Discussion Papers
Department of Economics
University of Copenhagen

No. 10-22

Stepwise Thinking in Strategic Games with Incomplete Information

Carsten S. Nielsen

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (E)

Stepwise Thinking in Strategic Games with Incomplete Information*

Carsten S. Nielsen[†]

First Draft: July 7, 2010
This Version: September 30, 2010

Abstract

This paper proposes a general incomplete information framework for studying behavior in strategic games with stepwise (*viz.* 'level-k' or 'cognitive hierarchy') thinking, which has been found to describe strategic behavior well in experiments involving players' initial responses to games. It is shown that there exist coherent stepwise beliefs, implied by step types, that have the potential to encode all relevant information. In the structure of stepwise beliefs, players are unaware of opponents doing at least as much thinking as themselves. As a result, there exists a Bayesian Nash equilibrium strategy profile in which any player at some step fixes the best responses of opponents at lower steps and then best responds herself.

JEL-Classifications: C70, C72, D80, D82.

Keywords: Game theory, interactive epistemology, unawareness, Bayesian Nash equilibrium, bounded rationality, level-k, cognitive hierarchy.

---

[*] I would like to thank Nagore Iriberri, Alexander Sebald, Peter Norman Sørensen, and seminar participants at the ESA World Meeting 2010 (Copenhagen), the QED Jamboree 2010 (Alicante), and Copenhagen for helpful comments and suggestions. All errors are mine.

[†] Department of Economics, University of Copenhagen, Øster Farimagsgade 5, Building 26, DK-1353, Copenhagen K, Denmark. Phone: (+45) 3532-3051. Fax: (+45) 3532-3064. E-mail: carsten.nielsen@econ.ku.dk. Web: http://www.econ.ku.dk/phdstudent/nielsen.

1 Introduction

The emergence of strategic (or non-cooperative) games has had a profound effect on economic theory. In these games it is, in general, assumed that when players strategize they correctly predict opponents' behavior and, by rationality, their actions in equilibrium must be best responses to opponents' actions. Despite the apparent success of strategic games, a large body of experimental research has come to question players' ability to make such accurate predictions in 'initial response games' (*viz.* games without learning, feedback or clear precedence).[1] One idea that has emerged from this research is that when players strategize, they think in steps: some players are nonstrategic and do not best respond (step 0), while others believe that their opponents are nonstrategic when they best respond (step 1), and yet others believe that they play against some distribution of opponents thinking in step 0 and step 1 when they best respond (step 2), and so on. The specification of steps of thinking typically assumes that step k thinkers know opponents think in $k-1$ steps (*viz.* level-k models), or believe that opponents think in $k-1$ or fewer steps with a frequency described by the Poisson distribution (*viz.* cognitive hierarchy models). The population of step thinkers in initial response games tend to be stable with most weight on step 1 and 2–regardless of the specification.

This paper proposes a general incomplete information framework for studying strategic games in which players, who think in steps, might have different information about the payoff relevant parameters. In doing so we take on a view of irrationality that is somehow different from what is conventional.[2] Here nonstrategic players are assumed to have well-defined beliefs, but fail to payoff maximize because they are unaware of the payoff relevant parameters. While strategic players are aware of the payoff relevant parameters, but are in general unaware of others doing at least as much thinking as themselves. Such a description does not concur with the standard belief system, as proposed by Harsanyi (1967–68) and constructed by Mertens and Zamir (1985), since it assumes that all players have beliefs about the payoff relevant parameters Θ, beliefs about opponents' beliefs about Θ, beliefs about opponents' beliefs about their beliefs about Θ, and so on *ad infinitum*.[3] If all players

---

[1] The experimental literature was initiated by Stahl and Wilson (1994, 1995) and Nagel (1995), and further developed and applied by Ho *et al.* (1998), Costa-Gomes *et al.* (2001), Bosch-Domenech *et al.* (2002), Camerer *et al.* (2004), Costa-Gomes and Crawford (2006), Crawford and Iriberri (2007b) and Crawford and Iriberri (2007a).

[2] Irrationality has typically been modeled as players' inability to payoff maximize (Aumann, 1992, 1997). It is assumed that players have well-defined probabilities over opponents' actions, yet it is permitted that players sometimes fail to maximize these payoffs. However, this raises a difficulty since subjective probabilities are usually defined via payoff maximization (Savage, 1954); non-maximizers of payoffs do not have probabilities.

[3] See also Brandenburger and Dekel (1993), Heifetz (1993), Mertens *et al.* (1994), and Heifetz and Samet

form such beliefs in the infinite, then the belief system is by construction commonly known. This observation motivates a reconsideration of how players thinking in steps form beliefs about each other.

The explicit description of the stepwise belief system begins with step 0 in which players are nonstrategic and unaware of the payoff relevant parameters, such that beliefs are defined on the empty set. The next step of thinking (step 1) is a bit more sophisticated, here beliefs are about Θ and opponents' nonstrategic beliefs. In step 2 players form beliefs about Θ, about opponents possible nonstrategic beliefs, and opponents' beliefs about Θ and their opponents nonstrategic beliefs, and so on. Explicitly describing 'epistemic' step types as interactive belief systems ensures that beliefs are constructed solely in terms Θ. Such a specification tells us what all stepwise belief configurations should look like. However, the disadvantage is that the entangled web of increasing steps of beliefs makes practical applications increasingly complex.

If we assume that players' stepwise belief system at each step can be summarize by a single entity, their step type, then the modeling becomes more 'manageable'. However, by introducing such step types it seems that a second 'level' of beliefs is required, wherein each player has beliefs over opponents' step types, over opponents' beliefs over their step type, and so on. Such beliefs about opponents' step types obtains naturally for any finite step type. The notion of a 'omniscient' player thinking in all stepwise belief configurations is however in this case not well define. Only considering finite step types may therefore be restrictive in the following sense: by modeling a specific strategic game with incomplete information using only finite step types we may miss some step types that are not present in the beliefs of a finite stepwise thinker, and can be found only in some higher step. If this is true for any finite step type, then the concept of step types is necessarily restrictive.

Our construction of infinite step types has two stages. First, it is shown that each coherent step type–compromising an explicit description–defines, in a natural way, a probability measure over the set of opponents' coherent lower step types, which can be extended to a unique probability measure associated with the infinite step type. (Coherency requires that a players' beliefs at different steps do not contradict each other–see Definition 1). This result (Proposition 1) is obtained in the broadest and most natural setup, that of probability (or measure) theory. Second, the model of stepwise beliefs is closed at each step by imposing, via a simple inductive definition, the requirement that each step type knows (belief with probability one) that opponents' lower step types are coherent, that each of these step types

---

(1998).

know that each of their opponents' lower step types are coherent, and so on. That is, the model is closed at each step by imposing stepwise mutual knowledge of coherency. Common knowledge is reached in the unrestricted situation where all players think in infinite steps. Related to this observation is Strzalecki (2009) who shows how stepwise thinking, when players put less weight on the types immediately below them as they think in more steps, can mitigate the discontinuity of predictions made by solution concepts when mutual knowledge comes close to common knowledge.[4]

To formalize the idea that there is noting intrinsically restrictive about the structure of stepwise thinking, the existence of a 'universal' step type that contains all step types is proven in Proposition 2. The existence of a universal step type guarantees that in principle any strategic game with incomplete information can be modeled using step types without any loss of generality. In most applications it is however convenient to use a smaller step type spaces than that implied by the universal step type space. A smaller set of step types which defines the strategic situation without loss of generality is therefore needed. For this purpose belief closed subsets are defined–see Definition 2. Belief closed subsets imply that when we choose only to consider a subset of players' step types, we also implicitly assume that all beliefs that are relevant for each player in a given strategic situation are included.

Having completed the construction of step types in Section 2, the idea of cognitive limitations is introduced. The reason for considering cognitive limitations is often justified by arguing that the brain has limits, and that it does not understand its own limitations.[5] Such an assumption imply that players are unaware of opponents doing at least as much thinking. However, conventional models of level-k and cognitive hierarchy thinking represent players' cognitive limitations by assuming that they believe that the event that opponents think in at least as many steps occurs with probability zero. Such an assumption lacks transparency: if players assigns probability zero to an event, then it is not clear whether they do so because they are unaware of the event, or because they are aware of the event but assigns probability zero to it occurring. The latter is not compatible with the common notion of cognitive limitations. The point made in Section 3 is that a nontrivial notion of unawareness (Dekel *et al.*, 1998; Modica and Rustichini, 1999), which does not imply modeling unawareness as zero probability events, naturally obtains in the definition of step types–see Proposition 3.[6]

---

[4] The quintessential illustration of such discontinuity is the email game of Rubinstein (1989), see Monderer and Samet (1989) and Dekel *et al.* (2006) for treatments developed in response to the email.

