

# Are the Players in an Interactive Belief Model Meta-certain of the Model Itself? (Extended Abstract)

Satoshi Fukuda

Department of Decision Sciences and IGIER, Bocconi University\*  
Milan, Italy

satoshi.fukuda@unibocconi.it

In an interactive belief model, are the players “commonly meta-certain” of the model itself? This paper formalizes such implicit “common meta-certainty” assumption. To that end, the paper expands the objects of players’ beliefs from events to functions defined on the underlying states. Then, the paper defines a player’s belief-generating map: it associates, with each state, whether a player believes each event at that state. The paper formalizes what it means by: “a player is (meta-)certain of her own belief-generating map” or “the players are (meta-)certain of the profile of belief-generating maps (i.e., the model).” The paper shows: a player is (meta-)certain of her own belief-generating map if and only if her beliefs are introspective. The players are commonly (meta-)certain of the model if and only if, for any event which some player  $i$  believes at some state, it is common belief at the state that player  $i$  believes the event. This paper then asks whether the “common meta-certainty” assumption is needed for an epistemic characterization of game-theoretic solution concepts. The paper shows: if each player is logical and (meta-)certain of her own strategy and belief-generating map, then each player correctly believes her own rationality. Consequently, common belief in rationality alone leads to actions that survive iterated elimination of strictly dominated actions.

## 1 Introduction

In an economic or game-theoretic model in which the players make their interactive reasoning about their strategies or rationality, the analysts “from outside of the model” assume that the players “commonly know” the model itself. Since the pioneering work of [1, 2, 3], how to model such assumption and what consequences such assumption has have been puzzling economic and game theorists.<sup>1</sup>

This paper has two objectives. The first aims at formalizing the “common knowledge” assumption of a model within the model itself. An interactive belief/knowledge model represents players’ beliefs/knowledge about its ingredients, that is, events. The model itself does not tell whether the players (commonly) believe/know the model itself, although the analysts assume that the players (commonly) believe/know the model in a meta-sense. I refer to the knowledge/belief of the model as the “meta-knowledge/meta-belief” of the model.<sup>2</sup>

The second objective is to examine the role that “meta-knowledge” of a model plays in game-theoretic analyses such as epistemic characterizations of solution concepts. For a given epistemic characterization of a game-theoretic solution concept such as iterated elimination of strictly dominated actions, do the outside analysts need to formally assume that the players “meta-know” an epistemic model of a game (that describes their interactive beliefs about their strategies and rationality)?

---

\*I would like to thank three anonymous referees of TARK 2021 for their helpful comments. This is an abbreviated technical summary of part of the full paper, which is available at the author’s website (<https://websfukuda.com/research/>).

<sup>1</sup>For this question, see also [4, 5], [6], [10, 11], [12], [13], [15], [16], [21], [25], [26], [30], [32], [33], [31], and [34].

<sup>2</sup>Since different epistemic models may feature different notions of qualitative or probabilistic beliefs or knowledge, I use the word the “(meta-)certainty” of a model to refer generically to the meta-knowledge or meta-belief of the model.

The first main result, presented in Section 4, characterizes the “common meta-certainty” assumption as follows. According to the formal test to be discussed, the players are commonly (meta-)certain of a model if and only if, for any event which some player  $i$  believes at some state, it is common belief that player  $i$  believes the event at that state.

In Section 2, I start with introducing a (belief) model. The model consists of the following three ingredients. The first is a measurable space of states of the world  $(\Omega, \mathcal{D})$ . Each state  $\omega \in \Omega$  is a list of possible specifications of the world, and the collection  $\mathcal{D}$  of events (i.e., subsets of  $\Omega$ ) are the objects of the players’ beliefs. The second is the players’ monotone belief operators  $(B_i)_{i \in I}$ . Player  $i$ ’s belief operator  $B_i$  associates, with each event  $E$ , the event that she believes  $E$ . Monotonicity means that if player  $i$  believes  $E$  at a state and if  $E$  implies (i.e., is included in)  $F$ , then she believes  $F$  at that state. The third is a common belief operator  $C$ , which associates, with each event  $E$ , the event that the players commonly believe  $E$ . Under certain assumptions on the players’ beliefs, an event  $E$  is common belief if and only if everybody believes  $E$ , everybody believes that everybody believes  $E$ , and so on *ad infinitum*.

The framework nests the following two standard models of belief or knowledge (or combinations thereof). The first is a possibility correspondence model of qualitative belief or knowledge (e.g., [1, 3], [15], [20], and [24]). The possibility correspondence associates, with each state, the set of states that she considers possible. The player believes an event  $E$  at a state when the possibility set at  $\omega$  is included in the event  $E$ . The second is a type space ([22]), where each player’s probabilistic beliefs are induced by her type mapping. The type mapping  $\tau_i$  associates, with each state  $\omega$ , her probability measure  $\tau_i(\omega)$  on the underlying states at that state. The type mapping  $\tau_i$  of player  $i$  induces her  $p$ -belief operator ([23]): it associates, with each event  $E$ , the event that player  $i$  believes  $E$  with probability at least  $p$ .

With the framework in mind, I formalize the (meta-)certainty of a model in two steps. In the first step, Section 3.1 expands the objects of the players’ beliefs from events to functions defined on the underlying states. Examples of such functions are random variables, strategies, and type mappings. Any such function  $x$  has to be defined on the state space  $\Omega$ , but the co-domain  $X$  can be any set such as the set  $\mathbb{R}$  of real numbers (a random variable), a set  $A_i$  of player  $i$ ’s actions (her strategy), and the set  $\Delta(\Omega)$  of probability measures on  $(\Omega, \mathcal{D})$  (a type mapping). I call the function  $x : \Omega \rightarrow X$  a signal if its co-domain  $X$  has “observational” contents  $\mathcal{X}$  (where “observation” is broadly construed as being an object of reasoning): it is a collection of subsets of  $X$  such that each  $F \in \mathcal{X}$  is deemed an event  $x^{-1}(F)$ . Formally, a *signal (mapping)* is a function  $x : (\Omega, \mathcal{D}) \rightarrow (X, \mathcal{X})$  such that each observation  $F \in \mathcal{X}$  is considered to be an event  $x^{-1}(F) \in \mathcal{D}$ . Player  $i$  is *certain* of the value of the signal  $x$  at a state  $\omega$  if, for any observational content  $F$  that holds at  $\omega$  (i.e.,  $\omega \in x^{-1}(F)$ ), player  $i$  believes the event  $x^{-1}(F)$  at  $\omega$  (i.e.,  $\omega \in B_i(x^{-1}(F))$ ). Player  $i$  is *certain* of  $x$  if she is certain of the value of  $x$  at every state. For example, let  $x : \Omega \rightarrow X$  be the strategy of player  $i$  and let every singleton action  $a \in X$  be observable to her; then, player  $i$  is certain of her own strategy if, wherever she takes an action  $a = x(\omega)$  at a state  $\omega$ , she believes at  $\omega$  that she takes action  $a$ . Having defined individual players’ (meta-)certainty, the players are *commonly certain* of the value of the signal  $x$  at a state  $\omega$  if, for any observational content  $F$  that holds at  $\omega$ , the event  $x^{-1}(F)$  is common belief at  $\omega$  (i.e.,  $\omega \in C(x^{-1}(F))$ ). The players are *commonly certain* of the signal  $x$  if they are commonly certain of its value at every state.

In the second step, Section 3.2 formulates a players’ “belief-generating map” as a signal that associates, with each state, her beliefs at that state. By the second step, I can apply the formalization of certainty and common certainty in the first step to the ingredients of a given model (i.e., players’ belief-generating maps). To that end, take player  $i$ ’s belief operator  $B_i$  from the model. I define a qualitative-type mapping  $t_{B_i}$ : it associates, with each state, whether player  $i$  believes each event or not at that state (formally, a binary set function from the collection of events to the binary values  $\{0, 1\}$  where 1 indicates the belief of an event). The qualitative-type mapping is a binary “type” mapping analogous to a type map-

ping  $\tau_i$  that represents player  $i$ 's probabilistic beliefs at each state in the context of probabilistic beliefs. Thus, the qualitative-type mapping  $t_{B_i}$  associates, with each state  $\omega$ , her qualitative belief  $t_{B_i}(\omega) \in \{0, 1\}$  (where  $t_{B_i}(\omega)(E) = 1$  if and only if  $\omega \in B_i(E)$ ) on  $(\Omega, \mathcal{D})$  at  $\omega$ . The qualitative-type mapping  $t_{B_i}$  is player  $i$ 's belief-generating mapping. Since the belief operator  $B_i$  and the qualitative-type mapping  $t_{B_i}$  are equivalent means of representing player  $i$ 's beliefs, a model means the profile of qualitative-type mappings. Thus, the formal test for whether the players are commonly certain of a given belief model is whether the players are certain of the profile of their qualitative-type mappings.

Before asking when a player is certain of all the players' qualitative-type mappings (i.e., the model), Section 3.3 characterizes when a player is certain of her own qualitative type-mapping in terms of her introspective properties of beliefs. Roughly, Proposition 1 shows that each player is certain of her own qualitative-type mapping if and only if her belief is introspective.

