WZB
Wissenschaftszentrum Berlin
für Sozialforschung

Terri Kneeland

Mechanism design with level-k types: Theory and an application to bilateral trade

Discussion Paper

SP II 2017–303

April 2017

Research Area
Markets and Choice

Research Unit
Economics of Change

Wissenschaftszentrum Berlin für Sozialforschung gGmbH
Reichpietschufer 50
10785 Berlin
Germany
www.wzb.eu

Affiliation of the authors:

Terri Kneeland, University College London (t.kneeland@ucl.ac.uk)

Abstract

Mechanism design with level-k types: Theory and an application to bilateral trade*

This paper studies mechanism design under the level-k solution concept. The first result gives a general necessary condition for a social choice rule to be level-k implementable. In some environments, this necessary condition is equivalent to Bayesian incentive compatibility, making level-k implementation more restrictive than Bayesian implementation. The second result shows that this is not a general implication. In the bilateral trade environment ex post efficient trade is always possible under level-k implementation. Further, ex post efficient trade is possible in a mechanism that is robust to different specifications of beliefs about the levels of reasoning of others and to any specification of beliefs about payoffs.

* 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, Eliaz Kfir, Wei Li, Rosemarie Nagel, Ariel Rubinstein, Sergei Severinov, and Rani Spiegler.

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' likely responses and adjust their beliefs by a finite number of iterated best responses.[2] The model is anchored in the behavior of level 0 types which is exogenously given and generally assumed to be uniformly random. 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 only 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 social choice rule is level-k implementable if for every profile of payoff types and levels, the actions played under the level-k model are consistent with the social choice rule.

Our first main result establishes a general necessary condition for level-k implementation (Proposition 1). The necessary condition is a set of level-k incentive constraints that are analogous to the Bayesian incentive constraints. The level-k incentive constraints ensure that an agent, with level of reasoning two or greater, will truthfully report his type given that he believes everyone else is truthfully reporting their type. There is an important distinction between the level-k incentive constraints and the Bayesian incentive constraints (agents truthfully report their payoff types under the beliefs that others truthfully report their payoff types). In the level-k model, a type is *not* equivalent to a payoff type. It is instead, a two-dimensional object consisting of both an agent's payoff type and her level

[1] For, pioneering works 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).

[2] Formally, the model in this paper most closely resembles Strzalecki (2014) which uses a type space approach to model the level-k solution concept under complete information. This paper adapts that approach to allow for incomplete information. However, this approach closely follows the spirit of the models of Crawford & Iriberri (2007), Crawford et al. (2009), and Crawford (2016) which adapted the level-k models to study auctions under incomplete information.

of reasoning. To see why this distinction matters notice that two agents with the same payoff type and different levels of reasoning may have different beliefs and hence, different incentives. Rather than focusing on mechanisms where agents truthfully report their payoff types, we consider augmented mechanisms where agents truthfully report their their payoff types and their levels of reasoning.[3,4] As such, there is a gap between the level-k and Bayesian incentive constraints. If the Bayesian incentive constraints hold, then the level-k incentive constraints also hold. But, it may be possible to ensure that the level-k incentive constraints hold without the Bayesian incentive constraints holding.

The ability to do so depends on the environment, most importantly, on the nature of the social choice rule. If the social planner is implementing a social choice function (a single-valued rule), it will not be possible. In this case, Bayesian implementation is a necessary condition for level-k implementation (Corollary 1). However, if the social planner is implementing a social choice correspondence (a multi-valued rule), it may be that the level-k incentive constraints hold without the Bayesian incentive constraints holding. The rest of the paper focuses on the case of ex post efficiency in a bilateral trade environment, which is an example of an environment with a multi-valued choice rule (the social planner desires efficient trade but does not care at which price the good is traded at). Our second main result is that, in the bilateral trade environment, *all* ex post efficient outcomes are level-k implementable (Proposition 2). This is in obvious contrast to Bayesian implementation where there is a conflict between ex post efficiency and incentive compatibility.

We then ask how robust are level-k mechanisms to alternative specifications of beliefs? We have two robustness results. Proposition 3 establishes that there exists a mechanism that implements ex post efficiency for any beliefs about the payoff types of others. This first robustness result is in stark contrast to standard Bayesian implementation results, where it is well known that the optimal mechanisms are sensitive to assumptions about beliefs about payoff types. Finding mechanisms that are robust to relaxing these strong common knowledge assumptions, typically known as the Wilson doctrine, can insure that a

[3] This mirrors the analysis in Bergemann & Morris (2005). They study implementation in incomplete information environments without the common prior assumption. In this environment, types with the same payoff type may hold different beliefs about the payoff types of others. Thus, a type represents not just the payoff type, but also higher-order beliefs about the payoff types of others. The authors use augmented mechanisms where agents report their types and not just their payoff types.

[4] The revelation principle does not hold in the level-k environment. This is true even if we allow augmented mechanisms where all types, with levels greater than one, truthfully report their type. This results from the 'menu effects' discussed in Crawford et al. (2009) due to the incentives of level 1 players. Level 1 agents best respond to Level 0 agents which play uniformly randomly. Thus, level 1 agents are not playing a best response to others truthfully reporting their type, and in fact, depending on the mechanism, may place positive weight on strategies that are never played by agents with levels greater or equal to one. There may be a role for these additional strategies, those played only by level 0 agents, in satisfying the incentives of level 1 agents.

social choice correspondence will be implemented even if the designer 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 show that ex post efficiency can be implemented in mechanisms that relax common knowledge of payoffs under the empirically plausible assumption of level-k reasoning.

Second, we show that allowing for different beliefs about the reasoning types of others does not affect implementation. The level-k model imposes very specific beliefs: a level k type believes her opponent is a level $k-1$ type. However, the spirit of limited depths of reasoning is maintained whenever a level k type has any beliefs over lower types $0, \ldots, k-1$. Proposition 4 shows that we can always find a mechanism that is robust to allowing a level k type agent to hold any beliefs over lower levels. Importantly, there exists a single mechanism that implements ex post efficiency that is robust to both beliefs about payoffs and beliefs about the levels of others.

There is a growing literature that focuses on behavioral mechanism design.[5] This paper adds to this literature by studying implementation under the level-k model. Four other papers study level-k implementation. The main contributions of the current work relative to these four are that this paper: (i) demonstrates that level-k implementation may be a weaker implementation concept than Bayesian implementation; (ii) studies the robustness of level-k implementation to common knowledge of payoff and level-k assumptions; and (iii) demonstrates the role of augmented mechanisms in level-k implementation.

Crawford (2016) studies level-k implementation in the continuous bilateral trade environment. Crawford restricts the space of mechanisms to be one where the set of messages is equivalent to the set of payoff types. Under the restricted mechanism, agents cannot signal both their payoff type and their level by their choice of message. As a result, Crawford finds that ex post efficiency is level-k implementable only when it is Bayesian implementable. We build on this work and show that under more general mechanisms ex post efficiency is always level-k implementable. In contemporaneous work, de Clippel et al. (2016) study level-k implementation in a general setting where the social planner aims to implement a single-value social choice rule. In their case, since they study single-valued choice rules, Bayesian incentive constraints are necessary for level-k implementation.[6] In contrast, we provide the

[5]Eliaz (2002) studies mechanism design when there is a proportion of 'faulty' agents that fail to act optimally. Cabrales & Serrano (2011) allow agents to learn in the direction of better replies. 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. Glazer & Rubinstein (1998), Eliaz & Spiegler (2006; 2007; 2008), Severinov & Deneckere (2006), and Wolitzky (2016) study behavioral mechanism design in individual decision problems.

