# Mechanism Design with Level-k Types: Theory and an Application to Bilateral Trade

Terri Kneeland*

October 7, 2021

**Abstract**

We develop necessary and sufficient conditions for level-k implementation that apply in independent private value environments. These conditions establish a set of level-k incentive constraints that are analogous to Bayesian incentive constraints. We show that in two special environments, the level-k incentive constraints collapse down to Bayesian incentive constraints. We then show, via a bilateral trade application, that this is not a general implication. Bilateral trade is ex post efficient under level-k implementation while it is not Bayesian implementable. We also address a robustness question concerning the common prior assumption embedded in level-k implementation by developing the concept of ex post level-k implementation. We develop necessary and sufficient conditions for ex post level-k implementation and show the relationship between ex post level-k and ex post implementation is analogous to the relationship between level-k and Bayesian implementation.

JEL codes: C72, D02, D90

*Department of Economics, University College London. Address: Gower Street, London, UK, WC1E 6BT. E-mail: t.kneeland@ucl.ac.uk. The project was supported by ERC Grant SUExp - 801635. I am grateful to Mike Peters for many invaluable discussions and guidance. For very helpful comments I would like to thank Vince Crawford, David Freeman, Yoram Halevy, Li Hao, Ryan Kendall, Kfir Eliaz, Wei Li, Rosemarie Nagel, Ariel Rubinstein, Sergei Severinov, Rani Spiegler, Alexander Wolitzky, and anonymous referees. The title of this paper is a play on Wolitzky's (2016) title "Mechanism Design with Maxmin Agents: Theory and an Application to Bilateral Trade".

# 1 Introduction

Laboratory experiments frequently find that behavior deviates from Nash and Bayesian equilibrium predictions when agents interact in novel environments. Non-equilibrium approaches, like level-k and cognitive hierarchy models, that relax the belief consistency assumptions of equilibrium models have been increasingly used to explain this behavior.[1] This empirical evidence prompts the need for extending the analysis of economic phenomena beyond an equilibrium analysis to other behaviorally plausible solution concepts.

This paper contributes to that end by analyzing mechanism design under the level-k solution concept. In the level-k model, agents anchor their beliefs in a naive model of others' behavior and adjust their beliefs by a finite number of iterated best responses. The model is anchored in the behavior of level 0 types, which is exogenously given. (The typical assumption in the literature specifies the anchor as uniformly random. The results in this paper hold for any arbitrary atomless anchor.) Level 1 types engage in one level of reasoning and best respond to level 0 behavior. Level 2 types engage in two levels of reasoning and best respond to level 1 types. And so on, with level $k$ types playing a best response to level $(k-1)$ types. This yields a tractable model of strategic behavior in which agents determine their optimal actions in a finite number of steps. The level-k solution concept relaxes the belief consistency assumption of equilibrium by allowing agents to hold (possibly) inaccurate beliefs about the levels of reasoning of their opponents. Our notion of level-k implementability is identical to the notion of Bayesian implementability, except our solution concept is the level-k solution concept: a (possibly multi-valued) social choice rule is level-k implementable if for every profile of payoff types and levels, there exist actions played under the level-k model which lead to outcomes that are consistent with the social choice rule.

Our main results establish general necessary and sufficient conditions for level-k implementation (Propositions 1 and 2). The level-k necessary conditions hold for general environments and the sufficient conditions hold in the case of independent private values. The conditions specify a set of level-k

[1] For pioneering work in the literature see Stahl & Wilson (1994; 1995), Nagel (1995), Costa-Gomes et al. (2001), and Camerer et al. (2004). For a recent survey of this literature, see Costa-Gomes et al. (2013).

incentive constraints that are analogous to standard Bayesian incentive constraints. The level-k incentive constraints require there to exist a function, for each agent, that maps payoff types to outcomes in such a way that truthfully reporting payoff types is optimal for that agent given everyone else is truthfully reporting their payoff type. Level-k incentive constraints allow these functions to differ across agents while Bayesian incentive constraints require these functions to be the same for all agents. The level-k incentive constraints are thus a weak relaxation of Bayesian incentive constraints. If Bayesian incentive constraints hold, then the level-k incentive constraints also hold. However, it may be possible to ensure that the level-k incentive constraints hold without the Bayesian incentive constraints holding.

The ability to satisfy the level-k incentive constraints independent of the Bayesian incentive constraints holding depends on the environment. We establish two restrictions on the environment where level-k incentive constraints collapse to Bayesian incentive constraints. The first is when the social planner is implementing a social choice function (a single-valued rule). And, the second is when the message space is restricted to that of the set of payoff types. In both these cases, Bayesian incentive constraints are necessary conditions for level-k implementation (Corollary 1 and Proposition 3). The results for these special cases mirror existing results in the literature. de Clippel et al. (2019) study single-valued social choice rules and Crawford (2021) studies implementation under the restriction to mechanisms where the message space is the set of payoff types in the bilateral trade environment. Both find that Bayesian incentive constraints are necessary conditions for level-k implementation. Both of these papers are discussed in detail in the related literature section below.

In contrast to the results in these two restricted environments, our sufficient conditions allow for the possibility that level-k implementation is *more* permissive than Bayesian implementation. We show that in a bilateral trade environment, ex post efficient trade is level-k implementable (Proposition 6). This is in obvious contrast to Bayesian implementation where there is a conflict between ex post efficiency and incentive compatibility. Thus, with this example, we show that the existing results in the literature - that bound level-k implementability to what is Bayesian implementable - arise purely from restrictions on either the environment or the mechanism, and do not hold in general.

Lastly the paper explores a robustness question concerned with relaxing the common prior assumption embedded in level-k implementation.[2] The definition of level-k implementation relies on the assumption of a common prior: agents' beliefs about others' levels are determined by the level-k model but agents' beliefs about the payoff types of others are determined by a common prior, as in Bayesian implementation. We develop the concept of ex post level-k implementation which effectively allows for any beliefs over payoff types. We establish general necessary and sufficient conditions for ex post level-k implementation (Propositions 4 and 5). As for level-k implementation, the necessary conditions hold for general environments and the sufficient conditions apply to environments of private values. The conditions specify a set of ex post level-k incentive constraints that are analogous to the standard ex post incentive constraints. The relationship between ex post level-k and ex post implementation mirrors the relationship between level-k and Bayesian implementation. If the ex post incentive constraints hold, then the ex post level-k incentive constraints hold. But, it may be possible to ensure that the ex post level-k incentive constraints hold without the ex post incentive constraints holding. We give an example and show that ex post efficient bilateral trade is ex post level-k implementable while it is not ex post implementable.

## Related literature

There is a growing literature that focuses on behavioral mechanism design. This paper adds to this literature by studying implementation under the level-k model. Four other papers study level-k implementation. Crawford et al. (2009) looks at setting optimal reserve prices in first and second price auctions when agents are level-k types. Gorelkina (2015) provides a level-k analysis of the expected externality mechanism.

[2] The need for the relaxation of strong common knowledge assumptions is known as the Wilson doctrine. Mechanisms that are robust to common knowledge assumptions can insure that a social choice rule will be implemented even if the planner does not know agents' beliefs about the payoffs of others. Much of this literature is due to Bergemann & Morris (2005), who investigate aspects of robust mechanism design (relaxing common knowledge of payoff assumptions) while maintaining the assumption of common knowledge of rationality. We investigate a version of robust implementation that relaxes common knowledge of payoffs under the empirically plausible assumption of level-k reasoning.

Crawford (2021) revisits Myerson & Satterthwaite's (1983) bilateral trade results under level-k implementation when the message set is restricted to the set of payoff types. Crawford considers two cases: (i) one where levels are unobservable and as such the social planner needs to screen both levels and payoff types (same environment as in this paper); and (ii) one where levels are observable, thus the social planner need only screen payoff types. In the first case, Crawford establishes a parallel result to Proposition 3 in this paper that shows that Bayesian incentive constraints are necessary for level-k implementation when the message set is restricted to the set of payoff types. And, hence shows the Myerson and Satterthwaite impossibility result for ex post efficient trade holds for level-k implementation when the message set is restricted. Crawford also explores what the 'second-best' level-k mechanisms look like in cases where full ex post efficient trade cannot be achieved. In the latter case, Crawford shows that when levels are observable, a setting not explored in this paper, the relationship between level-k and Bayesian implementation is ambiguous and that the Myerson and Satterthwaite impossibility result can break down. Crawford gives a complete Myerson-Satterthwaite-style characterization of the optimal (restricted) mechanism in this case.

The current paper is closest to de Clippel et al. (2019). They establish a set of necessary and sufficient conditions for level-k implementation in a general setting where the social planner aims to implement a single-valued social choice rule. Their main finding is that Bayesian incentive constraints are necessary conditions for level-k implementation. In contrast, we establish a set of necessary and sufficient conditions for the case of a (possibly) multi-valued social choice rule and find that level-k implementation is actually a weaker implementation requirement than Bayesian implementation: a social choice rule may be level-k implementable even though it is not Bayesian implementable.

de Clippel et al., however, use a slightly stronger definition of level-k implementation than the one used here - requiring a version of full implementation[3] where this paper allows partial implementation. One might wonder then, if the main result in our paper, that level-k implementation is weaker than Bayesian

[3]SIRBIC, which requires that the level-k incentive constraints hold with a strict inequality whenever the (single-valued) social choice rule is responsive, comes out as a necessary condition for full level-k implementation in de Clippel et al..

implementation, arises from (1) relaxing the environment from single-valued social choice rules to multi-valued rules or (2) relaxing the implementation requirement from full to partial implementation. We show that, (2), moving from full to partial implementation plays a minimal role under level-k implementation. First, Corollary 1 establishes that the necessary conditions for partial level-k implementation collapse to Bayesian incentive constraints when the social choice rule is single-valued (this replicates de Clippel et al.'s findings under partial implementation). Thus, demonstrating that moving from full to partial implementation when the social choice rule is single-valued gains nothing beyond Bayesian implementability. Second, we show that the only difference between the necessary and sufficient conditions for full and partial level-k implementation of multi-valued social choice rules is that the level-k constraints must hold with a strict inequality rather than a weak inequality. This, in theory, means that full level-k implementation may be possible when Bayesian implementation is not. And, lastly, we show that this is true in practice: we provide an example where bilateral trade is ex post efficient under full level-k implementation but it is not Bayesian implementable.

The previous two papers place restrictions on the implementation problem: Crawford restricts the message space to be equal to the set of payoff types and de Clippel et al. restrict attention to single-valued social choice rules. A general takeaway from these papers (when levels are unobservable) is that Bayesian incentive constraints determine the boundaries of what is level-k implementable. However, the current paper shows that this conclusion arises from their restrictions on the implementation problem. If the social planner is interested in multi-valued choice rules and willing to use general message spaces, then level-k implementation can be strictly less restrictive than Bayesian implementation. The bilateral trade application is one such example. Instead, one can take the weaker level-k necessary and sufficient conditions established in this paper, as defining boundaries for what is level-k implementable in independent, private value environments.

Both de Clippel et al. (2019) and Crawford (2021) use concepts of level-k implementation, similar to this paper, that requires a common prior assumption: agents' beliefs about the payoff types of others is determined by a common prior. The additional analysis of ex post level-k implementation in this paper relaxes the common prior assumption. This analysis makes a unique contribution to the literature.

There is also a literature on non-equilibrium design that does not employ the level-k model. Hagerty & Rogerson (1987), Bulow & Roberts (1989), Copic & Ponsati (2008; 2016), Mookherjee & Reichelstein (1992), Matsushima (2007; 2008), Bergemann & Morris (2005; 2009). Bergemann et al. (2011), de Clippel et al. (2015), Saran (2016), and Ollar & Penta (2017) all study implementation in dominant strategies, implementation in iterative dominance, implementation in rationalizable strategies, rationalizable implementation with an upper bound, or distribution-free implementation. Börgers & Li (2019) study implementation in strategically simple mechanisms that only require agents to use first-order beliefs. Healy (2006) studies implementation in public good games when agents are learning to play equilibrium strategies.[4]

This rest of the paper proceeds as follows. Section 2 sets up the general payoff environment and formalizes level-k implementation. Section 3 establishes necessary and sufficient conditions for level-k implementation. Section 4 looks at two examples of special environments where the level-k incentive constraints collapse down to Bayesian incentive constraints. Section 5 addresses what happens when we relax the common prior assumption by looking at ex post level-k implementation. Section 6 sets up the bilateral trade environment and shows that ex post efficient trade is both level-k and ex post level-k implementable. Section 7 concludes. Omitted proofs can be found in Appendix A.

[4]There is a literature that studies behavioral mechanism design that relies on equilibrium: Eliaz (2002) studies mechanism design when there is a proportion of 'faulty' agents that fail to act optimally. Glazer & Rubinstein (2012) allow the content and framing of the mechanism to play a role in behavior. de Clippel (2014) studies mechanism design when agents are not rational. Saran (2011a) shows that ex post efficient trade can be achieved under bilateral trade when there is a proportion of truthful traders. Saran (2011b) shows that incentive compatibility is not a necessary restriction for partial implementation of social choice functions when preferences are menu dependent. Wolitzky (2016) investigates mechanism design and bilateral trade when agents are maxmin expected utility maximizers. Glazer & Rubinstein (1998), Eliaz & Spiegler (2006; 2007; 2008), Severinov & Deneckere (2006) study behavioral mechanism design in individual decision problems.

# 2 Setup

## 2.1 General payoff environment

There is a finite set of agents $I = 1, 2, \ldots, n$. Agent $i$'s *payoff type* is $\theta_i \in \Theta_i$, where $\Theta_i$ is a finite set. There is a compact set of outcomes $Y$. Each agent has a continuous utility function $u_i : Y \times \Theta \to \mathbb{R}$. Note that we use the notation $X = X_1 \times \cdots \times X_n$ and $X_{-i} = X_1 \times \cdots \times X_{i-1} \times X_{i+1} \times \cdots \times X_n$ for sets $\{X_i\}_{i\in I}$ throughout this paper.