To the best of my knowledge, this is the first paper which systematically formalizes the statement that the players are (commonly) (meta-)certain of any given belief model within the model itself. The main result on this question, nevertheless, is related to [21]. He constructs a particular syntactic model in which the statement that the model is common knowledge is incorporated within itself. He formulates the sense in which the model is commonly known from Positive Introspection of common knowledge: if a statement is common knowledge then it is commonly known that the statement is common knowledge. In Theorem 1, in contrast, the players are commonly certain of a given model if and only if, at each state and for any event which some player believes at that state, it is common belief that the player believes the event at that state. Thus, in this paper, the key criteria is the positive introspective property of common belief with respect to each player's beliefs. Whenever some individual player believes some event, it is common belief that she believes it. [4, 5], in the context of partitional possibility correspondence models, formalizes the event that a player has an information partition by regarding it as a function.

Having characterized the common-certainty of the model, Section 5 examines the role that the "common meta-certainty" assumption plays in the epistemic characterization of iterated elimination of strictly dominated actions (IESDA) in a strategic game. Informally, it states: if the players are "logical," if they are commonly meta-certain of a game, and if they commonly believe their rationality, then the resulting actions survive any process (i.e., any order) of IESDA. Formally, it states: if the players commonly believe their rationality and if their common belief in their rationality is correct, then the resulting actions survive any process of IESDA.<sup>3</sup> Theorem 2 connects these two statements. If the players' beliefs are monotone (they believe any logical implication of their beliefs), consistent (i.e., they do not simultaneously believe an event and its negation), and finitely conjunctive (if they believe  $E$  and  $F$  then they believe its conjunction  $E \cap F$ ), and if each player is certain of her own strategy and the part of her own belief-generating process in the model (each player is not necessarily certain of how the opponents' beliefs are generated in the model), then each player correctly believes her own rationality, and hence they have correct common belief in their rationality. Thus, if the players are "logical" and each of them is meta-certain of the part of the model that governs her own beliefs, then common belief in rationality leads to actions that survive any process of IESDA.

On the one hand, Theorem 1 states that the players may not always be commonly certain of a belief model in the sense that each player is commonly certain of how the other players' beliefs are represented within the model. On the other hand, Theorem 2 asserts that common belief in rationality leads to IESDA even if the players may not be commonly certain of the belief model.

The paper is organized as follows. Section 2 defines a belief model. Section 3 characterizes the sense

---

<sup>3</sup>The formal statement is taken from [18, Theorem 3], which holds irrespective of the nature of beliefs. For seminal papers on implications of common belief in rationality, see, for example, [9], [29], and [30].

in which each player is certain of how her belief is generated in a model. Section 4 examines the sense in which the players are commonly certain of a model itself (i.e., how the players' beliefs are generated in the model). Section 5 studies how the assumption that the players are commonly certain of a model itself can make game-theoretic analyses coherent.

## 2 Framework

Throughout the paper, let  $I$  denote a non-empty finite set of *players*. The framework represents players' interactive beliefs by belief operators on a state space. Section 2.1 defines a belief model. Section 2.2 defines properties of beliefs.

### 2.1 A Belief Model

A *belief model* is a tuple  $\vec{\Omega} := \langle (\Omega, \mathcal{D}), (B_i)_{i \in I}, C \rangle$ , where: (i)  $(\Omega, \mathcal{D})$  is a non-empty measurable space of states of the world (call  $\Omega$  the *state space*); (ii)  $B_i : \mathcal{D} \rightarrow \mathcal{D}$  is player  $i$ 's (monotone) *belief operator*; and (iii)  $C : \mathcal{D} \rightarrow \mathcal{D}$  is a (monotone) *common belief operator* to be defined below.

While  $\Omega$  constitutes a non-empty set of *states* of the world, each element  $E$  of  $\mathcal{D}$  is an *event* about which the players reason. For each event  $E$ , the set  $B_i(E)$  denotes the event that (i.e., the set of states at which) a player  $i$  believes  $E$ . Thus, the player  $i \in I$  believes an event  $E \in \mathcal{D}$  at a state  $\omega \in \Omega$  if  $\omega \in B_i(E)$ . I assume that each player's belief operator satisfies *Monotonicity*:  $E \subseteq F$  implies  $B_i(E) \subseteq B_i(F)$ . It means that if player  $i$  believes some event then she believes any of its logical consequences.<sup>4</sup>

Since the players' beliefs are monotone, I introduce the common belief operator  $C : \mathcal{D} \rightarrow \mathcal{D}$  following [23]. Call an event  $E$  *publicly evident* if  $E \subseteq B_I(E) := \bigcap_{i \in I} B_i(E)$ . That is, everybody believes  $E$  whenever  $E$  is true. Denote by  $\mathcal{J}_{B_I}$  the collection of publicly-evident events. An event  $E$  is *common belief* at a state  $\omega$  if there is a publicly-evident event that is true at  $\omega$  and that implies the mutual belief in  $E$ : that is,  $\omega \in F \subseteq B_I(E)$  for some  $F \in \mathcal{J}_{B_I}$ . Now,  $C$  is assumed to satisfy that the set of states at which  $E$  is common belief is an event for each  $E \in \mathcal{D}$ :  $C(E) := \{\omega \in \Omega \mid \text{there is } F \in \mathcal{J}_{B_I} \text{ with } \omega \in F \subseteq B_I(E)\} \in \mathcal{D}$ .

Since players' beliefs are monotone and since  $\mathcal{D}$  is closed under countable intersection, if  $E$  is common belief, then everybody believes  $E$ , everybody believes that everybody believes  $E$ , and so forth *ad infinitum*:  $C(E) \subseteq \bigcap_{n \in \mathbb{N}} B_I^n(E)$ . The converse (set inclusion) holds, for example, when the mutual belief operator  $B_I$  (or every  $B_i$ ) satisfies *Countable Conjunction*:  $\bigcap_{n \in \mathbb{N}} B_I(E_n) \subseteq B_I(\bigcap_{n \in \mathbb{N}} E_n)$ , meaning that everybody believes the countable conjunction of events whenever everybody believes each of them.

I represent the players' beliefs on a measurable space  $(\Omega, \mathcal{D})$  so that I can analyze players' qualitative and probabilistic beliefs under the same framework. The full paper introduces the players' probabilistic beliefs on a measurable space  $(\Omega, \mathcal{D})$  by  $p$ -belief operators [23]. For each  $p \in [0, 1]$ , player  $i$ 's  $p$ -belief operator  $B_i^p$  associates, with each event  $E$ , the event that player  $i$  believes  $E$  with probability at least  $p$ . The full paper also introduces the common  $p$ -belief operator  $C^p$ .

### 2.2 Properties of Beliefs

Next, I introduce additional eight properties of beliefs. Various possibility correspondence models of qualitative beliefs and knowledge are represented as belief models that satisfy certain properties specified below. Fix a player  $i$ . I first introduce the following five logical properties of beliefs.

<sup>4</sup>The full paper dispenses with the monotonicity assumption, which enables one to define the meta-certainty of any belief model.

1. *Necessitation*:  $B_i(\Omega) = \Omega$ .
2. *Countable Conjunction*:  $\bigcap_{n \in \mathbb{N}} B_i(E_n) \subseteq B_i(\bigcap_{n \in \mathbb{N}} E_n)$  (for any events  $(E_n)_{n \in \mathbb{N}}$ ).
3. *Finite Conjunction*:  $B_i(E) \cap B_i(F) \subseteq B_i(E \cap F)$ .
4. The *Kripke property*:  $B_i(E) = \{\omega \in \Omega \mid b_{B_i}(\omega) \subseteq E\}$ , where  $b_{B_i}(\omega) := \bigcap \{E \in \mathcal{D} \mid \omega \in B_i(E)\}$  is the set of states player  $i$  considers *possible* at  $\omega$ .
5. *Consistency*:  $B_i(E) \subseteq (\neg B_i)(E^c)$ .

First, Necessitation means that the player believes a tautology such as  $E \cup E^c$ . Second, as discussed, Countable Conjunction means that if the player believes each of a countable collection of events, then she believes its conjunction. Third, Finite Conjunction is weaker than Countable Conjunction: if the player believes  $E$  and  $F$  then she believes its conjunction  $E \cap F$ . Fourth, to discuss the Kripke property, the player considers  $\omega'$  possible at  $\omega$  if, for any event  $E$  which she believes at  $\omega$ ,  $E$  is true at  $\omega'$ . The Kripke property provides the condition under which  $i$ 's belief is induced by her *possibility correspondence*  $b_{B_i} : \Omega \rightarrow \mathcal{P}(\Omega)$ : she believes  $E$  at  $\omega$  if and only if (hereafter, iff) her possibility set  $b_{B_i}(\omega)$  at  $\omega$  implies  $E$ . The Kripke property implies the previous three properties as well as Monotonicity. Fifth, Consistency means that the player cannot simultaneously believe an event  $E$  and its negation  $E^c$ .

Next, I move on to truth and introspective properties.