[6]They use a slightly stronger definition of level-k implementation than the one used here. Essentially,

level-k incentive constraints for the case of a general choice rule and show that they collapse to Bayesian incentive compatibility when the social choice rule is single-valued. They also provide sufficient conditions for level-k implementation in a number of specific single-valued choice rule environments. 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 Example

We illustrate the main results of the paper with a simple bilateral trade example with two types. The seller has one unit of a good to sell and the buyer wants to purchase one unit of the good. The seller has a payoff type, $c \in C = \{2, 6\}$, which represents his cost - the minimum value that the seller is willing to sell the good for. The buyer has a payoff type, $v \in V = \{3, 7\}$, which represents her value - the maximum value that the buyer is willing to pay for the good. Types are drawn from a uniform common prior, ρ (i.e. $\rho(v, c) = \frac{1}{4}$ for all $v, c \in V \times C$).

The set of outcomes is given by $Y = \mathbb{R} \cup \{\emptyset\}$. The outcome $\emptyset$ indicates that the good is not traded and the outcome $p \in \mathbb{R}$ indicates that the good is traded at price p. That is, the buyer pays the seller the amount p (if p is negative that would indicate that the seller pays the buyer). Agents have quasi-linear utility functions. If there is a trade at a price p, then the buyer has utility $v - p$ and the seller has utility $p - c$. If there is no trade, then the buyer and seller have utility 0.

We are interested in mechanisms that satisfy the following three properties:

1. Ex post efficiency: the mechanism transfers the good to the buyer if and only if the buyer's value is higher than the seller's cost.

2. Budget balance: the price paid by the buyer equals the price received by the seller.

3. Ex post individual rationality: both the buyer and seller prefers to participate in the trading institution.

In general, we are interested in the implementability of some social choice rule, in the case of bilateral trade, it is the ex post efficient social choice rule $F^*(v, c) = \{y | y \in \mathbb{R} \text{ if } v \geq c \text{ and } y = \emptyset \text{ otherwise}\}$. However, in the context of bilateral trade we are interested in the mechanism satisfying all of the above three properties and not just efficiency. To simplify

they require a version of strict implementation where this paper allows indifferences. This adds an additional strictness condition to their necessary conditions that is absent in this paper.

discussion, when we ask whether ex post efficiency is implementable, we do so under the condition that budget balance and ex post individual rationality are also satisfied.

Remark 1. Ex post efficiency is not Bayesian implementable.

We use the standard definition of Bayesian implementation. The ex post efficient social choice rule, F^*, is Bayesian implementable if there exists a mechanism, $\mathcal{M} = \langle M = M_b \times M_s, f : M \to Y \rangle$, such that in the game defined by that mechanism there is a Bayesian equilibrium $s = s_b \times s_s$, where $s_b : V \to M_b$ and $s_s : C \to M_s$ and the three conditions above hold: (i) ex post efficiency: $f\left((s_b(v), s_s(c))\right) \subset F^*\left((v, c)\right)$ for all $(v, c) \in V \times C$ where $F^*((v, c)) = \{y \in Y \,| y \in \mathbb{R}$ if $v \geq c$ and $y = \emptyset$ otherwise$\}$ is the ex post efficient social choice rule; (ii) budget balance (notice that budget balance is imposed automatically by our outcome space); and (iii) ex post individual rationality: $f((s_b(v), s_s(c))) \geq 0$ for all $(v, c) \in V \times C$.

Ex post efficiency 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 make zero rents in equilibrium as they never have an incentive to misreport. Thus any candidate mechanism must take the form as the one in Figure 1

		Seller	
		2	6
Buyer	7	p	6
	3	3	$\emptyset$

Figure 1: Structure of a Bayesian mechanism

A high value buyer believes the low and high valued seller types are equally likely (comes from the uniform prior assumption) and thus believes actions 2 and 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, a low valued seller believes the low and high valued buyer types are equally likely and thus believes actions 7 and 3 are equally likely For a low valued seller to truthfully report his 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 which will implement the ex post efficient choice rule under Bayesian implementation.

Remark 2. Ex post efficiency is level-k implementable.

Level-k implementation is defined similarly to Bayesian implementation except Bayesian equilibrium is replaced by the level-k solution concept.

Under the level-k model, player's have heterogeneous levels of reasoning. Level 0 types have zero depths of reasoning, they are non-strategic and we assume they play uniformly, randomly over the action space. Level 1 types have one level of reasoning, they best respond to level 0 types. Level 2 types have two levels of reasoning - they best respond to level 1 types. And, so on, with level k types having k levels of reasoning and best responding to level $k-1$ types.

Suppose, for the sake of this example, that there are only level 1 and level 2 types in the population. As such there are four different types for the buyer and the seller. There is both a high-valued and a low-valued payoff type for each of the two reasoning types. We formally model this using a type space structure in the next section.

A level 0 type plays according to the uniform distribution. A level k type plays a best response to her beliefs which are determined by the common prior and the actions each level $k-1$ type plays. An agent's level determines her beliefs about her opponent's level (i.e. a level k type believes her opponent is type level $k-1$). And, an agent's payoff type determines her beliefs about her opponent's payoff types (i.e. the beliefs of level k buyer with payoff type v about the seller's payoff type are determined by the common prior). Thus, in this example, a level k buyer believes it is equally likely that the seller is a type with level $k-1$ and payoff type 2 and a type with level $k-1$ and payoff type 6. A level k seller believes it is equally likely that the buyer is a type with level $k-1$ and payoff type 3 and a type with level $k-1$ and payoff type 7.

We use the following definition of level-k implementation. The ex post efficient social choice rule F^* is level-k implementable if there exists a mechanism, $\mathcal{M}$, such that in the game defined by that mechanism, it must be that the strategy profile $s = s_b \times s_s$, $s_b : V \times \{0, 1, 2\} \to M_b$ and $s_s : C \times \{0, 1, 2\} \to M_s$, is a level-k solution, and the following conditions hold: ex post efficiency: $f\left((s_b(v, j), s_s(c, l))\right) \subset F^*\left((v, c)\right)$ for all $(v, c) \in V \times C$ and all $j, l \in \{1, 2\}$, budget balance, and ex post individual rationality: $f\left((s_b(v, j), s_s(c, l))\right) \geq 0$ for all $(v, c) \in V \times C$ and all $j, l \in \{1, 2\}$.

First, notice that our notion of level-k implementability does not require the mechanism to satisfy ex post efficiency or ex post efficient individual rationality if there are level 0 types. This is for two reasons. The first is theoretical, level 0 agents are non-strategic and just play all actions randomly, hence the social planner cannot incentivize their behavior.[7]

[7] If this type of behavior is a concern we should consider an alternative form of implementability where the goal is to perhaps minimize the deviations from the social choice correspondence . See Eliaz (2002) for a possible way to model this form of implementability.

The second motivation is empirical. While there is mixed support, the estimated frequency of level 0 types is typically small (E.g. Arad & Rubinstein 2012; Costa-Gomes & Crawford 2006; Costa-Gomes et al. 2001; Brocas et al. 2014). Thus we interpret the existence of level 0 types in the model, but not in our implementability requirement, as types that exist only in the minds of the other types.

Second, notice that the ex post efficient social choice rule cannot be achieved with any direct mechanism from Figure 1. The level 1 type's beliefs are the same as the agents' beliefs in the Bayesian equilibrium. Instead, we consider augmented mechanisms. In these mechanisms, we allow agents with the same payoff type but different levels of reasoning to play different actions.

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

The mechanism in Figure 2 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 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 ($v = 3$) and playing $m_{(7,1)}$ is a best response for the high valued level 1 buyer ($v = 7$). Likewise, playing $m_{(6,-)}$ is a best response for the high valued level 1 seller ($c = 6$) and playing $m_{(2,1)}$ is a best response for the low valued level 1 seller ($c = 2$).

A level 2 buyer (of either payoff type) believes that high valued level 1 sellers play $m_{(6,-)}$ and low valued level 1 sellers play $m_{(2,1)}$. Since she believes high and low valued sellers are equally likely, she expects $m_{(6,-)}$ and $m_{(2,1)}$ to be played with equal probability. Thus, playing $m_{(3,-)}$ is the best response for a low valued level 2 buyer and playing $m_{(7,2)}$ is the best response for a high valued level 2 buyer. A level 2 seller (of either payoff type) believes that low valued level 1 buyers play $m_{(3,-)}$ and high valued level 1 buyers play $m_{(7,1)}$. Since he believes high and low valued buyers are equally likely, he expects $m_{(3,-)}$ and $m_{(7,1)}$ to be played with equal probability. Thus, playing $m_{(6,-)}$ is the best response for a high valued level 2 seller and playing $m_{(2,2)}$ is the best response for a low valued level 2 seller.

