Title
Essays in the Economics of Information and Epistemology

Author
Fukuda, Satoshi

Publication Date
2017

Essays in the Economics of Information and Epistemology

by

Satoshi Fukuda

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David S. Ahn, Chair
Professor William Fuchs
Professor Chris Shannon

Summer 2017

Essays in the Economics of Information and Epistemology

Abstract

Essays in the Economics of Information and Epistemology

by

Satoshi Fukuda

Doctor of Philosophy in Economics

University of California, Berkeley

Professor David S. Ahn, Chair

This dissertation consists of three chapters exploring the role that information plays in strategic situations. The first two chapters are devoted to modeling a decision maker's reasoning (knowledge, beliefs, and unawareness) about those of other decision makers in strategic environments. The third chapter, in turn, studies how information and such reasoning affect economic behaviors.

The first chapter, *The Existence of Universal Knowledge Spaces*, provides a formal framework which enables us to analyze players' interactive knowledge and beliefs in a strategic context with a number of desirable features. First, the framework can specify players' logical and introspective properties of knowledge as well as depth of their reasoning. In other words, the framework admits various forms of introspective and non-introspective knowledge as well as beliefs which may fail the truth axiom of knowledge. Second, the framework generalizes previous interactive knowledge and belief models. Especially, it proposes notions of common knowledge and common belief without assuming any property on individual knowledge and belief. The main result of the first chapter is to propose a canonical representation of players' knowledge and beliefs under this general framework. The canonical model is universal in the sense that it "contains" any other particular representation.

The second chapter, *Representing Unawareness on State Spaces*, examines notions of unawareness in terms of the lack of knowledge within the framework of a standard state space model. For example, if a notion of unawareness is defined as the two levels of the lack of knowledge, then a player is unaware of an event when she does not know it and she does not know that she does not know it. In this way, this chapter studies notions of unawareness by a layer of the lack of knowledge and by underlying properties of knowledge. The central questions of the chapter are as follows. When and how does a standard state space model have a sensible form of unawareness? How does unawareness relate to notions of ignorance and possibility? The results are as follows. First, any notion of unawareness reduces to the following two forms. A strong form of unawareness states that a player is unaware of an event when she is ignorant of the possibility that she knows it. A weak form of unawareness states that a player is unaware of an event when she is ignorant of the fact that she knows it.

Second, if a player is unaware of an event, then she is ignorant of being unaware of it. Third, if a player faces an infinite number of objects of knowledge, then it is possible that she knows that there is an event of which she is unaware, while she cannot know that she is unaware of any particular event. Fourth, unawareness is not necessarily monotone in knowledgeability in the sense that getting more information can lead a player to becoming unaware of some event.

The third chapter, *From Equals to Despots: The Dynamics of Repeated Decision Making in Partnerships with Private Information*, is a joint work with Vinicius Carrasco and William Fuchs. It considers the optimal dynamic renegotiation-proof mechanism among a group of privately informed agents who repeatedly take a joint common action but who are unable to resort to side-payments. The chapter provides a general framework which accommodates as special cases committee decision and collective insurance problems. Thus, it formally connects these separate strands of literature. In such a collective decision-making situation, the players may be tempted to exaggerate their preferred actions in order to manipulate the group action. While the first-best values can never be exactly attained in an incentive compatible way, the cost of incentives approximately disappears as the players become patient. In the optimal mechanism, a player who has a "strong" preference shock can influence a current joint action at the cost of forgoing continuation utilities. Such intertemporal trade-off in the optimal mechanism leads to the variations in the players' decision rights in a way such that they increase in expectation over time.

To my parents, for their fullest trust, encouragement, and support.

Table of Contents

Acknowledgments

In preparing this dissertation, I have been extremely fortunate to have received the tremendous professional and personal support from my advisers, colleagues, family, and friends. Thanks to all of them, I have been able to have a productive and wonderful time here at Berkeley. It is a real pleasure and honor acknowledging their support and companionship.

My advisers, David Ahn, William Fuchs and Chris Shannon, have been unstinting in their support during the entire duration of my graduate studies. David has always been giving me honest and invaluable advice, and he has been responsible for maintaining the active theory community at Berkeley. I am especially grateful for his advice on and help with my writing. Willie has introduced to me dynamic mechanism design/information economics through his second-year course, reading group, and personal interactions. Especially, I have learned a great deal from his comments and economic insights while working together. Whenever I meet and talk with Chris, it becomes always clear in which direction I should pursue in my research, as she always lets me understand both what is good and what needs to be improved in my research. I am especially grateful for her generosity and warm encouragement. All these experiences are gifts that my advisers have given me, and I hope I can keep producing ideas in information economics and epistemic game theory in order to better understand the role of information in economic and strategic interactions and in order to contribute to the profession. I believe this is one way that I can repay their invaluable support and kindness.