[5] Cognitive limitations have been studied (and found) in strategic games using selfreports, tests of memory, response times, measures of eye gaze and attention (Camerer *et al.*, 1994; Costa-Gomes *et al.*, 2001), and even brain imaging (Camerer *et al.*, 2005).

[6] Many different ways of modeling unawareness have been suggested, see for example, Fagin and Halpern

Given the characterization of the stepwise belief system, a Bayesian Nash equilibrium for games with stepwise thinking is proposed and its existence is proven for any finite game. From the definition of step types in the previous sections it may not be clear how one can analyze stepwise thinking using the standard tools of game theory. Section 4 starts by clarifying this relation. Thereafter a game with stepwise thinking is defined as a finite sequence of step $k \geq 1$ games, each describing a strategic situation at different steps of thinking. The step 0 game in which only nonstrategic players play against each other is omitted for obvious reasons. However, players thinking in more steps still take the nonstrategic players into account, since the actions of the nonstrategic players influence their expected payoffs. Because players are unaware of any situation which involves opponents doing at least as much thinking as themselves, they believe that the step k game they are confined to is the 'true' game, and does not change their beliefs in step games that demand more thinking than they are capable of. This implies that there exists a Bayesian Nash equilibrium strategy in the game with stepwise thinking in which players in any step k game fix the equilibrium strategies of opponents, who they belief do less thinking and thus are confined to some step $l < k$ game, and then choose their equilibrium based on this belief (Proposition 4 and 5). This observation suggest a procedure for constructing a Bayesian Nash equilibrium in a game with stepwise thinking; first we have to find an equilibrium in the step 0 game and then extend it step-by-step to 'higher' step games by fixing the equilibrium strategies of opponents in the respective 'lower' step games.

Finally, it it worth noticing that there exists a complementary literature which assumes that players can comprehend infinite hierarchies of beliefs (implicitly given by their types), but make systematic mistakes in equilibrium conjectures (first-order beliefs). For example, Eyster and Rabin (2005) propose a 'cursed' equilibrium where players have correct beliefs about the joint distribution of types, and also have correct beliefs about the aggregated distribution of opponents' play, conditional on each of their own types. However, instead of playing best response to the actual opponents' actions, each player chooses the best response to a convex combination of the actual actions and the aggregate distribution. Jehiel and Koessler (2008), building on Jehiel (2005), consider the 'analogy-based' equilibrium in which players group opponents' actions into analogy classes, with the player believing that actions in a given class are identical. Given this, the player's beliefs must correspond to the aggregate distribution of play across actions in an analogy class. However, in these 'behavioral' equilibrium models it is not obvious why it is fair to assume that players are

---

(1988), Modica and Rustichini (1999), Halpern (2001) and Heifetz *et al.* (2006). The relation between these models are understood from Halpern and Rêgo (2008) and Heifetz *et al.* (2008). The notion of unawareness used in this paper is closest to that of Heifetz *et al.* (2009).

sophisticated enough to consider infinite hierarchies of beliefs and at the same time fail to reason about first-order equilibrium beliefs.

The plan of the paper is as follows. Step types structures are constructed in Section 2. Section 3 shows why cognitive limitations induces a non-trivial notion of unawareness. Bayesian games with stepwise thinking and the associated Bayesian Nash equilibrium are developed in Section 4. Concluding remarks are made in Section 5, and proofs that does not appear in the paper can be found in Appendix A.

2 Construction of step types

In this section step k types are constructed. We begin with some mathematical preliminaries (2.1). Then the notion of a stepwise beliefs, coherency and mutual coherency are given (2.2-2.3), it is shown that the infinite step type is 'universal', and belief closed subspaces are defined (2.4).

2.1 Preliminaries

For a given topology space X and associated Borel sigma-algebra $\mathscr{B}_X$, let $\Delta(X)$ be the set of Borel probability measures $\mu : \mathscr{B}_X \to [0,1]$ on $(X, \mathscr{B}_X)$. A class $\mathscr{B}_X$ of subsets of X is a Borel sigma-algebra if it contains X itself and is closed under the formation of complements and countable unions. An element $\mu \in \Delta(X)$ satisfies $\mu(\emptyset) = 0$, $\mu(X) = 1$, $\mu(A) \in [0,1]$ for $A \in \mathscr{B}_X$, and if $A^1, A^2, \ldots$ is a countable disjoint sequence of sets in $\mathscr{B}_X$ then $\mu(\bigcup_{\alpha=1}^{\infty} A^\alpha) = \sum_{\alpha=1}^{\infty} A^\alpha$. The triplet $(X, \mathscr{B}_X, \mu)$ is a probability space. the support $\text{supp}(\mu) = \{x \in X : \mu(x) > 0\}$ is a set of points with positive probability, and the marginal of a measure μ on some set $A^\alpha \in \mathscr{B}_X$ is denoted $\text{marg}_{A^\alpha}\mu = \mu^\alpha$.

2.2 Stepwise beliefs

Let $N = \{1, \ldots, n\}$ players face a sequence $-i \in N \backslash \{i\}$ of opponents in a strategic game. In a n-player game of incomplete information the crucial elements governing strategic interaction–such as individual feasibility constraints, how actions are mapped into consequences and individual preferences over consequences–are represented by a vector of payoff relevant parameters θ which is (partially) unknown to some players. For the sake of simplicity, let us assume that θ in the finite set Θ determines the shape of each player's payoff function. The form of the parametric payoff functions $u_i(\cdot, \theta)$–or, more generally, the form of the mapping associating each conceivable parameter θ to the 'true' (but unknown) game $G(\theta)$–is assumed common knowledge. We may adopt a Bayesian approach by assuming that a player who has only partial knowledge about the payoff relevant parameters has some beliefs about the parameters which he does not know or he is uncertain about. However, unlike in a problem which involves a single decision maker, this is not enough in an interactive situation: as the decisions of other players are relevant, so are their beliefs, since they affect their decisions. Thus a player must have beliefs about the beliefs of other players. For the same reason, a player needs beliefs about the beliefs of other players about his beliefs and so on. In principle, a complete description of every relevant attribution of a player should include, not only her payoff relevant parameters, but also her epistemic type–that is, an infinite hierarchy of beliefs.

Stepwise beliefs are related to, but different from Harsanyi's beliefs. In particular, we assume that the nonsophisticated players are unaware of the payoff relevant parameters, while the sophisticated players have limited cognitive abilities in the sense that they do not believe that others think in at least as many steps as themselves. The iterated process of stepwise thinking begins with step 0 in which players are nonstrategic and beliefs are defined on the empty set. Players doing one or more steps of thinking are assumed strategic. In step 1 players think about the strategic situation Θ and opponents' beliefs in step 0. In step 2 players think about Θ and opponents' possible nonstrategic beliefs 0 and their beliefs about the payoff functions and their opponents' nonstrategic beliefs, and so on. Formally, define spaces

$$\begin{aligned} & X_i^0 = \emptyset; \\ & \text{for all } k \geq 1, \\ & X_i^k = \Theta \times [\cup_{l=0}^{k-1} \Delta(X_{-i}^l)]. \end{aligned} \tag{1}$$

An epistemic step k type t_i^k is just a belief in step k; $t_i^k = \mu \in \Delta(X_i^k)$. Let $T_{0,i}^k = \Delta(X_i^k)$ denote the set of all possible step k types of player i. Similar for $-i$. This characterization is explicit because it specifies the whole hierarchy of stepwise thinking.

By casual observation it seems that a second 'level' of step beliefs is required, wherein player i has beliefs over opponents' step types, over opponents' beliefs over her own step type, and so on. It can easily be verified that by imposing step types, Equation 1 becomes

$$
\begin{aligned}
&T_{0,i}^{0} = \Delta(\emptyset);\\
&\text{for all } k \geq 1,\\
&T_{0,i}^{k} = \Delta(\Theta \times [\cup_{l=0}^{k-1} T_{0,-i}^{l}]).
\end{aligned}
$$

This characterization of stepwise thinking introduce a step type space, which provides an implicit description of step types. Each point in the step type space is associated with a payoff relevant parameter, as well as opponents' belief at lower steps. A players' step type is thus an implicit description her beliefs about such points. This does however not necessarily mean that beliefs are well defined. The conditions under which the specification of beliefs is meaningful is defined next.

2.3 Coherent stepwise thinking

The step k types just constructed may not be meaningful. For example, if $t_i^1 \in T_{0,i}^1$, then for this to describe meaningful beliefs of player i (or $-i$), the marginal distribution of t_i^1 on X_i^0 must coincide with $t_i^0 \in \Delta(X_i^0)$. We therefore impose that the various step k types cannot contradict each other. In other words, different step k types should be coherent.