Given the strategies defined by the level-k model, 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 valued type, then regardless of whether the buyer and sellers are L1 or L2 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 L1 or L2 types, there will be trade. Ex post individual rationality is satisfied for all outcomes.

There exist actions in the mechanism that level 1 and level 2 types do not play. Neither level 1 or level 2 types play action m_0. Thus, we would not expect that action to be played by either the buyer or the seller (if there are no level 0 types in the population). However, the mechanism includes that action because level 1 types believe that action is being played with some positive probability by the level 0 type. Thus, these 'level 0' actions influence the behavior of level 1 types even though they are not played by types of higher levels.[8]

Next we illustrate the robustness of level-k implementation.

Remark 3. Robustness to limited depths of reasoning

The level-k model assumes a very particular form of beliefs: level k types believe there are only level $k-1$ types. But, it is reasonable to think that a level k type might put positive weight on other types as well. For example, the cognitive hierarchy model assumes that a level k type puts positive weight on all lower types, where the weight is determined according to the conditional Poisson distribution (Camerer et al. 2004).

The designer may believe that agents have finite depths of reasoning, but assuming he knows the beliefs of the agents about the levels of others is a strong assumption. Importantly, a level-k mechanism does not have to be sensitive to the specification of beliefs. The ex post efficient social choice rule is always implementable in a mechanism that is robust to the specification of beliefs about levels. Figure 3 provides an example of such a mechanism.

The beliefs of level 1 agents do not change (as they may only put weight on level 0 types). Thus, a level 1 type then believes that her opponent is playing each action with equal probability. Therefore, the low valued L1 buyer plays $m_{(3,-)}$ and the high valued L1 buyer plays $m_{(7,1)}$. Likewise, the high valued level 1 seller plays $m_{(6,-)}$ and the low valued level 1 seller plays $m_{(2,1)}$.

Now, a level 2 buyer may hold any beliefs over level 0 and level 1 types but still believes that high and low valued sellers are equally likely. Thus, for any beliefs over L0 and L1 types, playing $m_{(3,-)}$ is the best response for a low valued level 2 buyer and playing either $m_{(7,1)}$ or $m_{(7,2)}$ is the best response for a high valued L2 buyer. Similarly, a high valued level

[8]Strategies that are never played have been shown to still have an impact on strategic behavior in other experiments. Cooper et al. (1990) show that introducing dominated actions into coordination games changes behavior.

			Seller		
		m_0	$m_{(2,1)}$	$m_{(2,2)}$	$m_{(6,-)}$
Buyer	m_0	4.5	7	4.5	$\emptyset$
	$m_{(7,1)}$	2	4.5	5.5	6
	$m_{(7,2)}$	4.5	3.5	4.5	6
	$m_{(3,-)}$	$\emptyset$	3	3	$\emptyset$

Figure 3: A mechanism that is robust to limited depths of reasoning

2 seller will want to play $m_{(6,-)}$and a low valued level 2 seller will want to play either $m_{(2,1)}$ or $m_{(2,2)}$. Thus, each type for both the buyer and the seller still has an incentive to play an action that identifies their payoff type regardless of their belief structure. Therefore, the social planner does not need to know the specification of beliefs about levels of reasoning in order to implement the social choice correspondence.

Remark 4. Robustness to beliefs about payoffs

In the first three remarks, agents' beliefs about payoff types were determined by the common prior. This is a strong assumption - the social planner may not have this information when she designs the mechanism. Further, this assumption is problematic in the Bayesian implementation setting as optimal mechanisms are typically sensitive to the common prior. This is not true for level-k implementation. The ex post efficient correspondence is always implementable in a mechanism that is robust to the specification of beliefs about payoffs. Figure 4 provides an example of such a mechanism.

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

Figure 4: A mechanism that is robust different beliefs about payoffs

Again, the behavior of level 1 agents is unchanged as the beliefs of level 1 agents do not change (a level 0 behavior is independent of payoff types). 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 the best response for a low valued level 2 buyer and playing $m_{(7,2)}$ is the best response for a high valued level 2 buyer. Similarly, for any beliefs over payoff types, playing $m_{(6,-)}$ is the best response for a high valued level 2 seller and playing $m_{(2,2)}$ is the best response for a low valued level 2 seller.

Thus, for both the buyer and the seller, all types have an incentive to play an action that identifies their payoff type regardless of their belief structure. Therefore, the planner does not need to know the specification of beliefs about payoff types in order to implement the social choice correspondence.

This rest of the paper proceeds as follows. Section 3 sets up the general payoff environment and formalizes level-k implementation. Section 4 established necessary conditions for level-k implementation in the general environment. Section 5 sets up the bilateral trade environment and establishes the main result about ex post efficiency and level-k implementation. Section 6 establishes the robustness results.

3 Setup

3.1 General payoff environment

We first define level-k implementation and generate necessary conditions for a general payoff environment. Then, in section 5, we focus on the bilateral trade environment.

There is a finite set of agents $I = 1, 2, \dots, n$. Agent i's *payoff type* $\theta_i \in \Theta_i$, where Θ_i is a finite set. We write $\theta \in \Theta = \Theta_1 \times \cdots \times \Theta_N$. There is a compact set of outcomes Y. Each agent has a continuous utility function $u_i : Y \times \Theta \to \mathbb{R}$.

There is a social planner who is concerned with implementing a 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 θ.

3.2 Type spaces

We use the framework of a type space in order to formally define each agent's 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 constitute a Bayesian type space.

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

Given the common prior ρ, each payoff type forms her beliefs by conditioning on the common prior according to Bayes' rule. The belief of an agent with value θ_{-i} about the costs of the seller is given by $\rho(\theta_{-i}|\theta_i) = \frac{\rho(\theta)}{\sum_{\theta_{-i}\in\Theta_{-i}} \rho(\theta_i,\theta_{-i})}$.

Level-k models are designed to capture the idea that agents are often capable of performing only a finite number of levels of reasoning in order to figure out their optimal behavior. We use a type space approach to define the level-k model based on Strzalecki (2014) who develops the framework for games of complete information. We expand the framework here to account for games of incomplete information.

Definition. A **$\mathcal{B}$-based level-k type space** $\mathcal{L}$ is a structure $\mathcal{L} = \langle \mathcal{B}; \bar{k} \rangle$, where $\mathcal{B}$ is a Bayesian type space $\mathcal{B} = \langle \Theta; \rho \rangle$, $\bar{k} \geq 1 \in \mathbb{N}$

Given a $\mathcal{B}$-based level-k type space $\mathcal{L}$, we can define a set of types for each agent, $L^i = \Theta_i \times \{0, 1, \ldots, \bar{k}\}$. An agent's *type* $t_i = (\theta_i, k_i)$ is a 2-dimensional type representing both her payoff type, θ_i, and her level, k_i. An agent's level represents her depth 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. The type space contains all levels of reasoning from 0 to $\bar{k}$.

A agent's beliefs about the types of others are determined both by her payoff type and her level. The beliefs of a type $t_i = (\theta_i, k_i)$ about the types of others are determined by the function $\mu_i(t_{-i}|t_i)$:

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

An agent with a level k puts weight only on other types that have levels exactly equal to $k - 1$. This captures the core assumption of the level-k model. This assumption ensures that agents can calculate their optimal actions in a recursive fashion with a finite number of steps given the behavior of the level 0 types.

An agent's beliefs about the payoff types of other agents are determined by the common prior ρ. An agent with payoff type θ_i and level k_i believes that the payoff types of other agents are determined by $\rho(\theta_{-i}|\theta_i)$ and that others have level $k_i - 1$.

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

3.3 Solution concepts

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

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

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 $u^i(f(m_i, m_{-i}), \theta)$ for agent i.