6. *Truth Axiom*:  $B_i(E) \subseteq E$  (for all  $E \in \mathcal{D}$ ).
7. *Positive Introspection*:  $B_i(\cdot) \subseteq B_i B_i(\cdot)$  (i.e.,  $B_i(E) \subseteq B_i B_i(E)$  for all  $E \in \mathcal{D}$ ).
8. *Negative Introspection*:  $(\neg B_i)(\cdot) \subseteq B_i(\neg B_i)(\cdot)$ .

Sixth, Truth Axiom turns belief into knowledge in that knowledge has to be true while belief can be false. Truth Axiom implies Consistency. While knowledge satisfies Truth Axiom, qualitative and probabilistic beliefs are often assumed to satisfy Consistency. Seventh, Positive Introspection states that if the player believes some event then she believes that she believes it. Eighth, Negative Introspection states that if the player does not believe some event then she believes that she does not believe it. Truth Axiom and Negative Introspection yield Positive Introspection (e.g., [3]).

Three remarks are in order. First, the introspective properties will play important roles in whether a player is (meta-)certain of a belief model. Intuitively, Positive Introspection provides the sense in which the player believes her own belief (at least at face value) while Negative Introspection yields the sense in which the player believes the lack of her own belief. To see these points, an event  $E$  is *self-evident* to  $i$  if  $E \subseteq B_i(E)$ . That is,  $i$  believes  $E$  whenever  $E$  is true. Positive Introspection means that  $i$ 's belief in  $E$  is self-evident to  $i$ , and Negative Introspection means that  $i$ 's lack of belief in  $E$  is self-evident to  $i$ . Denote by  $\mathcal{J}_{B_i}$  the collection of self-evident events to  $i$ .

Second, the last four properties are restated in terms of  $b_{B_i}$  under the Kripke property:  $B_i$  satisfies Consistency iff  $b_{B_i}$  is serial (i.e.,  $b_{B_i}(\cdot) \neq \emptyset$ );  $B_i$  satisfies Truth Axiom iff  $b_{B_i}$  is reflexive (i.e.,  $\omega \in b_{B_i}(\omega)$  for all  $\omega \in \Omega$ );  $B_i$  satisfies Positive Introspection iff  $b_{B_i}$  is transitive (i.e.,  $\omega' \in b_{B_i}(\omega)$  implies  $b_{B_i}(\omega') \subseteq b_{B_i}(\omega)$ ); and  $B_i$  satisfies Negative Introspection iff  $b_{B_i}$  is Euclidean (i.e.,  $\omega' \in b_{B_i}(\omega)$  implies  $b_{B_i}(\omega) \subseteq b_{B_i}(\omega')$ ).

Third, various models of probabilistic and qualitative beliefs and knowledge take different sets of axioms. The framework accommodates possibility correspondence models of qualitative beliefs and knowledge when  $B_i$  satisfies the Kripke property. A partitional model of knowledge corresponds to the case when  $B_i$  satisfies Truth Axiom, Positive Introspection, and Negative Introspection.<sup>5</sup> A reflexive

<sup>5</sup>In fact, Truth Axiom, Negative Introspection, and the Kripke property yield all the other properties defined in this section.

and transitive (non-partitional) possibility correspondence model is characterized by Truth Axiom and Positive Introspection.<sup>6</sup> When it comes to fully-introspective qualitative beliefs,  $b_{B_i}$  is serial, transitive, and Euclidean iff  $B_i$  satisfies Consistency, Positive Introspection, and Negative Introspection.

### 3 When Is a Player Certain of Her Belief-Generating Mapping?

Section 3.1 extends an object of beliefs in a model from an event to a function (“signal”) defined on the state space. That is, the subsection formulates the statement that a player is certain of a function defined on the state space. Section 3.2 represents a player’s “belief-generating mapping” as a signal which associates, with each state, whether she believes each event or not. Section 3.3 asks the sense in which she is certain of her own belief-generating mapping in terms of the introspective properties.

#### 3.1 Functions as Objects of Players’ Beliefs

I start with defining a notion of a signal mapping. A signal mapping is any function  $x$  defined on the state space  $\Omega$  with “observational” contents. A signal is interpreted as a mapping from the underlying state space  $\Omega$  into a space of “observation”  $X$  endowed with “observational” contents  $\mathcal{X} \subseteq \mathcal{P}(X)$ . By observation, it means that each  $F \in \mathcal{X}$  is deemed an object of reasoning. That is, we call a mapping  $x : \Omega \rightarrow X$  a signal mapping if each “observational” content  $F \in \mathcal{X}$  can be regarded as an event  $x^{-1}(F) \in \mathcal{D}$  through inverting the mapping.

Formally, for a non-empty set  $X$  and a non-empty subset  $\mathcal{X}$  of  $\mathcal{P}(X)$ , call  $x : (\Omega, \mathcal{D}) \rightarrow (X, \mathcal{X})$  a *signal* (mapping) if  $x^{-1}(\mathcal{X}) \subseteq \mathcal{D}$ . Mathematically,  $x : (\Omega, \mathcal{D}) \rightarrow (X, \mathcal{X})$  is a signal if  $x : (\Omega, \mathcal{D}) \rightarrow (X, \sigma(\mathcal{X}))$  is measurable. Examples include strategies, random variables, and so on.

The main purpose of this subsection is to define the statement that a player is certain of a signal. A player  $i$  is *certain of the value of* a signal  $x : (\Omega, \mathcal{D}) \rightarrow (X, \mathcal{X})$  at  $\omega$ , if she believes any observational content  $F$  (i.e., believes  $x^{-1}(F)$ ) at  $\omega$  whenever it is true:  $x(\omega) \in F$ . She is *certain of* the signal  $x : (\Omega, \mathcal{D}) \rightarrow (X, \mathcal{X})$  if she is certain of its value at every  $\omega$ . Likewise, the players are *commonly certain of the value of* the signal  $x : (\Omega, \mathcal{D}) \rightarrow (X, \mathcal{X})$  at  $\omega$ , if the players commonly believe any observational content  $F$  at  $\omega$  whenever it is true. The players are *commonly certain of* the signal  $x : (\Omega, \mathcal{D}) \rightarrow (X, \mathcal{X})$  if they are certain of its value at every  $\omega$ . Formally:

**Definition 1.** Let  $\vec{\Omega}$  be a belief model, and let  $x : (\Omega, \mathcal{D}) \rightarrow (X, \mathcal{X})$  be a signal mapping.

1. (a) Player  $i$  is *certain of the value of the signal*  $x : (\Omega, \mathcal{D}) \rightarrow (X, \mathcal{X})$  at  $\omega$  if  $\omega \in B_i(x^{-1}(F))$  for any  $F \in \mathcal{X}$  with  $x(\omega) \in F$ .  
 (b) Player  $i$  is *certain of the signal*  $x : (\Omega, \mathcal{D}) \rightarrow (X, \mathcal{X})$  if she is certain of the value of the signal  $x$  at any state.
2. (a) The players are *commonly certain of the value of the signal*  $x : (\Omega, \mathcal{D}) \rightarrow (X, \mathcal{X})$  at  $\omega$  if  $\omega \in C(x^{-1}(F))$  for any  $F \in \mathcal{X}$  with  $x(\omega) \in F$ .  
 (b) The players are *commonly certain of* the signal  $x : (\Omega, \mathcal{D}) \rightarrow (X, \mathcal{X})$  if they are commonly certain of the value of the signal  $x$  is at every state.

For example, suppose that  $x : \Omega \rightarrow X$  is a decision function of a player which associates, with each state, the action taken at that state. Suppose the set of actions  $X$  is endowed with the collection of singleton actions  $\mathcal{X} = \{\{a\} \mid a \in X\}$ . Each action  $a$  corresponds to an observational content to the

<sup>6</sup>See, for example, [4], [6], [15], [20], [25], [27], and [28].

player, and  $x$  is a signal mapping if the set of states at which the player takes action  $a$  is an event:  $x^{-1}(\{a\}) = \{\omega \in \Omega \mid x(\omega) = a\} \in \mathcal{D}$  for each  $a \in X$ .

More specifically, let  $\Omega = \{\omega_1, \omega_2, \omega_3\}$  and  $X = \{a, b\}$ . For each  $i \in I = \{1, 2\}$ , let  $B_i$  be given by (i)  $B_i(E) = E \setminus \{\omega_3\}$  for each  $E \neq \Omega$ ; and (ii)  $B_i(\Omega) = \Omega$ . Suppose player 1's decision function  $x : (\Omega, \mathcal{P}(\Omega)) \rightarrow (X, \{\{a\}, \{b\}\})$  is given by  $(x(\omega))_{\omega \in \Omega} = (a, a, a)$ . Since  $B_1(\Omega) = \Omega$  and  $B_1(\emptyset) = \emptyset$ , whenever player 1 takes a certain action, she believes that she takes that action. Thus, player 1 is certain of  $x$ . If, instead, her decision function  $x : (\Omega, \mathcal{P}(\Omega)) \rightarrow (X, \{\{a\}, \{b\}\})$  is given by  $(x(\omega))_{\omega \in \Omega} = (a, b, a)$ , then at  $\omega_3$  at which she takes action  $a$ , she does not believe that she takes action  $a$ , because  $B_1(\{\omega_1, \omega_3\}) = \{\omega_1\}$ . Thus, player 1 is not certain of the value of  $x$  at  $x_3$ . Since  $C = B_1$ , the same arguments hold for the common certainty of  $x$ .

