max_i^{\theta'} A(\theta)$*, or*
2. *for all* $x \in F(\theta)$*:*
   (a) $L_i(x, \theta) \cap A(\theta) \subseteq L_i(x, \theta') \cap A(\theta)$ *and*
   (b) $SL_i(x, \theta) \cap A(\theta) \subseteq SL_i(x, \theta') \cap A(\theta)$.

Set-Comonotonicity is interpreted as follows. Suppose that there are two states of the world, θ and θ'. For every agent, consider *all* the alternatives selected by the SCC at θ. These alternatives must either be top-ranked at θ' *among those that are acceptable at* θ, or the following condition must hold: For every outcome selected by the SCC at θ, every acceptable outcome at θ that was considered strictly (or weakly) worse than an SCC-selected alternative must preserve that strict (or weak) ranking at θ'. If either of these conditions are met for each agent, then two conclusions follow. First, all outcomes implemented at θ must also be implemented at θ'. Second, all outcomes acceptable at θ must also be acceptable at θ'.

Note that this is a generalisation of the definition of Set-Monotonicity of Mezzetti and Renou (2012), and whenever the acceptability correspondence is trivial, $A(\theta) = X$ for all $\theta \in \Theta$, Set-Comonotonicity coincides with Set-Monotonicity. However, it is generally more demanding. In comparison to Gavan and Penta (2025)'s Comonotonicity (either strong or weak), instead, Set-Comonotonicity is weaker: To se this, note that weak Comonotonicity is weaker than strong Comonotonicity, and that Set-Comonotonicity implies weak Comonotonicity.

3.1 Necessity

We now turn to providing the main result on necessity. As Definition 1 becomes more demanding as the acceptability correspondence becomes finer, it is natural to consider the most stringent notion of Safety, and hence focus on *Maximal Safety*. The next result shows that *Set-Comonotonicity* is necessary for *Maximally Safe* Mixed Implementation:

Theorem 1. *If a social choice correspondence F is Maximally (A, k)-Safely Implementable in mixed Nash equilibrium, then (F, A) satisfies Set-Comonotonicity.*

This theorem is implied by the following proposition, which provides a necessary condition outside of the case of Maximal Safety.

Proposition 1. *If a social choice correspondence F is (non-maximally) (A, k)-Safely Implementable in mixed Nash equilibrium, then there exists some subcorrespondence of A, A', for which (F, A') satisfies Set-Comonotonicity.*

The proofs are relegated to the appendix.

It is worth pointing out that, due to nature of set inclusion, the following corollary applies, stating that condition (1) of Set-Comonotonicity, which puts restrictions on the SCC, is necessary without restricting attention to the case of maximal safety.

Corollary 1. *If F is (A, k)-Safe Mixed Implementable, then for any $\theta, \theta' \in \Theta$, it must be that $F(\theta) \subseteq F(\theta')$ whenever for all $i \in N$ one of the following holds:*

1. $F(\theta) \subseteq \max_i^{\theta'} A(\theta)$*, or*

2. *for all* $x \in F(\theta)$*:*

   (a) $L_i(x,\theta) \cap A(\theta) \subseteq L_i(x,\theta') \cap A(\theta)$ *and*

   (b) $SL_i(x,\theta) \cap A(\theta) \subseteq SL_i(x,\theta') \cap A(\theta)$*.*

Note when we have a social choice function where $|F(\theta)| = 1$ for all $\theta \in \Theta$, Set-Comonotonicity coincides with Comonotonicity. Therefore the following result is inherited from Gavan and Penta (2025), limiting the possibility of Safety concerns.

Corollary 2. *If* $|F(\theta)| = 1$ *for all* $\theta \in \Theta$*, and therefore we have a social choice function, and* $\exists\theta \in \Theta$ *such that* $A(\theta) = F(\theta)$*, then* $F(\theta') = F(\theta)$ *for all* θ'*.*

When preferences are strict, instead, then Set-Comonotonicity reduces to the following notion of *strong* Set-Comonotonicity, which is analogous to a notion of *Strong Set-Monotonicity* that was introduced in a working paper version of Mezzetti and Renou (2012), to cover the case of strict preferences:

Definition 6 (Strong Set-Comonotonicity)**.** (F, A) *satisfy* strong Set-Comonotonicity *if, for any* $\theta, \theta' \in \Theta$ *we have that (i)* $F(\theta) \subseteq F(\theta')$ *and (ii)* $A(\theta) \subseteq A(\theta')$ *whenever* $L_i(x,\theta) \cap A(\theta) \subseteq L_i(x,\theta') \cap A(\theta)$ *for all* $x \in F(\theta)$*.*

Strong Set-Comonotonicity simplifies the previous definition as follows: If, moving from state θ to θ', for every outcome x selected by the SCC at θ, all of the acceptable outcome at θ that are worse than $x$are still worse than x at θ', then it must be that all acceptable outcomes at θ are acceptable at θ' and all outcomes that are implemented at θ are also implemented at θ'. The simplification is made possible because, with strict preferences, there is no need to distinguish the weak and strict lower contour sets, and if $F(\theta) \subseteq \max_{\theta'} A(\theta)$, then it must be that $L_i(x,\theta') \cap A(\theta) = A(\theta)$, and hence therefore the first condition in Def. 5 no longer needs to be explicitly stated.

Using the results of theorem 1 and proposition 1, we can provide the following corollary, which shows that strong Set-Comonotonicity is necessary for Safe Mixed Implementation under strict preferences:

Corollary 3. *If preferences are strict and F is maximally (A, k)-Safely Implementable in mixed Nash equilibrium, then (F, A) satisfies strong Set-Comonotonicity.*

Furthermore, if F is (non-maximally) (A, k)-Safely Implementable in mixed Nash equilibrium, then there exists some sub-correspondence of A, $A' : \Theta \to 2^{\Theta} \setminus \{\emptyset\}$, such that (F, A') satisfy Strong Set-Comonotonicity.

3.2 Sufficiency

Next we show that Set-Comonotonicity is almost sufficient under strict preferences, when paired with the following No-Veto condition, from Gavan and Penta (2025).[5]

Definition 7. *F and A satisfy Safe No-Veto if, whenever $\exists j \in N$, $\theta, \theta' \in \Theta$ such that $x \succsim_i^{\theta} y$ for all $y \in A(\theta')$ and $i \neq j$ then $x \in F(\theta)$ and $A(\theta) = X$.*

Safe No-Veto requires that, when at a state θ, all but one agent agree that some outcome x is top ranked *within* the acceptability correspondence at some possibly different state θ', then that outcome must be implemented at θ, and the acceptability correspondence at state θ must be vacuous, in the sense that $A(\theta) = X$. Note that when the acceptability correspondence is vacuous *at all states*, then this condition coincide swith the standard notion of No-Veto by Maskin (1999).