For a given level-k type space, we can define the level-k solution concept. The level-k solution concept imposes that all types are rational (that is, they play a best response given their beliefs about the actions of other agents) and for beliefs about those actions to be consistent with what other types are actually doing in equilibrium. Level 0 types are not required to be rational.[9] Level 0 types are assumed to play uniformly randomly.

Definition. For a given game defined by a mechanism $\langle M, f \rangle$ and type space $\mathcal{L}$, a strategy profile $s : L \to M$ is th**e level-k solution** if and only if:

(i) $\sum_{t_{-i} \in L^{-i}} \mu_i(t_{-i}|t_i) \cdot u^i(f(s^i(t_i), s^{-i}(t_{-i})), \theta) \geq \sum_{t_{-i} \in L^{-i}} \mu_i(t_{-i}|t_i) \cdot u^i(f(m, s^{-i}(t_{-i})), \theta)$ for all $t_i \in L^i$ with $k_i > 0$, for all $m \in M^i$, and for all $i \in I$.

(ii) $s_i((\theta_i, 0))(m) = \frac{1}{|M_i|}$ for all $m \in M_i$ for all $\theta_i \in \Theta_i$, for all $i \in I$.

The level-k solution can be calculated recursively given the behavior of level 0 types. Level 1's actions are a best response to level 0's actions. Level 2's actions are a best response to level 1's actions, and so on.

We will be interested in comparing our level-k implementation results with the solution concept of Bayesian implementation. The two solution concepts differ in that the level-k equilibrium relaxes the belief consistency assumption imposed under Bayesian equilibrium by not requiring the incentive constraints to hold for level 0 types. We define Bayesian equilibrium for completeness.

Definition. For a given game defined by a mechanism $\langle M, f \rangle$ and type space $\mathcal{B}$, a strategy profile $s : \Theta \to M$ is a **Bayesian equilibrium** if and only if:

$\sum_{\theta_{-i} \in \Theta_{-i}} \rho(\theta_{-i}|\theta_i) \cdot u^i(f(s^i(\theta_i), s^{-i}(\theta_{-i})), \theta) \geq \sum_{\theta_{-i} \in \Theta_{-i}} \rho(\theta_{-i}|\theta_i) \cdot u^i(f(m, s^{-i}(\theta_{-i})), \theta)$ for all $\theta_i \in \Theta_i$, for all $m \in M^i$, and for all $i \in I$.

3.4 Implementation

A social choice correspondence, F, is level-k implementable on $\mathcal{L}$ if there exists a mechanism and a level-k solution that achieves F for every type profile in L with levels greater than 0.

[9] The behavior of level 0 types is specified outside of the model. Thus, level 0 types do not have to play a best response to their beliefs (and may play actions that are not a best response to any beliefs i.e. play dominated actions). In fact, there are no restrictions on a level 0's beliefs in a level-k type space and we do not formally define them.

Definition. A mechanism $\langle M, f\rangle$ and a message profile $m : L \to M$ **achieves F on $\boldsymbol{\mathcal{L}}$** if for all $\theta \times \boldsymbol{k} \in \Theta \times \cup_{\times i\in I}\{1, \ldots, \bar{k}\}$

$$f(m(\theta \times \boldsymbol{k})) \in F(\theta)$$

Note that for a mechanism to achieve F, we only require the messages sent by types with levels at least one ($k \geq 1$) be consistent with the social choice correspondence.

Definition. A social correspondence is **level-k implementable** on $\mathcal{L}$ if there exists a mechanism $\langle M, f\rangle$ and a message profile $m : L \to M$, such that m is the level-k solution and m achieves F on $\mathcal{L}$.

Notice that our notion of level-k implementation does not require knowledge of the *actual* distribution of types (and hence levels). This is because implementation requires that the outcome be consistent with the social choice correspondence for *all* type profiles and hence does not depend upon the distribution of types.[10]

For completeness, we give the definition of Bayesian implementation below.

Definition. A mechanism $\langle M, f\rangle$ and a message profile $m : V \times C \to M$ **achieves F on $\boldsymbol{\mathcal{B}}$** if for all $\theta \in \Theta$

$$f(m(\theta)) \in F(\theta)$$

The definition of level-k and Bayesian implementation differ only in that the former requires a level-k solution and the latter requires a Bayesian equilibrium that achieves the social choice correspondence.

Definition. A social choice correspondence F is **Bayesian implementable** on $\mathcal{B}$ if there exists a mechanism $\langle M, f\rangle$ and a message profile $m : V \times C \to M$, such that m is a Bayesian equilibrium and m achieves F on $\mathcal{B}$.

4 Necessary Conditions for Level-k Implementation

The following proposition gives a set of necessary conditions for level-k implementation in the form of level-k incentive constraints that are related to the Bayesian incentive constraints.

[10]This is not true for all mechanism design objectives. For example, it would not be true if the goal of the designer 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 types.

Proposition 1. *(Necessary Conditions) Let F be a social choice correspondence. Let $\mathcal{B}$ be a Bayesian type space and let $\mathcal{L}$ be a ($\mathcal{B}$-based) level-k type space. If F is level-k implementable, then there exists functions $f^i : \Theta \to Y$, for all $i \in I$, such that the following conditions hold:*

(i) $f^i(\theta) \in F(\theta)\ \forall \theta \in \Theta$, $\forall i \in I$

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

Proof:

Suppose that the social choice correspondence F is level-k implementable.

Then there exists some mechanism $\langle M, g \rangle$ and a function $l_i : L^i \to M_i$ such that $l = l_1 \times \cdots \times l_N$ is a level-k solution and achieves F.

Consider the behavior of an agent i with type $t_i = (\theta_i, 2)$:

$$\sum_{\theta_{-i} \in \Theta_{-i}} \rho(\theta_{-i}|\theta_i) \cdot u^i(g(l_i(\theta_i, 2), l_{-i}(\theta_{-i}, 1)), \theta)$$
$$\geq \sum_{\theta_{-i} \in \Theta_{-i}} \rho(\theta_{-i}|\theta_i) \cdot u^i(m', l_{-i}(\theta_{-i}, 1)), \theta) \quad \forall m' \in M_i$$

Define $f^i(\theta) = g(l_i(\theta_i, 2), l_{-i}(\theta_{-i}, 1))$

Thus, condition (ii) holds since playing $(\theta_{-i}, 2)$ was optimal in the original mechanism means it is at least as good as playing $m_i(\theta', 2)$ for any $\theta' \in \Theta_i$.

Condition (i) holds by definition of l achieving F.

□

The difference between the level-k incentive constraints and the Bayesian incentive constraints is that the level-k incentive constraints 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 relaxation of the cross-player restriction ($f^1 = \cdots = f^n$) arises because of the relaxation of consistent beliefs under the level-k model. 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 agents' incentive constraints can be satisfied by different f functions: a level 3 agent for player i with payoff type θ_i thinks she will receive $f^i(\theta)$ when playing against level 2 agents with payoff types' θ_{-i}, while a level 2 agent for player j with payoff type θ_j thinks she will receive $f^j(\theta)$ when playing against level 1 agents with payoff types' θ_{-j}.

Whether the cross-player restriction imposed under Bayesian implementation has bite depends on the environment. The following corollary shows that whenever the social planner wants to implement a social choice function (a single-valued choice rule), the level-k incentive constraints collapse down to the Bayesian incentive constraints. In this case, the social planner cannot improve upon the set of outcomes under level-k implementation. If a social choice correspondence is level-k implementable then it is also Bayesian implementable. (Notice that the level-k incentive constraints are only necessary conditions and in reality it may be harder to implement a social choice correspondence under level-k implementation than under Bayesian implementation.) In the next section, we show that this result does not extend to situations with social choice correspondences (such as in the bilateral trade environment). We show that the social planner can always implement the ex post efficient social choice correspondence under level-k implementation. The same cannot be said for Bayesian implementation.