For ease of terminology, player  $i$  is *certain of (the value of) the signal*  $x : \Omega \rightarrow X$  (at  $\omega$ ) *with respect to*  $\mathcal{X}$  if she is certain of (the value of) the signal  $x : (\Omega, \mathcal{D}) \rightarrow (X, \mathcal{X})$  (at  $\omega$ ). Likewise, the players are *commonly certain of (the value of) the signal*  $x : \Omega \rightarrow X$  (at  $\omega$ ) *with respect to*  $\mathcal{X}$  if they are commonly certain of (the value of) the signal  $x : (\Omega, \mathcal{D}) \rightarrow (X, \mathcal{X})$  (at  $\omega$ ).

Four remarks on Definition 1 are in order. First, I restate the fact that a player is certain of a signal in terms of self-evidence. Namely, player  $i$  is certain of a signal  $x : (\Omega, \mathcal{D}) \rightarrow (X, \mathcal{X})$  iff any observational content  $F \in \mathcal{X}$  (i.e., any event  $x^{-1}(F) \in \mathcal{D}$ ) is self-evident to  $i$ . Likewise, the players are commonly certain of the signal  $x : (\Omega, \mathcal{D}) \rightarrow (X, \mathcal{X})$  iff any observational content  $F \in \mathcal{X}$  is publicly-evident. Consequently, the players are commonly certain of a signal iff every player is certain of it.<sup>7</sup>

Second, when  $x : (\Omega, \mathcal{D}) \rightarrow (X, \mathcal{X})$  is a player's strategy, Definition 1 formalizes the statement that the player is certain of the strategy (e.g., [12] and [20]). To see this, assume that  $\mathcal{X}$  contains a singleton  $\{x(\omega)\}$  to reason about the action taken at  $\omega$ . That is, the set of states  $[x(\omega)] := x^{-1}(\{x(\omega)\}) = \{\omega' \in \Omega \mid x(\omega') = x(\omega)\}$  at which player  $i$  takes the same action as she does at  $\omega$  is an event. Then, player  $i$  is certain of her strategy iff  $[x(\omega)]$  is self-evident at every  $\omega \in \Omega$ . In words, player  $i$  is certain of her strategy  $x$  iff, whenever she takes action  $a = x(\omega)$  at  $\omega$ , she believes at  $\omega$  that she takes action  $a = x(\omega)$ .

Definition 1 also subsumes the formulation of the certainty of the strategy by [2] in the (countable) partitional state space model of knowledge. Let  $(b_{B_i}(\omega))_{\omega \in \Omega}$  be a countable partition on  $\Omega$ . In [2], the player "knows" her own strategy  $x$  iff the strategy  $x$  is measurable with respect to the partition (which turns out to be equivalent to  $b_{B_i}(\cdot) \subseteq [x(\cdot)]$ ). Since the partition is countable, the  $\sigma$ -algebra generated by the partition is equal to the self-evident collection:  $\mathcal{J}_{B_i} = \sigma(\{b_{B_i}(\omega) \in \mathcal{D} \mid \omega \in \Omega\})$ . Hence, player  $i$  is certain of her strategy  $x : (\Omega, \mathcal{D}) \rightarrow (X, \mathcal{X})$  iff  $x : (\Omega, \mathcal{J}_{B_i}) \rightarrow (X, \sigma(\mathcal{X}))$  is measurable.

Third, player  $i$  satisfies Necessitation iff she is certain of any constant signal. Likewise, the common belief operator  $C$  satisfies Necessitation (equivalently, every  $B_i$  satisfies Necessitation) iff the players are commonly certain of any constant signal.

Necessitation allows the players to be certain of any constant "random" variable that does not depend on the realization of a state. For example, consider whether player  $i$  is certain that an event  $B_j(E)$  is equal to an event  $F$  in a belief model. The outside analysts determine whether player  $i$  believes that player  $j$  believes an event  $E$  at a state  $\omega$  by examining whether  $\omega \in B_i B_j(E)$  since player  $j$ 's belief  $B_j(E)$  itself is an event. The (implicit) assumption in any (semantic) belief model is that  $E = F$  implies  $B_i(E) = B_i(F)$ . Thus, if two events are extensionally the same, then each player's belief in the two events are the same.<sup>8</sup> To assess player  $i$ 's belief about player  $j$ 's belief about  $E$ , how can the outside

<sup>7</sup>In contrast, it is not necessarily the case that each player is certain of a signal at a state  $\omega$  iff the players are commonly certain of the signal at  $\omega$ .

<sup>8</sup>Although such identification of events are implicitly assumed for any (semantic) belief model, one can construct a canonical ("universal") semantic model from a syntactic language which maximally distinguishes the denotations of events. In the canonical model, such identification of events can be minimized in a way such that two events are equated only when they are

analysts justify the fact that player  $i$  is able to equate  $B_j(E)$  with another event (say,  $F$ )? Since either  $B_j(E) = F$  or  $B_j(E) \neq F$ , player  $i$  is *certain that  $B_j(E)$  is an event  $F$*  if player  $i$  is certain of the indicator function  $\mathbb{I}_{B_j \leftrightarrow F}$ , where  $(B_j(E) \leftrightarrow F) := ((\neg B_j)(E) \cup F) \cap ((\neg F) \cup B_j(E))$ . If player  $i$ 's belief operator  $B_i$  satisfies Necessitation and if  $B_j(E) = F$ , then player  $i$  is certain of the constant indicator function  $\mathbb{I}_{B_j \leftrightarrow F}$ . Thus, under Necessitation, player  $i$  is certain that  $B_j(E) = F$  if it is indeed the case. This argument justifies that, under Necessitation, the outside analysts can say that the players are certain of equating two extensionally equivalent events (say,  $B_j(E)$  and  $F$ ) if they are indeed extensionally equivalent.

Fourth, it can be formally shown that player  $i$  is certain of a profile of signals (e.g., a strategy profile) iff she is certain of each of them. Thus, the players are commonly certain of a profile of signals iff every player is certain of every signal.

### 3.2 A Qualitative-Type Mapping that Represents a Player's Beliefs

In order to formulate a test under which the outside analysts can examine whether the players are commonly certain of a belief model, I define the “belief-generating map,” which I call the qualitative-type mapping ([17]), of a player. Given the belief operator of the player, the qualitative-type mapping associates, with each state, a binary value indicating whether the player believes each event in an analogous manner to the type mapping in the type-space literature.

To that end, recall that a (probabilistic-)type mapping associates, with each state  $\omega$ , the player's probabilistic beliefs  $\tau_i(\omega) \in \Delta(\Omega)$  at that state. With this in mind, let  $M(\Omega)$  be the set of binary set functions  $\mu : \mathcal{D} \rightarrow \{0, 1\}$  (i.e.,  $M(\Omega) \subseteq \{0, 1\}^{\mathcal{D}}$ ) that satisfy a given set of logical properties of beliefs defined in Section 2.2 (these properties will be shortly expressed in terms of  $\mu$ ). Call each  $\mu \in M(\Omega)$  a *qualitative-type*. Interpret  $\mu(E) = 1$  as the belief in an event  $E \in \mathcal{D}$ . Once  $M(\Omega) \subseteq \{0, 1\}^{\mathcal{D}}$  is defined as the set of qualitative-types that satisfy the given set of logical properties of beliefs, I represent player  $i$ 's beliefs by a *qualitative-type mapping*  $t_i : \Omega \rightarrow M(\Omega)$  satisfying a certain measurability condition specified below. It is a measurable mapping which associates, with each state  $\omega \in \Omega$ , player  $i$ 's qualitative-type  $t_i(\omega) \in M(\Omega)$  at  $\omega$ . Thus, player  $i$  believes an event  $E$  at  $\omega$  if  $t_i(\omega)(E) = 1$ .

Now, I define the logical properties of  $\mu$  in an analogous way to the corresponding logical properties of belief operators. Fix  $\mu \in \{0, 1\}^{\mathcal{D}}$ .

0. *Monotonicity*:  $E \subseteq F$  implies  $\mu(E) \leq \mu(F)$ .
1. *Necessitation*:  $\mu(\Omega) = 1$ .
2. *Countable Conjunction*:  $\min_{n \in \mathbb{N}} \mu(E_n) \leq \mu(\bigcap_{n \in \mathbb{N}} E_n)$ .
3. *Finite Conjunction*:  $\min(\mu(E), \mu(F)) \leq \mu(E \cap F)$ .
4. *The Kripke property*:  $\mu(E) = 1$  iff  $\bigcap \{F \in \mathcal{D} \mid \mu(F) = 1\} \subseteq E$ .
5. *Consistency*:  $\mu(E) \leq 1 - \mu(E^c)$ .

The interpretations of the above properties are similar to those in Section 2.2. Whether all of these properties are assumed or not depend on the model that the outside analysts study. For example, if the outside analysts examine a partitioned possibility correspondence model, then  $M(\Omega)$  is the set of qualitative-types that satisfy all the logical properties.