There is a social planner who is concerned with implementing a (possibly multi-valued) social choice rule $F : \Theta \to 2^Y \backslash \emptyset$. The planner would like the outcome to be an element of $F(\theta)$ whenever the true payoff type profile is $\theta$.

## 2.2 Type spaces

We use the framework of a type space in order to formally define agents' beliefs about the payoff types of others. The standard way to do this is to use a Bayesian type space. The set of payoff types along with a common prior over the set of payoff types constitutes a Bayesian type space.

**Definition 1.** A **Bayesian type space** $\mathcal{B}$ is a structure $\mathcal{B} = \langle \Theta; \rho \rangle$, where $\rho \in \triangle(\Theta)$.

Given the common prior $\rho$, each payoff type forms her beliefs by conditioning on the common prior according to Bayes' rule. The belief of an agent with payoff type $\theta_i$ about the payoff types of others is given by $\rho(\theta_{-i}|\theta_i) = \frac{\rho(\theta)}{\sum_{\theta_{-i}\in\Theta_{-i}} \rho(\theta_i,\theta_{-i})}$. We restrict attention to common priors that generate well-defined beliefs. Specifically, we assume that $\sum_{\theta_{-i}\in\Theta_{-i}} \rho(\theta_i, \theta_{-i}) > 0$ for all $\theta_i \in \Theta_i$ and $i \in I$.

Similarly, we use a type space approach to define the level-k model. The level-k type space generates types that differ both by their payoff type and their level of reasoning.[5]

**Definition 2.** A **$\mathcal{B}$-based level-k type space** $\mathcal{L}$ is a structure $\mathcal{L} = \langle \mathcal{B}; (T_i, \mu_i)_{i=1,\ldots,n}; \bar{k} \rangle$, such that $\mathcal{B}$ is a Bayesian type space $\mathcal{B} = \langle \Theta; \rho \rangle$, $T_i =$

[5] This approach follows the models of Crawford & Iriberri (2007a) and Crawford et al. (2009) which adapted the level-k models to incomplete information environments.

$\Theta_i \times \{0, 1, \ldots, \bar{k}\}$, and $\mu_i : \Theta_i \times \{1, \ldots, \bar{k}\} \to \triangle(T_{-i})$ such that

$$\mu_i((\theta_{-i}, k_{-i})|(\theta_i, k_i)) = \begin{cases} \rho(\theta_{-i}|\theta_i) & \text{if } k_j = k_i - 1 \; \forall j \neq i \\ 0 & \text{otherwise} \end{cases}.$$

In a level-k type space, an agent's *type,* $(\theta_i, k_i) \in T_i$, is 2-dimensional, representing both her payoff type, $\theta_i$, and her level, $k_i$. An agent's level represents her level of reasoning - an agent with a level $k$ uses only $k$ steps of reasoning in order to figure out her optimal behavior in any game.[6] An agent's beliefs about the types of others are determined both by her payoff type and her level. The beliefs of type, $(\theta_i, k_i)$, about the types of others are determined by the function $\mu_i((\theta_{-i}, k_{-i})|(\theta_i, k_i))$. An agent with a level $k$ puts weight only on types that have levels $(k-1)$. This captures the core assumption of the level-k model. An agent's beliefs about the payoff types of other agents are determined by the common prior $\rho$. Thus, an agent with payoff type $\theta_i$ and level $k$ believes that the payoff types of other agents are determined by $\rho(\cdot|\theta_i)$ and that others have level $k-1$.

We formally call this type space a Bayesian-based level-k type space because beliefs about payoff types are derived from a common prior. We simplify this terminology throughout the rest of this paper and refer to these type spaces simply as level-k type spaces.

## 2.3 Solution concepts

A mechanism specifies an action set for each agent and a mapping between action profiles and outcomes.

**Definition 3.** A **mechanism** $\langle M, f \rangle$ consists of a set of actions $M = M_1 \times \cdots \times M_n$ and a function $f : M \to \triangle(Y)$.

Given the payoff environment and (Bayesian or level-k) type space, a mechanism defines a $n$-agent incomplete information game with action set $M_i$ and payoffs defined by $u_i : Y \times \Theta \to \mathbb{R}$ and $f : M \to \triangle(Y)$ for agent $i$.

[6] The bound on the level of reasoning is not necessary, the results in this paper go through if $T_i = \Theta_i \times \{0, 1, 2, \ldots\}$, however bounding the depths of reasoning maintains the finiteness of the type space for simplicity.

For a given level-k type space, we can define the level-k solution concept. The level-k solution concept imposes that all types, $k \geq 1$, are rational (that is, they play a best response given their beliefs about the actions of other agents) and have consistent beliefs about the actions of other types. The behavior of level 0 types is specified outside of the model. Thus, level 0 types do not play a best response to their beliefs (and may play actions that are not a best response to any belief). To specify level 0 behavior, we define the notion of an anchor below.

**Definition 4.** For a given game defined by a mechanism $\langle M, f \rangle$ and a ($\mathcal{B}$-based) level-k type space $\mathcal{L} = \langle \mathcal{B}; (T_i, \mu_i)_{i=1,\ldots,n}; \bar{k} \rangle$, an **anchor** $\alpha = \alpha_1 \times \cdots \times \alpha_n$ is a mapping $\alpha_i : \Theta_i \times \{0\} \to \triangle(M_i)$ for all $i \in I$.

An anchor is a essentially a strategy for each level 0 type. Notice that anchors can be arbitrary and vary from agent to agent and payoff type to payoff type. A commonly applied assumption in the level-k literature is of a *uniformly random anchor*, which would restrict $\alpha_i$ to be the uniform random probability distribution over $M_i$ for all payoff types and agents. The results that follow will be proved either for an arbitrary anchor or under the assumption of an atomless anchor. If $M_i$ contains a continuum of messages for each $i$, then the anchor $\alpha$ is an *atomless anchor* if the distribution $\alpha_i(\theta_i, 0)$ of messages contains no atoms, for each $\theta_i$ and each $i$. If $M_i$ is a continuum, then the uniformly random anchor is an example of an atomless anchor.

We now define a level-k solution under a given anchor $\alpha$.

**Definition 5.** For a given game defined by a mechanism $\langle M, f \rangle$, a level-k type space $\mathcal{L} = \langle \mathcal{B}; (T_i, \mu_i)_{i=1,\ldots,n}; \bar{k} \rangle$, and an anchor $\alpha$, a strategy profile $s = s_1 \times \cdots \times s_n$, with $s_i : T_i \to \triangle(M_i)$ for all $i \in I$, is a **level-k solution under anchor $\boldsymbol{\alpha}$** if and only if:

**(i)** $s_i(\theta_i, 0) = \alpha_i(\theta_i, 0)$ for all $\theta_i \in \Theta_i$

**(ii)** $\sum_{\theta_{-i} \in \Theta_{-i}} \rho(\theta_{-i}|\theta_i) u_i(f(s_i(\theta_i, k), s_{-i}(\theta_{-i}, k-1)), \theta)$

$$\geq \sum_{\theta_{-i} \in \Theta_{-i}} \rho(\theta_{-i}|\theta_i) u_i(f(m_i', s_{-i}(\theta_{-i}, k-1), \theta) \ \forall m_i' \in M_i, \ \theta_i \in \Theta_i,$$

$k \in \{1, \ldots, \bar{k}\}$, $i \in I$

The level-k solution can be calculated recursively given the behavior of level 0 types (condition (i)). Level 1's actions are a best response to level 0's actions (condition (ii)). Level 2's actions are a best response to level 1's actions, and so on (condition (ii)). We have abused notation slightly in the above definition, as it may be the case that $f \circ s$ is a compound lottery. In this instance, $u_i(f(s(\cdot), \theta)$ in condition (ii) should be taken to be the corresponding expected utility representation.

## 2.4 Implementation

A social choice rule is level-k implementable if there exists a mechanism, an atomless anchor, and a level-k solution that achieves the social planners objective for every message sent. The formal definition is given below.

**Definition 6.** A social choice rule, $F$, is **level-k implementable** on $\mathcal{L} = \langle \mathcal{B}; (T_i, \mu_i)_{i=1,\ldots,n}; \bar{k} \rangle$ if there exists a mechanism $\langle M, f \rangle$, an anchor $\alpha$, and a strategy profile $s = s_1 \times \cdots \times s_n$ such that $s$ is a level-k solution under anchor $\alpha$ and $s$ achieves $F$: $f(s(\theta, \hat{k})) \in F(\theta)$ for all $(\theta, \hat{k}) \in \times \left\{ \Theta_i \times \{1, \ldots, \bar{k}\} \right\}_{i \in I}$.

First, notice that our notion of level-k implementability does not require the mechanism to satisfy the social choice rule for level 0 types. This is because level 0 agents are non-strategic, hence the social planner cannot incentivize their behavior.[7] There is also some empirical support that the proportion of level 0 agents is small (e.g. Arad & Rubinstein 2012; Costa-Gomes et al. 2001; Costa-Gomes & Crawford 2006; Brocas et al. 2014). Thus, we interpret level 0 types as types that exist only in the minds of other types.

Second, notice that our notion of level-k implementation does not require the social planner to have knowledge of the *actual* distribution of agents' levels. This is because implementation requires that the outcome be consistent with the social choice rule for *all* levels and hence does not depend upon the distribution.[8]

[7] If this type of behavior is a concern we should consider an alternative form of implementability. See Eliaz (2002) for one such possibility - the social planner tries to minimize the deviations from the social choice rule.

[8] This is not true for all mechanism design objectives. For example, it would not be true if the goal of the planner was to maximize expected revenue. If different levels (and payoff types) play different actions with different revenue consequences, then expected revenue will depend upon the actual distribution of both payoffs and levels.

We will be interested in comparing level-k implementation with that of Bayesian implementation. We define Bayesian implementation for completeness. In the below definition, we incorporate the results of the revelation principle, and hence define Bayesian implementation under a direct mechanism. Condition (ii) below states the standard Bayesian incentive constraints which will be contrasted with the level-k incentive constraints developed in the next section. Notice that when we compare level-k and Bayesian implementation throughout this paper we will be comparing Bayesian implementation given some Bayesian type space $\mathcal{B} = \langle \Theta; \rho \rangle$ and the related Bayesian-based level-k type space $\mathcal{L} = \langle \mathcal{B}; (T_i, \mu_i)_{i=1,\dots,n}; \bar{k} \rangle$. In other words, we will compare Bayesian and level-k implementation given a shared common prior $\rho$.

**Definition 7.** A social choice rule $F$ is **Bayesian implementable** on $\mathcal{B} = \langle \Theta; \rho \rangle$ if there exists a mechanism $\langle \Theta, f \rangle$ such that the following conditions hold:

**(i)** $f(\theta) \in F(\theta) \; \forall \theta \in \Theta$

**(ii)** $\sum_{\theta_{-i} \in \Theta_{-i}} \rho(\theta_{-i}|\theta_i) u_i(f(\theta), \theta) \geq \sum_{\theta_{-i} \in \Theta_{-i}} \rho(\theta_{-i}|\theta_i) u_i(f(\theta'_i, \theta_{-i}), \theta) \; \forall \theta'_i \in \Theta_i$, $\theta \in \Theta$ , $i \in I$

# 3 Necessary and sufficient conditions for level-k implementation

This section establishes necessary and sufficient conditions for level-k implementation. We provide the necessary and sufficient conditions separately as the necessary conditions hold in general environments, while the sufficient conditions only guarantee implementation in independent private value environments.

**Proposition 1.** *(Necessary Conditions) Let $F$ be a social choice rule. Let $\mathcal{B}$ be a Bayesian type space and let $\mathcal{L} = \langle \mathcal{B}; (T_i, \mu_i)_{i=1,\dots,n}; \bar{k} \rangle$ be a ($\mathcal{B}$-based) level-k type space with $\bar{k} \geq 2$. If $F$ is level-k implementable then there exists a function $f^i : \Theta \to \triangle(Y)$ for each $i \in I$ and a function $\bar{f} : \Theta \to \triangle(Y)$, such that the following conditions hold:*

**(i)** $f^i(\theta), \bar{f}(\theta) \in F(\theta)\ \forall \theta \in \Theta,\ i \in I$

**(ii)** $\sum_{\theta_{-i} \in \Theta_{-i}} \rho(\theta_{-i}|\theta_i) u_i(f^i(\theta), \theta) \geq \sum_{\theta_{-i} \in \Theta_{-i}} \rho(\theta_{-i}|\theta_i) u_i(f^i(\theta_i', \theta_{-i}), \theta)\ \forall \theta_i' \in \Theta_i \backslash \theta_i,$ $\theta \in \Theta,\ i \in I$

**(iii)** $\sum_{\theta_{-i} \in \Theta_{-i}} \rho(\theta_{-i}|\theta_i) u_i(f^i(\theta), \theta) \geq \sum_{\theta_{-i} \in \Theta_{-i}} \rho(\theta_{-i}|\theta_i) u_i(\bar{f}(\theta_i', \theta_{-i}), \theta)\ \forall \theta_i' \in \Theta_i,$ $\theta \in \Theta,\ i \in I$

The formal proof can be found in Appendix A. We will give the intuition here. If the social choice rule $F$ is level-k implementable, then there exists some mechanism $\langle M, g \rangle$, an anchor $\alpha$, and level-k solution under $\alpha$, $m = m_1 \times \cdots \times m_n$, that achieves $F$.

Consider the behavior of a level 2 agent $i$, specifically an agent with type $(\theta_i, 2)$. This agent sends message $m_i(\theta_i, 2)$ and believes that everyone else is of level 1. Define the notation $(\theta, k) = ((\theta_1, k), \ldots, (\theta_n, k))$ and $(\theta_{-i}, k) = ((\theta_1, k), \ldots, (\theta_{i-1}, k), (\theta_{i+1}, k), \ldots, (\theta_n, k))$. Then the agent believes others send the message profile $m_{-i}(\theta_{-i}, 1)$. This agent could send some other message $m_i(\theta_i', 2)$ for any $\theta_i' \in \Theta_i$. But, because $\langle M, g \rangle$ and $m$ level-k implement the social choice rule, it must be that she prefers to send $m_i(\theta_i, 2)$ given her beliefs. Define $f^i(\theta) = g(m_i(\theta_i, 2), m_{-i}(\theta_{-i}, 1))$ for all $\theta \in \Theta$. Thus, it must be that condition (ii) holds for agent $i$.