Definition 1. A step k type $t_i^k \in T_{0,i}^k$ is *coherent*[7] if for every $k \geq 1$

$$\text{marg}_{X_i^{k-1}} t_i^k = t_i^{k-1}.$$

Let $T_{1,i}^k$ denote the set of all coherent step k types belonging to player i. The following proposition shows that a coherent infinite step type exists and induces a belief over Θ and the space of all possible step types of opponents.

Proposition 1. For any $i \in N$ there exists an coherent infinite step type $t_i^\infty \in T_{1,i}^\infty$ which closes the hierarchy of stepwise thinking such that

$$T_{1,i}^{\infty} = \Delta(\Theta \times [\cup_{l=0}^{\infty} T_{0,-i}^{l}]).$$

Proof of Proposition 1 follows naturally from the following Lemma, which itself is essentially an adaptation of Kolmogorov's Extension Theorem due to Bochner (1960).

---

[7] What is here called coherency is usually refered to as (Kolmogorov) consistency. The term coherency is used to avoid confusion with Harsanyi's use of consistency, which means something different.

Lemma 1. Let $\{\mathscr{B}_{Z^k}\}_{k=0}^{\infty}$ be an increasing sequence of Borel sigma-algebras on $Z^0 \subset Z^1 \subset \cdots$. If each probability measure μ^k on $\mathscr{B}_{Z^k}$ is coherent such that

$$\text{marg}_{Z^0 \cup \cdots \cup Z^{k-1}} \mu^k = \mu^{k-1} \text{ for all } k \geq 1,$$

then there exists a unique extension of the sequence $\{\mu^0, \mu^1, \dots\}$ to μ on $\sigma(\cup_{k=0}^{\infty} \mathscr{B}_{Z^k})$ satisfying $\text{marg}_{Z^0 \cup \cdots \cup Z^k} \mu = \mu^k$ for all $k \geq 0$.

Proof. Let $Z = \cup_{k=0}^{\infty} Z^k$ and $\mathscr{B}_Z = \cup_{k=0}^{\infty} \mathscr{B}_{Z^k}$ be the corresponding collection of Borel sigma-algebras (note that $\emptyset, Z \in \mathscr{B}_Z$).

Define a measure μ on the field $\mathscr{B}_Z$ by $\mu(E) = \mu^k(E)$ for $E \in \mathscr{B}_{Z^k}$. Coherency guarantees that this is well defined. Now note that μ is nonnegative and $\mu(X^k) = \mu^k(X^k) = 1$ for all $k \geq 0$ (*viz.* μ is a probability measure on each $\mathscr{B}_{Z^k}$).

First we proof that μ is finitely additive. μ is finitely additive, for if a finite collection of sets belongs to $\mathscr{B}_Z$, then since $\{\mathscr{B}_{Z^k}\}$ is an increasing sequence of sigma-algebras, there is some k for which every member of the collection belongs to $\mathscr{B}_{Z^k}$. Consequently their union also belongs to $\mathscr{B}_{Z^k}$ and hence to $\mathscr{B}_Z$. The finitely additivity of μ is then guaranteed by that of each μ^k. This also proves that $\mathscr{B}_Z$ is a field (*viz.* algebraic structure).

Now observe that μ on $\mathscr{B}_Z$ is continuous from above because for any set $A^n \in \mathscr{B}_Z$ it is true that $A^n \downarrow \emptyset$, such that $\mu(A^n) \downarrow 0$ (Billingsley, 1995, Theorem 2.1).[8] If μ is a finitely additive probability measure on the field $\mathscr{B}_Z$, and if $A^n \downarrow \emptyset$ for sets $A^n \in \mathscr{B}_Z$ implies $\mu(A^n) \downarrow 0$, then μ is countably additive. To see this define $B^1 = A^1$ and $B^n = A^n \backslash A^{n-1}$. Then the B^n's are disjoint, $A^n = \cup_{k=0}^{n} B^k$, and $A^{\infty} = \cup_{k=0}^{\infty} B^k$. $Z = A^{\infty}$ since Z is the largest set in $\mathscr{B}_Z$. Indeed, if $Z = \cup_{k=0}^{\infty} B^k$ for disjoint sets B^k, then $C^n = \cup_{k>n}^{\infty} B^k = Z \backslash \cup_{k=0}^{n} B^k$ lies in the field $\mathscr{B}_Z$, and $C^n \downarrow \emptyset$. The hypothesis, together with finite additivity, gives $\mu(Z) - \sum_{k=0}^{n} \mu(B^k) = \mu(C^n) \downarrow 0$, and hence $\mu(Z) = \sum_{k=0}^{\infty} \mu(B^k)$. This also proves that $(Z, \mathscr{B}_Z, \mu)$ is a probability space.

We may now use Carathéodory's Lemma (Billingsley, 1995, Theorem 3.1) to extend the probability measure μ on the field $\mathscr{B}_Z$ uniquely to the generated Borel sigma-algebra $\sigma(\mathscr{B}_Z)$. ■

Proof of Proposition 1. In Lemma 1, set $Z^0 = \{\emptyset\}$ and $Z^k = \Theta \times [\cup_{l=0}^{k-1} T_{0,-i}^l]$ for $k \geq 1$, so $Z^0 \subset Z^1 \subset \cdots$, $Z^0 \cup \cdots \cup Z^k = X^k$ and $\cup_{k=0}^{\infty} Z^k = \Theta \times [\cup_{l=0}^{\infty} T_{0,-i}^l]$. A coherent infinite step type $t_i^{\infty} \in T_{1,i}^{\infty}$ is exactly $\mu \in \Delta(\cup_{k=0}^{\infty} Z^k)$. Lemma 1 thus implies that there exists a collection $T_{1,i}^{\infty} = \Delta(\Theta \times [\cup_{l=0}^{\infty} T_{0,-i}])$. ■

---

[8] By $A^n \downarrow \emptyset$ is meant $A^n \supset A^{n-1} \supset \cdots \supset A^0$ and $\cap_{n \geq 0} A^n = \emptyset$; $\mu(A^n) \downarrow 0$ means that $\mu(A^n) \geq \mu(A^{n-1}) \geq \cdots \geq \mu(A^0)$ and $\mu(A^n) \to 0$ (or $\lim_{n \to 0} \mu(A^n) = 0$).

Coherency implies that player i's step types determines i's beliefs over opponents' step types in a meaningful way. But player i's step type does not necessarily determine i's belief over opponents' beliefs over i's step types–in particular i might assign positive probability to opponents' possible step 0 to $k-1$ types being incoherent. For a step type to determine all step 0 to $k-1$ types of her opponents, mutual knowledge of coherency must be imposed. To do so, define a sequence of sets $\{T^k_{m,i}\}_{m\geq 2}$ by

$$T^k_{m,i} = \{t^k_i \in T^k_{1,i} : t^k_i(\Theta \times [\cup^{k-1}_{l=0} T^l_{m-1,-i}]) = 1\}.$$

Let the set of player i's mutual coherent step k types be given by $T^k_i = \cap^k_{m=1} T^k_{m,i}$, where T^k_i is the subset of $T^k_{1,i}$ obtained by requiring that i's step k type knows (belief with probability one) that opponents' step 0 to $k-1$ types are coherent, that i's step k type knows that opponents' step 0 to $k-1$ types knows that i's own step 0 to $k-2$ types are coherent, and so on. (Similar for $-i$.)

A question that often arises in the discussion of strategic situations is whether the information structure is common knowledge. The observation we wish to make is that the necessary rationality assumption made in strategic situations with stepwise thinking is not that of common knowledge, but rather the 'natural' assumption of stepwise mutual knowledge of coherency. To see this, consider, for example, the set $T^k_{2,i} = \{t^k_i \in T^k_{1,i} : t^k_i(\Theta \times [\cup^{k-1}_{l=0} T^l_{1,-i}]) = 1\}$ as defined above. The set $T^k_{2,i}$ is the set of step k types of player i that know that opponents' step 0 to $k-1$ types are coherent. So $T^k_{2,i}$ is the set of step k types of player i which can calculate beliefs about opponents' beliefs about i's own step 0 to $k-2$ types, that is the set of step k types that know of opponents' information structure. Similarly, $T^k_{3,i}$ is the set of step k types belonging to player i that can calculate beliefs about opponents' beliefs about i's beliefs over opponents' step 0 to $k-3$, that is, the set of step k types that know that opponents know of i's information structure, and so on in k inductive steps. The upshot is that the information structure relevant for a step k type is guaranteed by the assumption of mutual knowledge of coherency.