The third chapter of this dissertation is a joint work with Vinicius Carrasco and William Fuchs. I am delighted to incorporate our joint work in this dissertation. More than anything, I am very pleased with working together and sharing economic thoughts and intuitions.

I would also like to thank my co-author, Yuichiro Kamada, for wonderful time discussing and working on economic and game theory together. I am very lucky to have learned a lot from him through our discussions. During the first two years of coursework, I am very fortunate to have learned microeconomic and game theory from Chris Shannon, David Ahn, Haluk Ergin, John Morgan, Shachar Kariv, Benjamin Hermalin, and William Fuchs. I would also like to thank the graduate advisers, Patrick Allen, Victoria Lee, and Anna Cross for their support and efforts for all of us in the Economics graduate program.

I would like to cordially acknowledge financial supports from my advisers through research assistantship, the Economics department, the University, the 21st Century Cultural and Academic Foundation (Kikawada Foundation) in Japan, and the Japan-IMF Scholarship Program. The Japan-IMF Scholarship Program also gave me an opportunity to work at the IMF as a summer intern in 2014. I enjoyed very much working with Geert Almekinders and Alex Mourmouras as my supervisors there.

I have spent a tremendous amount of time in 668 (formerly known as 608-16) Evans Hall. Indeed, I am writing this acknowledgement here in 668. Among wonderful office-mates, I would especially like to thank my fellow theorists — Ivan Balbuzanov, Aluma Dembo, Matthew Leister, Michèle Müller-Itten, and Joseph Root for their companionship. I would also like to thank Takeshi Murooka, Aniko Öry, and Antonio Rosato of the theory student group.

My interests in information economics and epistemic game theory are definitely rooted in the six years of undergraduate and graduate studies at Chuo University in Japan. I am very pleased with acknowledging Hiroaki Hayakawa, Mitsuo Kono, Masaki Tomochi, and Akira Yokoyama for their encouragement and support in this dissertation.

Wonderful friendship with my colleagues at various departments (not only Economics but also Haas and ARE) of Berkeley is a precious asset for me. Also, I do appreciate friends of mine in this wonderful life journey for their companionship at various stages of my life so far. Although I do not name each of them here as it is a very difficult task, I am very grateful to all of you for their companionship. I hope I can convey my gratitude to each of you in some way or another.

Last but not least, I would like to thank my family members, Koichi, Junko, and Natsumi. It is extremely difficult, indeed impossible, to express how grateful I am. I would like instead to dedicate this dissertation to my parents, Junko and Koichi.

The Existence of Universal Knowledge Spaces*

Satoshi Fukuda

Abstract

We provide a formal framework capable of capturing players' interactive knowledge in a strategic context with a number of desirable features. First, we can specify players' logical and introspective abilities as well as the language that they can use in their reasoning. Second, our framework admits tractable representations of players' knowledge and common knowledge and nests previous interactive knowledge models. The main result of the paper is that the framework admits a canonical representation of players' knowledge. The canonical model is a "largest" model of knowledge to which any particular knowledge space can be mapped in a unique knowledge-preserving way.

JEL Classification: C70, D83
Keywords: Interactive Knowledge, Universal Knowledge Space, Hierarchies of Knowledge

*The author thanks David Ahn, William Fuchs, and Chris Shannon for encouragement, support, and guidance; Robert Anderson, Eric Auerbach, Yu Awaya, Ivan Balbuzanov, Paulo Barelli, Pierpaolo Battigalli, Benjamin Brooks, Christopher Chambers, Aluma Dembo, Eduardo Faingold, Amanda Friedenberg, Drew Fudenberg, Benjamin Golub, Sander Heinsalu, Ryuichiro Ishikawa, Yuichiro Kamada, Michihiro Kandori, Shachar Kariv, Ali Khan, Narayana Kocherlakota, Massimo Marinacci, Stephan Morris, Motty Perry, Herakles Polemarchakis, Joseph Root, Burkhard Schipper, Mikkel Sølvsten, Kenji Tsukada, and various seminar participants for comments and discussions. All remaining errors are mine.