Corollary 1. *Let F be a social choice function. If F is level-k implementable, then it is Bayesian implementable.*

Proof:

> From Proposition 1, it holds that there exists functions $f^i : \Theta \to Y$, for all $i \in I$ such that conditions (i) and (ii) hold. Since F is a social choice function, it must be that $f^1(\theta) = \cdots = f^n(\theta) = F(\theta)$ for all $\theta \in \Theta$. Therefore, it follows by definition that F is Bayesian implementable using the mechanism $\langle \Theta, F \rangle$.

□

The level-k incentive constraints are not sufficient conditions for level-k implementation because agents with only one level of reasoning must be incentivized differently. Level 1 agent's believe everyone else is a level 0, but a level 0's behavior is determined outside of the model and, as we assume throughout this paper, assumed to be uniformly random. This feature of level-k implementation is also the cause of the failure of the revelation principle. Level 1 types believe that others are placing some weight on all strategies. Thus, level 1 beliefs depend on the mechanism and, without pinning down further details of the environment, incentive constraints for level 1 implementation cannot be established. Because of this, we do not provide general sufficient conditions for level-k implementation. In the next section, we focus on a specific application - bilateral trade - where we can then say more about level-k implementation.

5 Bilateral Trade

5.1 Bilateral Trade Environment

The remainder of this paper focuses on the bilateral trade environment. There are two agents: a buyer and a seller. There is a set of outcomes $Y = \mathbb{R} \cup \{\emptyset\}$. The outcome y indicates whether trade has occurred ($y \in \mathbb{R}$) or if there is no trade ($y = \emptyset$). If there is trade, y represents the price of trade ($y \in \mathbb{R}$ is the price paid by the buyer to the seller).

Agents have quasi-linear utility functions. Buyer's values are represented by $v \in V$ where V is some finite set, $V = \{\underline{\mathrm{v}}, \ldots, \bar{v}\}$. Seller's values are represented by $c \in C$ where C is some finite set, $C = \{\underline{\mathrm{c}}, \ldots, \bar{c}\}$. For any outcome y, the utility of a buyer with a valuation v is

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

and the utility of a seller with a cost c is

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

We make three assumptions on the payoff space. The first two assumptions ensure a non-triviality to the decision environment. The third assumption imposes a uniformity condition on the environment, simplifying the analysis.

A1: $\underline{\mathrm{c}} \leq \underline{v} < \bar{c}$

A2: For any $v, v' \in V$ with $v' < v$, there exists a $c \in C$ such that $v' < c \leq v$. And, for any $c, c' \in C$ with $c < c'$, there exists a $v \in V$ such that $c < v \leq c'$.

A3: There exists a $d \in \mathbb{R}_{++}$ such that for any $v \in V$ or $c \in C$ there exists an $n, m \in \mathbb{Z}_+$ such that $v = \underline{v} + nd$ and $c = \underline{\mathrm{c}} + md$.

The social planner is concerned with finding a mechanism that implements the following three conditions:

1. Ex post efficiency: the mechanism transfers the good to the buyer if and only if the buyer's valuation of the good is higher than the seller's. In other words, the social planner is concerned with implementing the ex post efficient social choice correspondence $F^* : \Theta \to 2^Y \backslash \emptyset$:

$$F^*(\theta) = \{y \in Y \,|y \neq \emptyset \text{ if } v \geq c\}.$$

2. Budget balance: the price paid by the buyer equals the price received by the seller. This requirement is satisfied automatically by our restriction of the outcome space Y.

3. Ex post individual rationality: each trader always prefers to participate in the trading institution, rather than not participate.

Our main focus is on whether the social planner can implement ex post efficiency under level-k implementation (condition 1). However, conditions 2 and 3 are sensible to impose in the bilateral trade environment and are easily satisfied here.

5.2 Ex Post Efficiency

This section contains the main result.

We first show (Lemma 1) that the necessary conditions set out in Proposition 1 are satisfied in this environment. Proposition 2 then gives the main result: the ex post efficient correspondence is level-k implementable. The proof is constructive, and uses the necessary conditions for level-k implementation to build an appropriate mechanism.

Lemma 1. *Let $\mathcal{B}$ be a Bayesian type space and let $\mathcal{L}$ be a ($\mathcal{B}$-based) level-k type space. Then there exists functions $f^b : V \times C \to Y$ and $f^s : V \times C \to Y$ that satisfy the conditions of Proposition 1 for $F = F^*$.*

PROOF:

Define

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

Define

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

First, it is easy to see that condition (i) holds for both f^b and f^s as both assign the outcome $\emptyset$ when $v < c$.

Now consider the utility of the buyer under f^b with value v when seller reports costs c. There are two cases to consider:

(i) $v < c$: In this case utility of 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 $v - c < 0$ since $v' \geq c > v$.

(ii)$c \leq v$: In this case 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 -1 and utility 0.

Thus, the buyer has (weakly) higher utility when reporting v than reporting any other value v'regardless of the beliefs of the buyer about the reports c. In other words, condition (ii) is satisfied for the buyer.

Now consider the utility of the seller under f^s with cost c when the buyer reports value v. There are two cases to consider:

(i) $v < c$: In this case utility of seller is 0 when reporting c and reporting any other cost c' either has no effect (if $c' > v$) or achieves trade (if $c' \leq v$) with utility $v - c < 0$ since $c' \leq v < c$.

(ii)$c \leq v$: In this case utility of seller is $v - c \geq 0$ when reporting c and reporting any other cost c'either has no effect (if $c' \leq v$) or achieves outcome -1 and utility 0 (if $c' > v$).

Thus, the seller has (weakly) higher utility when reporting c than reporting any other cost c' regardless of the beliefs of the seller about the reports v. In other words, condition (ii) is satisfied for the seller.

□

Now we state our main result.

Proposition 2. *Let $\mathcal{B}$ be a Bayesian type space and let $\mathcal{L}$ be a ($\mathcal{B}$-based) level-k type space. The ex post efficient correspondence, F^*, is level-k implementable on $\mathcal{L}$, under a mechanism that satisfies budget balance and ex post individual rationality.*

Proof:

This proof is constructive. We will build a mechanism where each type has an incentive to truthfully reveal both their payoff type (value or cost) and their level. A key behavioral difference between level-k and Bayesian implementation will be that level 1 types do not believe others are truthfully revealing their type (value or cost and level), rather they best respond to a uniform distribution over all actions. Besides f^b and f^s which will be used to ensure that level k types, with $k > 1$ truthfully report their type, we need some additional components of the mechanism to ensure that level 1 will want to truthfully report their type. To do this we need to define functions f^{b*} and f^{s*}which help incentivize level 1 types.

Since type sets are finite, we can order $V = \{v_0 = \underline{v}, v_1, \ldots, v_N = \bar{v}\}$ and $V = \{c_0 = c, c_1, \ldots, c_N = \bar{c}\}$ where $v_n = \underline{v} + nd$ and $c_n = \underline{c} + nd$.

Define

$$f^{b*}(v_i, c_j) = \begin{cases} c_j - (\bar{k} - \frac{1}{2})jd & \text{if } v_i \geq c_j \\ \emptyset & \text{otherwise} \end{cases}$$

and

$$f^{s*}(v_i, c_j) = \begin{cases} v_i + (\bar{k} - \frac{1}{2})id & \text{if } v_i \geq c_j \\ \emptyset & \text{otherwise} \end{cases}.$$

For the purpose of this proof, it is only necessary for there to be one level 0 action for each type i.e. $V \times \{0\}$ which would reward level 1 buyers and sellers with f^{b*}or f^{s*} respectively. However, in following section we show that there is one mechanism that is robust to different belief specifications. We build those features into the mechanism now, using multiple level 0 actions: $V \times \{0_1\}, \ldots, V \times \{0_{\bar{k}}\}$.

To fill in the 'gaps' in the mechanism, we use the function, f^m, which compromises between the buyer's and seller's interests. And, a function, $f^{\emptyset}$, which never trades regardless of the messages sent.