Corollary 1. For all $i \in N$ there exists a well defined hierarchy of stepwise thinking such that

$$\begin{aligned} &T^0_i = \Delta(\{\emptyset\}); \\ &\text{for all } k \geq 1, \\ &T^k_i = \Delta(\Theta \times [\cup^{k-1}_{l=0} T^l_{-i}]). \end{aligned}$$

Proof. From mutual knowledge of coherency we have that $T^k_i \subseteq T^k_{1,i}$. Since Proposition 1 implies that $T^{\infty}_{1,i}$ exists, $T^{\infty}_{i} \subseteq T^{\infty}_{1,i}$ also exists. ■

An obvious question to ask at this point is why the particular information structure in Corollary 1 is a natural (or 'canonical') description of stepwise thinking. The reason is that the marginal probability assigned by t^{∞}_{i} to a given event in $\cup^{k}_{l=0} X^{l}_{i}$ is equal to the probability that t^{k} assigns to that same event. That is, in deriving probability on the space $\Theta \times [\cup^{k-1}_{l=0} T^{l}_{-i}] = \cup^{k}_{l=0} X^{l}_{i}$ from t^{k}, the measure t^{∞}_{i} preserves the probabilities specified by t^{k}_{i} for all $k \geq 0$.

2.4 The universal belief space and belief closed subspaces

The structure developed in the previous subsections generated a set T^{∞}_{i} of infinite step types into which the infinite stepwise hierarchies of beliefs regarding the payoff relevant parameters can be embedded. We can rightfully deem the infinite step type set 'universal' if it contains all possible hierarchies of possible stepwise hierarchies.

Proposition 2. There exist a belief preserving embedding

$$T^{k}_{i} \subseteq T^{\infty}_{i} \text{ for any } k \geq 0,$$

such that the belief system can rightfully be deemed universal T^{U}_{i} for all $i \in N$.

Proof. See Appendix.

The coherency assumption (Definition 1) ensures that there is one and only one way to assign to any player of any step type t^{k}_{i} a corresponding step type t^{U}_{i} in T^{U}_{i} so that the same probability is assigned by t^{k}_{i} and t^{U}_{i} to the same event $E \subseteq \Theta \times [\cup^{k-1}_{l=0} T^{l}_{-i}]$. By definition the universal belief space T^{U}_{i} includes all possible beliefs over the payoff relevant parameters in the strategic situation as well as all the stepwise beliefs of opponents. That is, there exists a stepwise thinking model that describes all the relevant information in a given strategic situation with incomplete information, such that we can be confident that there is nothing intrinsically restrictive about the structure of stepwise thinking.

In most applications we would however typically only consider some subset of the universal type space. A smaller set of step types which defines the strategic situation without loss of generality is therefore needed, and for this purpose belief closed subsets are defined. For example, in most applications of interest player i does not assign positive probabilities to all points in $\Theta \times [\cup^{\infty}_{l=0} T^{l}_{-i}]$. That is, player i does not consider as possible any point not in $\text{supp}(t^{k}_{i}) \subseteq \Theta \times [\cup^{k-1}_{l=0} T^{l}_{-i}]$. Note that the finiteness of $\Theta \times [\cup^{k-1}_{l=0} T^{l}_{-i}]$ implies that t^{k}_{i} must be a probability measure with finite support. Now let $t_i^k \in Y_i^k$ denote the set of all points player i believes possible. Clearly all step types for which $\text{supp}(t_i^k) \neq \emptyset$ are relevant, but this is not all because some opponent j may not consider all points in $\Theta \times [\cup_{m=0}^{l-1} T_{-j}^m]$ (where $1 < l < k$) as possible, that is they may only consider possible points in $\text{supp}(t_j^l) \subseteq \Theta \times [\cup_{m=0}^{l-1} T_{-j}^m]$. Let $t_j^l \in Y_j^l$ denote the set of all points opponent j belief possible. Similar for all other opponents. This observation motivates the following definition:

Definition 2. A belief closed subspace is a closed subset $Y_i^k \subset T_i^U$ for which each $i \in N$, any $k \geq 1$, and all $t_i^k \in Y_i^k$ satisfies

$$\text{supp}(t_i^k) \subset \Theta \times [\cup_{l=0}^{k-1} Y_{-i}^l].$$

A belief closed subspace is a closed subset of T_i^U which is also closed under stepwise beliefs. In any $t_i^k \in Y_i^k$, it contains all the epistemic features which are relevant in the mind of player i who thinks in step k. If $t_i^k \notin Y_i^k$, then player i thinking in step k do not believe that t_i^k is possible, she does not belief that any of her opponents believes it is possible, she does not believe that any of her opponents believes that their opponents believes it is possible, and so on. Belief closed subsets implies that when we choose only to consider a subset of players' step types, we also assumes that all beliefs that are relevant for each player in a given strategic situation are included.

3 Cognitive limitations and unawareness

In this section the definitions necessary to talk about cognitive limitations are introduced (3.1), and it is shown that a nontrivial notion of unawareness arises naturally in our definition of such limitations (3.2).

3.1 States, events, and belief operators

A state specifies, for each player what she would do and believe if the state obtains. Note the subjunctive conditional; stepwise thinking with incomplete information does not only concern what actually happens, but also considers what could have happened in states that did not actually occur. Let Ω be the set of states, every element $\omega \in \Omega$ corresponds to a complete description of all the relevant aspects of the strategic situation, including what each player beliefs. The information structure on Ω is specified in terms of sigma-algebras. Let $\mathscr{B}_\Omega$ denote the Borel sigma-algebra on Ω. Each subset $E \in \mathscr{B}_\Omega$ is an event; its negation is denoted $\neg E = \Omega \backslash E$.

States are related to step types by introducing a mapping $\tau_I : \Omega \rightarrow [\cup_{l=0}^{\infty} T_I^l]$, where $I \subseteq N$ is a group of players and $\cup_{l=0}^{\infty} T_I^l$ is the set of induced step types related to I. The mapping $\tau_I(\omega) = t_I^k$ is the profile of step k types assigned to group I, when the state is ω. If the group I observes $\tau_I(\omega)$ then they deduces that the state must be in the set $\tau_I^{-1}(t_I^k)$. For example, if $I = \{i\}$ then $\tau_i(\omega) \in \Delta(\Theta \times [\cup_{l=0}^{k-1} T_{-i}^l])$ defines i's beliefs about her opponents' 'limited cognitive' abilities (see Section 2), and $\tau_i^{-1}(t_i^k)$ denotes the event 'the step type of player i is t_i^k'. Finally, let $\varphi : \Omega \rightarrow \Theta$ specify the payoff relevant parameter $\varphi(\omega) = \theta$ corresponding to any state ω. We need to relate any event $E \subseteq \Omega$ to any event in our universal type space. Since it is assumed that i knows her own step type t_i^k, she will consider elements (θ, t_{-i}^k) such that $(\theta, t_i^k, t_{-i}^k) \in E$. Slightly abusing notation let $E_{-i} = \{(\varphi(\omega), \tau_{-i}(\omega)) : (\varphi(\omega), \tau_i(\omega), \tau_{-i}(\omega)) \in E\}$ denote the event $E_{-i} \subseteq \Theta \times [\cup_{l=0}^{\infty} T_{-i}^l]$ that correspond to the event $E \subseteq \Omega$.

At state ω player i believes event $E \subseteq \Omega$, conditional on observing $\tau_i(\omega)$, with probability $\tau_i(\omega)(E_{-i})$. Thus $\{\omega : \tau_i(\omega)(E_{-i}) = 1\}$ is the event 'i belief event E conditional on observing $\tau_i(\omega)$ at ω.' We use belief operators to represent events about interactive beliefs:

Definition 3. For any $i \in N$ and some event E, the belief operator for i is defined by:

$$\boldsymbol{B}_i(E) = \{\omega : \tau_i(\omega)(E_{-i}) = 1\}.$$

Clearly, $\boldsymbol{B}_i(E) \in \mathscr{B}_\Omega$ is itself an event. Note also that $\boldsymbol{B}_i(\cdot)$ satisfies monotonicity [$E \subseteq F$ implies $\boldsymbol{B}_i(E) \subseteq \boldsymbol{B}_i(F)$], conjecture [$\boldsymbol{B}_i(E \cap F) = \boldsymbol{B}_i(E) \cap \boldsymbol{B}_i(F)$], and consistency $\boldsymbol{B}_i(\emptyset) = \emptyset$.

3.2 Nontrivial unawareness

Players are in models with stepwise thinking assumed overconfident and limited in their beliefs about their opponents–that is, players are in general unaware of opponents doing at least as much thinking as themselves. We now consider whether the hierarchy of stepwise thinking allows for a 'nontrivial' notion of such unawareness. By nontrivial we mean that the state space Ω can have states ω in which players do not know an event, and do not know that they do not know. Dekel *et al.* (1998) show that standard state spaces allow only for a trivial notion of unawareness. Namely, if a player is unaware of something then she is unaware of everything and knows nothing. More generally, they show that no standard state space can capture adequately the notion of unawareness.