Alternatively, this agent could send messages of the form $m_i(\theta_i', 1)$ for any $\theta_i' \in \Theta_i$. But again, because $\langle M, g \rangle$ and $m$ level-k implements the social choice rule, it must be that she prefers to send $m_i(\theta_i, 2)$. Define $\bar{f}(\theta) = g(m(\theta, 1))$ for all $\theta \in \Theta$. Thus, it must be that condition (iii) holds for agent $i$. The same argument extends to all agents. Further, it must be the case that $g(m_i(\theta_i, 2), m_{-i}(\theta_{-i}, 1)), g(m(\theta, 1)) \in F(\theta)$ for all $\theta \in \Theta$ since $g$ achieves $F$. Thus, condition (i) must hold. Therefore, it is possible to find functions $\{f^i\}_{i \in I}$ and $\bar{f}$ such that conditions (i)-(iii) hold.

The necessary conditions are generated from the incentive requirements of a level 2 agent. The incentive requirements for higher levels mimic those for the level 2 type. However, the incentive requirements for the level 1 type may be quite different. This is because level 1 types believe all others are level 0 types, and the behavior of the level 0 types is exogenously given. But, it turns out that in independent private value environments it is possible to satisfy the level 1 incentive constraints without any additional conditions beyond (i)-(iii) under atomless anchors. Thus, within independent private value environments the conditions in Proposition 1 are both necessary and sufficient for level-k implementation.

We first give the intuition for why the necessary conditions are sufficient. We will use the functions $\{f^i\}_{i\in I}$ and $\bar{f}$ to construct a mechanism that will level-k implement the social choice rule. Suppose that agents could send messages about both their payoff types and their levels i.e. the message space for agent $i$ is $\Theta_i \times \{0, \ldots, \bar{k}\}$. Consider agent $i$ and suppose that all other agents are truthfully reporting levels and payoff types (putting aside level 1 agents for now). Let the function $f^i$ determine the outcomes that agent $i$ believes will be implemented when agent $i$ truthfully reports her level ( i.e. the outcomes that will occur when agent $i$ reports level $k$ and everyone else reports levels $k-1$). And, let the function $\bar{f}$ determine the outcomes that agent $i$ believes will be implemented when agent $i$ reports any other level (e.g. when a level 2 agent reports level 1 when all other agents report levels 1). If an agent believes that all other types are truthfully reporting levels and payoff types, then under condition (ii) and (iii) agents will want to truthfully report their payoff types and levels as well. Thus, the problem for the social planner is really how to incentivize level 1 agents to truthfully report their payoff types and levels for some exogenously given behavior (anchor) of level 0 agents.

To resolve this issue, we will extend the mechanism developed in de Clippel et al. (2019)[9] (proof of Proposition 3 in their paper). This mechanism satisfies the level 1 incentive constraints by effectively manipulating the beliefs of level 1 agents to mimic that of truthful reporting of payoff types and levels. To understand how the mechanism works, consider the case where level 0 behavior is uniformly random and consider the following. Suppose the planner augmented the message space with an additional set, $Z_i$, for each agent, i.e. the message set for agent $i$ would then be $\Theta_i \times \{0, \ldots, \bar{k}\} \times Z_i$. And, suppose the designer could use the message sent from $Z_i$ to screen the agents into those that were level 0 and those that had higher levels, i.e. define the set $Z_i^+$ to be such

[9] Earlier drafts of this paper only provided the sufficient conditions in bilateral trade environments. After de Clippel et al. (2019) was written in 2016, we were able to adapt their proof technique to show our necessary conditions were also sufficient in general independent private value environments.

that any message sent in $Z_i^+$ would indicate the agent had a level of at least 1 and any other message would suggest a level of 0. The planner could then do the following: if he received a message in $Z_i^+$ he would take the type reports (payoff types and levels) at face value, but if he received any other message, he would modify the reports by randomly choosing a payoff type according to the common prior distribution $\rho_i$ and set a level of 0. The planner would then assign outcomes under these (potentially modified) type reports via the $\{f^i\}_{i \in I}$ and $\bar{f}$ functions as described above. If it was also the case that $Z_i$ is an uncountable set and $Z_i^+$ is countable, then a level 1 agent who believed a level 0 agent $j$ is uniformly randomly sending messages, would believe that the planner would almost surely ignore the type report of agent $j$ and use the modified report where the payoff type is randomly chosen from $\rho_j$ and set a level 0. This means that a level 1 agent would have the same beliefs about the outcomes that occur when she believes level 0 behavior is uniformly random as she would if she believed level 0 agents were truthfully reporting their level and payoff type. As such, level 1 agents would have an incentive to truthfully report their payoff type and level.

Given this intuition it is easy to see why such a proof only works for the case of independent, private value environments and atomless anchors. First, the social planner needs to be able to mimic the beliefs of the agents under the common prior by drawing the payoff types from a random distribution. He can do this only in the case of independent values when he draws a payoff type for agent $j$ randomly from $\rho_j$. Second, agents should only care about the modified payoff types used by the social planner and not the actual payoffs types of the other agents. In other words, agents need to only care about the payoff types of others to the extent that it tells them about the messages sent and not because it impacts utility directly. Thus, the environment needs to be one of private values. Third, agent $i$ needs to believe that the modified payoff types are drawn from $\rho_{-i}$ and levels are set to 0 almost surely. This will happen in the case of atomless anchors, but cannot be guaranteed with this mechanism otherwise.

Proposition 2 gives the formal result. Note that the proposition is stated only for the case when $n \geq 3$. The case when $n = 2$ is discussed in Remark 1 below. Define an environment of private values to be one where utility functions are such that $u_i : Y \times \Theta_i \to \mathbb{R}$ for all $i \in I$. And, define an environment of independent values to be one where the Bayesian type space, $\mathcal{B} = \langle \Theta; \rho \rangle$, is such that $\rho = \prod_i \rho_i$ for some $\rho_1 \times \cdots \times \rho_n \in \triangle(\Theta_1) \times \cdots \times \triangle(\Theta_n)$. Notice that $\rho(\theta_{-i}|\theta_i) = \rho(\theta_{-i}|\theta_i')$ for any $\theta_i, \theta_i' \in \Theta_i$ in the case of independent value environments. We will use the notation $\rho(\theta_{-i}) = \rho(\theta_{-i}|\theta_i)$ in this case.

**Proposition 2.** *(Sufficient Conditions) Let $F$ be a social choice rule, the environment be one of independent private values, and $n \geq 3$. Let $\mathcal{B}$ be a Bayesian type space and let $\mathcal{L} = \langle \mathcal{B}; (T_i, \mu_i)_{i=1,\ldots,n}; \bar{k} \rangle$ be a ($\mathcal{B}$-based) level-k type space. If there exists a function $f^i : \Theta \to \triangle(Y)$ for each $i \in I$ and a function $\bar{f} : \Theta \to \triangle(Y)$ such that the conditions (i)-(iii) hold in Proposition 1 then $F$ is level-k implementable. Further, there exists a mechanism $\langle M, f \rangle$ such that for any atomless anchor $\alpha$ there is a level-k solution under $\alpha$ that achieves $F$.*

Proof[10]:

Consider the following mechanism where the message space for agent $i$ is equal to $M_i = \Theta_i \times \{0, 1, \ldots \bar{k}\} \times [-1, 1]$ and consists of a report of her payoff type $\theta_i \in \Theta_i$, level $k_i \in \{0, 1, \ldots, \bar{k}\}$, and a real number, $z_i \in [-1, 1]$. Let $\alpha$ be any atomless anchor.

Let the indicator function $I_i : [-1, 1]^n \to \{0, 1\}$ be defined as follows:

$$I_i(z) \quad = \quad \begin{cases} 1 & \text{if } z_i = \kappa z_j \text{ for some } \kappa \in \mathbb{Z},\ \forall j \in I \\ 0 & \text{otherwise} \end{cases}$$

Define $\tilde{\theta}_i : M \to \Theta_i$ and $\tilde{k}_i : M \to \{0, 1, \ldots, \bar{k}\}$ in the following way. For a given message profile $m = (\theta, k, z)$, if $I_i(z) = 1$ the planner takes the reports as given and sets $\tilde{\theta}_i(m) = \theta_i$ and $\tilde{k}_i(m) = k_i$; otherwise the planner sets $\tilde{\theta}_i(m)$ to some randomly chosen $\Theta_i$ according to the prior $\rho_i$ and $\tilde{k}_i(m) = 0$. The planner then assigns outcomes based on the reports $(\tilde{\theta}, \hat{\tilde{k}})$ according to the function $g : \Theta \times \{0, \ldots, \bar{k}\}^n \to \triangle(Y)$ defined by

[10]While we have maintained finiteness of the type space for simplicity, this proof relies on a message space that contains a continuum. A complete proof requires showing that such a mechanism and strategies are measurable. We direct the reader to de Clippel et al. (2019) for such a proof.

$$g(\theta,\hat{k}) \quad = \quad \begin{cases} f^i(\theta) & \text{if } \hat{k}_j = \hat{k}_i - 1 \text{ for all } j \neq i \in I \\ \bar{f}(\theta) & \text{otherwise} \end{cases}.$$

First, consider an agent with payoff $\theta_j$ and level 0. This agent plays an atomless strategy $\alpha_j$. Thus, the probability that the realized value of $\kappa z_i$ is equal to $z_j$ is zero for any $z_i \in [-1, 1]$. (For example, if $z_i = 1/2$ then the set of $z_j$ for which there exists some $\kappa \in \mathbb{Z}$ such that $z_j = \frac{1}{2}\kappa$ is finite: $\{-1, -\frac{1}{2}, 0, \frac{1}{2}, 1\}$.)

Now, consider an agent $i$ with payoff $\theta_i$ and level 1. She believes that all $j \neq i$ are level 0 agents playing an atomless strategy $\alpha_j$. Hence, our level 1 agent believes that the planner will almost surely use a payoff type for agent $j$ that is picked at random according to the prior $\rho_j$ and a level report equal to 0. Further, if our level 1 agent sends a non-zero report, $z_i \neq 0$, she will expect the planner to disregard her payoff type report and choose randomly according to $\rho_i$ with probability 1. However, if our level 1 agent sends a zero report, $z_i = 0$, she will expect the planner to use her payoff type and level as reported.

Thus, if she sends the message $(\theta_i', k_i, z_i)$ with $z_i = 0$ and $k_i = 1$, she will expect to receive the following lottery over outcomes

$$\sum_{\theta_{-i}\in\Theta_{-i}} \rho(\theta_{-i}) \cdot f^i(\theta_i', \theta_{-i}).$$

If she sends the message $(\theta_i', k_i, z_i)$ with $z_i = 0$ and $k_i \neq 1$, she will expect to receive the following lottery over outcomes

$$\sum_{\theta_{-i}\in\Theta_{-i}} \rho(\theta_{-i}) \cdot \bar{f}(\theta_i', \theta_{-i}).$$

And, if she sends the message $(\theta_i', k_i, z_i)$ with $z_i \neq 0$, she will expect to receive the following lottery over outcomes

$$\sum_{\theta\in\Theta} \rho(\theta) \cdot \bar{f}(\theta).$$

By condition (ii) we have that

$$\sum_{\theta_{-i}\in\Theta_{-i}} \rho(\theta_{-i}) \cdot u_i(f^i(\theta), \theta_i) \geq \sum_{\theta_{-i}\in\Theta_{-i}} \rho(\theta_{-i}) \cdot u_i(f^i(\theta_i', \theta_{-i}), \theta_i)$$

for all $\theta_i' \in \Theta_i$ .

By condition (iii) we have that

$$\sum_{\theta_{-i} \in \Theta_{-i}} \rho(\theta_{-i}) \cdot u_i(f^i(\theta), \theta_i) \geq \sum_{\theta_{-i} \in \Theta_{-i}} \rho(\theta_{-i}) \cdot u_i(\bar{f}(\theta_i', \theta_{-i}), \theta_i)$$

for all $\theta_i' \in \Theta_i$ .

It must then also be true that

$$\sum_{\theta_{-i} \in \Theta_{-i}} \rho(\theta_{-i}) \cdot u_i(f^i(\theta), \theta_i) \geq \sum_{\theta \in \Theta} \rho(\theta) \cdot u_i(\bar{f}(\theta), \theta_i).$$

Thus, for agent $i$ with payoff type $\theta_i$ and level 1, reporting $(\theta_i, 1, 0)$ is a best response.

We now prove that an agent with payoff type $\theta_i$ and level $k$ will send the message $(\theta_i, k_i, 0)$ by induction on the following statement: Let $k \geq 1$ and assume that if for all $l \in \{1, \ldots, k-1\}$, $\theta_j \in \Theta_j$, and $j \in I$ an agent $j$ with payoff type $\theta_j$ and level $l$ will report $(\theta_j, l, 0)$, then an agent $i$ with payoff type $\theta_i$ and level $k$ will report $(\theta_i, k, 0)$.

The result is true for $k = 1$ by the above argument. Now consider an agent $i$ with payoff type $\theta_i$ and level $k \in \{2, \ldots, \bar{k}\}$. She expects that all other agents will send reports $z_j = 0$, $k_j = k - 1$, and truthfully report their payoff type. So, she expects that the social planner will always take their payoff and level reports as given.

Thus, if she sends the message $(\theta_i', k, 0)$ she will expect to receive the following lottery over outcomes

$$\sum_{\theta_{-i} \in \Theta_{-i}} \rho(\theta_{-i}) \cdot f^i(\theta_i', \theta_{-i}).$$

If she sends the message $(\theta_i', k_i, 0)$ with $k_i \neq k$, she will expect to receive the following lottery over outcomes

$$\sum_{\theta_{-i} \in \Theta_{-i}} \rho(\theta_{-i}) \cdot \bar{f}(\theta_i', \theta_{-i})$$

for some $j \in I$.