With this, the following result can be seen as a generalisation of the sufficiency result in Mezzetti and Renou (2012), which showed that Set-Monotonicity and No-Veto are sufficient when Safety concerns are not present:

Theorem 2. *If Θ is domain of strict preferences, $n \geq 3$ and (F, A) satisfy (strong) Set-Comonotonicity and Safe No-Veto, then F is (A, k)-Safely Implementable in mixed Nash equilibrium for all $1 \leq k < \frac{n}{2}$.*

[5] Recall that when preferences are strict, there is no distinction between strong Set-Comonotonicity and Set-Comonotonicity.

4 Applications

In this section we study three applications to gain further insights on Safe Mixed Nash Implementation. First, we discuss two important instances where the addition of Safety considerations make Mixed Nash Implementation impossible to attain. Then, we discuss an economically relevant setting where reasonable Safety concerns can be accommodated within Mixed Nash Implementation, but not when only pure Nash equilibria are considered (with or without Safety considerations).

4.1 The Strong Pareto Correspondence

An important social choice correspondence is the strong Pareto correspondence, which, at each state of the world, selects the outcomes where no other feasible outcome makes someone better off without making anyone worse off.[6] On the global domain of *single-top* preferences, where for each state $\theta \in \Theta$ and for each individual $i \in N$, there is a unique outcome $x \in X$ which is the top alternative, $x \succ_i^\theta y$ for all $y \in X$, Mezzetti and Renou (2012) show that, in the domain of single-top preferences, the strong Pareto correspondence is implementable in mixed Nash equilibrium, but not Nash implementable. Here we study how this SCC interacts with Safety concerns. To do so, let us first define the strong Pareto correspondence, F^{PO}, formally:

$$F^{PO}(\theta) = \left\{ \begin{array}{c} x \in X | \nexists y \in X \text{ such that } \forall i \in N, x \in L_i(y, \theta) \\ \text{and } \exists i \in N \text{ such that } x \in SL_i(y, \theta) \end{array} \right\}$$

Although this SCC is known to be mixed Nash implementable on the domain of single-top preferences (Mezzetti and Renou, 2012), we will show that it is *not* Safe mixed Nash Implementable for *any* non-vacuous acceptability correspondence.

Proposition 2. *On the global domain of single-top preferences, F^{PO} is mixed-Nash im-*

[6] This is in a sense the standard notion of Pareto efficiency, where the term 'strong' is added to stress the comparison with the *weak* Pareto correspondence, which selects outcomes where there is no feasible alternative that makes everyone strictly better off.

plementable but is not (A,k)*-Safe mixed Nash implementable for any* A *such that* $\exists\theta \in \Theta$ *where* $A(\theta) \neq X$.

To see this, let A denote an acceptability correspondence such that $\exists y \in X$ and $\theta \in \Theta$ s.t. $y \notin A(\theta)$. Then, if $y \notin F^{PO}(\theta)$, it is certainly not the case that $\{y\} = F^{PO}(\theta)$. Now let us consider a state θ' where, for all $x, z \in A(\theta)$, $x \succsim_i^\theta z$ implies $x \succsim_i^{\theta'} z$ and $x \succ_i^\theta z$ implies $x \succ_i^{\theta'} z$. That is, all strict and weak preferences of any alternatives in $A(\theta)$ between θ and θ' are preserved. By Corollary 1, it must be that $F^{PO}(\theta) \subseteq F^{PO}(\theta')$. However, we can select θ' such that $y \succ_i^{\theta'} x$ for all $x \in X$, which is a single-top preference and therefore in the domain. Therefore, $F^{PO}(\theta') = \{y\}$ by the definition of F^{PO}. However, it cannot be the case that $F^{PO}(\theta) \neq \{y\}$, $F^{PO}(\theta) \subseteq F^{PO}(\theta')$ and $F^{PO}(\theta') = \{y\}$. With this, we reach a contradiction and the strong Pareto correspondence cannot be Safely Implemented when there is some state where the acceptability correspondence does not include all alternatives.

4.2 The Top-Cycle Correspondence

Next we consider the *top-cycle correspondence.* In words, say that x dominates y at state θ, which we denote by $x \gg^\theta y$, if the number of agents who strictly prefer x to y is larger than the number who strictly prefer y to x. At each state, let $F^{TC}(\theta)$ denotes the smallest set of alternatives such that any alternative in the set dominates any alternative outside the set. Formally:

$$F^{TC}(\theta) = \bigcap \{X' | x' \in X', x \in X \backslash X' \text{ implies } x' \gg^\theta x\}$$

Mezzetti and Renou (2012) show that, while the top-cycle correspondence is *not* implementable in *pure* Nash equilibrium, and therefore not Safe Nash Implementable for any acceptability concerns, it is implementable in *mixed* Nash equilibrium when preferences are strict. Yet, an identical reasoning to that which gave us the impossibility result for Safe mixed Implementation of the Strong Pareto Correspondence, also implies the impossibility of safety conerns for the top-cycle correspondence:

Proposition 3. *On the global domain of strict preferences, the top-cycle correspondence* *is implementable in mixed Nash implementation but is not (A, k)-Safe mixed Nash implementable for any A such that $\exists\theta \in \Theta$ where $A(\theta) \neq X$.*

4.3 An Expansion of the Top Cycle Correspondence

In contrast to the results in the previous two subsections, here we present a possibility result for Safe Mixed Implementation. More specifically, in environments with strict and single-peaked preferences, we consider a modification of the top-cycle correspondence that selects not only the Condorcet winner (which would be the unique outcome of in the top-cycle correspondence in this domain), but also its immediate neighbours. We show that, under natural acceptability constraints, this social choice correspondence is Safely Implementable in mixed Nash equilibrium.

Formally, we consider the case where the number of agents n is odd, $n > 3$, and all agents have single-peaked preferences. To this end, let $X = \{x^1, x^2, \ldots, x^m\}$ denote the set of allocations, with $2n - 2 > m > n + 1$, and with the alternatives ordered so that $x^1 < x^2 < \cdots < x^m$. A preference relation $\succsim_i^\theta$ is *single-peaked* if, for each agent $i \in N$ and each state $\theta \in \Theta$, there exists a peak $x^{p_{i,\theta}} \in X$ such that: (a) for all $l \geq l' \geq p_{i,\theta}$, $x^l \succsim_i^\theta x^{l'}$ implies $x^{l'} \precsim_i^\theta x^{l-1}$ and (b) for all $l \leq l' \leq p_{i,\theta}$, $x^l \succsim_i^\theta x^{l'}$ implies $x^{l'} \precsim_i^\theta x^{l+1}$. We will assume that preferences are strict (i.e., no indifference) and that, at each state, no two agents share the same top-ranked alternative.