We will in the following clarify why cognitive limitations are nontrivial. In other words, why stepwise thinking models imply a nontrivial notion of unawareness of opponents doing at least as much thinking as herself. Intuitively, there exists events which are unmeasurable in the mind of a stepwise thinker, these are events which demand that opponents do at least as much thinking as the player (see Corollary 1). Formally, the set of states in which player i is aware of E is given by an awareness operator:

Definition 4. For any $i \in N$, $\pi \in [0,1]$ and some event E, the awareness operator for i is defined as:

$$\boldsymbol{A}_i(E) = \{\omega : \tau_i(\omega)(E_{-i}) \geq \pi\}.$$

It follows that $\boldsymbol{A}_i(E) \in \mathscr{B}_\Omega$. A player is at ω aware of E if and only if her step type as defined by ω is concentrated on a space in which the event is 'expressible'. That is, a player is aware of any event to which she can assign some probability. The unawareness operator is naturally defined as the negation of awareness:

Definition 5. For any $i \in N$ and some event E, the unawareness operator for i is defined as:

$$\boldsymbol{U}_i(E) = \neg \boldsymbol{A}_i(E).$$

Again, clearly $\boldsymbol{U}_i(E) \in \mathscr{B}_\Omega$. By showing that the unawareness operator complies with the properties that any appealing concept of unawareness should satisfy, as suggested by Dekel *et al.* (1998), the following proposition shows that unawareness of opponents doing as much or more thinking is nontrivial in stepwise thinking models.

Proposition 3. Let E be an event. In stepwise thinking the following properties of unawareness obtains:

(i) Plausibility: $\boldsymbol{U}_i(E) \subseteq \neg \boldsymbol{B}_i(E) \cap \neg \boldsymbol{B}_i \neg \boldsymbol{B}_i(E)$,

(ii) *BU* introspection: $\boldsymbol{B}_i \boldsymbol{U}_i(E) = \emptyset$,

(iii) *AU* introspection: $\boldsymbol{U}_i(E) \subseteq \boldsymbol{U}_i \boldsymbol{U}_i(E)$.

(iv) Weak necessitation: $\neg \boldsymbol{U}_i(E) \subseteq \boldsymbol{B}_i(\Omega)$.

Proof. See Appendix.

Plausibility implies that a player is unaware of E if she does not have any beliefs about E, and does not have any beliefs about not having any beliefs about E. *BU* introspection states that a player cannot have any beliefs about her own unawareness. *AU* introspection is the property that if a player is unaware of an event E, then she must be unaware of being unaware. Finally, weak necessitation says that if a player is not unaware of E, then she knows any tautology involving E. The four properties together preclude unawareness in any standard state space model. In other words, the state space of stepwise thinking models is nonstandard. In particular, stepwise thinking rule out 'strong' necessitation ($\boldsymbol{B}_i(\Omega) = \Omega$). That is, a player in our state space does not need to be certain of all tautologies. This is fundamental for stepwise thinking; a player need not to know the 'true' state if it involves opponents doing as much or more thinking as herself.

4 Bayesian games with stepwise thinking

The Bayesian Nash equilibrium has become a benchmark for the analysis of 'standard games' with incomplete information, this concept is now applied to the class of games with stepwise thinking. First the main definitions are given (4.1), then it is shown how we can represent our Bayesian games as a 'random vector model' and a 'prior lottery model' (4.2), we then consider how we can analyze such models using the standard tools of game theory (4.3)-(4.4), and finally an example is considered (4.5).

4.1 Strategic forms with stepwise thinking

A strategic form game with stepwise thinking consists of different layers of strategic form step k games each describing the strategic situation at a step of thinking. We will say that a step k type is confined to a step k game. Remember that a belief closed subset Y_i^k contains all the epistemic features which are relevant in the mind of player i who is thinking in step k. This implies that if some player i is confined to a step $l < k$ game, then because of unawareness her belief closed subspace will be the same in any step $k \geq l$ game (that is, $Y_i^l = Y_i^k$). Situations in which only nonstrategic players play against each other will for obvious reasons be omitted. The nonstrategic step 0 types will however still be in the beliefs of step $k \geq 1$ types, since step 0 types actions influence their expected payoff. These observations motivates the following definition:

Definition 6. A Bayesian game with stepwise thinking Γ is a finite ordered set of step $k \geq 1$ games $G^k \in \Gamma$ defined by

$$G^k = \left\langle A_1, \ldots, A_n; Y_1^k, \ldots, Y_n^k; u_1, \ldots, u_n \right\rangle. \tag{2}$$

The notation A_i denotes the finite set of player i's actions; Y_i^k is the (belief closed) set of player i's step k types, each representing different information that player i can have in the step k game; and player i's payoff function $u_i : A \times Y_i^k \times [\cup_{l=0}^{k-1} Y_{-i}^l] \to \mathbb{R}$ depends on the set of action profiles $A = \prod_{i \in N} A_i$, player i's own step k types, and the step types of opponents she is aware of.

4.2 Induced stepwise belief systems

The induced beliefs of any player i is given by the conditional probability $p_i(\cdot|t_i^k) \in \Delta(\cup_{l=0}^{k-1} Y_{-i}^l)$, and can be derived from her 'signal' $\tau_i(\omega)$ at state $\omega \in \Omega$. Remember that the only information a player has about the payoff relevant parameters Θ is given by $\tau_i(\omega) = t_i^k$, we can therefore without loss of generality impose that Θ is Y_i^k-measurable. That is, there is a mapping $\alpha_i : \cup_{k=0}^{\infty} Y_i^k \to \Theta$ which relates any step k type of player i to an element in Θ. We can now relate the two concepts by having $\tau_i(\omega) \in \Delta(\Theta \times [\cup_{l=0}^{k-1} Y_{-i}^l])$ satisfying:

$$\tau_i(\omega)(\theta, t_{-i}^l) = \begin{cases} p_i(t_{-i}^l | t_i^k) & \text{if } \theta = \alpha_{-i}(t_{-i}^l), \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

A system of induced beliefs is an array $(p_i(\cdot|t_i^k))_{t_i^k \in Y_i^k}$ for which: (i) the system is projective–the marginal of $p_i(\cdot|t_i^k)$ on $\Delta(\cup_{l=0}^{k-2} T_{-i}^l)$ is $p_i(\cdot|t_i^{k-1})$ for all $k \geq 2$, and (ii) for player i of step k type the priors $p_i(\cdot|t_i^m)$ for $m > k$ are undefined. Note that by assuming that there is a one-to-one correspondence between the payoff relevant parameters and step types, we restrict our intention to a smaller class of Bayesian models.

In defining the conditional probability $p_i(\cdot|t_i^k)$ we took on an interim point of view. The strategic situation was implicitly assumed to be analyzed at a stage subsequent to the player knowing her step k type. That is, we rendered any prior stage meaningless. However, most applications of incomplete information games assumes an ex ante point of view before the player knows her type. In this view players have prior beliefs over a common set of types. At the interim stage players are given their types, update their priors, and make appropriate adjustments in their beliefs. This interpretation can however be misleading when players think in steps; if a player has beliefs over all step k types ex ante she should also have beliefs over all step k types interim, that is, after learning her own step k type (her own limited ability to think about opponents' even more limited thinking).[9]

---

[9] The plausibility and justification of the ex ante versus the interim view of information models has been extensively discussed in the literature, see Harsanyi (1967–68), Dekel and Gul (1997), Gul (1998), and Aumann (1998).

However, an array of conditional probabilities $(p_i(\cdot|t_i^k))_{t_i^k \in Y_i^k}$ can always be derived from some 'prior' by imposing consistency through a convex combination $p_i = \sum_{t_i^k \in Y_i^k} \beta_{t_i^k} p_i(\cdot|t_i^k)$ (where $\sum_{t_i^k \in Y_i^k} \beta_{t_i^k} = 1$, $\beta_{t_i^k} \geq 0$ for all $t_i^k \in Y_i^k$) for each step of thinking. Such a 'prior' does not represent i's beliefs in a hypothetical ex ante stage, it is only a technical device to express the belief $p_i(\cdot)$. If we moreover assume that these priors are the common for all players, then $p = p_i$ for all $i \in N$. In this case, a game with incomplete information simply corresponds to a game with imperfect information about a fictitious chance move selecting the vector of step types according to probability measure p. This is the so called 'random vector model' of the Bayesian game. The random vector model can also be interpreted as a population model in which for each player/role $i \in N$ there is a population of potential players characterized by the different step types. An actual player is drawn at random from each population i to play the game. This is the 'prior lottery model' of the Bayesian game.