I formally define the measurability condition of a qualitative-type mapping. A *qualitative-type mapping* is a measurable mapping  $t_i : (\Omega, \mathcal{D}) \rightarrow (M(\Omega), \mathcal{D}_M)$  which satisfies given (logical and) introspective properties of beliefs, where  $\mathcal{D}_M$  is the  $\sigma$ -algebra generated by the sets of the form  $\beta_E := \{\mu \in M(\Omega) \mid$

---

explicitly assumed to be equivalent by the outside analysts (see [19] for a formal assertion).



$\mu(E) = 1\}$  for all  $E \in \mathcal{D}$ . The set  $\beta_E$  is the set of types under which  $E$  is believed. Thus,  $\beta_E$  is an informational content indicating that event  $E$  is believed. Note that  $t_i : \Omega \rightarrow M(\Omega)$ , by construction, satisfies given logical properties because any element in  $M(\Omega)$  satisfies them. For example, if every  $\mu \in M(\Omega)$  satisfies the Kripke property, then every  $t_i(\omega)$  satisfies it. Denote  $b_{t_i}(\omega) := \bigcap \{E \in \mathcal{D} \mid t_i(\omega)(E) = 1\}$  for each  $\omega \in \Omega$ .

The measurability condition of  $t_i$  requires each  $t_i^{-1}(\beta_E) = \{\omega \in \Omega \mid t_i(\omega)(E) = 1\}$  to be the event that player  $i$  believes  $E$ . Next, I define Truth Axiom and the introspective properties of  $t_i$ .

6. *Truth Axiom*:  $t_i(\omega)(E) = 1$  implies  $\omega \in E$ .
7. *Positive Introspection*:  $t_i(\omega)(E) = 1$  implies  $t_i(\omega)(t_i^{-1}(\beta_E)) = 1$ .
8. *Negative Introspection*:  $t_i(\omega)(E) = 0$  implies  $t_i(\omega)(\neg t_i^{-1}(\beta_E)) = 1$ .

### 3.3 Certainty of Own Type Mapping

I apply the certainty of a signal to a qualitative-type mapping. Proposition 1 below roughly states that a player is certain of her own qualitative-type mapping iff her beliefs are introspective.

**Proposition 1.** *Let  $\vec{\Omega}$  be a belief model, and let  $t_{B_i} : \Omega \rightarrow M(\Omega)$  be player  $i$ 's qualitative-type mapping.*

1. (a) *Player  $i$  is certain of  $t_{B_i}$  with respect to  $\{\beta_E \mid E \in \mathcal{D}\}$  iff  $B_i$  satisfies Positive Introspection.*  
 (b) *Player  $i$  is certain of  $t_{B_i}$  with respect to  $\{\neg\beta_E \mid E \in \mathcal{D}\}$  iff  $B_i$  satisfies Negative Introspection.*  
 (c) *If player  $i$  is certain of  $t_{B_i} : (\Omega, \mathcal{D}) \rightarrow (M(\Omega), \mathcal{D}_M)$ , then  $B_i$  satisfies Positive Introspection and Negative Introspection.*
2. (a) *Let  $B_i$  satisfy Truth Axiom. Player  $i$  is certain of  $t_{B_i} : (\Omega, \mathcal{D}) \rightarrow (M(\Omega), \mathcal{D}_M)$  iff  $B_i$  satisfies (Positive Introspection and) Negative Introspection.*  
 (b) *Let  $B_i$  satisfy Consistency and Countable Conjunction. Player  $i$  is certain of  $t_{B_i} : (\Omega, \mathcal{D}) \rightarrow (M(\Omega), \mathcal{D}_M)$  iff  $B_i$  satisfies Positive Introspection and Negative Introspection.*

Part (1) characterizes the certainty of the qualitative-type mapping  $t_{B_i}$  with respect to the possession or lack of beliefs. In contrast, Parts (2a) and (2b), respectively, examine the sense in which player  $i$  is certain of her qualitative-type mapping  $t_{B_i} : (\Omega, \mathcal{D}) \rightarrow (M(\Omega), \mathcal{D}_M)$  in a model of knowledge and belief.

Part (1a) states that player  $i$  is certain of her qualitative-type mapping  $t_{B_i}$  with respect to the possession of beliefs iff her belief operator  $B_i$  satisfies Positive Introspection. Parts (1a) and (1b) jointly state that  $B_i$  satisfies Positive Introspection and Negative Introspection iff player  $i$  is certain of her qualitative-type mapping  $t_{B_i}$  with respect to  $\{\beta_E \mid E \in \mathcal{D}\} \cup \{\neg\beta_E \mid E \in \mathcal{D}\}$ .

I discuss two implications of Proposition 1. First, it sheds light on the literature of non-partitional knowledge models without Negative Introspection (see footnote 6). Part (1) implies that, without imposing Negative Introspection, player  $i$  is not certain of her own qualitative-type mapping with respect to  $\mathcal{D}_M$  (or  $\{\beta_E \mid E \in \mathcal{D}\} \cup \{\neg\beta_E \mid E \in \mathcal{D}\}$ ). Rather, she takes her own information at face value in the sense that she is only certain of her qualitative-type mapping with respect to her own beliefs  $\{\beta_E \mid E \in \mathcal{D}\}$ . Proposition 1 formalizes the sense in which “she takes her own information at face value.”

In contrast, Proposition 1 (2a) shows that, in a partitional possibility correspondence model of knowledge, a player is fully certain of her possibility correspondence when Truth Axiom, (Positive Introspection) and Negative Introspection hold. While the proposition does not necessarily require  $B_i$  to satisfy the Kripke property, consider a model of knowledge in which  $B_i$  satisfies Truth Axiom and the Kripke property, i.e.,  $B_i$  is induced by the reflexive possibility correspondence  $b_{B_i}$ . Then, player  $i$  is certain of her “knowledge-generating” mapping iff  $B_i$  satisfies (Positive Introspection and) Negative Introspection.

Proposition 1 (2b) shows that, for a serial possibility correspondence, a player is fully certain of her possibility correspondence when her beliefs satisfy Positive Introspection and Negative Introspection.

Second, suppose a player has qualitative belief and knowledge. Consider a model  $\langle (\Omega, \mathcal{D}), (K_i)_{i \in I}, C \rangle$  where  $K_i : \mathcal{D} \rightarrow \mathcal{D}$  is player  $i$ 's (monotone) knowledge operator. Now, for each player  $i$ , let  $B_i : \mathcal{D} \rightarrow \mathcal{D}$  be her (monotone) qualitative-belief operator. Let  $t_{B_i}$  be player  $i$ 's qualitative-type mapping that represents  $B_i$ , and ask whether player  $i$  is certain of her qualitative-type mapping  $t_{B_i} : (\Omega, \mathcal{D}) \rightarrow (M(\Omega), \mathcal{D}_M)$ . Proposition 1 (2a) implies that player  $i$  is certain of  $t_{B_i} : (\Omega, \mathcal{D}) \rightarrow (M(\Omega), \mathcal{D}_M)$  iff  $K_i$  satisfies *Positive Certainty* (with respect to  $B_i$ ):  $B_i(\cdot) \subseteq K_i B_i(\cdot)$  and *Negative Certainty* (with respect to  $B_i$ ):  $(\neg B_i)(\cdot) \subseteq K_i(\neg B_i)(\cdot)$ . Whenever player  $i$  believes an event, she knows that she believes it. Whenever player  $i$  does not believe an event, she knows that she does not believe it. In fact, these two properties are often assumed in a model of belief and knowledge. Proposition 1 (2a) justifies the assumptions in terms of the certainty of one's knowledge about her own beliefs.

## 4 When are the Players Commonly Certain of a Belief Model?

I formalize the sense in which the players are commonly certain of a belief model itself: the players are commonly certain of the profile of their qualitative-type mappings. As discussed, it is sufficient to ask when every player  $i$  is certain of each player  $j$ 's qualitative-type mapping.

To that end, observe that Proposition 1 applies to the case in which player  $i$  is certain of player  $j$ 's qualitative-type mapping. For example, if player  $i$  is certain of player  $j$ 's qualitative-type mapping  $t_j : (\Omega, \mathcal{D}) \rightarrow (M(\Omega), \mathcal{D}_M)$ , then  $B_{t_j}(\cdot) \subseteq B_i B_{t_j}(\cdot)$  and  $(\neg B_{t_j})(\cdot) \subseteq B_i(\neg B_{t_j})(\cdot)$  hold. Proposition 1 implies:

**Remark 1.** Let  $\vec{\Omega}$  be a belief model, and let  $t_{B_j} : \Omega \rightarrow M(\Omega)$  be player  $j$ 's qualitative-type mapping.