We will study an adaptation of the top-cycle correspondence. First note that, in the case of single-peaked preferences, the top-cycle correspondence that we defined in the previous subsection coincides with the allocation rule that selects, at every state, the unique Condorcet winner, which is the allocation corresponding to the median peak. The extension of the top-cycle correspondence that we consider adds to this sole allocation, also the two outcomes that are on "either" side of the median peak.

One can think of this rule as capturing the following story: The designer has a view that the Condorcet winner should be implemented, but they also want to accommodate some fairness considerations. Specifically, they are concerned that the Condorcet winner could be criticized as being unfair as it is pinned down by the alternative that is maximal for the individual with the median peak, and therefore can be seen as a 'tyranny of the median'. To address this criticism, the planner decides to design a rule that also allows for slightly better allocations to those with peaks to the left and to the right. This way, it cannot be that only the most preferred outcome of the median peaked individual is implemented, but they provide some 'concessions' to the groups to the left and to the right.

We coin this rule as the *neighbourhood* of Condorcet rule. Specifically, say that $w(\theta) \in X$ is a Condorcet winner at θ if:

$$|\{i \in N | w(\theta) \succ_i^\theta y\}| > |\{i \in N | y \succ_i^\theta w(\theta)\}| \qquad \forall y \neq w(\theta)$$

Hence, letting $m(\theta)$ denote the index of the median peak at state θ, the Condorcet winner selects the median peak allocation $x^{m(\theta)} = w(\theta)$. Our neighbourhood of the Condorcet rule selects not only $x^{m(\theta)}$, but also the allocations on either side of it. Formally:

$$F^\dagger(\theta) = \{x^{m(\theta)-1}, x^{m(\theta)}, x^{m(\theta)+1}\}$$

Note that, given our domain restriction, all agents have a different peak at each state, which ensures that $w(\theta) \neq x^1, x^m$ for all θ and therefore $|F(\theta)| = 3$ for all θ.

We first note that this rule is not Nash implementable and, therefore, not Safely Implementable for *any* acceptability correspondence.[7]

Lemma 1. *On the global domain of single-peaked and strict preferences where no two agents share the same peak at any state, $F^\dagger$ is not Nash implementable, and hence not (A, k)-Safely Nash Implementable for any A or k.*

We will now consider the possibility of implementing this neighbourhood Condorcet correspondence in mixed Nash equilibrium with reasonable Safety concerns. Specifically,

[7]It is worth noting that on this domain the Condorcet winner is implementable (Maskin, 1999). Interested readers are pointed to Healy and Peress (2015), who show when the set of states includes all of those when the Condorcet winner exists, but includes at least more state than the rule which selects the Condorcet winner when possible is not Nash implementable.

we consider an acceptability correspondence that, at each state θ, includes all the alternatives that are between the most extreme peaks possible across all states θ' that share the same median peak as θ:

$$A^{\dagger}(\theta) = \bigcup_{\{\theta'|m(\theta')=m(\theta)\}} \{x^{l'}, x^{l'+1}, ..., x^{l''-1}, x^{l''} \in X | l' = \min_{i\in N} p_{i,\theta'}, l'' = \max_{i\in N} p_{i,\theta'}\}$$

Note that $F^{\dagger}(\theta) \subseteq A^{\dagger}(\theta)$.

The next proposition shows that this rule is Safe Mixed Nash Implementable under the above acceptability correspondence. This is the case for *any* set of preferences that satisfy the restrictions above.[8]

Proposition 4. *On any domain of single peaked and strict preferences where no two agents share the same peak, $F^{\dagger}$ is $(A^{\dagger}, k)$-Safe mixed Nash Implementable with $1 \leq k < \frac{n}{2}$.*

5 Related Literature

The notion of mixed Nash implementation is taken from Mezzetti and Renou (2012), who provide an *ordinal* approach. They show that Set-Monotonicity is necessary and together with No-Veto is sufficient when $n \geq 3$. Thus, our results can be seen as a generalisation of theirs, accounting for Safety concerns. Their paper shows that many popular social choice functions that are not implementable in Nash equilibrium are in fact implementable in mixed Nash equilibrium. For instance, the Pareto correspondence and top-cycle correspondence. We show that when preferences are rich enough this is not the case if *any* Safety concerns are adopted. However, in line with their results, we show some interesting rules could be implemented in mixed Nash, even accounting for Safety concerns, when they could not be implemented in Nash.

[8]Note that this is not immediate as $A(\theta)$ is pinned down by conditions of *all* states that have the same Condorcet winner. Therefore, Safe Implementation on a larger space does not imply Safe Implementation on the smaller space as the acceptability correspondences across the two could differ.

Gavan and Penta (2025), study (pure) Nash implementation with Safety concerns. Here, instead, we study this with mixed Nash equilibrium, following the ordinal approach of Mezzetti and Renou (2012). They introduce a generalisation of Maskin Monotonicity called *Comonotonicity*. Comonotonicity imposes joint restrictions on both the acceptability correspondence and the social choice correspondence. It is shown that Comonotonicity is necessary and, together with Safe No-Veto, is sufficient when $n \geq 3$. In a similar sense, this paper shows that Set-Comonotonicity, a generalisation of Mezzetti and Renou (2012)'s Set-Monotonicity, is necessary and, together with the same Safe No-Veto condition, is sufficient when preferences are strict and $n \geq 3$. Gavan and Penta (2025) relate their concept to Eliaz (2002)'s fault tolerant implementation, Shoukry (2019)'s outcome robust implementation, Jackson and Palfrey (2001)'s voluntary implementation, Hayashi and Lombardi (2019)'s constrained implementation, amongst others.[9] We point interested readers towards the discussion in Gavan and Penta (2025).