The following two examples illustrate how the two most frequently used applied stepwise thinking models–the level-k model (Stahl and Wilson (1994, 1995) and Nagel (1995)) and the cognitive hierarchy model (Camerer *et al.* (2004)) are special cases of the 'prior lottery model' presented above. In both models it is typically assumed that there is no uncertainty about the payoff such that Θ is a singleton. We can therefore ease notation and let step k types be identified by the corresponding $k \in \mathbb{N}$.

Example 1 (Level-k models)**.** In level-k models step k types believe with probability one that opponents are step $k-1$ thinkers, that is, $p(k-1|k) = 1$.

Example 2 (Cognitive Hierarchy models)**.** In Cognitive Hierarchy models step k types believe that they play against a distribution of step $l < k$ opponents, that is, $p(l|k) = \frac{\lambda(l)}{\sum_{l=0}^{k-1} \lambda(l)}$ where $\lambda \in \Delta(\mathbb{N})$ is assumed to be a Poisson distribution.

4.3 Mixed strategies and expected payoffs

We interpret players' 'plan of play' not as deterministic, but rather regulated by probabilistic rules. Denoted by $\Delta(A_i)$ the set of probability distributions over A_i and refer to the mapping $\sigma_i : Y_i^k \to \Delta(A_i)$ as a mixed strategy of player i of step k type. A mixed strategy $\sigma_i(a_i|t_i^k)$ for player i thus specifies the conditional probability that player i of step k type plays action a_i. Let $A_{-i} = \prod_{j \neq i} A_i$ be the set of action profiles of opponents and $\sigma_{-i} : Y_{-i}^l \to \Delta(A_{-i})$ be a mixed strategy profile of player i's opponents, where $\sigma_{-i}(a_{-i}|t_{-i}^l)$ is the conditional probability that opponents $-i$ of step l types plays action profile a_{-i}.

The specification of nonstrategic players 'plan of play' is key. It is often assumed that any given player i think that nonstrategic opponents choose their actions uniformly, such that $\sigma_{-i}(a_{-i}|t^0_{-i}) \in U(A_{-i})$. However, we do not here restrict ourselves to any interpretation of how nonstrategic players behave, but instead leave it to applications.

We assume that the probability distribution p puts positive weight on each $t_i^k \in Y_i^k$ and fully determines the probability distribution $p(t^l_{-i}|t^k_i)$; player i's conditional belief about the step types $[\cup_{l=0}^{k-1} Y^l_{-i}]$ of opponents given her own step k type $t_i^k \in Y_i^k$. The expected payoff of player i thinking in step k conditioned on the strategy profile $\sigma = (\sigma_i, \sigma_{-i}) \in \Sigma$ can now be expressed as

$$\mathbb{E}_{t_i^k}[u_i|\sigma] := \sum_{t^l_{-i} \in [\cup_{l=0}^{k-1} Y^l_{-i}]} p(t^l_{-i}|t^k_i) \sum_{a_{-i} \in A_{-i}} \sigma_{-i}(a_{-i}|t^l_{-i}) u_i(a_i, a_{-i}; t^k_i, t^l_{-i}), \tag{4}$$

where $\sigma_{-i}(a_{-i}|t'_{-i})$ is player i thinking in step k's mixed strategy about the action profile a_{-i} of opponents t^l_{-i} which induce payoff $u_i(a_i, a_{-i}; t^k_i, t'_{-i})$. Evaluating this term gives the expected utility from the actions of opponents. However, player i thinking in step k does, in general, not know the profile of opponents facing her and thus evaluates her expected utility with respect to her beliefs $p(t^l_{-i}|t^k_i)$.

4.4 Equilibrium and existence

Although a stepwise game is not a 'standard game', the Bayesian Nash equilibrium concept based on the notion of best response can be adapted yielding a solution concept.

Definition 7. A Bayesian Nash equilibrium of a stepwise thinking game Γ is a profile $\sigma^* \in \Sigma$ of strategies with the property that for every $i \in N$, $t^k_i \in Y^k_i$, $G^k \in \Gamma$ we have

$$\text{supp}(\sigma_i) \subseteq \arg\max_{a_i \in A_i} \mathbb{E}_{t_i^k}[u_i|\sigma_i(a_i), \sigma^*_{-i}]. \tag{5}$$

Thus, a Bayesian Nash equilibrium of a stepwise thinking game specifies a behavior for each player which is a best response to what she believes is the behavior of her opponents, that is, a best response to the mixed strategies of her opponents given his step type.

Proposition 4. There exists a Bayesian Nash equilibrium in each step k game $G^k \in \Gamma$.

Proof. See Appendix.

Step k types are unaware of opponents doing at least as much thinking as themselves and do not consider these players when calculating their expected payoffs. However, step k types are aware of opponents thinking in steps $l < k$ and take these types in to account. When choosing an equilibrium mixed strategy a step k type who is confined to G^k thus maximizes her expected payoffs based on the mixed strategies of opponents thinking in step $l < k$, but does not consider the mixed strategies of opponents doing more thinking than her. An equilibrium mixed strategy profile in a game with stepwise thinking Γ is thus a profile in which step types who do more thinking fix the equilibrium strategies of step types doing less thinking, before they find their own equilibrium strategy. The following proposition helps clarify the existence of such an equilibrium mixed strategy profile.

Proposition 5. Consider any two step games $G^l, G^k \in \Gamma$ where $l < k$. There exists an equilibrium strategy profile σ^* in G^k, in which step l types play their equilibrium strategies in G^l and step k types play theirs in G^k.

Proof. See Appendix.

This proposition suggests a procedure for constructing a Bayesian Nash equilibrium in a game with stepwise thinking. We start with an equilibrium in the step 1 game, and then extend it step-by-step to 'higher' step games by taking the equilibrium strategies of opponents in the respective 'lower' step games as given. This formulation suggest a method of finding a Bayesian Nash equilibrium in a game with stepwise thinking: (i) First calculate for each player thinking in step 1 the best response to the nonstrategic players thinking in step 0, then (ii) extend it step-by-step to players thinking in higher steps by fixing the best response of players doing less thinking.

4.5 Example

Stepwise thinking and the solution concept just characterized is illustrated by the following trivial example of a run on the bank (Diamond and Dybvig, 1983). First we consider a 'prior lottery model' with complete information, and then extent it to a situation with asymmetric information.

(i) *Complete information*

Let there be two (almost) equally sized populations of depositors A and B. Each depositor is small in that her stake is negligible as a proportion of the whole. If a population of depositors withdraws their money from the bank, then they obtain a guaranteed payoff of $r > 0$. If they leave their money in, and the other population of depositors leave their money in as well, they get a payoff of R, where $r < R < 2r$. But if they leave their money in, and the other population depositors withdraws, then the bank will go bankrupt and they get a payoff of zero. The payoffs in each of these situations are:

		B	
		In	Out
A	In	R, R	$0, r$
	Out	$r, 0$	r, r

Figure 1: Payoff matrix

If this was a ‘standard game’ it would have two equilibria: [In,In] and [Out,Out]. We will however here be interested in a scenario in which population A thinks in one step and population B in two, and each population is certain that the other thinks in one less step.[10] For simplicity let the nonstrategic depositors choose their actions according to an uniform distribution. The two populations chooses simultaneously.

First, assume that the actions of step 0 thinkers are perfectly correlated. Population A believes that depositors in population B are nonstrategic. Depositors in A thus theorize that depositors in B withdraw their money with probability $\frac{1}{2}$ and stays in otherwise. Population A therefore expects that the payoff is r if they withdraw and $\frac{R}{2}$ if they stay in. Since $r > \frac{R}{2}$, population A’s best response is to withdraw their money. Population B believes that depositors in A are thinking in step 1 and expect that they will withdraw their money. The best response of population B is therefore to withdraw their money as well since $r > 0$. That is, we have a run on the bank since [Out,Out] is the unique (inferior) equilibrium in the game with stepwise thinking just described.

Now assume that the actions of step 0 thinkers are independent. This implies that depositors in B withdraw their money with probability $\frac{1}{2^n}$, where n is the number of depositors in population B. It follows that $r < (1 - \frac{1}{2^n})R$ for $n \geq 2$ such that depositors in A thinking in step 1 expects that depositors in B thinking in step 0 stays in, and therefore choose to stay in themselves. Foreseeing this line of events the depositors in B thinking in step 2 will also stay in since $R > r$. The unique (superior) equilibrium is in this case [In,In].

Notice that the coordination on equilibrium is not determined only by the fundamentals (money), nor is it determined by some payoff irrelevant variable that has nothing to do with the fundamentals (‘sunspots’). Rather, what matters are depositors steps of thinking. Especially, the beliefs of depositors in population A about the actions of the (in their mind) nonstrategic depositors in B.