Contents

1 Introduction

Consider a group of players who reason interactively about unknown parameter values, such as the payoffs and strategies in a game. The value of such exogenous parameters is called a state of nature. Each player must reason about states of nature and about each other's knowledge about states of nature, and so on. In this paper, we construct a first formal framework general enough to represent any conceivable form of such interactive knowledge. An arbitrary structure will capture some possible aspects of players' interactive reasoning but will generally exclude others. We construct a sufficiently rich knowledge space that includes all possible forms of reasoning about knowledge. The construction enables analysis of formal questions regarding knowledge without ad hoc restrictions on the nature of reasoning. Then claims regarding knowledge can be logically disassociated from extraneous structure that is not immediately related to interactive knowledge.

A model of knowledge consists of the following three ingredients. The first ingredient is a set Ω. Each element $\omega \in \Omega$ is a list of possible specifications of the realization of the exogenous parameters (i.e., the prevailing state of nature $s \in S$, where S is the set of unknown parameter values) and players' interactive knowledge regarding states of nature S (i.e., their knowledge about states of nature S, their knowledge about their knowledge about S, and so on). Call each specification ω a state of the world.

The second ingredient is the set of statements about which the players can reason. These statements, which we call events of the world, are specified as subsets of states of the world Ω. Thus a model of knowledge requires a description of the language available to the players, modeled as a collection of subsets of Ω which we call the domain.

The third ingredient is players' knowledge operators defined on the domain. For each event of the world E, player i's knowledge operator assigns the set of states at which she knows E, i.e., the event that i knows E. By iterative application of this operator, we can unpack higher orders of interactive reasoning.

In sum, a model of knowledge (a knowledge space) consists of the description of states of the world, the description of events of the world, and players' knowledge operators.

In applications, typically a specific model of knowledge is assumed a priori. This leaves open the possibility that some relevant aspects of reasoning are excluded. To address this, we construct a canonical representation of players' knowledge. This space is universal in the sense that any other knowledge space is embedded in it in a unique manner that maintains all the structure of that smaller space. That is, any form of reasoning in the smaller space can be retrieved in the universal space in a unique way. At the same time, we prove that the space is complete in the sense of including all possible forms of reasoning. It reveals what form of reasoning is indeed lost in the specific smaller space.

The main result of the paper (Theorem 1 in Section 3) is to demonstrate the existence of a universal knowledge space. The existence of a universal knowledge space ensures that players' interactive knowledge in a strategic situation can be modeled by the knowledge space approach without neglecting any form of reasoning.

We also demonstrate the applicability of our framework. First, we establish the existence of a universal knowledge space under a variety of assumptions on players' logical and introspective abilities. Our result is theoretically interesting in that the existence of a universal knowledge space is unrelated to assumptions on players' knowledge. At the same time, it is substantively interesting because we establish the canonical representation of knowledge even when players are less than "perfectly rational" in terms of their logical and introspective abilities.

Second, we show (in Section 4) that our framework subsumes alternative representations of knowledge and common knowledge.[1] We demonstrate that our framework nests previous frameworks. Thus, our approach explicitly clarifies what implicit assumptions on players' knowledge are imposed in a specific previous framework. We can also reproduce existent theorems regarding knowledge and common knowledge obtained in a specific previous framework.

We identify the conditions on players' knowledge operators under which their knowledge is represented as information sets on the underlying states of the world, where each player's information set associated with a state of the world represents the set of states of the world she considers possible at that state. That is, our framework nests the most standard model of knowledge, known as a partition model of knowledge (Aumann [3]).

In relation to the first point, we can also accommodate other previous models of knowledge. For example, the players may not fully be introspective about what they do not know. Our framework nests non-partitional models (also called possibility correspondence models) of knowledge which attempt to relax this introspective property called Negative Introspection.[2] There, each player's knowledge at each state of the world is represented as a general information set, which may not necessarily constitute a partition of the states of the world. The study of non-partitional models ranges from implications of common knowledge (e.g., Agreement theorems (Aumann [3])) to solution concepts in game theory.[3]

Moreover, our framework also nests other forms of possibility correspondence models. One such example is a situation where players' knowledge may not necessarily be true.[4]

---

[1] Pioneering attempts to model common knowledge include, but are not limited to, Aumann [3], Friedell [29], Lewis [48], and McCarthy, Sato, Hayashi, and Igarashi [53]. We show that various previous formalizations of common knowledge can be nested in our framework. Hence, we can analyze, within our framework, implications of common knowledge on strategic situations.

[2] Non-partitional models are motivated in part by notions of unawareness. The pioneering studies of unawareness include Fagin and Halpern [27], Modica and Rustichini [62, 63], and Pires [70]. See also Schipper [79] for a recent overview of the field. While Modica and Rustichini [62] and Dekel, Lipman, and Rustichini (hereafter, DLR) [21] demonstrate that standard state space models (including non-partitional models) have limitations in representing unawareness, in the second chapter of this dissertation (Fukuda [30]), we study how standard state space models can (and cannot) represent notions of unawareness within our framework.