1. (a) Player  $i$  is certain of  $t_{B_j}$  with respect to  $\{\beta_E \mid E \in \mathcal{D}\}$  iff  $B_j(\cdot) \subseteq B_i B_j(\cdot)$ .  
 (b) Player  $i$  is certain of  $t_{B_j}$  with respect to  $\{\neg\beta_E \mid E \in \mathcal{D}\}$  iff  $(\neg B_j)(\cdot) \subseteq B_i(\neg B_j)(\cdot)$ .  
 (c) If player  $i$  is certain of  $t_{B_j} : (\Omega, \mathcal{D}) \rightarrow (M(\Omega), \mathcal{D}_M)$ , then  $B_j(\cdot) \subseteq B_i B_j(\cdot)$  and  $(\neg B_j)(\cdot) \subseteq B_i(\neg B_j)(\cdot)$ .
2. (a) Let  $B_i$  satisfy Truth Axiom. Player  $i$  is certain of  $t_{B_j} : (\Omega, \mathcal{D}) \rightarrow (M(\Omega), \mathcal{D}_M)$  iff  $(B_j(\cdot) \subseteq B_i B_j(\cdot))$  and  $(\neg B_j)(\cdot) \subseteq B_i(\neg B_j)(\cdot)$ .  
 (b) Let  $B_i$  satisfy Consistency and Countable Conjunction. Player  $i$  is certain of  $t_{B_j} : (\Omega, \mathcal{D}) \rightarrow (M(\Omega), \mathcal{D}_M)$  iff  $B_j(\cdot) \subseteq B_i B_j(\cdot)$  and  $(\neg B_j)(\cdot) \subseteq B_i(\neg B_j)(\cdot)$ .

Roughly, Remark 1 states that player  $i$  is certain of player  $j$ 's qualitative-type mapping  $t_{B_j}$  if and only if (i) whenever player  $j$  believes an event  $E$  at  $\omega$ , player  $i$  believes player  $j$  believes  $E$  at  $\omega$ ; and (ii) whenever player  $j$  does not believe an event  $E$  at  $\omega$ , player  $i$  believes player  $j$  does not believe  $E$  at  $\omega$ .

Now, I ask when the players are commonly certain of the qualitative-type mappings in a belief model.

**Theorem 1.** Let  $\vec{\Omega}$  be a belief model, and let  $t_{B_i} : \Omega \rightarrow M(\Omega)$  be player  $i$ 's qualitative-type mapping.

1. Assume Truth Axiom for every  $B_i$ . The players are commonly certain of the profile of qualitative-type mappings  $t_{B_i} : (\Omega, \mathcal{D}) \rightarrow (M(\Omega), \mathcal{D}_M)$  iff  $B_i = B_j$  for every  $i, j \in I$ , (Positive Introspection), and Negative Introspection. In particular,  $B_i = C$  for each  $i \in I$ .
2. Assume Consistency and Countable Conjunction for every  $B_i$ . The players are commonly certain of the profile of qualitative-type mappings  $t_{B_i} : (\Omega, \mathcal{D}) \rightarrow (M(\Omega), \mathcal{D}_M)$  iff  $B_i(\cdot) \subseteq C B_i(\cdot)$  and  $(\neg B_i)(\cdot) \subseteq C(\neg B_i)(\cdot)$  for every  $i \in I$ . In particular,  $C = B_I$ .

While Part (1) studies a knowledge model, Part (2) does a belief model. I start with discussing implications of Part (2). This part states that the players are commonly certain of their qualitative-type mappings iff (i) for any event  $E$  which some player  $i$  believes at some state  $\omega$ , it is commonly believed that player  $i$  believes  $E$  at  $\omega$ ; and (ii) for any event  $E$  which some player  $i$  does not believe at some state  $\omega$ , it is commonly believed that player  $i$  does not believe  $E$  at  $\omega$ .

This part imposes a strong requirement that, under Consistency and Countable Conjunction, the mutual belief and common belief operators coincide if the players are commonly certain of their qualitative-type mappings.<sup>9</sup> For any event  $E$  which everybody believes at some state  $\omega$ , it is commonly believed that everybody believes  $E$  at  $\omega$ :  $B_I(\cdot) \subseteq CB_I(\cdot)$ . Intuitively, in a model of which the players are commonly certain, if everybody believes an event  $E$  then it is common belief that everybody believes  $E$ . Thus, if everybody believes  $E$  then everybody believes that everybody believes  $E$ . Hence, the first-order mutual belief itself implies any higher-order mutual beliefs, and thus the mutual and common beliefs coincide.

Next, Part (1) provides a contrast between knowledge and belief. In a knowledge model with Truth Axiom, for the players to be commonly certain of the model, it is *necessary* that their knowledge coincides with each other. In contrast, in a belief model without Truth Axiom, there exists a model in which the players' beliefs are different but they are commonly certain of their qualitative-type mappings.

Yet, Theorem 1 is an impossibility result in the following sense. In Part (1), every player's knowledge operator coincides. In Part (2), the mutual and common belief operators coincide. In this regard, informally, Theorem 1 has some similarity with the impossibility of agreeing-to-disagree [1]: (under a common prior) if two players have common knowledge of their posteriors then the posteriors coincide. Here, if players are commonly certain of their knowledge operators, then their knowledge operators coincide.

Finally, as an implication of Theorem 1, suppose that the players are commonly certain of a belief model. If player  $i$  is certain of a signal  $x$ , then is player  $j$  certain of the signal  $x$ , too? While the players' beliefs may not be homogeneous, the proposition below shows that this is the case.

**Proposition 4.** *Let  $\vec{\Omega}$  be a belief model such that each  $B_i$  satisfies Consistency. Let  $x : (\Omega, \mathcal{D}) \rightarrow (X, \mathcal{X})$  be a signal such that, for any  $F \in \mathcal{X}$ , there exists a sub-collection  $(F_\lambda)_{\lambda \in \Lambda}$  of  $\mathcal{X}$  with  $F^c = \bigcup_{\lambda \in \Lambda} F_\lambda$ .*

1. *If player  $i$  is certain of  $x : (\Omega, \mathcal{D}) \rightarrow (X, \mathcal{X})$  and if player  $j$  is certain of player  $i$ 's qualitative-type mapping  $t_{B_i} : (\Omega, \mathcal{D}) \rightarrow (M(\Omega), \mathcal{D}_M)$ , then player  $j$  is certain of  $x : (\Omega, \mathcal{D}) \rightarrow (X, \mathcal{X})$ .*
2. *Suppose that the players are commonly certain of the profile of their qualitative-type mappings  $t_{B_i} : (\Omega, \mathcal{D}) \rightarrow (M(\Omega), \mathcal{D}_M)$ . Then, player  $i$  is certain of  $x : (\Omega, \mathcal{D}) \rightarrow (X, \mathcal{X})$  iff player  $j$  is certain of  $x : (\Omega, \mathcal{D}) \rightarrow (X, \mathcal{X})$ .*

The meta-common-certainty assumption states that if player  $i$  is certain of her own strategy and if player  $j$  is certain of player  $i$ 's type mapping then player  $j$  is certain of player  $i$ 's strategy. In particular, if the players are commonly certain of the profile of their type mappings and if each player is certain of her own strategy, then it follows that the players are commonly certain of the strategy profile. The next section examines the role of such meta-certainty assumptions on game-theoretic solution concepts.

---

<sup>9</sup>The converse does not hold, i.e.,  $C = B_I$  does not necessarily imply that the players are certain of the profile of their qualitative-type mappings.

## 5 What Role Does the “Meta-Certainty” of a Model Play in Game-theoretic Analyses?

This section studies the role that the “meta-certainty” assumption plays in game-theoretic analyses of solution concepts. Specifically, it considers the solution concept of iterated elimination of strictly dominated actions (IESDA) in a strategic game. Informally, an epistemic characterization of IESDA states that, in a strategic game, if the (i) “logical” players are (ii) “commonly (meta-)certain of the game” and if they (iii) commonly believe their rationality, then their resulting actions survive IESDA. Formally, in the context of the framework of this paper, [18] shows that if the players commonly believe each player’s rationality and if each of them correctly believes their own rationality, then their resulting actions survive IESDA, without assuming any property on individual players’ beliefs. This paper connects these two statements as follows: first, suppose that the players are logical in that their beliefs satisfy Consistency and Finite Conjunction in addition to Monotonicity. Second, suppose that each of them is certain of their own qualitative-type mapping and strategy. Third, suppose that the players commonly believe their rationality. Then, their resulting actions survive IESDA.

Here I show that the certainty (of her own strategy and type mapping) allows her to correctly believe her own rationality. In other words, if a player is able to reason about informativeness of her own beliefs, she is able to correctly believe her own rationality.

### 5.1 A Strategic Game, a Model of a Game, and Rationality

To define the notion of rationality in a game, define a (*strategic*) *game* as a tuple  $\Gamma = \langle (A_i)_{i \in I}, (\succsim_i)_{i \in I} \rangle$ :  $A_i$  is a non-empty at-most-countable set of player  $i$ ’s actions, and  $\succsim_i$  is  $i$ ’s (complete and transitive) preference relation on  $A := \times_{i \in I} A_i$ .<sup>10</sup> Denote by  $\sim_i$  and  $\succ_i$  the indifference and strict relations, respectively.