Implementation in mixed Nash equilibrium departs from Nash implementation in three ways. Firstly, as the name suggests, the solution concept used is mixed Nash equilibrium rather than pure Nash equilibrium. A more subtle difference is that this notion of implementation allows for the use of a stochastic mechanism. As agents are assumed to possess vNM preferences, it is natural to allow the designer to randomise between outcomes. This is further emphasised in the third difference; any element of the SCC must be in the *support* of an equilibrium, but said equilibrium may put positive probability on other elements in the SCC. It is worth pointing out that allowing for stochastic mechanisms does in fact make implementation easier as it strictly increases the set of mechanisms that can be used. Bochet (2007) and Benoît and Ok (2008) show that, under mild domain restrictions, Maskin Monotonicity is both necessary and sufficient for Nash implementation with stochastic mechanisms. A similar result, allowing for stochastic mechanisms while maintaining pure strategies, is considered with Safety concerns in Gavan and Penta (2025).

[9] Safe Implementation can also be seen as a notion of robustness of the mechanism with respect to possible misspecification of the agents' strategic interaction. In this sense, it is also related to Bochet and Tumennasan (2023a,b). For a distinct but related approach, that seeks implementation with respect to a wide range of solution concept, see Jain et al. (2024). Clearly, these notions of robustness are quite different from the those in the literature inspired by the sa called *Wilson doctrine* (e.g., Bergemann and Morris (2005, 2009a,b), Penta (2015), Müller (2016), Ollár and Penta (2017, 2023, 2024), etc.)

It is shown that, under mild domain restrictions, Safe No-Veto can be dropped from the sufficient conditions.

Bochet and Maniquet (2010) studied virtual implementation, where stochastic mechanisms are permitted, with support restrictions. They provide a necessary and sufficient condition, extended monotonicity, which restricts the joint behaviour of the SCC and the (state dependent) support, in a similar fashion to the joint restriction we provide on the SCC and the acceptability correspondence.

Xiong (2022) provided a full characterisation of implementation in mixed Nash equilibrium, providing a condition based on a construction of the correct set of alternatives used to sustain equilibria, in a similar fashion to the "C_i" sets of Moore and Repullo (1990) and the constructive approach of Sjöström (1991) for Nash implementation. (See Korpela (2010), for an insightful connection between the two approaches). Here, we do not attempt to provide a full characterisation for Safe Implementation in mixed Nash equilibrium, and rather take the approach of providing simple, easy to interpret, necessary and sufficient conditions, that better highlight the role of the more novel aspects brought about by the Safety concerns.

6 Conclusion

This paper extends the framework of Safe Implementation (Gavan and Penta, 2025) to environments where agents may use mixed strategies, so as to gain a deeper understanding of the restrictions that Safety considerations impose on Implementation. We do this by following the 'ordinal approach' of Mezzetti and Renou (2012), which imposes an extra robustness desideratum on the way that risk preferences are embedded in the baseline setting, since we think it provides the best way to distil the effects that allowing for mixed strategies has on the implementation toolkit, independent of the impact of the stronger assumptions on agents' preferences that are needed to deal with the possibility of randomization.

Our main results identify *Set-Comonotonicity* as the key condition that balances the expanded equilibrium set, due to the introduction of randomisation, with the robustness requirements that are due to Safety considerations. We showed that Set-Comonotonicity is necessary for (Maximal) Safe Mixed Nash Implementation and, under strict preferences and a Safe No-Veto condition from Gavan and Penta (2025), it is also sufficient for Safe Mixed Nash Implementation. These findings generalise the classical results of Mezzetti and Renou (2012) by embedding them within the Safety context, and they build directly on the Safe Implementation framework of Gavan and Penta (2025), thereby deepening our understanding of the impact of Safety concerns on implementation.

Our results indicate that while randomisation can expand the designer's implementation toolkit, Safety concerns impose nontrivial restrictions on which social choice correspondences are implementable, especially when the acceptability of off-equilibrium outcomes is constrained. Through our applications, we demonstrate both the power and the limitations of this extended framework, illustrating that some correspondences that are implementable under mixed strategies alone that become infeasible when even minimal Safety concerns are introduced. In contrast, economically interesting rules can be implemented with natural Safety concerns even when Nash implementation (and therefore Safe Nash Implementation) is not possible.

References

Benoît, J.-P. and Ok, E. A. (2008). Nash implementation without no-veto power. *Games and Economic Behavior*, 64(1):51–67.

Bergemann, D. and Morris, S. (2005). Robust mechanism design. *Econometrica*, 73(6):1771–1813.

Bergemann, D. and Morris, S. (2009a). Robust implementation in direct mechanisms. *The Review of Economic Studies*, 76(4):1175–1204.

Bergemann, D. and Morris, S. (2009b). Robust virtual implementation. *Theoretical Economics*, 4(1).

Bochet, O. (2007). Nash implementation with lottery mechanisms. *Social Choice and Welfare*, 28(1):111–125.

Bochet, O. and Maniquet, F. (2010). Virtual Nash implementation with admissible support. *Journal of mathematical economics*, 46(1):99–108.

Bochet, O. and Tumennasan, N. (2023a). Defaults and benchmarks in mechanism design. *Working Paper*.

Bochet, O. and Tumennasan, N. (2023b). Resilient mechanisms. *Working Paper*.

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

Gavan, M. J. and Penta, A. (2025). Safe Implementation. *Theoretical Economics*. Forthcoming.

Hayashi, T. and Lombardi, M. (2019). Constrained implementation. *Journal of Economic Theory*, 183:546–567.

Healy, P. J. and Peress, M. (2015). Preference domains and the monotonicity of Condorcet extensions. *Economics Letters*, 130:21–23.

Jackson, M. O. and Palfrey, T. R. (2001). Voluntary implementation. *Journal of Economic Theory*, 98(1):1–25.

Jain, R., Lombardi, M., and Penta, A. (2024). Strategically robust implementation. *Working Paper*.

Korpela, V. (2010). Nash implementation theory—a note on full characterizations. *Economics Letters*, 108(3):283–285.

Maskin, E. (1977). Nash equilibrium and welfare optimality. *mimeo, M.I.T.*

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

Mezzetti, C. and Renou, L. (2012). Implementation in mixed Nash equilibrium. *Journal of Economic Theory*, 147(6):2357–2375.

Moore, J. and Repullo, R. (1990). Nash implementation: A full characterization. *Econometrica*, 58(5):1083–1099.

Müller, C. (2016). Robust virtual implementation under common strong belief in rationality. *Journal of Economic Theory*, 162:407–450.

Ollár, M. and Penta, A. (2017). Full implementation and belief restrictions. *American Economic Review*, 107(8):2243–77.

Ollár, M. and Penta, A. (2023). A network solution to robust implementation: the case of identical but unknown distributions. *Review of Economic Studies*, 90(5):2517–2554.