[3] See, for example, Bacharach [6], Binmore and Brandenburger [11], Brandenburger, Dekel, and Geanakoplos (hereafter, BDG) [17], Dekel and Gul [20], Geanakoplos [31], Morris [64], Rubinstein and Wolinsky [72], Samet [75, 76], and Shin [80].

[4] In the literature, knowledge is qualitatively distinguished from belief in that a player can only know

Thus we can deal with qualitative belief as well as knowledge. Indeed, we can endow each player with both (qualitative) belief and knowledge.[5] Our general model clarifies the assumptions regarding logical sophistication that are formally implicit in possibility correspondence models. We can further relax players' logical abilities in that they can fail to know a logical consequence of their knowledge.

Third, while players' interactive knowledge is an important consideration in strategic contexts, epistemic analyses of strategic situations call for both knowledge and probabilistic belief.[6] Our framework admits a probability (precisely, a measurable) space as a domain of players' knowledge, and thus we can incorporate players' probabilistic assessments into the framework. Indeed, the consideration of the domain of a knowledge space has often been neglected. In standard models of knowledge in the previous literature, any subset of states of the world Ω is considered to be an object of knowledge. Under such specification, knowledge and probabilistic beliefs may be incompatible with each other when the knowledge of an event is not in the domain of the probability space. This incompatibility has been one of the issues that have hampered the epistemic analyses of players' knowledge and belief.

Fourth, we pose the following conceptual question: can we formalize the sense in which the players know (or commonly know) the structure of a model of knowledge itself?[7] We provide a formal test (in Section 5) according to which we (i.e., the outside analysts) can say that the players know (indeed, commonly know) the structure of a model of knowledge itself, provided that they are introspective about their own knowledge and that assumptions on their knowledge are homogeneous. In this test, we regard each player's knowledge as a signal and ask whether each player knows the signal that generates each other's knowledge. We show that each player knows her own signal when her knowledge is introspective. In the universal knowledge space, moreover, no relevant aspect of players' interactive knowledge is left unspecified. If assumptions on players' knowledge are homogeneous, each player has an identical knowledge operator. We show that the players (in a formal sense) know (and commonly know) each other's signal that generates their knowledge.

---

what is true while she can believe something false. Since we deal with a wide variety of assumptions on "knowledge" at the same time, by abusing the terminologies, we refer to such a belief as knowledge in a generic context. In a specific application, however, the distinction between knowledge and belief should be made.

[5]Such consideration would be needed if, for example, we analyze each player's knowledge about her own strategy and her belief about her opponents' strategies (e.g., Dekel and Gul [20]). The analysis of an extensive form game would also call for knowledge and belief if we analyze players' knowledge about their past observed moves and their beliefs about past unobserved moves and future moves (e.g., Battigali and Bonanno [9]). See also Dekel and Gul [20, Section 5].

[6]Once we incorporate both knowledge and probabilistic belief, a distinction between knowledge and probability-one belief would emerge in a continuous model. For example, an agent believes with probability one that a random number drawn from the interval $[0, 1]$ is irrational while she does not know it (Monderer and Samet [61]).

[7]This question has been puzzling a number of economists and game theorists. See such discussions by Aumann [3, 4, 5], Bacharach [6, 7], Binmore and Brandenburger [11], Brandenburger and Dekel [16], BDG [17], Dekel and Gul [20], Fagin, Geanakoplos, Halpern, and Vardi (hereafter, FGHV) [26], Gilboa [32], Pires [70], Roy and Pacuit [71], Tan and Werlang [82].

In conclusion, we provide a framework capable of capturing players' interactive knowledge in the following ways: (i) the framework contains a universal knowledge space; (ii) the framework admits a wide variety of assumptions on players' logical and introspective abilities; and (iii) the framework allows us to specify the language that the players can use through selecting feasible domains. Moreover, the framework is amenable to tractable representations of players' knowledge and common knowledge in the sense that it nests and generalizes previous models of interactive knowledge.

The paper is organized as follows. The rest of this section reviews the previous literature and provides a technical overview of the main result. Section 2 defines a knowledge space, properties of knowledge, and a universal knowledge space. Section 3 demonstrates that our framework admits a universal knowledge space (Theorem 1).

Section 4 provides tractable representations of players' knowledge and common knowledge. In Section 5, we formally ask the sense in which the players know the structure of a model of knowledge.