And, if she sends the message $(\theta_i', k_i, z_i)$ with $z_i \neq 0$, she will expect to receive the following lottery over outcomes

$$\sum_{\theta \in \Theta} \rho(\theta) \cdot \bar{f}(\theta).$$

By the same argument above, conditions (ii) and (iii) then imply that our agent will send the report $(\theta_i, k, 0)$.

Therefore, if we define $m_i(\theta_i, k_i) = (\theta_i, k_i, 0)$ for all $\theta_i \in \Theta_i$, $k_i \in \{1, \ldots, \bar{k}\}$ and $i \in I$, then $m$ is a level-k solution under $\alpha$ and achieves $F$ by condition $(i)$. Hence, $F$ is level-k implementable. Further, since $\alpha$ was an arbitrary atomless anchor, $m$ acheives $F$ for all atomless anchors.

□

*Remark* 1. *The case when $n = 2$.* The case when $n = 2$ requires an extra sufficient condition, in addition to conditions (i)-(iii). The extra condition is needed because when $n = 2$ it is possible to deviate unilaterally to mimic other level-k belief structures. For example, consider agent 1 with level 2 (and take $\bar{k} = 2$) and the mechanism used in Proposition 2. This agent believes everyone else is level 1. When $n = 2$, this means she believes she could receive outcomes under either $f^2$, $\bar{f}$, or $f^1$ when reporting level 0, level 1, or level 2 respectively. However, when $n \geq 3$, she believes she will receive outcomes under $\bar{f}$, $\bar{f}$, or $f^1$ when reporting level 0, level 1, or level 2 respectively. Thus, we need an additional condition when $n = 2$: truthfully reporting payoff type under $f^i$ must be preferable to reporting any other payoff type under $f^j$.

*Remark* 2. *Relationship to Bayesian incentive constraints.* The level-k necessary and sufficient conditions generalize the standard Bayesian incentive constraints. The difference between the two is that the level-k conditions can be satisfied with a different function, $f^i$, for each agent, whereas the Bayesian incentive constraints must hold using the same function, $f$, for all agents. The intuition for this is straightforward. The relaxation of the cross-player restriction ($f^1 = \cdots = f^n$) arises because of the relaxation of consistent beliefs under the level-k model, i.e. a level 3 agent believes she is facing level 2 agents while a level 2 agent believes she is facing level 1 agents. Thus, all incentive constraints can be satisfied by different $f$ functions for each agent: an agent $\boldsymbol{i}$ with payoff type profile $\theta_i$ and level $k$ thinks she will receive $f^{\boldsymbol{i}}(\theta)$

when playing against level $k-1$ agents with payoff type profile $\theta_{-i}$, while an agent $\boldsymbol{j}$ with payoff type $\theta_j$ and level $k$ thinks she will receive $f^{\boldsymbol{j}}(\theta)$ when playing against level $k-1$ agents with payoff type profile $\theta_{-j}$. Agents with level $k$ never believe they are playing against other agents with level $k$. Because of these (potentially) inconsistent beliefs, the planner can promise different agents outcomes derived from different functions, $f^1, \ldots, f^n$.

Further, notice that if the Bayesian incentive constraints are satisfied for some function $f$, then conditions (i)-(iii) are automatically satisfied when $f^i = f$ for all $i \in I$ and $\bar{f} = f$. Therefore, if a social choice rule is Bayesian implementable then it will also be level-k implementable in independent private value environments. However, it is not necessarily the case that if a social choice rule is level-k implementable then it will be Bayesian implementable. We illustrate this with a bilateral trade application in Section 6.

*Remark* 3. *Full vs partial implementation.* Throughout this paper we use the notion of *partial* level-k implementation.[11] This arises because a message a type sends in a level-k solution needs not be a strict best response. This allows for the possibility of multiple level-k solutions where some of these solutions are not consistent with the social choice rule (notice, that multiple level-k solutions only arise through indifferences). We could achieve full implementation by requiring strict inequalities in conditions (ii) and (iii). Simple arguments can then be applied in the proof of Proposition 2 to show that requiring strict inequalities in conditions (ii) and (iii) is sufficient for full level-k implementation.[12] Thus, at least in independent private value environments,

[11] Technically, we use a notion of *weak-partial* level-k implementation and are contrasting this with a notion of *weak-full* implementation. *Partial* refers to the fact we only require there to exist one level-k solution that is consistent with the social choice rule, whereas *full* implementation would require all level-k solutions to be consistent. The *weak* aspect refers to the fact we do not require every selection of the social choice rule to be obtained as an outcome of some level-k strategy profile. More specifically, let $\langle M, g\rangle$ be a mechanism and $s$ be a level-k solution, then weak implementation requires $g(s(\theta, \hat{k})) \subseteq F(\theta)$ for all $(\theta, \hat{k}) \in \times \left\{\Theta_i \times \{1, \ldots, \bar{k}\}\right\}_{i \in I}$. One could also require a second type of full implementation where $g(s(\theta, \hat{k})) = F(\theta)$ . Any reference to full implementation in this paper is a reference only to the first type of full implementation and not the second.

[12] Condition (iii) holding with strict inequalities is not a necessary condition for full level-k implementation. Consider the case where $f^i = \bar{f} = f$ for all $i \in I$ and suppose condition (ii) holds with strict inequalities. In this case condition (iii) does not hold with strict

the relationship between partial and full level-k implementation is straightforward. A similar relationship may hold in general environments but this has not been shown here.

de Clippel et al. (2019) study full level-k implementation under the restriction to single-valued choice rules. They find that Bayesian incentive constraints are necessary conditions for full level-k implementation. In the next section, we show that if we restrict attention to single-valued choice rules, we replicate de Clippel et al.'s finding for weak level-k implementation. But this result does not hold necessarily for multi-valued choice rules under either partial or full level-k implementation. In section 6 we give an example that shows that ex post efficient trade is fully level-k implementable while it is not Bayesian implementable.[13]

*Remark* 4. *Level* 0 *anchor.* First off, notice that the necessary conditions specified in Proposition 1 do not place any assumptions on the anchor. Thus, the necessary conditions are independent of any level 0 behavior assumptions. This arises because the necessary conditions are generated only from the incentive requirements of a level 2 (or higher) agent which are not directly affected by the anchor specification.

Second, notice that the sufficient conditions depend on the anchor. Proposition 2 holds only for the case of atomless anchors. However, we can also consider alternative anchor assumptions such as truthful reporting. The idea that agents truthfully report their payoff types is a common consequence in the literature that follows from the revelation principle. It is also an assumption that has been applied in behavioral mechanism design (for examples see Crawford & Iriberri (2007b) and Saran (2011a)). The language of agents truthfully reporting their payoff types and levels is regularly used throughout this paper. But, that need not be interpreted literally as an agent having an understanding of their payoff type and level, rather it was simply a label put on a message

inequalities, however, one could easily adapt the mechanism in Proposition 2 to acheive full level-k implementation. In particular, define $g = f$ regardless of the (modified) level reports.

[13]de Clippel et al. (2019) also have an example, Example 2, where a multi-valued social choice rule is fully level-k implementable. But, in their example, there is a selection of the social choice rule that is Bayesian incentive compatible. This is in contrast to the bilateral trade example, where ex-post efficient trade is not Bayesian incentive compatible and yet it is level-k implementable in the weak-full sense.

that the agent had an incentive to send (i.e. truth telling is a consequence of the mechanism, but not an assumption). However, when we discuss whether agents truthfully report their types, when they are not incentivized to do so, as is the case for level 0 agents, we must take the idea of what it means to truthfully report payoff types and levels much more seriously. It might be unreasonable for a planner to think agents will truthfully report their levels as it seems likely that agents will have no real conceptualization of what their levels of reasoning are. This is in contrast to the idea of a payoff type, where an agent is likely able to conceptualize how she much she values a good in an auction, for example, and may feel inclined to truthfully report because she prefers not to lie.

Given this, we may want to consider the possibility that level 0 agents truthfully report their payoff type, but may not truthfully report with respect to other dimensions of the message space. For example, one can interpret the mechanism underlying the proof of Proposition 2 as asking subjects to report their payoff type, an integer between 0 and $\bar{k}$, and a real number in $[-1, 1]$. It may be valuable to think about level 0 subjects here truthfully reporting their payoff type, but then randomly choosing integers and real numbers. Notice the mechanism in the above proof would easily extend to capture this notion of truthful reporting as long as the anchor is atomless on $[-1, 1]$.

*Remark* 5. *Cognitive hierarchy.* We can extend the results beyond the simple level- k model to more general limited depth of reasoning models like cognitive hierarchy.[14] Specifically, under the level-k model, if an agent is level $k$, she believes that others have levels exactly equal to $(k-1)$. In general, we might allow an agent with level $k$ to hold beliefs over all lower levels. As long as a level $k$ type only puts weight on lower levels, the spirit of limited depth of reasoning is maintained with each type being able to calculate her optimal action recursively, in a finite number of steps.

Appendix C addresses the question of whether we can design a mechanism that is robust to relaxing the level-k belief assumption. In this appendix we generalize our type space and solution concept to the limited depth of reasoning

[14]In the cognitive hierarchy model, a level $k$ type has beliefs over all lower levels determined by a conditional Poisson distribution. See Camerer et al. (2004) for specifics.

(LDoR) concept to relax beliefs about the depths of reasoning of others and consider a form of robust implementation where we ask whether there exists a mechanism that will implement a social choice rule for any LDoR type space. We show that if you strengthen the sufficient conditions as in the $n = 2$ case (i.e. truthfully reporting payoff type under $f^i$ is preferable to reporting any other payoff type under $f^i$, $\bar{f}$ or $f^j$ for all $i, j \neq i \in I$), then these conditions are sufficient to implement a social choice rule for any LDoR type space.

# 4 Special environments

In this section we look at two restricted environments. In the first, we restrict attention to single-valued social choice rules and in the second we restrict attention to mechanisms where the message set is equal to the set of payoff types. These special environments are the environments studied in de Clippel et al. (2019) and Crawford (2021) respectively. In both of these cases, we show that Bayesian incentive constraints are necessary conditions for level-k implementation. This establishes parallel results to those found by de Clippel et al. (Proposition 1) and Crawford (Lemma 1).

Corollary 1 formalizes this result for the restriction to social choice functions.

**Corollary 1.** *Let $F$ be a single-valued social choice rule. Let $\mathcal{B}$ be a Bayesian type space and let $\mathcal{L} = \langle \mathcal{B}; (T_i, \mu_i)_{i=1,\ldots,n}; \bar{k}\rangle$ be a ($\mathcal{B}$-based) level-k type space with $\bar{k} \geq 2$. Then $F$ is level-k implementable only if it is Bayesian implementable*

PROOF:

> From Proposition 1, there exists functions $\bar{f} : \Theta \to \triangle(Y)$ and $f^i : \Theta \to \triangle(Y)$, for all $i \in I$ such that conditions (i)-(iii) hold. Since $F$ is a single-valued social choice rule, it must be that $\bar{f} = f^1(\theta) = \cdots = f^n(\theta) = F(\theta)$ for all $\theta \in \Theta$. Therefore, it follows that $F$ is Bayesian implementable using the mechanism $\langle \Theta, F \rangle$.

□

The following proposition demonstrates that restricting the message space to the set of payoff types has the same effect as restricting our social choice rule to a social choice function - Bayesian incentive constraints are necessary conditions for level-k implementation,.

In order to show this we need to assume a richness condition on the environment, Assumption A*:

**A*:** For every $i \in I$ and for any two payoff types $\theta_i \neq \theta_i' \in \Theta_i$ there exists a $\theta_{-i} \in \Theta_{-i}$ such that $F(\theta_i, \theta_{-i}) \cap F(\theta_i', \theta_{-i}) = \emptyset$.

Assumption A* is likely to hold in many environments. For example, as we'll see in Section 6, it holds in the bilateral trade environment with the ex post efficient social choice rule as long as for every two values of the buyer $v < v' \in V$ there exists a cost for the seller, $c \in C$, that falls between, $v \leq c \leq v'$. And, similarly for the seller. Proposition 3 establishes the result.

**Proposition 3.** *Let $F$ be a social choice rule. Let $A^*$ hold. Let $\mathcal{B}$ be a Bayesian type space and let $\mathcal{L} = \langle \mathcal{B}; (T_i, \mu_i)_{i=1,\ldots,n}; \bar{k} \rangle$ be a ($\mathcal{B}$-based) level-k type space with $\bar{k} \geq 2$. If we restrict mechanisms to only allow messages about payoff types, $M_i = \Theta_i$ for all $i \in I$, then $F$ is level-k implementable only if it is Bayesian implementable.*

PROOF:

Suppose $F$ is level-k implementable. Then there exists some mechanism $\langle \Theta, g \rangle$, anchor $\alpha$, and function $m_i : \Theta_i \times \{0, \ldots, \bar{k}\} \to \triangle(\Theta_i)$ for each $i \in I$ such that $m = m_1 \times \cdots \times m_n$ is a level-k solution under $\alpha$ and achieves $F$.

Suppose that there exists some agent $i \in I$, and two types for agent $i$ with $\theta_i \neq \theta_i' \in \Theta_i$ such that $\text{supp}(m_i(\theta_i, k)) \cap \text{supp}(m_i(\theta_i', j)) \neq \emptyset$ for some $j, k \in \{1, \ldots, \bar{k}\}$. Since these two types send the same message as part of a level-k solution that achieves $F$, the social planner must be satisfied with them receiving the same outcome under every payoff profile i.e. it must follow that for $m_i \in \text{supp}(m_i(\theta_i, k)) \cap \text{supp}(m_i(\theta_i', j))$, $g(m_i, m_{-i}(\theta_{-i}, 1)) \in F(\theta_i, \theta_{-i}) \cap F(\theta_i', \theta_{-i})$ for every $\theta_{-i} \in \Theta_{-i}$. This contradicts assumption $A^*$. Therefore, it must be true that for any two types for agent $i$ with $\theta_i \neq \theta_i' \in \Theta_i$ that $\text{supp}(m_i(\theta_i, k)) \cap \text{supp}(m_i(\theta_i', j)) = \emptyset$ for any $j, k \in$

$\{1,\ldots,\bar{k}\}$. Thus, it follows that we can define the function $\psi_k^i : \Theta_i \to \Theta_i$ by $\psi_k^i(\theta_i) = m_i(\theta_i, k)$ and that it is both 1-1 and onto.