Ollár, M. and Penta, A. (2024). Incentive compatibility and belief restrictions. *BSE working paper*.

Penta, A. (2015). Robust dynamic implementation. *Journal of Economic Theory*, 160:280–316.

Postlewaite, A. and Wettstein, D. (1989). Feasible and continuous implementation. *The Review of Economic Studies*, 56(4):603–611.

Shoukry, G. F. (2019). Outcome-robust mechanisms for Nash implementation. *Social Choice and Welfare*, 52(3):497–526.

Sjöström, T. (1991). On the necessary and sufficient conditions for Nash implementation. *Social Choice and Welfare*, 8(4):333–340.

Xiong, S. (2022). Nash implementation by stochastic mechanisms: a simple full characterization. *arXiv preprint arXiv:2211.05431*.

A Proofs

Proof of Proposition 1 and Theorem 1. For the mechanism $\mathcal{M}$ that (A, k)-safely mixed Nash implements F take $A'(\theta) = \bigcup_{u^\theta \in \mathcal{U}^\theta} \bigcup_{\sigma^* \in \mathcal{C}^{\mathcal{M}(\theta,u)}} \bigcup_{\sigma \in B_k(\sigma^*)} supp(P(\sigma, g))$ that is, the support of all k deviations from equilibrium, across all possible equilibria, across all possible cardinal representations. Clearly by definition $A'(\theta) \subseteq A(\theta)$ and for any maximally (A, k)-safely implemented in mixed Nash equilibrium F it must be that $A'(\theta) = A(\theta)$. Therefore assume that A is maximally safe, understanding that if it is not then it represents the sub-correspondence that is.

Claim 1. *Suppose $L_i(x, \theta) \cap A(\theta) \subseteq L_i(x, \theta') \cap A(\theta)$ and $SL_i(x, \theta) \cap A(\theta) \subseteq SL_i(x, \theta')$ for all $x \in F(\theta)$. Then, given any cardinal representation $u_i(\cdot, \theta)$ of $\succsim_i^\theta$, there exists a cardinal representation $u_i(\cdot, \theta')$ of $\succsim_i^{\theta'}$ such that $u_i(x, \theta') \leq u_i(x, \theta)$ for all $x \in A(\theta)$ and $u_i(x, \theta') = u_i(x, \theta)$ for all $x \in F(\theta)$.*

The proof follows the logic of Mezzetti and Renou's claim C. We have included this for completeness and to ensure safety concerns do not interfere with the logic.

Proof of Claim 1. For all $x \in F(\theta)$ let $u_i(x, \theta') = u_i(x, \theta)$. To see this has no contradiction notice that as $L_i(x, \theta) \cap A(\theta) \subseteq L_i(x, \theta') \cap A(\theta)$ if $y \in F(\theta) \cap L_i(x, \theta)$ then $u_i(y, \theta') = u_i(y, \theta) \leq u_i(x, \theta) = u_i(x, \theta')$, as required as $y \in L_i(x, \theta')$ by the conditions laid out. Similarly, as $SL_i(x, \theta) \cap A(\theta) \subseteq SL_i(x, \theta') \cap A(\theta)$ if $y \in F(\theta) \cap L_i(x, \theta)$ then $u_i(y, \theta') = u_i(y, \theta) < u_i(x, \theta) = u_i(x, \theta')$, as required as $y \in F(\theta) \cap SL_i(x, \theta)$.

As inequalities are strict when required, there are enough open set to represent $\succsim_i^{\theta'}$ with $u_i(\cdot, \theta')$.

To see that at least one such representation satisfies $u_i(x, \theta') \leq u_i(x, \theta)$ for all $x \in A(\theta)$ first note that for any $x \in A(\theta)$ such that $x \sim_i^{\theta'} y$ for some $y \in F(\theta)$ we can set $u_i(x, \theta') = u_i(y, \theta') = u_i(y, \theta)$. Note that $x \notin SL_i(y, \theta') \cap A(\theta)$ and therefore $x \notin SL_i(y, \theta) \cap A(\theta)$. Therefore, $u_i(x, \theta) \geq u_i(y, \theta)$. Therefore, in this selection $u_i(x, \theta) \geq u_i(y, \theta) = u_i(y, \theta') = u_i(x, \theta')$, satisfying the condition. Now consider some $x \in A(\theta)$ such that $x \succ_i^{\theta'} y$ for some $y \in F(\theta)$. For this alternative, note that $x \notin L_i(y, \theta) \cap A(\theta)$ and therefore $x \notin L_i(y, \theta) \cap A(\theta)$. Therefore $u_i(x, \theta) > u_i(y, \theta) = u_i(y, \theta')$. Therefore

$u_i(x,\theta')$ can be selected in the open set of $(u_i(y,\theta), u_i(x,\theta))$ and maintaining all strict preferences in θ'. Finally, consider some $x \in A(\theta)$ such that $y \succ_i^{\theta'} x$ for all $y \in F(\theta)$. It must be that $y \in SL_i(y,\theta) \cap A(\theta)$ for all $y \in F(\theta)$. Therefore $u_i(x,\theta')$ must be strictly less than $u_i(y,\theta') = u_i(y,\theta)$ for all $y \in F(\theta)$. Selecting $u_i(x,\theta')$ to be strictly lower than all $u_i(y,\theta)$ with $y \in F(\theta)$ and $u_i(x,\theta)$ yields no contradictions, as open sets are available between all strict preferences this allows for such a representation to exist. □

The proof will now proceed by the contrapositive. Suppose that $F(\theta) \not\subset F(\theta')$ while the conditions of Set-Comonotonicity holds. Take any cardinal representation $u(\cdot,\theta)$ of $\succsim^\theta$ and any equilibrium $\sigma^* \in \mathcal{C}^{\mathcal{M}}(u,\theta)$. If σ^* is an equilibrium for some cardinal representation $u(\cdot,\theta') \in \mathcal{U}^{\theta'}$ then it follows that the support of $P(\sigma^*, g)$ must also be included in $F(\theta')$ by definition of implementation. Further, the support of all k deviations of σ^* must also be included in $A(\theta')$ by the definition of maximal acceptability.

As $F(\theta) \nsubseteq F(\theta')$ for any $x \in F(\theta)\backslash F(\theta')$ there is some σ^* and $u(\cdot,\theta)$ such that σ^* selects x with positive probability and is a mixed Nash equilibrium. However, σ^* cannot be an equilibrium at *any* cardinal representation $u(\cdot,\theta') \in \mathcal{U}^{\theta'}$. In particular, for any one such that $u(\cdot,\theta')$ satisfies the condition of the previous claim.