Define

$$f^m(v, c) = \begin{cases} \frac{v+c}{2} & \text{if } v \geq c \\ \emptyset & \text{otherwise} \end{cases}$$

and

$$f^{\emptyset}(v, c) = \emptyset.$$

Now construct the mechanism, $\langle M, g \rangle$, in the following way:

Define

$$
\begin{aligned}
g((v,0_j)(c,0_k)) &= f^{\emptyset}(v,c)\ \forall(v,c)\in V\times C,\ \forall j,k\in\{1,\dots,\bar{k}\}\\
g((v,k)(c,0_k)) &= f^{b*}(v,c)\ \forall(v,c)\in V\times C,\ \forall k\in\{1,\dots,\bar{k}\}\\
g((v,k)(c,0_j)) &= f^{s}(v,c)\ \forall(v,c)\in V\times C,\ \forall j,k\in\{1,\dots,\bar{k}\}\text{ with } j<k\\
g((v,k)(c,0_j)) &= f^{s}(v,c)\ \forall(v,c)\in V\times C,\ \forall j,k\in\{1,\dots,\bar{k}\}\text{ with } j>k\\
g((v,0_k)(c,k)) &= f^{s*}(v,c)\\
g((v,0_j)(c,k)) &= f^{b}(v,c)\ \forall(v,c)\in V\times C,\ \forall j,k\in\{1,\dots,\bar{k}\}\text{ with } j<k\\
g((v,0_j)(c,k)) &= f^{s}(v,c)\ \forall(v,c)\in V\times C,\ \forall j,k\in\{1,\dots,\bar{k}\}\text{ with } j>k\\
g((v,k)(c,k)) &= f^{m}(v,c)\ \forall(v,c)\in V\times C\\
g((v,k)(c,j)) &= f^{b}(v,c)\ \forall(v,c)\in V\times C\ \forall j,k\in\{1,\dots,\bar{k}\}\text{ with } j<k\\
g((v,j)(c,k)) &= f^{s}(v,c)\ \forall(v,c)\in V\times C\ \forall j,k\in\{1,\dots,\bar{k}\}\text{ with } j>k
\end{aligned}
$$

For example, if $\bar{k}=3$, the mechanism would take the following shape:

	$C\times\{0_1\}$	$C\times\{0_2\}$	$C\times\{0_3\}$	$C\times\{1\}$	$C\times\{2\}$	$C\times\{3\}$
$V\times\{0_1\}$	$f^{\emptyset}$	$f^{\emptyset}$	$f^{\emptyset}$	f^{s*}	f^{b}	f^{b}
$V\times\{0_2\}$	$f^{\emptyset}$	$f^{\emptyset}$	$f^{\emptyset}$	f^{s}	f^{s*}	f^{b}
$V\times\{0_3\}$	$f^{\emptyset}$	$f^{\emptyset}$	$f^{\emptyset}$	f^{s}	f^{s}	f^{s*}
$V\times\{1\}$	f^{b*}	f^{b}	f^{b}	f^{m}	f^{s}	f^{s}
$V\times\{2\}$	f^{s}	f^{b*}	f^{b}	f^{b}	f^{m}	f^{s}
$V\times\{3\}$	f^{s}	f^{s}	f^{b*}	f^{b}	f^{b}	f^{m}

We need to check that l defined by $l_b(v,k)=(v,k)$ for all $k\geq 1$ and $l_s(c,k)=(c,k)$ for all $k\geq 1$ forms a level-k solution. We do so by induction on the statement: If $l_b(v,k-1)=(v,k-1)$ and $l_s(c,k-1)=(c,k-1)$, then $l_b(v,k)=(v,k)$ and $l_s(c,k)=(c,k)$ for all $k\in\{2,\dots,\bar{k}\}$

First we start by check the incentives for a buyer with value v and level $k\in\{2,\dots,\bar{k}\}$ and assume $l_b(v,k-1)=(v,k-1)$. From the proof of Lemma 1 we know that

$$u_b(f^b((v,c)),v)\ \geq\ u_b(f^b(v',c),v)\ \forall v'\in V,\ \forall(v,c)\in V\times C$$

So, we simply need to show the following three conditions:

(1) $u^b(f^b((c,v)),v) \geq u^b(f^s(v',c),v)\ \forall v' \in V,\ \forall (v,c) \in V \times C$

(2) $u^b(f^b((c,v)),v) \geq u^b(f^m(v',c),v)\ \forall v' \in V,\ \forall (v,c) \in V \times C$

(3) $u^b(f^b((c,v)),v) \geq u^b(f^{s*}(v',c),v)\ \forall v' \in V,\ \forall (v,c) \in V \times C$

For condition (1) there are two cases to consider:

(i) $v < c$: In this case utility of buyer is 0 when reporting v under f^b and under f^s reporting value v' gives utility 0 (if $v' < c$) or achieves trade (if $v' \geq c$) but at a price $v' \geq c$ giving utility of $v - v' < 0$.

(ii)$c \leq v$: In this case utility of buyer is $v - c \geq 0$ when reporting v under $f^b$and under f^s reporting value v' either gives a utility of 0 (if $v' < c$), or a utility of $v - v' \leq v - c$ (if $c \leq v'$)

For condition (2) there are two cases to consider:

(i) $v < c$: In this case utility of buyer is 0 when reporting v under f^b and under f^m reporting any other value v' either gives a utility of 0 (if $v' < c$) or achieves trade (if $v' \geq c$) at a price $(v' + c)/2 > v$ giving utility of $v - (v' + c)/2 < 0$.

(ii)$c \leq v$: In this case utility of buyer is $v - c \geq 0$ when reporting v under $f^b$and under f^sreporting any value v' either gives a utility of 0 (if $v' < c$) or a utility of $v - (v' + c)/2 \leq v - c$ (if $v' \geq c$)

For condition (3) there are two cases to consider:

(i) $v < c$: In this case utility of buyer is 0 when reporting v under f^b and under f^{s*} reporting any other value v_m either gives a utility of 0 (if $v_m < c$) or achieves trade (if $v_m \geq c$) at a price $v_m + (\bar{k} - \frac{1}{2})md > v$ giving utility of $v - v_m - (\bar{k} - \frac{1}{2})md < 0$.

(ii)$c \leq v$: In this case utility of buyer is $v - c \geq 0$ when reporting v under f^b and under f^{s*} reporting any value v_m either gives a utility of 0 (if $v_m < c$) or a utility of $v - v_m - (\bar{k} - \frac{1}{2})md \leq v - c$ (if $v_m \geq c$)

Therefore, for any buyer with type (v, k) where $k > 1$, playing (v, k) is a best response to $(c, k - 1)$ for any $c \in C$. Therefore, $l_b(v, k) = (v, k)$ is consistent with the requirements of a level-k solution for all $(v, k) \in V \times \{2, \ldots, \bar{k}\}$. The same holds for the seller by an analogous argument. Thus, the result follows by induction.

Now, it must be shown that $l_b(v, 1) = (v, 1)$ is consistent with the requirements of a level-k solution for all $v \in V$. That is, it must be shown that for a buyer with a value of

v playing against a seller with a cost c, playing $(v, 1)$ is a best response to the strategy: $l_{s0}(m) = l_s((c, 0))(m) = \frac{1}{|M_s|} = \frac{1}{2\bar{k}|C|}$ for all $m \in M_s$.

If type $(v_n, 1)$ reports (v_m, k), $k \geq 1$, her payoff is

$$\begin{aligned} u_b(((v_m, k), l_{s0}), v_n) &= \frac{1}{2\bar{k} \cdot |C|} \sum_{c \leq v_m} \left[(v_n - c + b_c) + (\bar{k} - 1)(v_n - c) + (\bar{k} - 1)(v_n - v_m) + \left(v_n - \frac{1}{2}(v_m + c) \right) \right] \\ &= \frac{1}{2\bar{k} \cdot |C|} \sum_{c \leq v_m} \left[b_c + (\bar{k} + \frac{1}{2})(v_n - c) + (\bar{k} - \frac{1}{2})(v_n - v_m) \right] \end{aligned}$$

where $b_{c=c_j} = (\bar{k} - \frac{1}{2}) jd$.