**Claim**: $m_i(\theta_i, k) = m_i(\theta_i, j)$ for all $j, k \in \{1,\ldots,\bar{k}\}$, $\theta_i \in \Theta_i$.

To see this, suppose not. Then there exists a $\theta_i \in \Theta_i$ and some $j, k \in \{1,\ldots,\bar{k}\}$ with $j \neq k$ such that $m_i(\theta_i, k) \neq m_i(\theta_i, j)$.

Because, $\psi_j^i$ is 1-1 and onto there must exist some $\theta_i' \neq \theta_i \in \Theta_i$ such that $\psi_j^i(\theta_i') = m_i(\theta_i, k)$. But, then $m_i(\theta_i', j) = m_i(\theta_i, k)$ which is a contradiction.

Therefore, it follows that $m_i(\theta_i, k) = m_i(\theta_i, j)$ for all $j, k \in \{1,\ldots,\bar{k}\}$.

Define $\tilde{g} : \Theta \to \triangle(Y)$ by $\tilde{g}(\theta) = g(m(\theta, 2))$ for all $\theta \in \Theta$.

Then, for any $i \in I$, $\theta \in \Theta$, and $\theta' \in \Theta_i$

$$\begin{aligned}
\sum_{\theta_{-i}\in\Theta_{-i}} \rho(\theta_{-i}|\theta_i)\cdot u_i(\tilde{g}(\theta),\theta) - \sum_{\theta_{-i}\in\Theta_{-i}} \rho(\theta_{-i}|\theta_i)\cdot u_i(\tilde{g}(\theta',\theta_{-i}),\theta) \\
&= \sum_{\theta_{-i}\in\Theta_{-i}} \rho(\theta_{-i}|\theta_i)\cdot u_i(g(m_i(\theta_i,2), m_{-i}(\theta_{-i},2)),\theta) \\
&\qquad - \sum_{\theta_{-i}\in\Theta_{-i}} \rho(\theta_{-i}|\theta_i)\cdot u_i(m_i(\theta',2), m_{-i}(\theta_{-i},2)),\theta) \\
&= \sum_{\theta_{-i}\in\Theta_{-i}} \rho(\theta_{-i}|\theta_i)\cdot u_i(g(m_i(\theta_i,2), m_{-i}(\theta_{-i},1)),\theta) \\
&\qquad - \sum_{\theta_{-i}\in\Theta_{-i}} \rho(\theta_{-i}|\theta_i)\cdot u_i(m_i(\theta',2), m_{-i}(\theta_{-i},1)),\theta) \\
&\geq 0
\end{aligned}$$

The inequality follows from the fact $m$ is a level-k solution. Thus, the Bayesian incentive constraints hold for $\tilde{g}$. Further, $\langle\Theta, \tilde{g}\rangle$ achieves $F$ because $\tilde{g}(\theta) = g(m((\theta, 2))) \in F(\theta)$. Therefore, it follows by definition that $F$ is Bayesian implementable using the mechanism$\langle\Theta, \tilde{g}\rangle$.

□

*Remark* 6. *Equivalence between level-k and Bayesian implementation.* Under the conditions of Corollary 1 we get an equivalence between Bayesian implementation and level-k implementation in independent private value environments. This is because we can show that if a single-valued social choice rule $F$ is Bayesian implementable then it is level-k implementable by setting $\bar{f} = F$

and $f^i = F$ for all $i \in I$ and applying Proposition 2. However, we cannot guarantee an equivalence between Bayesian and level-k implementation under the conditions of Proposition 3 because Proposition 2 will not apply if we place restrictions on the message space. Therefore, it may be the case that a social choice rule is Bayesian implementable while it is not level-k implementable under mechanisms where the message space is restricted to the set of payoff types.

# 5 Ex post level-k implementation

This section addresses the question of whether we can design a mechanism that is robust to relaxing the common prior assumption that is present in the level-k type space - the assumption that beliefs about payoffs are determined by a specific common prior. To address this, we generalize our solution concept to that of ex post level-k implementation. This effectively relaxes the common prior assumption that has been maintained so far. This section shows that the relationship between ex post level-k and ex post implementation is analogous to the relationship between level-k and Bayesian implementation. The ex post level-k solution concept is defined below.

**Definition 8.** For a given game defined by a mechanism $\langle M, f \rangle$ and a $\bar{k} \in \mathbb{N}_+$, a strategy profile $s = s_1 \times \cdots \times s_n$, with $s_i : \Theta_i \times \{0, \ldots, \bar{k}\} \to \triangle(M_i)$ for all $i \in I$, is the **ex post level-k solution under anchor $\boldsymbol{\alpha}$** if and only if:

**(i)** $s_i(\theta_i, 0) = \alpha_i(\theta_i, 0)$ for all $\theta_i \in \Theta_i$

**(ii)** $u_i(f(s_i(\theta_i, k), s_{-i}(\theta_{-i}, k-1)), \theta) \geq u_i(f(m_i', s_{-i}(\theta_{-i}, k-1)), \theta)\ \forall\ m_i' \in M_i$, $\theta \in \Theta$, and $k \geq 1$, $i \in I$.

This solution concept allows us to define an ex post level-k implementation concept that is an analogue to ex post implementation: there must exist a mechanism, an anchor, and an ex post level-k solution which is consistent with the social choice rule for any realization of payoff types, $\theta \in \Theta$.

**Definition 9.** Fix a $\bar{k} \in \mathbb{N}_+$. A social choice rule is **ex post level-k implementable** if there exists a mechanism $\langle M, f \rangle$, an anchor $\alpha$, and a strategy profile $s_i : \Theta_i \times \{0, \ldots, \bar{k}\} \to \triangle(M_i)$ for all $i \in I$, such that $s = s_1 \times \cdots \times s_n$ is an ex post level-k solution under $\alpha$ and $f(s(\theta, \hat{k})) \in F(\theta)$ for all $(\theta, \hat{k}) \in \times \left\{\Theta_i \times \{1, \ldots, \bar{k}\}\right\}_{i \in I}$.

We will be interested in comparing ex post level-k implementation with that of ex post implementation. We define ex post implementation for completeness. Condition (ii) below states the standard ex post incentive constraints which will be contrasted with the ex post level-k incentive constraints developed in the next subsection.

**Definition 10.** A social choice rule $F$ is **ex post implementable** if there exists a mechanism $\langle \Theta, f \rangle$ such that

**(i)** $f(\theta) \in F(\theta)\ \forall \theta \in \Theta$

**(ii)** $u_i(f(\theta), \theta) \geq u_i(f(\theta', \theta_{-i}), \theta)\ \forall \theta' \in \Theta,\ \theta \in \Theta,\ i \in I$

## 5.1 Necessary and sufficient conditions for ex post level-k implementation

This section gives the necessary and sufficient conditions for ex post level-k implementation. The conditions are stated separately as the necessary conditions hold in general environments and the sufficient conditions hold in private value environments under atomless anchors.

The necessary conditions for ex post level-k implementation are given in Proposition 4 and are analogous to those in Proposition 3. The proof follows analogously to the proof of Proposition 3 and can be found in Appendix A.

**Proposition 4.** *(Ex post Necessary Conditions) Let $F$ be a social choice rule. Let $\bar{k} \geq 2$. If $F$ is ex post level-k implementable, then there exists a function $f^i : \Theta \to \triangle(Y)$ for each $i \in I$ and a function $\bar{f} : \Theta \to \triangle(Y)$, such that the following conditions hold:*

**(i)** $f^i(\theta), \bar{f}(\theta) \in F(\theta)\ \forall \theta \in \Theta, \forall i \in I$

**(ii)** $u_i(f^i(\theta), \theta) \geq u_i(f^i(\theta'_i, \theta_{-i}), \theta)\ \forall \theta'_i \in \Theta_i \setminus \theta'_i,\ \theta \in \Theta,\ i \in I$

**(iii)** $u_i(f^i(\theta), \theta) \geq u_i(\bar{f}(\theta'_i, \theta_{-i}), \theta)\ \forall \theta'_i \in \Theta_i,\ \theta \in \Theta,\ i \in I$

Similar to the case for level-k implementation, the necessary conditions are sufficient for ex post level-k implementation in private value environments when $n \geq 3$ (the case when $n = 2$ can be found in Appendix B). The proof follows analogously to that of Proposition 2 and can be found in Appendix A.

**Proposition 5.** *(Ex post Sufficient Conditions) Let $F$ be a social choice rule. Let the environment be one of private values and let $n \geq 3$. If there exists a function $f^i : \Theta \to \triangle(Y)$ for each $i \in I$ and a function $\bar{f} : \Theta \to \triangle(Y)$ such that conditions (i)-(iii) hold in Proposition 4 then $F$ is ex post level-k implementable. Further, there exists a mechanism $\langle M, f \rangle$ such that for any atomless anchor $\alpha$ there is an ex post level-k solution under $\alpha$ that achieves $F$.*

*Remark* 7. *Relationship to ex post implementation.* The conditions in Proposition 4 generate a set of ex post level-k incentive constraints that generalize the standard ex post incentive constraints. If the social choice rule is ex post implementable then it is ex post level-k implementable in private value environments.

*Remark* 8. *Special Environments.* In Section 4 we considered two special environments which lead to the cross-player restrictions in the level-k incentive constraints being automatically imposed. Simple arguments extend these results to the ex post environment as well. In these two special environments ex post incentive constraints are necessary conditions for ex post level-k implementation. Again, this need not hold in general. In Section 6 we show that ex post efficient bilateral trade is ex post level-k implementable while it is not ex post implementable.

*Remark* 9. *Ex post level-k implementability implies level-k implementability.* Just as ex post implementation implies Bayesian implementation, it is also true that ex post level-k implementation implies level-k implementation. This is easy to see because the ex post level-k sufficient conditions imply that the level-k sufficient conditions will hold for any common prior.

# 6 Bilateral trade

## 6.1 Bilateral trade environment

The remainder of this paper focuses on the bilateral trade environment. Buyer's values are given by a finite set $V$. Seller's costs are given by a finite set $C$. The set of outcomes is given by $Y = \mathbb{R} \cup \{\emptyset\}$, where outcome $\emptyset$ indicates the good is not traded and outcome $p \in \mathbb{R}$ indicates the good is traded at price $p$. Agents have quasi-linear utility functions, $u_b : Y \times V \to \mathbb{R}$ and $u_s : Y \times C \to \mathbb{R}$. For any outcome $p$, the utility of a buyer with a valuation $v$ is

$$u_b(p, v) \quad = \quad \begin{cases} v - p & \text{if } p \in \mathbb{R} \\ 0 & \text{otherwise} \end{cases}$$

and the utility of a seller with a cost $c$ is

$$u_s(p, c) \quad = \quad \begin{cases} p - c & \text{if } p \in \mathbb{R} \\ 0 & \text{otherwise} \end{cases}.$$

We are interested in mechanisms that satisfy the ex post efficient social choice rule $F^*(v, c) = \{y | y \in \mathbb{R} \text{ if } v \geq c \text{ and } y = \emptyset \text{ otherwise}\}$. The ex post efficient choice rule requires trade whenever the buyer's value is above the seller's cost. We are also interested in the mechanism satisfying two additional properties: budget balance (the price paid by the buyer equals the price received by the seller - this is already imposed by the description of the environment) and ex post individual rationality (both the buyer and seller prefer to participate in the trading institution than receive the utility of 0).[15]

## 6.2 Ex post efficient trade - general environment

This section contains the main result: ex post efficient trade is ex post level-k implementable. As a result, ex post efficient trade is also level-k implementable.

[15] Formally, given a level-k type space $\mathcal{L} = \langle \mathcal{B}; (T_i, \mu_i)_{i=1,\ldots,n}; \bar{k} \rangle$, a $\langle M, f \rangle$, and a strategy profile $s$, ex post individual rationality requires that $u_i(f(s(\theta, \hat{k})), \theta_i) \geq 0$ for all $(\theta, \hat{k}) \in \times \left\{ \Theta_i \times \{1, \ldots, \bar{k}\} \right\}_{i \in I}$ and $i \in I$.

Notice that Corollary 1 does not apply in this environment. This is because the ex post efficient choice rule, $F^*$, is a multi-valued choice rule. If the buyer's value is above the seller's cost, ex post efficiency requires trade, but the planner does not care at what price the good is traded. This means it may be possible to level-k implement ex post efficient trade even if it is not Bayesian implementable.

Also, notice that if we impose an additional assumption, A1, on the decision environment then the conditions for Proposition 3 are satisfied. This means that in order to implement ex post efficient trade under level-k implementation, we will need to use mechanisms with messages spaces larger than the set of payoff types.

**A1:** For any $v, v' \in V$ with $v' < v$, there exists a $c \in C$ such that $v' \leq c \leq v$. And, for any $c, c' \in C$ with $c < c'$, there exists a $v \in V$ such that $c \leq v \leq c'$.

Proposition 6 establishes the result.