Thus, assuming M is countable, it must be that there exists an $i \in N$ and message m_i^* in the support of σ_i^* such that:

$$\sum_{m_{-i}} [U_i(g(m_i^*, m_{-i}), \theta) - U_i(g(m_i', m_{-i}), \theta)]\sigma_{-i}^*(m_{-i}) \geq 0 \qquad \text{(for all } m_i')$$

$$\sum_{m_{-i}} [U_i(g(m_i^*, m_{-i}), \theta') - U_i(g(m_i', m_{-i}), \theta')]\sigma_{-i}^*(m_{-i}) < 0 \qquad \text{(for some } m_i')$$

Therefore

$$\sum_{m_{-i}} [U_i(g(m_i^*, m_{-i}), \theta) - U_i(g(m_i^*, m_{-i}), \theta')]\sigma_{-i}^*(m_{-i}) >$$
$$\sum_{m_{-i}} [U_i(g(m_i', m_{-i}), \theta) - U_i(g(m_i', m_{-i}), \theta')]\sigma_{-i}^*(m_{-i})$$

Consider the agents who satisfy $F(\theta) \subseteq \max_i^{\theta'} A(\theta)$. As deviations can only lead to allocations in $A(\theta)$, it is clear that they cannot be the agents with a profitable deviation as any deviation leads to alternatives in $A(\theta)$.

Instead consider the agents for which, for all $x \in F(\theta)$, $L_i(x, \theta) \cap A(\theta) \subseteq L_i(x, \theta') \cap$ Therefore, as $u_i(x, \theta) = u_i(x, \theta')$ for all $x \in F(\theta)$ (which are the only allocation in the support of σ^*), and $u_i(x, \theta') \leq u_i(x, \theta)$ for all $x \in A(\theta)$, which are the only reachable allocations in the support of up to k player deviations from σ^* (and therefore one). We can therefore conclude that $U_i(g(m_i^*, m_{-i}), \theta) = U_i(g(m_i^*, m_{-i}), \theta')$ for all m_{-i} in the support of σ_{-i}^*. Therefore $\sum_{m_{-i}}[U_i(g(m_i^*, m_{-i}), \theta) - U_i(g(m_i^*, m_{-i}), \theta')]\sigma_{-i}^*(m_{-i}) = 0$. Further, $U_i(g(m_i^*, m_{-i}), \theta') \leq U_i(g(m_i', m_{-i}), \theta')$ for all m_{-i} in the support of σ_{-i}^*. However, together this implies that

$$\sum_{m_{-i}}[U_i(g(m_i^*, m_{-i}), \theta) - U_i(g(m_i^*, m_{-i}), \theta')]\sigma_{-i}^*(m_{-i}) = 0 \leq \sum_{m_{-i}}[U_i(g(m_i^*, m_{-i}), \theta') - U_i(g(m_i', m_{-i}), \theta')]\sigma_{-i}^*(m_{-i})$$

A contradiction.

For an arbitrary cardinal representation $u(\cdot, \theta)$ and arbitrary equilibrium σ^* at such a cardinal representation, we conclude that it must also be an equilibrium for some cardinal representation $u(\cdot, \theta')$. With this, we can conclude that a) the support of such an equilibrium must be in $F(\theta')$ (therefore $F(\theta) \subseteq F(\theta')$) and b) the outcomes in the support of a k deviation from σ^* must be included in the acceptability correspondence at θ' and therefore for a maximally safe A; $A(\theta) \subseteq A(\theta')$. □

Proof of theorem 2. Let $\mathcal{U} = \bigcup_{\theta \in \Theta} \mathcal{U}^\theta$ and define $\Theta\mathcal{U}$ as $\{(\theta, u) \in \Theta \times \mathcal{U} : u \in \mathcal{U}^\theta\}$. Consider the following mechanism $\langle M, g \rangle$. For each player $i \in N$ let $M_i = \Theta\mathcal{U} \times \{\alpha^i : \alpha^i : X \times \Theta^2 \to X\} \times \{\tilde{x}^i : \Theta \cup \{1\} \to \bigcup_{\theta \in \Theta} A(\theta) | \tilde{x}^i(\theta) \in A(\theta) \forall \theta \in \Theta\} \times \mathbb{N}$. That is, each player announces a state an associated cardinal utility function, a function from alternatives and pairs of states into outcomes, an outcome for each state, an acceptable outcome for each state and one additional outcome, and a positive integer. A typical element $m_i \in M_i$

is given by $((\theta^i, u^i), \alpha^i, \tilde{x}^i, n^i)$. Let $\mathbf{1}[x]$ be the degenerate lottery leading to the outcome x.

The allocation rule is as follows:

1. If $m_i = ((\theta, u), \alpha, \tilde{x}, 1)$ for all $i \in N$ and $\alpha(x, \theta, \theta) = x$ for all $x \in F(\theta)$ then:

$$g(m) = \frac{1}{|F(\theta)|} \sum_{x \in F(\theta)} \mathbf{1}[x]$$

2. If there exists some $j \in N$ such that $m_i = ((\theta, u), \alpha, \tilde{x}, 1)$ for all $i \in N \backslash \{j\}$ with $\alpha(x, \theta, \theta) = x$ for all $x \in F(\theta)$ and $m_j = ((\theta^j, u^j), \alpha^j, \tilde{x}^j, z^j) \neq m_i$ then

$$g(m) = \frac{1}{|F(\theta)|} \sum_{x \in F(\theta)} \delta_x(m)\mathbf{1}[\alpha^j(x, \theta, \theta')] + (1 - \delta_k(x))\mathbf{1}[x]$$

   where

$$\delta_x(m) = \begin{cases} \delta \in (0,1) & \text{if } \alpha^j(x, \theta, \theta^j) \in L_j(x, \theta) \cap A(\theta) \\ 0 & \text{otherwise} \end{cases}$$

3. If $\exists D \subset N$ such that $1 < |D| \leq k$ and $m_i = ((\theta, u), \alpha, \tilde{x}, 1)$ for all $i \notin D$, then the allocation in $\tilde{x}^j(\theta) \in A(\theta)$ where $j \in D$ is such that $z^j \geq z^{j'}$ for all $j' \in D$. If more than one has the highest integer select uniformly.

4. Otherwise, $g(m) = \tilde{x}^i(1)$ where i is such that $n^i > n^j$ for all $j \neq i$. If there is a tie for the highest integer uniformly randomise.