Let $h(v_m, k) = 2\bar{k} \cdot |C| \left(u_b(((v_n, 1), l_{s0}), v_n) - u_b(((v_m, k), l_{s0}), v_n) \right)$.

Reporting $(v_n, 1)$ is better than over-reporting (v_m, k) $(m > n)$, $k \geq 1$:

$$\begin{aligned} h(v_m, k) &= \sum_{c \leq v_n} \left[b_c + (\bar{k} + \frac{1}{2})(v_n - c) \right] - \sum_{c \leq v_m} \left[b_c + (\bar{k} + \frac{1}{2})(v_n - c) + (\bar{k} - \frac{1}{2})(v_n - v_m) \right] \\ &= -\sum_{c \leq v_m} \left[(\bar{k} - \frac{1}{2})(v_n - v_m) \right] - \sum_{v_n < c \leq v_m} \left[b_c + (\bar{k} + \frac{1}{2})(v_n - c) \right] \\ &\geq (m+1)(\bar{k} - \frac{1}{2})(m - n)d - \left[\sum_{0=j \leq m} j - \sum_{0=j \leq n} j \right] (\bar{k} - \frac{1}{2})d \\ &= (m+1)(\bar{k} - \frac{1}{2})(m - n)d - \frac{1}{2} \left[(m+1)m - (n+1)n \right] (\bar{k} - \frac{1}{2})d \\ &= (\bar{k} - \frac{1}{2})d \left[(n + l + 1)l - \frac{1}{2} \left((n + l + 1)(n + l) - (n+1)n \right) \right], \text{ since } m = n + l \text{ for some } l \geq 1 \\ &= (\bar{k} - \frac{1}{2})d \left[nl + l^2 + l - \frac{1}{2}n^2 - \frac{1}{2}nl - \frac{1}{2}nl - \frac{1}{2}l^2 - \frac{1}{2}n - \frac{1}{2}l + \frac{1}{2}n^2 + \frac{1}{2}n \right] \\ &= \frac{1}{2}(\bar{k} - \frac{1}{2})dl \left[l + 1 \right] \\ &\geq 0 \end{aligned}$$

Reporting $(v_n, 1)$ is better than under-reporting (v_m, k) $(m \leq n)$, $k \geq 1$:

$$
\begin{aligned}
h(v_m,k) &= \sum_{c\le v_n}\left[b_c+(\bar{k}+\frac{1}{2})\,(v_n-c)\right]-\sum_{c\le v_m}\left[b_c+(\bar{k}+\frac{1}{2})\,(v_n-c)+(\bar{k}-\frac{1}{2})\,(v_n-v_m)\right]\\
&= \sum_{v_m<c\le v_n}\left[b_c+(\bar{k}+\frac{1}{2})\,(v_n-c)\right]-\sum_{c\le v_m}(\bar{k}-\frac{1}{2})\,(v_n-v_m)\\
&\ge \sum_{0=i\le n}\left[i(\bar{k}-\frac{1}{2})d\right]-\sum_{0=j\le m}\left[i(\bar{k}-\frac{1}{2})d\right]-(m+1)(n-m)(\bar{k}-\frac{1}{2})d\\
&= \frac{1}{2}\left[n(n+1)-m(m+1)\right](\bar{k}-\frac{1}{2})d-(m+1)(n-m)(\bar{k}-\frac{1}{2})d\\
&= \frac{1}{2}\left[n(n+1)-(n-l)(n-l+1)\right](\bar{k}-\frac{1}{2})d-(n-l+1)l(\bar{k}-\frac{1}{2})d \text{ since } m=n-l \text{ for some } l\ge 1\\
&= (\bar{k}-\frac{1}{2})d\left[\frac{1}{2}n^2+\frac{1}{2}n-\frac{1}{2}n^2+\frac{1}{2}nl-\frac{1}{2}n+\frac{1}{2}nl-\frac{1}{2}l^2+\frac{1}{2}l-nl+l^2-l\right]\\
&\quad (\bar{k}-\frac{1}{2})d\left[\frac{1}{2}l^2-\frac{1}{2}l\right]\\
&= \frac{1}{2}(\bar{k}-\frac{1}{2})dl\,[l-1]\\
&\ge 0
\end{aligned}
$$

It is left to show that reporting $(v_n, 1)$ is better than reporting $(v_m, 0_k)$ for all $v_m \in V$ and $k \in \{1, \dots, \bar{k}\}$.

There is no incentive to over-report $(v_m, 0_k)$ $(m > n)$.

Reporting $(v_n, 1)$ is better than under-reporting $(v_m, 0_k)$ $(m \le n)$:

$$
\begin{aligned}
h(v_m,0_k) &= \sum_{c\le v_n}\left[b_c+(\bar{k}+\frac{1}{2})\,(v_n-c)\right]\\
&\quad -\sum_{c\le v_m}\left[k(v_n-v_m)+(\bar{k}-k)(v_n-c)-s_{v_m}\right]\\
&= \sum_{c\le v_n}\left[b_c+(\bar{k}+\frac{1}{2})\,(v_n-c)\right]+\sum_{c\le v_m}s_{v_m}-\sum_{c\le v_m}\left[k(v_n-v_m)+(\bar{k}-k)(v_n-c)\right]\\
&\ge \frac{1}{2}n(n+1)(\bar{k}-\frac{1}{2})d+\frac{1}{2}n(n+1)(\bar{k}+\frac{1}{2})d+m(m+1)(\bar{k}-\frac{1}{2})d\\
&\quad -(m+1)(n-m)kd-(m+1)(n+1)(\bar{k}-k)d+\frac{1}{2}m(m+1)(\bar{k}-k)d\\
&= \frac{1}{2}n(n+1)\bar{k}d-\frac{1}{4}n(n+1)d+\frac{1}{2}n(n+1)\bar{k}d+\frac{1}{4}n(n+1)d+m(m+1)\bar{k}d-\frac{1}{2}m(m+1)d\\
&\quad -(m+1)(n-m)kd-(m+1)(n+1)\bar{k}d+(m+1)(n+1)kd+\frac{1}{2}m(m+1)\bar{k}d-\frac{1}{2}m(m+1)kd\\
&= \left[n(n+1)-n(m+1)+\frac{3}{2}m(m+1)-(m+1)\right]\bar{k}d\\
&\quad +\frac{1}{2}m(m+1)kd-\frac{1}{2}m(m+1)d+(m+1)kd\\
&\ge 0
\end{aligned}
$$

Budget balance is imposed by the our definition of the mechanism. Ex post individual rationality is guaranteed by our choice of f^b, f^s, and f^m. All agent's are ensured a weakly positive payoff under any outcome. Negative payoffs are possible under f^{b*} and f^{s*} but no types (with $k > 0$) ever play actions that lead to f^{b*} and f^{s*}. Lastly, notice that the level-k solution l achieves F^*. This is easy to see because f^{b*}, f^{s*}, f^b, f^s, and f^m all ensure trade when $v \geq c$ and no trade when $v < c$ and all level-k types (with the exception when $k = 0$) truthfully report their payoff types.

□

6 Robustness of level-k mechanisms

6.1 Relaxing beliefs about payoffs

This subsection shows that we can design a level-k mechanism that is robust to relaxing the assumption that beliefs about payoffs are determined by a specific common prior. Proposition 3 shows that the level-k mechanism that implements ex post efficient choice correspondence is robust to different beliefs about payoff types. It is possible to design a mechanism such that each payoff type wants to truthfully report her payoff type regardless of her beliefs about the payoff types of others.

Proposition 3. *There exists a mechanism $\langle M, f\rangle$ that level-k implements the ex post efficient correspondence, F^*, for any $\mathcal{B}$-based level-k type space $\mathcal{L} = \langle \mathcal{B}, \bar{k}\rangle$.*

PROOF

Use the same mechanism as in proof of Proposition 2. The behavior of any level 1 type is unaffected because level 1 beliefs about opponent's behavior (that level 0 types randomize across action set) - is independent of beliefs about payoff types. Behavior of all higher types is also unchanged as a buyer with a type (v, k) would want to truthfully report even if he knew the cost of the seller holds for any realization of (v, c) and hence holds for any beliefs ρ (shown in proof of Proposition 2).