**Proposition 6.** *The ex post efficient social choice rule, $F^*$, is ex post level-k implementable under a mechanism that satisfies budget balance and ex post individual rationality.*

The proof of Proposition 6 follows by showing that there exists functions $f^b : V \times C \to \triangle(Y)$, $f^s : V \times C \to \triangle(Y)$ and $\bar{f} : V \times C \to \triangle(Y)$ that satisfy sufficient conditions for level-k implementation when $n = 2$ (Proposition 8 in Appendix B). The formal proof can be found in Appendix A, but, the basic intuition works as follows. Consider three ways to determine the price if there is trade: (1) give the buyer all the surplus (i.e. set the price to seller's reported cost); (2) give the seller all the surplus (i.e. set the price to buyer's reported value); (3) have the buyer and seller split the surplus (i.e. set the price equal to the average of reported cost and value). Use these to define the functions $f^b$, $f^s$, and $\bar{f}$ respective. Formally, define

$$f^b(v,c) = \begin{cases} c & \text{if } c \leq v \\ \emptyset & \text{otherwise} \end{cases},$$

$$f^s(v,c) = \begin{cases} v & \text{if } c \leq v \\ \emptyset & \text{otherwise} \end{cases},$$

and

$$\bar{f}(v,c) = \begin{cases} \frac{v+c}{2} & \text{if } c \leq v \\ \emptyset & \text{otherwise} \end{cases}.$$

Now, let every agent report their cost or value and their level. If a buyer reports a level exactly one higher than the seller than let trade and prices be determined according to $f^b$. If a seller reports a level exactly one higher than the buyer, then let trade and prices be determined according to $f^s$. Otherwise, let trade and prices be determined according to $\bar{f}$. In other words, trade occurs if and only if the reported value is above the reported cost and at the price that is most favorable to the agent that reports exactly one level higher than the other agent; otherwise they split the surplus. Suppose now that all agents truthfully report their payoff type and level. Given that, both buyers and sellers have an incentive to truthfully report their levels as they will then receive the most favorable prices for themselves. Further, if an agent reports their level truthfully, then they also have an incentive to truthfully report their payoff type, since messages sent will only affect the likelihood of trade but not the price. This ensures that truthfully reporting one's own payoff type and level is a best response given that everyone else is truthfully reporting their level, regardless what the payoff type realization is. The only thing left is for the planner to incentivize level 1 agents; this can be done as in the proof of Proposition 2, by having the planner screen for level 0 types by using the reported real number from $[-1, 1]$.

## 6.3 Ex post efficient trade - a 2 type example

In this subsection, we go through a simple 2-type example to give a concrete illustration of a mechanism that is level-k implementable. We will also use this example to demonstrate that it is possible to design a mechanism that fully level-k implements ex post efficient trade, i.e. under the mechanism, any level-k solution will be consistent with ex post efficient trade.

In this example, the seller has two possible costs: $C = \{2, 6\}$, and the buyer has two possible values: $V = \{3, 7\}$. Types are drawn from a uniform common prior, $\rho$ (i.e. $\rho(v,c) = \frac{1}{4}$ for all $(v,c) \in V \times C$).

*Claim* 1. Ex post efficient trade is not Bayesian implementable.

This was shown by Matsuo (1989) who gives sufficient conditions for ex post efficiency in the two type bilateral trade environment. To see the intuition, recall that the revelation principle ensures that we need only consider mechanisms where all agents truthfully report their type. Further, the low valued buyer and the high valued seller should receive zero utility in equilibrium. Thus any candidate mechanism must take the form of the one in Figure 1 for some $p \in \mathbb{R}$. This mechanism should be understood in the following way: the buyer chooses the row message, the seller chooses the column message, and the corresponding element in the table is the outcome that occurs. For example, if the buyer sends message $m_7$ and the seller sends message $m_6$ then there is trade at a price of 6. Alternatively, if the buyer sends message $m_3$ and the seller sends message $m_6$ then there is no trade.

| | | Seller: $m_2$ | Seller: $m_6$ |
|---|---|---|---|
| Buyer | $m_7$ | $p$ | 6 |
| | $m_3$ | 3 | $\emptyset$ |

Figure 1: Structure of a Bayesian mechanism

A high valued buyer believes the low and high cost seller types are equally likely (comes from the uniform prior assumption) and thus believes actions $m_2$ and $m_6$ are equally likely. Thus, for a high valued buyer to truthfully report her payoff type the trading price must be less than or equal to 4, i.e. $p \leq 4$. Similarly, for a low cost seller to truthfully report her payoff type the trading price must be greater than or equal to 5, i.e. $p \geq 5$. These two conditions are incompatible. There is no mechanism that will implement the ex post efficient choice rule under Bayesian implementation.

*Claim* 2. Ex post efficient trade is ex post level-k implementable

Suppose, for the sake of this example, that there are only level 0, level 1 and level 2 types in the population and consider the uniformly random anchor.

The mechanism in Figure 2 ex post level-k implements the ex post efficient choice rule. To see this notice that level 0 agents (regardless of their payoff type) are assumed to play each action with equal probability. A level 1 type

| | | Seller | | | |
|---|---|---|---|---|---|
| | | $m_0$ | $m_{(2,1)}$ | $m_{(2,2)}$ | $m_{(6,-)}$ |
| | $m_0$ | 4.5 | 7 | 3 | $\emptyset$ |
| Buyer | $m_{(7,1)}$ | 2 | 4.5 | 6 | 6 |
| | $m_{(7,2)}$ | 6 | 3 | 4.5 | 6 |
| | $m_{(3,-)}$ | $\emptyset$ | 3 | 3 | $\emptyset$ |

Figure 2: A level-k mechanism

then believes that her opponent is playing each action with equal probability. Therefore, playing $m_{(3,-)}$ is a best response for the low valued level 1 buyer and playing $m_{(7,1)}$ is a best response for the high valued level 1 buyer. Likewise, playing $m_{(6,-)}$ is a best response for the high cost level 1 seller and playing $m_{(2,1)}$ is a best response for the low cost level 1 seller.

Effectively, a level 2 buyer may hold any beliefs over payoff types but believes her opponent is a level 1 type. For any beliefs about the payoff types of the seller, playing $m_{(3,-)}$ is a best response for a low valued level 2 buyer and playing $m_{(7,2)}$ is a best response for a high valued level 2 buyer. Similarly, for any beliefs over payoff types of the buyer, playing $m_{(6,-)}$ is a best response for a high cost level 2 seller and playing $m_{(2,2)}$ is a best response for a low cost level 2 seller.

Given the strategies defined by the ex post level-k solution, for any pair of level 1 or level 2 types, the outcome will be consistent with the ex post efficient social choice rule. In other words, if the buyer is the low valued type and the seller is the high cost type, then regardless of whether the buyer and sellers are level 1 or level 2 types, there will not be trade. For any other payoff type profile $(v, c) \neq (3, 6)$, regardless of whether the buyer and sellers are level 1 or level 2 types, there will be trade. Ex post individual rationality is also satisfied.

*Remark* 10. *Full level-k implementation.* Notice that there is a unique level-k solution under a uniform random anchor in the above mechanism in the level-k type space with the uniform common prior, i.e. all types with levels at least 1 are playing strict best responses. Thus, this mechanism level-k implements ex post efficient trade in both the partial and full sense.

# 7 Conclusion

This paper explores the theoretical implications of level-k implementation by defining the boundaries of what is level-k implementable. It gives necessary and sufficient conditions for level-k implementation and establishes the relationship to Bayesian implementation. It relaxes the common prior assumption that underlies level-k and Bayesian implementation by defining the concept of ex post level-k implementation. It gives necessary and sufficient conditions for ex post level-k implementation and shows that the relationship between ex post level-k and ex post implementation mirrors that between level-k and Bayesian implementation.

This paper is not, however, a practical guide to level-k implementation. And, there are many questions that need to be answered to address the practical relevance of level-k mechanisms. The first question being, do level-k mechanisms work? This is an empirical question and one that can potentially be answered with experiments. Will mechanisms of the type underlying the sufficiency proofs in this paper achieve the desired social choice outcomes in the lab? Is the assumption about atomless anchors the right assumption? Level-k models are motivated by the explanation of behavior in novel situations. But, what is a novel situation? Is it a situation where an agent truly has no experience? Or, will level-k mechanisms work even when agents have (limited) experience? Then there is the question of what is the real world analogue of a level-k mechanism? This paper shows that common mechanisms like auctions may not be able to achieve all level-k implementable outcomes as the message space is restricted to the set of payoff types.

While this paper does not provide answers to these questions, it does provide a solid framework to start investigating these questions by supplying tight necessary and sufficient conditions for level-k implementation and providing some insight into the types of mechanisms that might be necessary to achieve level-k implementable outcomes.

# References


Arad, A. & Rubinstein, A. (2012). The 11-20 Money Request Game: Evaluating the Upper Bound of Level-k Reasoning. *American Economic Review*, 102(7), 3561–73.

Bergemann, D. & Morris, S. (2005). Robust Mechanism Design. *Econometrica*, 73(6), 1771–1813.

Bergemann, D. & Morris, S. (2009). Robust Implmentationin Direct Mechanisms. *Review of Economic Studies*, 76, 1175–1204.

Bergemann, D., Morris, S., & Tercieux, O. (2011). Rationalizable Implementation. *Journal of Economic Theory*, 146.

Börgers, T. & Li, J. (2019). Strategically Simple Mechanisms. *Econometrica*, 87, 2003–2035.

Brocas, I., Carrillo, J. D., Wang, S. W., & Camerer, C. F. (2014). Imperfect Choice or Imperfect Attention? Understanding Strategic Thinking in Private Information Games. *Review of Economic Studies*, 81(3), 944–970.

Bulow, J. & Roberts, J. (1989). The Simple Economics of Optimal Auctions. *Journal of Political Economy*, 97, 1060–1090.

Camerer, C. F., Ho, T.-H., & Chong, J.-K. (2004). A Cognitive Hierarchy Model of Games. *Quarterly Journal of Economics*, 119(3), 861–898.

Copic, J. & Ponsati, C. (2008). Robust Bilateral Trade and Mediated Bargaining. *Journal of the European Economic Association*, 6, 570–589.

Čopič, J. & Ponsati, C. (2016). Optimal Robust Bilateral Trade: Risk Neutrality. *Journal of Economic Theory*, 163, 276–287.

Costa-Gomes, M. & Crawford, V. P. (2006). Cognition and Behavior in Two-Person Guessing Games: An Experimental Study. *American Economic Review*, 96(5), 1737–1768.

Costa-Gomes, M., Crawford, V. P., & Broseta, B. (2001). Cognition and Behavior in Normal-Form Games: An Experimental Study. *Econometrica*, 69(5), 1193–1235.

Costa-Gomes, M. A., Crawford, V. P., & Iriberri, N. (2013). Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications. *Journal of Economic Literature*, 51, 5–62.

Crawford, V., Kugler, Neeman, & Pauzner (2009). Behaviorally optimal auction design: examples and observations. *Journal of the European Economic Association*, 7, 377–387.

Crawford, V. P. (2021). Efficient mechanisms for level-k bilateral trading. *Games and Economic Behavior*, 127(3), 80–101.

Crawford, V. P. & Iriberri, N. (2007a). Fatal Attraction: Salience, Naivete, and Sophistication in Experimental Hide-and-Seek Games. *American Economic Review*, 97(5), 1731–1750.

Crawford, V. P. & Iriberri, N. (2007b). Level-k Auctions: Can a Non-Equilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions. *Econometrica*, 75(6), 1721–1770.

de Clippel, G. (2014). Behavioral implementation. *American Economic Review*, 104(10), 2975–3002.

de Clippel, G., Saran, R., & Serrano, R. (2015). Mechanism Design with Bounded Depth of Reasoning and Small Modeling Mistakes. *working paper*.

de Clippel, G., Saran, R., & Serrano, R. (2019). Level-k mechanism design. *Review of Economic Studies*, 86(3), 1207–1227.

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

Eliaz, K. & Spiegler, R. (2006). Contracting with diversely naive agents. *Review of Economic Studies*, 73(3), 689–714.

Eliaz, K. & Spiegler, R. (2007). A mechanism design approach to speculative trade. *Econometrica*, 75(3), 875–884.

Eliaz, K. & Spiegler, R. (2008). Optimal speculative trade among large traders. *Review of Economic Design*, 12(1), 45–74.

Glazer, J. & Rubinstein, A. (1998). Motives and implementation: on the design of mechanisms to elicit opinions. *Journal of Economic Theory*, 79, 157–173.

Glazer, J. & Rubinstein, A. (2012). A model of persuasion with a boundedly rational agent. *Journal of Political Economy*, 120(6), 1057–1082.

Gorelkina, O. (2015). The expected externality mechanism in a level-k environment. *International Journal of Game Theory*, 47, 103–131.

Hagerty, K. M. & Rogerson, W. P. (1987). Robust Trading Mechanisms. *Journal of Economic Theory*, 42, 94–107.

Healy, P. J. (2006). Learning Dynamics for Mechanism Design: An Experimental Comparison of Public Goods Mechanisms. *Journal of Economic Theory*, 129, 114–149.

Matsuo, T. (1989). On incentive compatible, individually rational, and ex post efficient mechanisms for bilateral trading. *Journal of Economic Theory*, 48, 189–194.

Matsushima, H. (2007). Mechanism Design with Side Payments: Individual Rationality and Iterative Dominance. *Journal of Economic Theory*, 133, 1–30.

Matsushima, H. (2008). Detail-free Mechanism Design in Twice Iterative Dominance: Large Economies. *Journal of Economic Theory*, 141, 134–151.

Mookherjee, D. & Reichelstein, S. (1992). Dominant Stragey Implementation of Bayesian Incentive Compatible Allocation Rules. *Journal of Economic Theory*, 56, 378–399.

Myerson, R. B. & Satterthwaite, M. A. (1983). Efficient mechanisms for bilateral trading. *Journal of Economic Theory*, 29, 265–281.

Nagel, R. (1995). Unraveling in Guessing Games: An Experimental Study. *American Economic Review*, 85(5), 1313–1326.

Ollar, M. & Penta, A. (2017). Full Implementation and Belief Restrictions. *American Economic Review*, 107, 2243–2277.

Saran, R. (2011a). Bilateral trading with naive traders. *Games and Economic Behavior*, 72, 544–557.

Saran, R. (2011b). Menu-dependent Preferences and Revelation Principle. *Journal of Economic Theory*, 146, 1712–1720.