Step 1. Fix a state θ and any cardinal representation $u \in \mathcal{U}^\theta$. First it will be shown that for any $x \in F(\theta)$ there is an equilibrium $\sigma^* \in \mathcal{C}^{\mathcal{M}}(\theta, u)$ such that the support of $P(\sigma^*, g)$ contains x. Consider a set of strategies σ^* such hat $\sigma_i^* = ((\theta, u), \alpha, \tilde{x}, 1)$ for all $i \in N$. That is, a strategy that lies in rule 1. Player i deviating can only lead to rule 2, in which either the same allocation is given or one that is strictly worse at state θ via the inclusion of positive probability on $\alpha^j(x, \theta, \theta^j) \in L_j(x, \theta) \cap A(\theta)$, i.e. it shifts weight from $x \in F(\theta)$ to a less preferred option $\alpha^j(x, \theta, \theta^j)$. Therefore σ^* is a Nash equilibrium that has $x \in F(\theta)$ in it's support at state θ.

Step 2. Next it must be shown that for any $\theta \in \Theta$, $u \in \mathcal{U}^\theta$ that for any $\sigma^* \in \mathcal{C}^{\mathcal{M}}(\theta, u)$ the support of $P(\sigma^*, g)$ is included in $F(\theta)$. Further, it needs to be shown that safety is satisfied. Let $g^O(m) = \{x \in X | g(m)[x] > 0\}$ be the outcomes that occur with positive probability when the message m is played. Partition the set of messages into four sub-cases corresponding to the four rules of the mechanism.

1. $R_1 = \{m \in M | m_i = ((\theta, u), \alpha, \tilde{x}, 1)$ for all $i \in N$ and $\alpha(x, \theta, \theta) = x$ for all $x \in F(\theta)\}$.

2. $R_2^i = \{m \in M|\ m_j = ((\theta, u), \alpha, \tilde{x}, 1)$ for all $j \in N \backslash \{i\}$ and $\alpha(x, \theta, \theta) = x$ for all $x \in F(\theta)$ while $m_i \neq m_j\}$. Let $R_2 = \bigcup_{i \in N} R_2^i$.

3. $R_3^D\{m \in M|\ m_j = ((\theta, u), \alpha, \tilde{x}, 1)$ for all $j \in N \backslash D$ and $\alpha(x, \theta, \theta) = x$ for all $x \in F(\theta)$ while $m_i \neq m_j$ for all $i \in D\}$. Let $R_3 = \bigcup_{D \in \{D | 1 < |D| \leq k\}} R_3^D$.

4. $R_4 = M \backslash [R_1 \cup R_2 \cup R_3]$.

Consider $\sigma^* \in \mathcal{C}^{\mathcal{M}}(u^*, \theta^*)$ and let M^* be the set of message profiles that occur with positive probability under σ^*. It will be shown that $g(m^*) \subseteq F(\theta^*)$ and additionally any k-deviation from σ^* would lead to $A(\theta^*)$, which is implied by $g(m_D, m^*_{-D}) \in A(\theta^*)$ for all D such that $|D| \leq k$ and $m_D \in M_D$, $m^*_{-D} \in M^*_{-D}$.

To do so, we will consider a candidate deviation from this equilibrium σ_i^D, which will be constructive. It will be shown that the only occasion in which this is not a profitable deviation (and therefore it is possible that σ^* is a Nash equilibrium) is exactly when strong Set-Comonotonicity or Safe No-Veto can be applied. With this, we will show that for this mechanism the premise of sufficiency is upheld.

To this end, for any $i \in N$ and $m_i^* = ((\theta^i, u^i), \alpha^i, x^i, z^i) \in M_i^*$ define a deviation message $m_i^D(m_i^*) = ((\theta^i, u^i), \alpha^D, \tilde{x}^D, z^D)$ where 1) $\alpha^D(x, \theta, \theta) \in L_i(x, \theta) \cap A(\theta)$ is such that $\alpha^D(x, \theta, \theta) \succsim_i^{\theta^*} y$ for all $y \in L_i(x, \theta) \cap A(\theta)$. 2) $\tilde{x}^D(1)$ is such that $\tilde{x}^D(1) \succsim_i^\theta y$ for all $y \in \bigcup_{\theta' \in \Theta} A(\theta')$ while $\tilde{x}(\theta')$ is such that $\tilde{x}(\theta') \succsim_i^\theta y$ for all $y \in A(\theta')$. 3) $z^D > z^j$ for all but a mass of $1 \geq 1 - \mu > 0$ messages of other players played in equilibrium. Note this can be arbitrarily small but cannot be 0 due to the possibility of randomising over infinitely many strategies. Now let

$$\sigma_i^D(m_i) = \begin{cases} \sigma_i^*(m_i^*) & \text{if } m_i = m_i^D(m_i) \text{ for some } m_i^* \in M_i^* \\ 0 & \text{otherwise} \end{cases}$$

Notice that now no message in the support of $\sigma_i^D, \sigma_{-i}^*$ falls under rule 1. They must fall under either R_2^i, R_3^D such that $i \in D$, or R_4.

Suppose that $(m_i^*, m_{-i}^*) \in R_1$ would be realised with positive probability. That is, for all $j \neq i$ $m_j^* = m_i^* = ((\theta, u), \alpha, \tilde{x}, 1)$. if $\alpha^D(x, \theta, \theta) \neq x$ for some $\theta \in \Theta$ and $x \in F(\theta)$, by the strict preferences it must be that this is strictly profitable upon this realisation m_{-i}^* as it will ensure strictly more probability is put on $\alpha^D(x, \theta, \theta)$. By construction If $\alpha^D(x, \theta, \theta) = x$ for all $\theta \in \Theta$ and $x \in F(\theta)$ then the deviation does not change the allocation. We conclude that σ_i^D does not reduce the payoff for i in this case. In the case where it is not profitable for any agent to make sure a deviation, it must be that $L_i(x, \theta) \cap A(\theta) \subseteq L_i(x, \theta^*) \cap A(\theta)$ for all $x \in F(\theta)$ and therefore it must be that this deviation is not profitable for this particular message. As this is true for all i, as any player could make such a deviation, it also implies that, by strong Set-Comonotonicity we have that $F(\theta) \subseteq F(\theta^*)$, and therefore there would be no contradiction to m^* being in the support of the mixed Nash equilibrium. Further, as $A(\theta) \subseteq A(\theta^*)$ by the same condition we conclude that there is no contradiction to safety, as only allocations in $A(\theta)$ are reachable in k deviations from m^* (including mixed deviations).