A (belief) *model* of the game  $\Gamma$  is a tuple  $\langle (\Omega, \mathcal{D}), (B_i)_{i \in I}, C, (\sigma_i)_{i \in I} \rangle$  (abusing the notation, denote it by  $\vec{\Omega}$ ) with the following two properties. First,  $\langle (\Omega, \mathcal{D}), (B_i)_{i \in I}, C \rangle$  is a belief model. Second,  $\sigma_i : \Omega \rightarrow A_i$  is a *strategy* of player  $i$  satisfying the measurability condition that  $\sigma_i^{-1}(\{a_i\}) \in \mathcal{D}$  for all  $a_i \in A_i$ . Denote  $[\sigma_i(\omega)] := \sigma_i^{-1}(\{\sigma_i(\omega)\})$  for each  $\omega \in \Omega$ .

Denote by  $[a'_i \succsim_i a_i] := \{\omega' \in \Omega \mid (a'_i, \sigma_{-i}(\omega')) \succsim_i (a_i, \sigma_{-i}(\omega'))\} \in \mathcal{D}$  for any  $a_i, a'_i \in A_i$ . In words,  $[a'_i \succsim_i a_i]$  is the event that player  $i$  prefers taking action  $a'_i$  to  $a_i$  given the opponents’ strategies  $\sigma_{-i}$ . The set  $[a'_i \succsim_i a_i]$  is an event because  $[a'_i \succsim_i a_i] = \sigma_{-i}^{-1}(\{a_{-i} \in A_{-i} \mid (a'_i, a_{-i}) \succsim_i (a_i, a_{-i})\}) \in \mathcal{D}$ . Define  $[a'_i \succ_i a_i]$  and  $[a'_i \sim_i a_i]$  analogously.

Denote by  $\text{RAT}_i$  the event that player  $i$  is *rational* (see, e.g., [7, 8, 14]):

$$\text{RAT}_i := \{\omega \in \Omega \mid \omega \in B_i([a'_i \succ_i \sigma_i(\omega)]) \text{ for no } a'_i \in A_i\}.$$

It can be seen that  $\text{RAT}_i$  is indeed an event. Let  $\text{RAT}_I := \bigcap_{i \in I} \text{RAT}_i$ . Player  $i$  is *rational* at  $\omega \in \Omega$  if there is no action  $a'_i \in A_i$  such that she believes that playing  $a'_i$  is strictly better than playing  $\sigma_i(\omega)$  given the opponents’ strategies  $\sigma_{-i}$ . In other words, player  $i$  is rational at  $\omega$  if, for any action  $a'_i$ , she always considers it possible that playing  $\sigma_i(\omega)$  is at least as good as playing  $a'_i$  given the opponents’ strategies  $\sigma_{-i}$ :  $\omega \in (\neg B_i)(\neg[\sigma_i(\omega) \succsim_i a'_i])$  for any  $a'_i \in A_i$ .

Now, the epistemic characterization of IESDA is stated as follows. Suppose that each player  $i$  correctly believes her own rationality:  $B_i(\text{RAT}_i) \subseteq \text{RAT}_i$ . If every player’s rationality is common belief at

<sup>10</sup>The assumption on the cardinality of each action set  $A_i$  is to simplify the analysis. It guarantees that each player is able to reason about any subset of action profiles and that the rationality of each player is an event.

$\omega$ , i.e.,  $\omega \in \bigcap_{i \in I} C(\text{RAT}_i)$ , then the resulting actions  $(\sigma_i(\omega))_{i \in I} \in A$  survive any process of IESDA.<sup>11</sup>

Finally, player  $i$  is *certain of her own strategy*  $\sigma_i$  if she is certain of  $\sigma_i : (\Omega, \mathcal{D}) \rightarrow (A_i, \{\{a_i\} \mid a_i \in A_i\})$ , equivalently,  $[\sigma_i(\cdot)] \subseteq B_i([\sigma_i(\cdot)])$ . Note that, under Consistency in addition to Monotonicity, if player  $i$  is certain of her own strategy then  $B_i([\sigma_i(\cdot)]) = [\sigma_i(\cdot)]$ ,  $[\sigma_i(\cdot)]^c = B_i([\sigma_i(\cdot)]^c)$ , and  $B_i(\Omega) = \Omega$ .<sup>12</sup>

## 5.2 The Role of Meta-certainty in Correctly Believing One's Own Rationality

I ask under what conditions player  $i$  correctly believes her own rationality:  $B_i(\text{RAT}_i) \subseteq \text{RAT}_i$ . For qualitative belief, the standard assumptions on qualitative belief (i.e., Consistency, Positive Introspection, Negative Introspection, and the Kripke property) guarantee that  $B_i(\text{RAT}_i) = \text{RAT}_i$  (e.g., [7, 8]).<sup>13</sup> Here, I provide a compatibility condition on belief with informativeness, under which a player correctly believes her own rationality. The compatibility condition does not hinge on a particular form of belief, i.e., whether it is qualitative or probabilistic.

To that end, a state  $\omega$  is *at least as informative as* another state  $\omega'$  to  $i$  (precisely, according to  $t_{B_i}$ ) iff  $t_{B_i}(\omega')(\cdot) \leq t_{B_i}(\omega)(\cdot)$ . Fix  $\omega \in \Omega$ , and let  $(\uparrow t_{B_i}(\omega)) := \{\omega' \in \Omega \mid t_{B_i}(\omega)(\cdot) \leq t_{B_i}(\omega')(\cdot)\}$  be the set of states that are at least as informative to  $i$  as  $\omega$ . Under the Kripke property,  $\omega' \in (\uparrow t_{B_i}(\omega))$  iff  $b_{B_i}(\omega') \subseteq b_{B_i}(\omega)$ . The full paper extensively studies the notion of informativeness. Now:

**Definition 2.** Player  $i$ 's belief (operator  $B_i$ ) is *compatible with informativeness* if  $(\uparrow t_{B_i}(\omega)) \cap E \neq \emptyset$  for any  $E \in \mathcal{D}$  with  $\omega \in B_i(E)$ .

In words, player  $i$ 's beliefs are compatible with informativeness if, for any event  $E$  which player  $i$  believes at some  $\omega$ , there exists a state  $\omega'$  in  $E$  which is at least as informative as  $\omega$ . In the context of qualitative beliefs, if player  $i$ 's belief operator  $B_i$  satisfies the Kripke property, Consistency, and Positive Introspection, then  $B_i$  is compatible with informativeness. The compatibility with informativeness does not necessarily imply the Kripke property (and vice versa). If player  $i$ 's belief operator  $B_i$  is compatible with informativeness, then it satisfies  $B_i(\emptyset) = \emptyset$ . Thus, under Finite Conjunction, if  $B_i$  is compatible with informativeness, then it satisfies Consistency.

The following proposition states that the compatibility of beliefs with informativeness is implied by the certainty of a type mapping.

**Proposition 5.** Let  $\vec{\Omega}$  be a belief model. Assume: (i)  $(\uparrow t_{B_i}(\cdot)) \in \mathcal{D}$ ; (ii)  $B_i$  satisfies Consistency and Finite Conjunction; and that (iii) player  $i$  is certain of  $t_{B_i} : \Omega \rightarrow M(\Omega)$  with respect to  $\{\{\mu \in M(\Omega) \mid \mu(\cdot) \geq t_{B_i}(\omega)(\cdot)\} \mid \omega \in \Omega\}$ . Then,  $B_i$  is compatible with informativeness.

The proposition states that, under the regularity condition (i), if player  $i$  is logical (in that her belief operator satisfies Consistency and Finite Conjunction) and if she is certain of her qualitative-type mapping, then her beliefs are compatible with informativeness. Theorem 2 below establishes that if player  $i$ 's beliefs are compatible with informativeness then she correctly believes her rationality, which is a part of the preconditions of the epistemic characterization of IESDA.

<sup>11</sup>Since each player's belief operator  $B_i$  satisfies Monotonicity,  $B_i(\text{RAT}_I) \subseteq \bigcap_{i \in I} B_i(\text{RAT}_i)$  and  $C(\text{RAT}_I) \subseteq \bigcap_{i \in I} C(\text{RAT}_i)$ . Thus, if every player  $i$  correctly believes the rationality of the players, then each player correctly believes her own rationality. Likewise, if it is common belief that the players are rational, then, for every  $i \in I$ , it is common belief that player  $i$  is rational. Hence, I examine the weaker condition that each player  $i$  correctly believes her own rationality.

<sup>12</sup>Thus, under Consistency and Monotonicity of  $B_i$ , the certainty of own strategy implies that if player  $i$  is rational at  $\omega$ , then she never takes a strictly dominated action at  $\omega$  (if she takes a strictly dominated action, then her belief violates Necessitation).

<sup>13</sup>It can be seen that Consistency, Positive Introspection, and the Kripke property in addition to the certainty of  $i$ 's own strategy yield  $B_i(\text{RAT}_i) \subseteq \text{RAT}_i$ . Likewise, Negative Introspection and the Kripke property in addition to the certainty of  $i$ 's own strategy yield  $\text{RAT}_i \subseteq B_i(\text{RAT}_i)$ .

Now, the main result of this section is as follows: a player correctly believes her own rationality if: (i) she is certain of her own strategy; (ii) her belief is compatible with the informativeness; and if (iii) her belief is (finitely) conjunctive so that she can simultaneously reason about her own strategy and her own rationality.