Saran, R. (2016). Bounded Depths of Rationality and Implementation with Complete Information. *Journal of Economic Theory*, 165, 517–564.

Severinov, S. & Deneckere, R. (2006). Screening when some agents are non-strategic: does a monopoly need to exclude? *RAND Journal of Economics*, 37, 816–840.

Stahl, D. O. & Wilson, P. W. (1994). Experimental Evidence on Player's Models of Other Players. *Journal of Economic Behavior and Organization*, 25(3), 309–327.

Stahl, D. O. & Wilson, P. W. (1995). On Player's Models of Other Players: Theory and Experimental Evidence. *Games and Economic Behavior*, 10(1), 218–254.

Strzalecki, T. (2014). Depth of Reasoning and Higher Order Beliefs. *Journal of Economic Behavior and Organization*, 108, 108–122.

Wolitzky, A. (2016). Mechanism design with maxmin agents: Theory and an application to bilateral trade. *Theoretical Economics*, 11(3), 971–1004.

# Appendix A Omitted proofs

## Proof of Proposition 1:

Suppose that the social choice rule $F$ is level-k implementable. Then there exists some mechanism $\langle M, g\rangle$, an anchor $\alpha$, and a function $m_i : T_i \rightarrow$

$\triangle(M_i)$ for each $i \in I$ such that $m = m_1 \times \cdots \times m_N$ is a level-k solution under $\alpha$ and achieves $F$.

Consider the behavior of an agent $i$ with type $t_i = (\theta_i, 2)$. Then, it must be true that:

$$\begin{aligned}\sum_{\theta_{-i}\in\Theta_{-i}} &\rho(\theta_{-i}|\theta_i)\cdot u_i(g(m_i(\theta_i,2),m_{-i}(\theta_{-i},1)),\theta)\\ &\geq \sum_{\theta_{-i}\in\Theta_{-i}} \rho(\theta_{-i}|\theta_i)\cdot u_i(m_i(\theta_i',2),m_{-i}(\theta_{-i},1)),\theta) \quad \forall\theta_i'\in\Theta_i\end{aligned} \tag{1}$$

As well, it must be true that:

$$\begin{aligned}\sum_{\theta_{-i}\in\Theta_{-i}} &\rho(\theta_{-i}|\theta_i)\cdot u_i(g(m_i(\theta_i,2),m_{-i}(\theta_{-i},1)),\theta)\\ &\geq \sum_{\theta_{-i}\in\Theta_{-i}} \rho(\theta_{-i}|\theta_i)\cdot u_i(m_i(\theta_i',1),m_{-i}(\theta_{-i},1)),\theta) \quad \forall\theta_i'\in\Theta_i\end{aligned} \tag{2}$$

Define $f^i(\theta) = g(m_i(\theta_i, 2), m_{-i}(\theta_{-i}, 1))$ and $\bar{f}(\theta) = g(m_i(\theta_i, 1), m_{-i}(\theta_{-i}, 1))$ for all $\theta \in \Theta$ and for all $i \in I$.

Condition (ii) holds from (1). Condition (iii) holds from (2). Condition (i) holds by definition of $\langle M, g\rangle$ and $m$ level-k implementing $F$.

□

## Proof of Proposition 4:

Suppose that the social choice rule $F$ is ex post level-k implementable. Then there exists some mechanism $\langle M, g\rangle$, an anchor $\alpha$, and function $m_i : T_i \to \triangle(M_i)$ for each $i$ such that $m = m_1 \times \cdots \times m_N$ is an ex post level-k solution under $\alpha$ and achieves $F$.

Define $f^i(\theta) = g(m_i(\theta_i, 2), m_{-i}(\theta_{-i}, 1))$ for all $\theta \in \Theta$ and for all $i \in I$. And, define $\bar{f}(\theta) = g(m(\theta, 1))$ for all $\theta \in \Theta$ . By similar arguments made in the proof of Proposition 1 it can be shown that conditions (i)-(iii) hold.

□

## Proof of Proposition 5:

Fix $\bar{k} \in \mathbb{N}_+$.

Choose some $\bar{\theta}_i \in \Theta_i$ for each $i \in I$.

Consider the following mechanism where the message space for agent $i$ is equal to $M_i = \Theta_i \times \{0, 1, \ldots \bar{k}\} \times [-1, 1]$ and consists of a report of her payoff type $\theta_i \in \Theta_i$, level $k_i \in \{0, 1, \ldots, \bar{k}\}$, and a real number, $z_i \in [-1, 1]$. Let $\alpha$ be an atomless anchor.

Let the indicator function $I_i : [-1, 1]^n \to \{0, 1\}$, be defined as in the proof of Proposition 2.

Define $\tilde{\theta}_i : M \to \Theta_i$ and $\tilde{k}_i : M \to \{0, 1, \ldots, \bar{k}\}$ in the following way. For a given message profile $m = (\theta, k, z)$, if $I_i(z) = 1$ then $\tilde{\theta}_i(m) = \theta_i$ and $\tilde{k}_i(m) = k_i$; otherwise the planner sets $\tilde{\theta}_i(m) = \bar{\theta}_i$ and $\tilde{k}_i(m) = 0$. The planner then assigns outcomes based on the reports $(\tilde{\theta}, \hat{\tilde{k}})$ according to the function $\tilde{g} : \Theta \times \{0, \ldots, \bar{k}\}^n \to \triangle(Y)$ defined by

$$\tilde{g}(\theta, \hat{k}) \quad = \quad \begin{cases} f^i(\theta) & \text{if } \hat{k}_j = \hat{k}_i - 1 \text{ for all } j \neq i \in I \\ \bar{f}(\theta) & \text{otherwise} \end{cases} .$$

Analogous arguments to those in the proof of Proposition 2 will show that for agent $i$ with payoff type $\theta_i$ and level $k$, reporting $(\theta_i, k, 0)$ is a best response for any $\theta_{-i} \in \Theta_{-i}$.

Therefore, if we define $m_i(\theta_i, k_i) = (\theta_i, k_i, 0)$ for all $\theta_i \in \Theta_i$ with $k_i \in \{1, \ldots, \bar{k}\}$, then $m$ is an ex post level-k solution and $m$ achieves $F$ by condition $(i)$. Therefore, $F$ is ex post level-k implementable. Further, since $\alpha$ was an arbitrary atomless anchor, $m$ acheives $F$ for all atomless anchors.

$\square$

## Proof of Proposition 6

Define

$$f^b(v,c) = \begin{cases} c & \text{if } c \leq v \\ \emptyset & \text{otherwise} \end{cases}$$

and

$$f^s(v,c) = \begin{cases} v & \text{if } c \leq v \\ \emptyset & \text{otherwise} \end{cases}$$

and

$$\bar{f}(v,c) = \begin{cases} \frac{v+c}{2} & \text{if } c \leq v \\ \emptyset & \text{otherwise} \end{cases}.$$

First, it is easy to see that condition (i) in Proposition 8 holds for $f^b$, $f^s$, and $\bar{f}$ as they assign the outcome $\emptyset$ only when $v < c$.

Now consider the utility of the buyer with value $v$ when the seller reports cost $c$. Consider first the comparison of the outcomes $f^b(v,c)$ to outcomes $f^b(v',c)$ for some value $v' \in V$. There are two cases to consider:

(i)$v < c$: The utility of the buyer is 0 when reporting $v$ and reporting any other value $v'$ either has no effect (if $v' < c$) or achieves trade (if $v' \geq c$) with a utility of $v - c \leq 0$.

(ii)$c \leq v$: The utility of buyer is $v - c \geq 0$ when reporting $v$ and reporting any other value $v'$ either has no effect (if $v' > c$) or achieves outcome $\emptyset$ and utility 0.

Next, consider the comparison of the outcomes $f^b(v,c)$ to outcomes $f^s(v',c)$ for some value $v' \in V$. There are two cases to consider:

(i)$v < c$: The utility of the buyer under $f^b$ is 0 when reporting $v$ and reporting any other value $v'$ under $f^s$ either has no effect (if $v' < c$) or achieves trade (if $v' \geq c$) with a utility of $v - v' \leq 0$.

(ii)$c \leq v$: The utility of buyer is $v - c \geq 0$ under $f^b$ when reporting $v$ and reporting any other value $v'$ under $f^s$ either achieves outcome $\emptyset$ (if $v' < c$) and utility 0 or achieves trade (if $v' \geq c$) with a utility of $v - v' \leq v - c$.

Last, consider the comparison of the outcomes $f^b(v,c)$ to outcomes $\bar{f}(v',c)$ for some value $v' \in V$. There are two cases to consider:

(i)$v < c$: The utility of the buyer under $f^b$ is 0 when reporting $v$ and reporting any other value $v'$ under $\bar{f}$ either has no effect (if $v' < c$) or achieves trade (if $v' \geq c$) with a utility of $v - \frac{v'+c}{2} \leq 0$.

(ii)$c \leq v$: The utility of buyer is $v - c \geq 0$ under $f^b$ when reporting $v$ and reporting any other value $v'$ under $\bar{f}$ either achieves outcome $\emptyset$ (if $v' < c$) and utility 0 or achieves trade (if $v' \geq c$) with a utility of $v - \frac{v'+c}{2} \leq v - c$.

Thus, the buyer has (weakly) higher utility when reporting $v$ under $f^b$ than reporting any other value $v'$ under $f^b$, $f^s$, or $\bar{f}$ regardless of the cost of the seller, $c$. In other words, conditions (ii)-(iv) in Proposition 8 is satisfied for the buyer. Analogously, conditions (ii)-(iv) are satisfied for the seller.

All outcomes assigned in the mechanism are determined by $f^b$, $f^s$, and $\bar{f}$, which satisfy ex post individual rationality whenever types are truthfully reporting their payoff type. Budget balance is satisfied automatically given the specification of the environment. The environment is one of private values, thus the result follows from Proposition 8 in Appendix B.

□

# Appendix B Implementation when n=2

**Proposition 7.** *(Sufficient Conditions n=2) Let $F$ be a social choice rule and let the environment be one of independent private values and let $n = 2$. Let $\mathcal{B}$ be a Bayesian type space and let $\mathcal{L} = \langle \mathcal{B}; (T_i, \mu_i)_{i=1,\dots,n}; \bar{k} \rangle$ be a ($\mathcal{B}$-based) level-k type space. If there exists a function $f^i : \Theta \to \triangle(Y)$ for each $i \in I$ and a function $\bar{f} : \Theta \to \triangle(Y)$ such that the conditions (i)-(iii) hold in Proposition 1 and condition (iv) holds below, then $F$ is level-k implementable. Further, there exists a mechanism $\langle M, f \rangle$ such that for any atomless anchor $\alpha$ there is a level-k solution under $\alpha$ that achieves $F$.*

**(iv)** $\sum_{\theta_{-i} \in \Theta_{-i}} \rho(\theta_{-i}|\theta_i) u_i(f^i(\theta), \theta_i) \geq \sum_{\theta_{-i} \in \Theta_{-i}} \rho(\theta_{-i}|\theta_i) u_i(f^j(\theta_i', \theta_{-i}), \theta_i)$ $\forall \theta_i' \in \Theta_i$, $\theta \in \Theta$, $i, j \neq i \in I$

Proof:

Consider the same mechanism in the proof of Proposition 2. Let $\alpha$ be an atomless anchor.

Consider an agent $i$ with payoff $\theta_i$ and level 1.

If she sends the message $(\theta_i', k_i, z_i)$ with $z_i = 0$ and $k_i = 1$, she will expect to receive the following lottery over outcomes

$$\sum_{\theta_{-i}\in\Theta_{-i}} \rho(\theta_{-i}) \cdot f^i(\theta_i', \theta_{-i}).$$

If she sends the message $(\theta_i', k_i, 0)$ with $k_i \neq k$, she will expect to receive one of the two following lottery over outcomes

$$\sum_{\theta_{-i}\in\Theta_{-i}} \rho(\theta_{-i}) \cdot f^j(\theta_i', \theta_{-i})$$

for $j \neq i$ or

$$\sum_{\theta_{-i}\in\Theta_{-i}} \rho(\theta_{-i}) \cdot \bar{f}(\theta_i', \theta_{-i}).$$

And, if she sends the message $(\theta_i', k_i, z_i)$ with $z_i \neq 0$, she will expect to receive the following lottery over outcomes

$$\sum_{\theta\in\Theta} \rho(\theta) \cdot \bar{f}(\theta).$$

Thus, by conditions (ii)-(iii) it must be that reporting $(\theta_i, 1, 0)$ is a best response.

We now prove that an agent with payoff type $\theta_i$ and level $k$ will send the message $(\theta_i, k_i, 0)$ by induction on the following statement: Let $k \geq 1$ and assume that if for all $l \in \{1, \ldots, k-1\}$, $\theta_j \in \Theta_j$, and $j \in I$ an agent $j$ with payoff type $\theta_j$ and level $l$ will report $(\theta_j, l, 0)$, then an agent $i$ with payoff type $\theta_i$ and level $k$ will report $(\theta_i, k, 0)$.

The result is true for $k = 1$ by the above argument. Now consider an agent $i$ with payoff type $\theta_i$ and level $k \in \{2, \ldots, \bar{k}\}$.

If she sends the message $(\theta_i', k, 0)$ she will expect to receive the following lottery over outcomes

$$\sum_{\theta_{-i}\in\Theta_{-i}} \rho(\theta_{-i}) \cdot f^i(\theta_i', \theta_{-i}).$$

If she sends the message $(\theta_i', k_i, 0)$ with $k_i \neq k$, she will expect to receive one of the two following lottery over outcomes

$$\sum_{\theta_{-i} \in \Theta_{-i}} \rho(\theta_{-i}) \cdot f^j(\theta_i', \theta_{-i})$$

for $j \neq i$ or

$$\sum_{\theta_{-i} \in \Theta_{-i}} \rho(\theta_{-i}) \cdot \bar{f}(\theta_i', \theta_{-i}).$$

And, if she sends the message $(\theta_i', k_i, z_i)$ with $z_i \neq 0$, she will expect to receive the following lottery over outcomes

$$\sum_{\theta \in \Theta} \rho(\theta) \cdot \bar{f}(\theta).$$

By the same argument above, conditions (ii)-(iv) imply that our agent will send the report $(\theta_i, k, 0)$.