Suppose that $(m_i^*, m_{-i}^*) \in R_2^i$ occurs. $j \neq i$ $m_i^* \neq m_j^* = ((\theta, u), \alpha, \tilde{x}, 1)$ Then for some $j \in N$, $j \neq i$ it must be that σ_j^D is a profitable deviation as they can induce their most preferred allocation in $A(\theta)$ when they announce the highest integer, which occur with sufficiently high probability to ensure that the strict preferences ensure their most preferred provides more utility than implemented by m^*, regardless of the outcome that occurs with the remaining probability. Given this, it is strictly profitable to deviate unless $g(m_i^*, m_{-i}^*) = y$ is the most preferred outcome at state θ^* within $A(\theta)$. In which case, the deviation does not decrease the utility. If it were not to change the outcome for any such j, by Safe No-Veto, it must be that $y \in F(\theta^*)$ and $A(\theta^*) = X$. Therefore, if the overall deviation is not strictly profitable (and this event occurs), then the condition of sufficiency is not violated. Otherwise, the deviation strictly increases the utility for some agent i with a sufficiently high probability in this event.

By a similar logic, either the deviation is strictly profitable when $(m_i^*, m_{-i}^*) \in R_3 \cup R_4$ for some agents or the condition of Safe No-Veto is used to show that the outcome would not violate the notion of safety or implementation.

As all cases are a weak improvement, and a strict improvement unless it would be permissible for σ^* to be an equilibrium given the condition of Safe No-Veto and strong Set-Comonotonicity, we have shown that implementation is not violated. □

Proof of Lemma 1. Consider a state θ with $F(\theta) = \{x^{l-1}, x^l, x^{l+1}\}$ where $l \geq 3$, x^{l-1} is not top ranked for any agent, and $x^{l-1} \succ_i^\theta x^{l+1}$ for the median peaked agent i. Consider an alternative state of the world θ' such that all preferences are the same for all agents, bar i, the median peaked agent. At θ', $x^{l-1} \succ_i^{\theta'} x^l$. Note this implies that, by the definition of F, $F(\theta') = \{x^{l-2}, x^{l-1}, x^l\}$. However, for all agents we have that $L_j(x^{l+1}, \theta) = L_j(x^{l+1}, \theta')$, and therefore by Maskin monotonicity we require that $x^{l+1} \in F(\theta')$. A contradiction. □

Proof of Proposition 4. The proof will rely on showing the conditions of Theorem 2 hold. Note that preferences are strict and $n \geq 3$.

We will now show that Set-Comonotonicity holds. Suppose that there are some θ, θ' such that for all $i \in N$, for all $x \in F(\theta)$, $L_i(x, \theta) \cap A(\theta) \subseteq L_i(x, \theta') \cap A(\theta)$. As this is the case, it must be that for $x^l = w(\theta)$ we have that for all agent's that if $x^l \succ_i^\theta y$ then $x^l \succ_i^{\theta'} y$ for $y \in \bigcup_{\theta' | w(\theta) = w(\theta')} \{x^{l'}, x^{l'+1}, ..., x^{l''-1}, x^{l''} | l' = \min_{i \in N} p_{i,\theta}, l'' = \max_{i \in N} p_{i,\theta}\}$. We will argue the median voter remains the same, and therefore $F(\theta') = F(\theta)$. It is clear that the median peaked individual, m, cannot have changed their preferences, as $L_m(x^{p_{m,\theta}}, \theta) \cap A(\theta) = A(\theta)$, and therefore $L_j(x^{p_{m,\theta}}, \theta') \cap A(\theta) = A(\theta)$. As $x^{p_{m,\theta}}$ must be interior in $A(\theta)$, as all agents have different peaks, it is only possible if $x^{p_{m,\theta}} = x^{p_{m,\theta'}}$. Therefore, if it were not the case, there must be some agent i such that $\text{sign}(p_{i,\theta} - p_{m,\theta}) \neq \text{sign}(p_{i,\theta'} - p_{m,\theta})$, to ensure the median voter changes. If $x^{p_{i,\theta'}} \in A(\theta)$, then this is a contradiction as $x^{p_{i,\theta'}} \in L_i(x, \theta) \cap A(\theta)$ while $x^{p_{i,\theta}} \notin L_i(x, \theta) \cap A(\theta)$, a contradiction. Therefore it must be that $x^{p_{i,\theta'}} \notin A(\theta)$. Suppose that $p_{i,\theta'} > \max_{\{\theta'|w(\theta')=w(\theta)\}} \max_{j \in N} p_{j,\theta'}$. Then it must be that $L_i(x, \theta') \cap A(\theta) = \{x^h \in A(\theta) | h \geq x^l, s.t. x^l = x\}$. However, this would require that $p_{i,\theta} < p_{m,\theta}$. If $p_{i,\theta} < p_{m,\theta} - 1$ then $\exists x^{h'} \in L_i(x, \theta) \cap A(\theta)$ such that $h' < p_{m,\theta}$, and therefore $L_i(x, \theta) \cap A(\theta) \not\subseteq L_i(x, \theta') \cap A(\theta)$. Therefore it must be that $p_{i,\theta} = p_{m,\theta} - 1$. However, as $x^{p_{m,\theta}-1} \in F(\theta)$, we can conclude that $x^{p_{m,\theta}} \in L_i(x^{p_{m,\theta}+1}, \theta) \cap A(\theta)$ while $x^{p_{m,\theta},\theta} \notin L_i(x^{p_{m\theta}+1}, \theta') \cap A(\theta)$, a contradiction as $x^{p_{m,\theta}} \in F(\theta)$. By analogy, if $p_{i,\theta'} < \min_{\{\theta'|w(\theta')=w(\theta)\}} \min_{j \in N} p_{j,\theta'}$ we conclude that it cannot be that Set-Comonotonicity holds. Therefore, whenever there are some θ, θ' such that for all $i \in N$, for all $x \in F(\theta)$, $L_i(x, \theta) \cap A(\theta) \subseteq L_i(x, \theta') \cap A(\theta)$ the median voter is the same and $F(\theta) = F(\theta')$. Further, by construction, $A(\theta) = A(\theta')$. Therefore the condition is satisfied.

All that is left to show is that Safe No-Veto is satisfied. To see this, recall that each agent has a different peak. Therefore there are n alternatives that must be the peak at any state $\theta \in \Theta$. As $|X| < 2n - 2$, the peaks are such that and $|A(\theta)| \geq n$ for any $\theta \in \Theta$, it must be that at least three peaks at θ' are in $A(\theta)$. As they have different peaks, and therefore maximal alternatives, we conclude that Safe No-Veto is satisfied. □