(ii) Asymmetric information

---

[10] Evidence of such level-k thinking between a population of step 1 and 2 types, in which step types are certain that opponents think in one less step, has been found in many experiments (see, for example, Costa-Gomes and Crawford, 2006).

Now imagine that if the bank goes bankrupt, then there exists some institution which can either do nothing or bail out the bank ($\Theta = \{\text{Nothing, Bailout}\}$). If the bank is bailed out, then the depositors who does not withdraw their money will get a payoff of R instead of zero. Assume for simplicity that population A, thinking in step 1, is uncertain about whether or not the bailout is going to happen ($Y_A^1 = \{t_A^1[\text{Nothing}], t_A^1[\text{Bailout}]\}$). Whereas population B, thinking in two steps, knows it will ($Y_B^2 = \{t_B^2[\text{Bailout}]\}$). The payoffs in each of these situations are:

		B	
		In	Out
A	In	R, R	$0, r$
	Out	$r, 0$	r, r

Nothing

		B	
		In	Out
A	In	R, R	R, r
	Out	r, R	r, r

Bailout

Figure 2: Two payoff relevant parameters: $\Theta = \{\text{Nothing, Bailout}\}$

Consider again the situation in which the actions of step 0 thinkers are perfectly correlated. Population A reasons that the payoff is r if they withdraw, and $p(t_B^0|t_A^1[\text{Bailout}])R + (1 - p(t_B^0|t_A^1[\text{Bailout}]))\,\frac{R}{2}$ if they stay in. That is, if $p(t_B^0|t_A^1[\text{Bailout}]) > \frac{2r}{R} - 1$ then the depositors in population A will stay in, otherwise they will withdraw their money. Being certain that the bailout is going to happen, the depositors in population B will always stay in. The unique equilibrium is therefore [In,In] if $p(t_B^0|t_A^1[\text{Bailout}]) > \frac{2r}{R} - 1$, and [Out,In] otherwise.

Intuitively, a run on the bank can (in this simple example) be prevented if depositors assigns a 'high enough' probability to the bank being bailed out. That is, the higher the payoff is from staying in relative to withdrawing, the lower the probability of a bailout has to be.

5 Conclusion

This paper introduced a general incomplete information framework for studying stepwise thinking. The framework we considered was general enough to: (*i*) analyze players abilities to predict opponents' behavior at the most fundamental level, (*ii*) cover payoff relevant uncertainty; and (*iii*) allow for the examination of situations involving stepwise thinking separate from the solution concept.

Along the way it was also shown that there exists a coherent universal step type space which contains all step types such that there is nothing intrinsically restrictive about the proposed structure. Such universality was obtained in the broadest and most natural setup, that of probability (or measure) theory. The structure of stepwise thinking implied that players were unaware of opponents doing at least as much thinking as themselves. Within this structure, we could admit as much uncertainty as might seem appropriate in any situation, by enlarging the set of step types which represents uncertainty about the payoff relevant parameters.

Onto this structure we appended the orthodox Bayesian Nash equilibrium concept. Because players are unaware of any situation involving opponents doing at least as much thinking as themselves, they believe that the game they are confined to is the ‘true’ game. This implies that there exists a Bayesian Nash equilibrium strategy in the game with stepwise thinking in which players in any step game fix the equilibrium strategies of opponents, who they believe do less thinking, and choose their own equilibrium strategy based on this belief. This suggest a procedure for constructing an equilibrium; first we have to find an equilibrium in the step 0 game and then extend it step-by-step to ‘higher’ step games by fixing the equilibrium strategies of opponents in the respective ‘lower’ step games.

A Appendix

A.1 Proof of Proposition 2

First note that $T_i^k \subseteq T_i^\infty$ is equivalent to $\Delta(\cup_{l=0}^k X_i^l) \subseteq \Delta(\cup_{l=0}^\infty X_i^l)$. We need to show that it is true that for all $k \geq 0$, $\Delta(\cup_{l=0}^k X_i^l) \subseteq \Delta(\cup_{l=0}^\infty X_i^l)$.

The construction of hierarchies of stepwise thinking implies that for all $k \geq 0$, $\cup_{l=0}^k X_i^l \subseteq \cup_{l=0}^\infty X_i^l$. We know from the proof of Proposition (5) that the Borel sigma-algebra on $\cup_{l=0}^\infty X_i^l$ is well defined. Since the generated Borel sigma-algebra is the smallest algebra containing all open sets (or, equivalently, all closed sets), it holds true that for all $k \geq 0$, $\mathscr{B}_{[\cup_{l=0}^k X_i^l]} \subseteq \mathscr{B}_{[\cup_{l=0}^\infty X_i^l]}$ and thus $\Delta(\cup_{l=0}^k X_i^l) \subseteq \Delta(\cup_{l=0}^\infty X_i^l)$. The coherency assumption ensures that $\Delta(\cup_{l=0}^k X_i^l) = \mathrm{marg}_{X^0 \cup \cdots \cup X^k} \Delta(\cup_{l=0}^\infty X_i^l)$, such that beliefs are preserved. ■

A.2 Proof of Proposition 3

With slight abuse of notation we write $\boldsymbol{B}_i(E_{-i})$, $\boldsymbol{A}_i(E_{-i})$ and $\boldsymbol{U}_i(E_{-i})$ for the events in the step type space which corresponds to events $\boldsymbol{B}_i(E)$, $\boldsymbol{A}_i(E)$ and $\boldsymbol{U}_i(E)$, respectively, in the state space. For example, $\boldsymbol{B}_i(E_{-i}) = \{(\varphi(\omega), \tau_{-i}(\omega)) : (\varphi(\omega), \tau_i(\omega), \tau_{-i}(\omega)) \in \boldsymbol{B}_i(E)\}$. Now to the proofs of the four properties:

(i) *Plausibility*: This property is equivalent to $\boldsymbol{B}_i(E) \cup \boldsymbol{B}_i \neg \boldsymbol{B}_i(E) \subseteq \boldsymbol{A}_i(E)$. By Definition 3 and 4 we have that $\boldsymbol{B}_i(E) \subseteq \boldsymbol{A}_i(E)$. To see that $\boldsymbol{B}_i \neg \boldsymbol{B}_i(E) \subseteq \boldsymbol{A}_i(E)$, note that $\omega \in \boldsymbol{B}_i \neg \boldsymbol{B}_i(E)$ *iff* $\tau_i(\omega)(\neg \boldsymbol{B}_i(E_{-i})) = 1$. This implies that $\neg \boldsymbol{B}_i(E) \subseteq \boldsymbol{A}_i(E)$. Hence $\omega \in \boldsymbol{A}_i(E)$.

(ii) *BU introspection*: $\boldsymbol{B}_i \boldsymbol{U}_i(E) = \emptyset$. To see that this is true consider that some $\omega \in \boldsymbol{B}_i \boldsymbol{U}_i(E)$ *iff* $\tau_i(\omega)(\boldsymbol{U}_i(E_{-i})) = 1$, which can only be true if $\boldsymbol{U}_i(E) \subseteq \boldsymbol{A}_i(E)$. By Definition 5 this is impossible and $\omega \notin \boldsymbol{B}_i \boldsymbol{U}_i(E)$.

(iii) *AU introspection*: $\boldsymbol{U}_i(E) \subseteq \boldsymbol{U}_i \boldsymbol{U}_i(E)$ is equivalent to $\boldsymbol{A}_i \boldsymbol{U}_i(E) = \boldsymbol{A}_i(E)$. Then $\omega \in \boldsymbol{A}_i \boldsymbol{U}_i(E)$ *iff* $\tau_i(\omega)(\boldsymbol{U}_i(E_{-i})) \geq \pi$. Hence $\omega \in \boldsymbol{A}_i \boldsymbol{U}_i(E)$ *iff* $\omega \in \boldsymbol{A}_i(E)$ by Definition 4.

(iv) *Weak necessitation*: $\neg \boldsymbol{U}_i(E) \subseteq \boldsymbol{B}_i(\Omega)$ is equivalent to $\boldsymbol{A}_i(E) \subseteq \boldsymbol{B}_i(\Omega)$. $\omega \in \boldsymbol{A}_i(E)$ *iff* $\tau_i(\omega)(E_{-i}) \geq \pi$ (Definition 4), and $\omega \in \boldsymbol{B}_i(\Omega)$ *iff* $\tau_i(\omega)(\Theta \times [\cup_{l=0}^\infty T_{-i}^l]) = 1$ (Definition 3). Since $E_{-i} \subseteq \Theta \times [\cup_{l=0}^\infty T_{-i}^l]$ and $\pi \leq 1$ (awareness if a weaker condition than belief) then it hold true that $\omega \in \boldsymbol{A}_i(E)$ *iff* $\omega \in \boldsymbol{B}_i(\Omega)$. ■

A.3 Proof of Proposition 4

Proposition 4 follows naturally from the following Theorem:

Theorem 1 (Kakutani, 1941)**.** If D is a nonempty compact and convex subset of Euclidean space, and ϕ is an upper hemicontinuous, nonempty, and convex valued correspondence $\phi : D \to D$, then ϕ has a fixed point, that is, there is a $d \in D$ such that $d \in \phi(d)$.