□

There exists a single mechanism that level-k implements the social choice correspondence regardless of the common prior. In other words, there is a mechanism that implements the ex post efficient social choice correspondence under level-k implementation for any $\mathcal{B}$-based level-k type space $\mathcal{C} = \langle \mathcal{B}, \bar{k}\rangle$, where $\mathcal{B} = \langle \Theta_1, \ldots, \Theta_n, \rho\rangle$. We refer to this type of implementation, that does not depend on the underlying common prior assumption, as **payoff robust level-k** implementation.

The example in Section 2 illustrated an example which was payoff robust level-k implementable but not Bayesian implementable.

6.2 Relaxing beliefs about levels

The analysis up to this point has been based on the level-k model. This is a type of limited depth of reasoning model where agents have particular beliefs about the levels of others. Specifically, if an agent is of level k, he believes that others have levels exactly equal to $k-1$. In general, we might allow a agent with level k to hold beliefs over all lower levels.[11] The following definition of limited depth of reasoning models generalize the level-k type space in exactly that way.

Definition. A **(Bayesian based) limited depth of reasoning type space (LDoR type space)** is a type space $\mathcal{L}^{LDoR} = \langle \mathcal{B}, \bar{k}, \lambda^1, \ldots, \lambda^{\bar{k}} \rangle$ with $\lambda^k \in \triangle(\{0, \ldots, k-1\})$.

As in the level-k type space, an agent's beliefs about the cognitive types of others are determined both by her payoff type and her level. The beliefs of a type $t_i = (\theta_i, k_i)$ about the types of others, $c_{-i} = (\theta_{-i}, \boldsymbol{v}_{-i})$, is determined by the function $b_i(c_{-i}|c_i)$:

$$b_i(c_{-i}|c_i) = \left(\lambda^k_{v_1} \times \cdots \times \lambda^k_{v_{I-1}}\right) p(\theta_{-i}|\theta_i)$$

where the notation $\theta \times \boldsymbol{v}$ represents a payoff type-level profile $(\theta_1, v_1) \times \cdots \times (\theta_I, v_I)$, where $\boldsymbol{v} \in \{0, \ldots, \bar{k}\}^I$.

A agent with a level k puts weight only on other types that have a level less than or equal to $k-1$. This captures the core assumption of the limited depth of reasoning literature which is that a agent of level k believes that other agents have levels strictly less than k. This assumption ensures that agents can calculate their optimal actions in a recursive fashion with a finite number of steps given the behavior of level 0 types. As in the level-k model, a agent's beliefs about the payoff types of others are determined by the common prior ρ.

Given the definition of an LDoR type space, we can apply the level-k solution concept and level-k implementation to this generalized type space, by simply using the belief function $b_i(c_{-i}|c_i)$ defined for the LDoR type space in the definition of a level-k solution. We refer to this as LDoR implementation.

[11] Cognitive hierarchy models relax the level-k belief structure in this way. 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.

Given that the LDoR type space generalizes level-k type spaces it could be more difficult to implement a particular social choice correspondence under this generalization. But, this is actually not the case. In the example from Section 2, the social choice correspondence can be implemented under LDoR. Further, this can be done in a robust way. There exists a single mechanism that can implement the social choice correspondence under *any* LDoR type space.

This can be done more generally in the bilateral trade environment. This result is formalized in the following proposition.

Proposition 4. *There exists a mechanism that LDoR implements the ex post efficient social choice correspondence, F^*, in any ($\mathcal{B}$-based) LDoR type space $\mathcal{C}^{LDoR} = \langle \mathcal{B}, \bar{k}, \lambda^1, \ldots, \lambda^{\bar{k}} \rangle$.*

PROOF

Use the same mechanism as in proof of Proposition 2.

The proof follows by induction on the statement: If types (v, i) and (c, j) report (v, i) and (c, j) respectively for all $(v, i) \in V \times \{1, \ldots, k-1\}$ and $(c, j) \in C \times \{1, \ldots, k-1\}$, then type (v, k) type will report (v, k) for all $v \in V$ and (c, k) type will report (c, k) for all $C \in C$.

The behavior of any level 1 type is unchanged, thus the result holds for $k = 1$.

Now assume the (c, j) type reports (c, j) for all $C \in C$ and $j \in \{1, \ldots, k-1\}$.

Consider the behavior of a buyer with type (v_n, k). Suppose she is aware that she is playing against a seller with a cost c. Then, she believes the seller is playing some mixture, σ, over $\{(c, 1), (c, 2), \ldots, (c, k-1)\}$ with probability $1 - \lambda_0^k$ and the strategy l_{0s} with probability λ_0^k.

Suppose $\lambda_0^k = 1$. It is easy to see that the strategy (v_n, k) gives the same payoff as $(v_n, 1)$ and since $(v_n, 1)$ is a best response to l_{0s}, (v_n, k) must be as well.

Now, suppose that $\lambda_0^k = 0$.

Reporting (v_n, k) gives utility

$$
\begin{aligned}
u_b((v_n, k), \sigma), v_n) &= (v_n - c) \\
&\geq \mu_1(v_n - c) + \frac{1}{2}\mu_2(v_n - c) + \frac{1}{2}\mu_2(v_n - v_m) + \mu_3(v_n - v_m) + (1 - \mu_1 - \mu_2 - \mu_3)(v_n - v_m) \\
&\quad -(1 - \mu_1 - \mu_2 - \mu_3)(\bar{k} - \frac{1}{2})md \text{ for any } \mu_1, \mu_2, \mu_3 \in [0, 1], \text{ with } \mu_1 + \mu_2 + \mu_3 = 1,\ v_m \geq c \\
&= u_b((v_m, k'), \sigma), v_n) \text{ for } v_m \geq c,\ k' \geq 0 \text{ and some } \mu_1, \mu_2, \mu_3 \in [0, 1] \text{ with } \mu_1 + \mu_2 + \mu_3 = 1 \\
&\geq u_b((v_m, k'), \sigma), v_n) \text{ for } v_m < c,\ k' \geq 0
\end{aligned}
$$

Therefore, reporting (v_n, k) is the best response when both $\lambda_0^k = 0$ and when $\lambda_0^k = 1$, thus it must be the best response when $\lambda_0^k \in [0, 1]$.

Therefore, $l_b(v, k) = (v, k)$ is a level-k solution. The analogous argument can be made to show that the seller truthfully reports his type, i.e. $l_s(c, k) = (c, k)$

□

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.

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.

Cabrales, A. & Serrano, R. (2011). Implementation in adaptive better-response dynamics: Towards a general theory of bounded rationality in mechanisms. *Games and Economic Behavior*, 73, 360–374.

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

Cooper, R. W., Dejong, D. V., Forsythe, R., & Ross, T. W. (1990). Selection Criteria in Coordination Games: Some Experimental Results. *American Economic Review*, 80, 218–233.

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. (2016). Efficient mechanisms for level-k bilateral trading.

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

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

de Clippel, G., Saran, R., & Serrano, R. (2016). Level-k mechanism design.

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.

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

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

Severinov, S. & Deneckere, R. (2006). Screening when some agents are nonstrategic: 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.

Discussion Papers of the Research Area Markets and Choice 2017

Research Unit: **Market Behavior**

Dorothea Kübler, Julia Schmid, Robert Stüber SP II 2017-201
Be a man or become a nurse: Comparing gender discrimination by employers across a wide variety of professions

Research Unit: **Economics of Change**

Jannis Engel, Nora Szech SP II 2017-301
The political economy of multilateral aid funds

Maja Adena, Jeyhun Alizade, Frauke Bohner, Julian Harke, Fabio Mesners SP II 2017-302
Quality certifications for nonprofits, charitable giving, and donor's trust: experimental evidence

Terri Kneeland SP II 2017-303
Mechanism design with level-k types: Theory and an application to bilateral trade

All discussion papers are downloadable:
http://www.wzb.eu/en/publications/discussion-papers/markets-and-choice