**Theorem 2.** *Suppose that player  $i$  is certain of her own strategy (i.e.,  $[\sigma_i(\cdot)] \subseteq B_i([\sigma_i(\cdot)])$ ). Also, let  $B_i$  be compatible with informativeness and satisfy Finite Conjunction. Then, player  $i$  correctly believes her own rationality:  $B_i(\text{RAT}_i) \subseteq \text{RAT}_i$ .*

Proposition 5 and Theorem 2 imply that player  $i$  correctly believes her own rationality if she is logical in that her belief operator satisfies Consistency and Finite Conjunction and if she is certain of her own type mapping and strategy. Theorem 2 states that, for the role of the meta-certainty assumption of a belief model on IESDA, it is not necessary that each player is certain of the profile of type mappings but it is sufficient that each player is certain of her own type mapping. One can incorporate the assumptions that each player is certain of her own qualitative type-mapping and strategy into the condition that she is certain of the part of the model of a game  $\langle (\Omega, \mathcal{D}), (t_{B_i}, \sigma_i) \rangle$  that dictates her beliefs and strategy.

## References

- [1] Robert J. Aumann (1976): *Agreeing to Disagree*. *The Annals of Statistics* 4(6), pp. 1236–1239, doi:10.1214/aos/1176343654.
- [2] Robert J. Aumann (1987): *Correlated Equilibrium as an Expression of Bayesian Rationality*. *Econometrica* 55(1), pp. 1–18, doi:10.2307/1911154.
- [3] Robert J. Aumann (1999): *Interactive Epistemology I, II*. *International Journal of Game Theory* 28(3), pp. 263–300, 301–314. doi:10.1007/s001820050111, doi:10.1007/s001820050112.
- [4] Michael Bacharach (1985): *Some Extensions of a Claim of Aumann in an Axiomatic Model of Knowledge*. *Journal of Economic Theory* 37(1), pp. 167–190, doi:10.1016/0022-0531(85)90035-3.
- [5] Michael Bacharach (1990): *When Do We Have Information Partitions?* In M. Bacharach & M. Dempster, editors: *Mathematical Models in Economics*, Oxford University Press, pp. 1–23.
- [6] K. Binmore & A. Brandenburger (1990): *Common Knowledge and Game Theory*. In K. Binmore, editor: *Essays on the Foundations of Game Theory*, Basil Blackwell, pp. 105–150.
- [7] Giacomo Bonanno (2008): *A Syntactic Approach to Rationality in Games with Ordinal Payoffs*. In Giacomo Bonanno, Wiebe van der Hoek & Michael Wooldridge, editors: *Logic and the Foundations of Game and Decision Theory (LOFT 7)*, Amsterdam University Press, pp. 59–86, doi:10.5117/9789089640260.
- [8] Giacomo Bonanno (2015): *Epistemic Foundations of Game Theory*. In Hans van Ditmarsch, Joseph Y. Halpern, Wiebe van der Hoek & Barteld Pieter Kooi, editors: *Handbook of Epistemic Logic*, College Publications, pp. 443–487.
- [9] Adam Brandenburger & Eddie Dekel (1987): *Rationalizability and Correlated Equilibria*. *Econometrica* 55(6), pp. 1391–1402, doi:10.2307/1913562.
- [10] Adam Brandenburger & Eddie Dekel (1989): *The Role of Common Knowledge Assumptions in Game Theory*. In Frank Hahn, editor: *The Economics of Missing Markets, Information, and Games*, Oxford University Press, pp. 46–61.
- [11] Adam Brandenburger & Eddie Dekel (1993): *Hierarchies of Beliefs and Common Knowledge*. *Journal of Economic Theory* 59(1), pp. 189–198, doi:10.1006/jeth.1993.1012.
- [12] Adam Brandenburger, Eddie Dekel & John Geanakoplos (1992): *Correlated Equilibrium with Generalized Information Structures*. *Games and Economic Behavior* 4(2), pp. 182–201, doi:10.1016/0899-8256(92)90014-J.

- [13] Adam Brandenburger & H. Jerome Keisler (2006): *An Impossibility Theorem on Beliefs in Games*. *Studia Logica* 84(2), pp. 211–240, doi:10.1007/s11225-006-9011-z.
- [14] Yi-Chun Chen, Ngo Van Long & Xiao Luo (2007): *Iterated Strict Dominance in General Games*. *Games and Economic Behavior* 61(2), pp. 299–315, doi:10.1016/j.geb.2007.02.002.
- [15] Eddie Dekel & Faruk Gul (1997): *Rationality and Knowledge in Game Theory*. In David M. Kreps & Kenneth F. Wallis, editors: *Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress*, 1, Cambridge University Press, pp. 87–172, doi:10.1017/CCOL521580110.005.
- [16] Ronald Fagin, John Geanakoplos, Joseph Y. Halpern & Moshe Y. Vardi (1999): *The Hierarchical Approach to Modeling Knowledge and Common Knowledge*. *International Journal of Game Theory* 28(3), pp. 331–365, doi:10.1007/s001820050114.
- [17] Satoshi Fukuda (2017): *The Existence of Universal Knowledge Spaces*. In: *Essays in the Economics of Information and Epistemology*, Ph.D. Dissertation, the University of California at Berkeley, pp. 1–113.
- [18] Satoshi Fukuda (2020): *Formalizing Common Belief with No Underlying Assumption on Individual Beliefs*. *Games and Economic Behavior* 121, pp. 169–189, doi:10.1016/j.geb.2020.02.007.
- [19] Satoshi Fukuda (2021): *The Existence of Universal Qualitative Belief Spaces*.
- [20] John Geanakoplos (1989): *Game Theory without Partitions, and Applications to Speculation and Consensus*. Cowles Foundation Discussion Paper No. 914, Yale University.
- [21] Itzhak Gilboa (1988): *Information and Meta-Information*. In: *Proceedings of the 2nd Conference on Theoretical Aspects of Reasoning about Knowledge*, Morgan Kaufmann Publishers Inc., pp. 227–243.
- [22] John C. Harsanyi (1967-1968): *Games with Incomplete Information Played by “Bayesian” Players, I-III*. *Management Science* 14, pp. 159–182, 320–334, 486–502. doi:10.1287/mnsc.14.3.159, doi:10.1287/mnsc.14.5.320, doi:10.1287/mnsc.14.7.486.
- [23] Dov Monderer & Dov Samet (1989): *Approximating Common Knowledge with Common Beliefs*. *Games and Economic Behavior* 1(2), pp. 170–190, doi:10.1016/0899-8256(89)90017-1.
- [24] Stephen Morris (1996): *The Logic of Belief and Belief Change: A Decision Theoretic Approach*. *Journal of Economic Theory* 69(1), pp. 1–23, doi:10.1006/jeth.1996.0035.
- [25] Cesaltina Pacheco Pires (1994): *Do I know  $\Omega$ ? An Axiomatic Model of Awareness and Knowledge*.
- [26] Olivier Roy & Eric Pacuit (2013): *Substantive assumptions in interaction: a logical perspective*. *Synthese* 190(5), pp. 891–908, doi:10.1007/s11229-012-0191-y.
- [27] Dov Samet (1990): *Ignoring Ignorance and Agreeing to Disagree*. *Journal of Economic Theory* 52(1), pp. 190–207, doi:10.1016/0022-0531(90)90074-T.
- [28] Hyun Song Shin (1993): *Logical Structure of Common Knowledge*. *Journal of Economic Theory* 60(1), pp. 1–13, doi:10.1006/jeth.1993.1032.
- [29] Robert Stalnaker (1994): *On the Evaluation of Solution Concepts*. *Theory and Decision* 37(1), pp. 49–73, doi:10.1007/BF01079205.
- [30] Tommy Chin-Chiu Tan & Sérgio Ribeiro da Costa Werlang (1988): *The Bayesian Foundations of Solution Concepts of Games*. *Journal of Economic Theory* 45(2), pp. 370–391, doi:10.1016/0022-0531(88)90276-1.
- [31] Tommy Chin-Chiu Tan & Sérgio Ribeiro da Costa Werlang (1992): *On Aumann’s Notion of Common Knowledge: an Alternative Approach*. *Revista Brasileira de Economia - RBE* 46(2), pp. 151–166.
- [32] Spyros Vassilakis & Shmuel Zamir (1993): *Common Belief and Common Knowledge*. *Journal of Mathematical Economics* 22(5), pp. 495–505, doi:10.1016/0304-4068(93)90039-N.
- [33] Sérgio Ribeiro da Costa Werlang (1987): *Common Knowledge*. In John Eatwell, Murray Milgate & Peter Newman K., editors: *The New Palgrave: A Dictionary of Economics*, 2nd edition, Macmillan, pp. 74–85, doi:10.1007/978-1-349-20181-5\_5.

- [34] Robert Wilson (1987): *Game-Theoretic Analyses of Trading Processes*. In Truman Fassett Bewley, editor: *Advances in Economic Theory: Fifth World Congress*, Cambridge University Press, pp. 33–70, doi:10.1017/CCOL0521340446.002.