Proof of Proposition 4. By using Definition 7 define $\beta^k : \Delta(A) \to \Delta(A)$ by $\beta^k(\sigma^*) = \prod_{i \in N} \beta_i^k(\sigma_{-i}^*)$ where

$$\beta_i^k(\sigma_{-i}^*) = \{\sigma_i^* \in \Sigma_i : \mathbb{E}_{t_i^k}[u_i|\sigma^*] \geq \mathbb{E}_{t_i^k}[u_i|(\sigma_i, \sigma_{-i}^*)] \text{ for all } t_i^k \in T_i^k, \sigma_i \in \Sigma_i\}. \tag{6}$$

$\Delta(A)$ is by definition a nonempty compact and convex subset of Euclidean space. $\beta^k(\sigma^*)$ is upper hemicontinuous because $\mathbb{E}_{t_i^k}[u_i|\sigma^*]$ is continuous for each (finite) $t_i^k \in Y_i^k$ and $i \in N$, nonempty since each $\mathbb{E}_{t_i^k}[u_i|\sigma^*]$ is continuous and $\Delta(A)$ is compact, and convex valued because each $\mathbb{E}_{t_i^k}[u_i|\sigma^*]$ is quasi-concave on $\Delta(A)$ ($\beta_i^k(\sigma_{-i}^*) = \{\sigma_i : \mathbb{E}_{t_i^k}[u_i|(\sigma_i, \sigma_{-i}^*)] \geq \mathbb{E}_{t_i^k}[u_i|\sigma^*]$ for each $t_i^k \in T_i^k\}$). Therefore, by Theorem 1, $\beta^k(\sigma^*)$ has a fixed point, that is, there is some $\sigma^* \in \beta^k(\sigma^*)$. By definition, σ^* is a fixed point of $\beta^k(\sigma^*)$ *iff* it is a Bayesian Nash equilibrium in $G^k \in \Gamma$. ■

A.4 Proof of Proposition 5

We need to show that σ^* is an equilibrium strategy profile in Γ^k in which step $l < k$ types play their equilibrium strategies in Γ^l and step k types play theirs in Γ^k. Suppose not, then there would be a profitable deviation

$$\mathbb{E}_{t_i^\chi}[u_i|(\sigma_i, \sigma_{-i}^*)] > \mathbb{E}_{t_i^\chi}[u_i|\sigma^*] \tag{7}$$

for some $t_i^\chi \in T_i^\chi$, $\chi \in \{l, k\}$ and $i \in N$.

(i) For $\chi = k$, a player's strategy σ_i is not an equilibrium strategy in Γ^k by Definition 7–a contradiction.

(ii) For $\chi = l$, since a player's expected payoff is (due to unawareness) identical in Γ^l and Γ^k, her strategy σ_i is not an equilibrium strategy in Γ^l by Definition 7–a contradiction.

Hence σ^* must be an equilibrium strategy profile in Γ^k. ■

References

AUMANN, R. (1992). *Irrationality in Game Theory.* MIT Press.

— (1997). Rationality and bounded rationality. *Games and Economic Behavior*, **21**, 2–14.

— (1998). Common priors: A reply to gul. *Econometrica*, **66**, 929–938.

BILLINGSLEY, P. (1995). *Probability and Measure.* John Wiley & Sons.

BOCHNER, S. (1960). *Harmonic Analysis and the Theory of Probability.* Berkeley: University of California Press.

BOSCH-DOMENECH, A., MONTALVO, J., NAGEL, R. and SATORRA, A. (2002). One, two,(three), infinity...: Newspaper and lab beauty-contest experiments. *American Economic Review*, **92**, 1687–1701.

BRANDENBURGER, A. and DEKEL, E. (1993). Hierarchies of beliefs and common knowledge. *Journal of Economic Theory*, **59**, 189–198.

CAMERER, C., HO, T.-H. and CHONG, J.-K. (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics*, **119**, 861–898.

—, JOHNSON, E., RYMON, T. and SEN, S. (1994). Cognition and framing in sequential barganing for gains and losses. In *Frontiers of Game Theory, eds. K. Binmore, A. Kirman and P. Tani.*

—, LOEWENSTEIN, G. and PRELEC, D. (2005). Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Perspectives*, **43**, 9–64.

COSTA-GOMES, M. and CRAWFORD, V. (2006). Cognition and behavior in two-person guessing games: An experimental study. *American Economic Review*, **96**, 1737–1768.

—, — and BROSETA, B. (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica*, **69**, 1193–1235.

CRAWFORD, V. and IRIBERRI, N. (2007a). Fatal attraction: Salience, naivete, and sophistication in experimental "hide-and-seek games. *American Economic Review*, **97**, 1731–1750.

— and — (2007b). Level-k auctions: Can a nonequilibrium model of strategic thinking explain the winner's curse and overbidding in private-value auctions? *Econometrica*, **75**, 1721–1770.

DEKEL, E., FUDENBERG, D. and MORRIS, S. (2006). Topologies on types. *Theoretical Economics*, **1**, 275–309.

— and GUL, F. (1997). Rationality and knowledge in game theory. In *Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress.*

—, Lipman, B. and Rustichini, A. (1998). Standard state-space models preclude unawareness. *Econometrica*, **66**, 159–173.

Diamond, D. and Dybvig, P. (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*, **91**, 401–419.

Eyster, E. and Rabin, M. (2005). Cursed equilibrium. *Econometrica*, **73**, 1623–72.

Fagin, R. and Halpern, J. (1988). Beliefs, awareness, and limited reasoning. *Artificial Intelligence*, **34**, 39–76.

Gul, F. (1998). A comment on aumann's bayesian view. *Econometrica*, **66**, 923–927.

Halpern, J. (2001). Alternative semantics for unawareness. *Games and Economic Behavior*, **37**, 321–339.

— and Rêgo (2008). Interactive unawareness revisited. *Games and Economic Behavior*, **62**, 323–262.

Harsanyi, J. (1967–68). Games with incomplete information played by bayesian players i-iii. *Management Science*, **14**, 159–82, 320–34, 486–502.

Heifetz, A. (1993). The bayesian formulation of incomplete information, the non-compact. *International Journal of Game Theory*, **21**, 329–338.

—, Meier, M. and Schipper, B. (2006). Interactive unawareness. *Journal of Economic Theory*, **130**, 78–94.

—, — and — (2008). A cannonical model for interactive unawareness. *Games and Economic Behavior*, **62**, 304–324.

—, — and — (2009). Bayesian games with unawareness and unawareness perfection. *Mimeo.*

— and Samet, D. (1998). Topology-free typology of beliefs. *Journal of Economic Theory*, **82** (2), 324–341.

Ho, T., Camerer, C. and Weigelt, K. (1998). Iterated dominance and iterated best response in experimental "p-beauty contests". *American Economic Review*, **88**, 947–969.

Jehiel, P. (2005). Analogy-based expectations equilibrium. *Journal of Economic Theory*, **123**, 81–104.

— and Koessler, F. (2008). Revisiting games of incomplete information with analogy-based expectations equilibrium. *Games and Economic Behavior*, **62**, 533–57.

Kakutani, S. (1941). A generalization of brouwer's fixed point theorem. *Duke Mathematical Journal*, **8**, 457–459.

Mertens, J.-F., Sorin, S. and Zamir, S. (1994). Repeated games, part a: Background material. *CORE Discussion Paper No. 9420.*

— and Zamir, S. (1985). Formulation of bayesian analysis for games with incomplete information. *International Journal of Game Theory*, **14**, 1–29.

Modica, S. and Rustichini, A. (1999). Unawareness and partitional information structures. *Games and Economic Behavior*, **27**, 265–298.

Monderer, D. and Samet, D. (1989). Approximating common knowledge with common beliefs. *Games and Economic Behavior*, **1**, 170–190.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American Economic Review*, **85**, 1313–1326.

Rubinstein, A. (1989). The electronic mail game: Strategic behavior under "almost common knowledge". *American Economic Review*, **79**, 385–291.

Savage, L. (1954). *The Foundations of Satistics.* John Wiley & Sons.

Stahl, D. and Wilson, P. (1994). Experimental evidence on players' models of other players'. *Journal of Economic Behavior and Organization*, **25**, 309–327.

— and — (1995). On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*, **10**, 218–254.

Strzalecki, T. (2009). Dept of reasoning and higher order beliefs. *Unpublished, Department of Economics, Harvard University.*