Therefore, if we define $m_i(\theta_i, k_i) = (\theta_i, k_i, 0)$ for all $\theta_i \in \Theta_i$, $k_i \in \{1, \ldots, \bar{k}\}$ and $i \in I$, then $m$ is a level-k solution and achieves $F$ by condition $(i)$. Hence, $F$ is level-k implementable. Further, since $\alpha$ was an arbitrary atomless anchor, $m$ acheives $F$ for all atomless anchors.

□

**Proposition 8.** *(Ex Post Sufficient Conditions n=2) Let $F$ be a social choice rule and let the environment be one of private values and let $n = 2$. If there exists a function $f^i : \Theta \to \triangle(Y)$ for each $i \in I$ and a function $\bar{f} : \Theta \to \triangle(Y)$ such that the conditions (i)-(iii) hold in Proposition 4 and condition (iv) holds below, then $F$ is ex post level-k implementable. Further, there exists a mechanism $\langle M, f \rangle$ such that for any atomless anchor $\alpha$ there is an ex post level-k solution under $\alpha$ that achieves $F$.*

**(iv)** $u_i(f^i(\theta), \theta) \geq u_i(f^j(\theta_i', \theta_{-i}), \theta)\ \forall \theta_i' \in \Theta_i,\ \theta \in \Theta,\ i, j \neq i \in I$

PROOF:

Fix $\bar{k} \in \mathbb{N}_+$. Consider the same mechanism in the proof of Proposition 5. Let $\alpha$ be an atomless anchor.

By analogous arguments to those in the proof of Proposition 7 it can be show that conditions (ii)-(iv) imply that an agent with payoff type $\theta_i$ and level $k_i > 0$ sending the report $(\theta_i, k, 0)$ is a best response for any $\theta_{-i} \in \Theta_{-i}$. Therefore, if we define $m_i(\theta_i, k_i) = (\theta_i, k_i, 0)$ for all $\theta_i \in \Theta_i$ with $k_i \in \{1, \ldots, \bar{k}\}$, then $m$ is an ex post level-k solution and $m$ achieves $F$ by condition $(i)$. Therefore, $F$ is ex post level-k implementable. Further, since $\alpha$ was an arbitrary atomless anchor, $m$ acheives $F$ for all atomless anchors.

□

# Appendix C LDoR implementation

This appendix addresses the question of whether we can design a mechanism that is robust to relaxing one of the belief assumptions that is present in the level-k type space. Specifically, under the level-k model, if an agent is level $k$, she believes that others have levels exactly equal to $(k-1)$. In general, we might allow an agent with level $k$ to hold beliefs over all lower levels.

The following definition of a limited depth of reasoning (LDoR) type space generalizes the level-k type space by allowing the agent to hold any arbitrary beliefs over lower levels of others. This approach is based on Strzalecki (2014) who develops the framework for games of complete information. We expand the framework here to allow for incomplete information.

**Definition 11. $\mathcal{B}$-based limited depth of reasoning type space (LDoR type space)** is a type space $\mathcal{L}^{LDoR} = \langle \mathcal{B}; (T_i, k_i, \theta_i, b_i)_{i=1,\ldots,n}; \bar{k} \rangle$, such that $\mathcal{B}$ is a Bayesian type space $\mathcal{B} = \langle \Theta; \rho \rangle$, $T_i$ is a finite set for all $i \in I$, $k_i : T_i \to \{0, \ldots, \bar{k}\}$, $\theta_i : T_i \to \Theta_i$, and $b_i : \Theta_i \times \{1, \ldots, \bar{k}\} \to \triangle(T_{-i})$ such that for all $t_i \in T_i$:

$$b_i(t_i)(\{t_{-i} \in T_{-i} | \text{ such that } k_j(t_j) < k_i(t_i) \; \forall j \neq i \in I\}) = 1$$

and for all $l \in \{0, \ldots, k_i(t_i) - 1\}$ with $b_i(t_i)(\{t_{-i} \in T_{-i} | k_j(t_j) = l \text{ for some } j \neq i \in I\}) > 0$ then

$$b_i(t_i)(\{t_{-i} \in T_{-i} | \theta_{-i}(t_{-i}) = \theta_{-i} \text{ and } k_j(t_j) = l \text{ for all } j \neq i \in I\}) = \rho(\theta_{-i} | \theta_i(t_i))$$

for all $\theta_{-i} \in \Theta_{-i}$.

As in the level-k type space, an agent's type, $t_i$, represents both her payoff type, $\theta_i$, and her level, $k_i$. We abuse notation here and also let $\theta_i$ and $k_i$ be functions that map types to payoff types and levels, respectively. Thus, for a type, $t_i$, her payoff type is given by $\theta_i(t_i)$ and her level is given by $k_i(t_i)$. The belief function, $b_i$, specifies, for each type, her beliefs about the types of others. We impose two belief restrictions. The first restriction requires that each type only puts positive weight on types that have strictly lower levels. The second restriction maintains the common prior assumption.

Given the definition of an LDoR type space, we can define the solution and implementation concepts: the LDoR solution and LDoR implementation.

**Definition 12.** For a given game defined by a mechanism $\langle M, f\rangle$ and type space $\mathcal{L}^{LDoR} = \langle \mathcal{B}; (T_i, k_i, \theta_i, b_i)_{i=I}\, ; \bar{k}\rangle$, a strategy profile $s = s_1 \times \cdots \times s_n$, with $s_i : T_i \to \triangle(M_i)$ for all $i \in I$, is th**e LDoR solution under anchor $\boldsymbol{\alpha}$** if and only if:

**(i)** $s_i(t_i) \sim \alpha(t_i)$ for all $t_i \in \{t \in T_i | k_i(t) = 0\}$, $i \in I$

**(ii)** $\sum\limits_{t_{-i}\in T_{-i}} b_i(t_{-i}|t_i)u_i(f(s(t), \theta(t)) \geq \sum\limits_{t_{-i}\in T_{-i}} b_i(t_{-i}|t_i)u_i(f(m_i', s_{-i}(t_{-i})), \theta(t))$
$\forall\ m_i' \in M_i$, $t_i \in T_i$ with $k_i(t_i) \geq 1$, $i \in I$.

The LDoR solution is similar to the level-k solution. To define LDoR implementation however, we go a step further here and apply a further robustness criterion - that the social choice rule be implementable under *any* LDoR type space.

**Definition 13.** Fix a $\bar{k}$ and a Bayesian type space $\mathcal{B}$. A social rule $F$ is **LDoR implementable** if there exists a mechanism $\langle M, f\rangle$ and an anchor $\alpha$ such that for any $\mathcal{B}$-based LDoR type space $\mathcal{L}^{LDoR} = \langle \mathcal{B}; (T_i, k_i, \theta_i, b_i)_{i=1,\ldots,n}\, ; \bar{k}\rangle$, there exists a strategy profile $s_i : T_i \to \triangle(M_i)$ for all $i \in I$, such that $s = s_1 \times \cdots \times s_n$ is an LDoR solution under anchor $\alpha$ and $f(s(t)) \in F(\theta(t))$ for all $t \in \{t \in T | k_i(t_i) \geq 1, \forall i \in I\}$.

Proposition 9 gives the sufficient conditions for LDoR implementation. The sufficient conditions are the same as those required for level-k implementation when $n = 2$. These results are stated below.

**Proposition 9.** *(LDoR Sufficient Conditions) Let $F$ be a social choice rule. Let the environment be one of independent private values. If conditions (i)-(iii) hold in Proposition 1 plus condition (iv) below then $F$ is LDoR implementable.*

**(iv)** $\sum_{\theta_{-i}\in\Theta_{-i}} \rho(\theta_{-i}|\theta_i)u_i(f^i(\theta),\theta_i) \geq \sum_{\theta_{-i}\in\Theta_{-i}} \rho(\theta_{-i}|\theta_i)u_i(f^j(\theta_i',\theta_{-i}),\theta_i)$ $\forall\theta_i' \in \Theta_i$, $\theta \in \Theta$, $i, j \neq i \in I$

Proof:

Consider the following mechanism where the message space for agent $i$ is equal to $M_i = \Theta_i \times \{0,1,\ldots\bar{k}\} \times [-1,1]$ . Let $\alpha$ be an atomless anchor.

Define the functions $I_i : [-1,1]^n \to \{0,1\}$, $\tilde{\theta}_i : M \to \Theta_i$, and $\tilde{k}_i : M \to \{0,1,\ldots,\bar{k}\}$ as in Proposition 2. Then, let the planner assigns outcomes based on the reports $(\tilde{\theta},\hat{\tilde{k}})$ according to the function $\tilde{g} : \Theta \times \{0,\ldots,\bar{k}\}^n \to \triangle(Y)$ defined by

$$\tilde{g}(\theta,\hat{k}) \;=\; \begin{cases} f^i(\theta) & \text{if } k_i > \max\{k_1,\ldots,k_{i-1},k_{i+1},\ldots k_n\} \\ \bar{f}(\theta) & \text{otherwise} \end{cases}.$$

Let $\mathcal{L}^{LDoR} = \langle \mathcal{B}, (T_i,k_i,\theta_i,b_i)_{i\in I}, \bar{k}\rangle$ be an LDoR type space.

Consider an agent $t_i \in T_i$ with $\theta_i(t_i) = \theta_i$ and level $k_i(t_i) = 1$. The beliefs and incentives for level 1 agents are unchanged relative to the level-k type space and mechanism in Proposition 2. Thus, for any agent $i$ with payoff type $\theta_i$ and level 1, reporting $(\theta_i,1,0)$ is a best response.

We now prove that an agent with payoff type $\theta_i$ and level $k$ will send the message $(\theta_i,k,0)$ by induction on the following statement: Let $k \geq 1$ and assume that if for all $l \in \{1,\ldots,k-1\}$, $\theta_j \in \Theta_j$, and $j \in I$ an agent $j$ with payoff type $\theta_j$ and level $l$ will report $(\theta_j,l,0)$, then an agent $i$ with payoff type $\theta_i$ and level $k$ will report $(\theta_i,k,0)$.

The result is true for $k = 1$ by the above argument. Now, consider an agent $t_i$ with payoff type $\theta_i(t_i) = \theta_i$ and level $k_i(t_i) = k \in \{2,\ldots,\bar{k}\}$. She expects other agents that have strictly positive levels to always send reports $z_j = 0$. Thus she expects that the social planner will always take their payoff and level reports as given. For agents with levels of 0, she expects the planner to almost surely use a payoff type randomly chosen according to $\rho_j$ and level report of 0 for those agents.

Thus, if she sends the message $(\theta_i', k, 0)$ she will expect to receive the following lottery over outcomes

$$\sum_{\theta_{-i}\in\Theta_{-i}} \rho_i(\theta_{-i}) \cdot f^i(\theta_i', \theta_{-i})$$

If she sends the message $(\theta_i', k_i, 0)$, $k_i \neq k$, she will expect to receive the following lottery over outcomes

$$\sum_{j\in I}\beta^j \cdot \sum_{\theta_{-i}\in\Theta_{-i}} \rho_i(\theta_{-i}) \cdot f^j(\theta_i', \theta_{-i}) + \left(1 - \sum_{j\in I}\beta^j\right) \sum_{\theta_{-i}\in\Theta_{-i}} \rho_i(\theta_{-i}) \cdot \bar{f}(\theta_i', \theta_{-i})$$

for some $\beta = (\beta^j)_{j\in I}$ such that $\beta^j \in [0,1]$ and $\sum_{j\in I}\beta^j \leq 1$ .

And, if she sends the message $(\theta_i, k_i, z_i)$ with $z_i \neq 0$, then she will expect to receive the following lottery over outcomes

$$\sum_{j\in I}\beta^j \cdot \sum_{\theta\in\Theta}\rho(\theta) \cdot f^j(\theta) + \left(1 - \sum_{j\in I}\beta^j\right) \sum_{\theta\in\Theta}\rho(\theta) \cdot \bar{f}(\theta)$$

for some for $\beta = (\beta^j)_{j\in I}$ such that $\beta^j \in [0,1]$ and $\sum_{j\in I}\beta^j \leq 1$ .

By condition (ii) and (iv) it must follow that

$$\begin{aligned} &\sum_{\theta_{-i}\in\Theta_{-i}} \rho_i(\theta_{-i}) \cdot f^i(\theta_i, \theta_{-i}) \\ \geq &\sum_{j\in I}\beta^j \cdot \sum_{\theta_{-i}\in\Theta_{-i}} \rho_i(\theta_{-i}) \cdot f^j(\theta_i', \theta_{-i}) + \left(1 - \sum_{j\in I}\beta^j\right) \cdot \sum_{\theta_{-i}\in\Theta_{-i}} \rho_i(\theta_{-i}) \cdot \bar{f}(\theta_i', \theta_{-i}) \\ \geq &\sum_{j\in I}\beta^j \cdot \sum_{\theta\in\Theta}\rho(\theta) \cdot f^j(\theta) + \left(1 - \sum_{j\in I}\beta^j\right) \cdot \sum_{\theta\in\Theta}\rho(\theta) \cdot \bar{f}(\theta) \end{aligned}$$

for any $\beta = (\beta^j)_{j\in I}$ such that $\beta^j \in [0,1]$ and $\sum_{j\in I}\beta^j \leq 1$ and for all $\theta_i' \in \Theta_i$.

Thus, for agent $i$ with payoff type $\theta_i$ and level $k$, reporting $(\theta_i, k, 0)$ is a best response. Therefore, if we define $m_i(t_i) = (\theta_i(t_i), k_i(t_i), 0)$ for all $t_i \in T_i$ with $k_i(t_i) \in \{1, \ldots, \bar{k}\}$, then $m$ is a LDoR solution and our given mechanism implements $F$.

Since the above holds for an arbitrary LDoR type space, $F$ is LDoR implementable.

$\square$