Section 6 carries out an alternative construction of a universal knowledge space (Theorem 3) by formalizing a notion of hierarchies of knowledge. Section 7 provides applications to richer settings involving dynamics and other epistemic notions. Throughout the paper, all the proofs are relegated to Appendix A.

1.1 Literature Review

The original economic model of knowledge is Aumann's [3] partitional model. While the partitional approach has been prevalent, this approach poses the following three concerns.

First, a standard partitional knowledge space has, as its domain, the power set of Ω. This class of standard partitional knowledge spaces cannot be closed in the following sense: there is no standard partitional knowledge space that includes all standard partitional knowledge spaces. In the next subsection, we will discuss these negative results and how we circumvent them in more detail.

The second issue is the sophisticated logical and introspective abilities that the standard partitional models render to players. A player whose knowledge is dictated by a partition is logically omniscient. Moreover, she is fully introspective about what she knows and what she does not know.

Third, partitions and probabilistic beliefs may be incompatible with each other. While players' probabilistic beliefs are defined on measurable events, the domain of knowledge is all subsets. Thus, players' knowledge may not be an object of their beliefs.

We can identify conditions on knowledge operators under which each player's knowledge is represented as a partition (more generally, a possibility correspondence). Thus, our framework nests partitional and non-partitional knowledge spaces in a way such that there is a universal partitional (non-partitional) knowledge space defined on a general domain.

How does the existence of a universal knowledge space relate to the various strands of the universal type space literature? Harsanyi [35] proposed the notion of type. Each player's type summarizes her probabilistic belief about exogenous parameter values and players' types (usually, types of her opponents) themselves. An arbitrary type space induces hierarchies of probabilistic beliefs. The question arises as to whether there exists a canonical (universal) type space which contains any other type space.

In their pioneering work, Armbruster and Böge [2], Böge and Eisele [12], and Mertens and Zamir [59] establish a universal belief/type space consisting of coherent hierarchies of beliefs with certain topological assumptions on unlderlying states of nature.[8]

Heifetz and Samet (hereafter, HS) [39], whose approach we follow, establish a universal type space without topological assumptions on states of nature. Their logical approach is extended to establishing a universal finitely additive belief space and a universal knowledge-belief space by Meier [55, 56].[9]

A special kind of a type structure is a possibility structure (Brandenburger [14] and Brandenburger and Keisler [18]) used in epistemic analyses of games. Possibility structures model players' probabilistic and non-probabilistic interactive beliefs. Mariotti, Meier, and Piccione [51] show the existence of a universal possibility structure by imposing topological restrictions on players' non-probabilistic beliefs. As a related model, Salonen [74] establishes a universal belief hierarchies where each agent's type induces a filter over basic propositions (which constitute a Boolean algebra) and the set of opponents' types.

Partly in order to make a formal connection between the knowledge space and type space approaches, we define in Section 5.1 a notion that we call a knowledge-type. Each knowledge-type is a mapping from events into the true-or-false statements about the knowledge of an event. Thus, each player's knowledge-type at each state describes her knowledge of events at that state. A knowledge-type mapping is then a mapping from a given set of states

[8]Their results have been extended under weaker topological assumptions by Brandenburger and Dekel [16], Heifetz [36], Mertens, Sorin, and Zamir [58], and Pintér [69]. Pintér [69] dispenses with a topological assumption on states of nature in his construction of the universal type space consisting of coherent hierarchies. Instead, he imposes a suitable topology on each higher order space consisting of probability measures. Extensions to richer structures include conditional probability systems (Battigalli and Siniscalchi [10]) and ambiguous beliefs (Ahn [1]).

[9]The following two strands of literature also extend HS's [39] probabilistic universal type space. First, Moss and Viglizzo [65, 66] reformulate and generalize σ-additive type spaces as coalgebras for the endofunctor which we simply denote by F (see Moss and Viglizzo [65, Definition 2.1] for the precise definition of F). Then, a universal type space is reformulated as a final (terminal) coalgebra. They show that the set of descriptions of each point (type profile together with a state of nature) in all coalgebras, endowed with measurable and coalgebra structures, constitutes a final coalgebra. Furthermore, by invoking the Lambek lemma in category theory (Lambek [47]), there is an isomorphism between a final coalgebra T and its image of the endofunctor $F(T)$, which establsihes the "(belief-)completeness" of T (see Brandenburger [14] and Brandenburger and Keisler [18] for (belief-)completeness). Second, Meier [57] axiomatizes classes of belief/type spaces and shows that the space of all maximally consistent sets of formulas of his infinitary probability logic (i.e., the canonical space) is a universal space, which is isomorphic to the universal type space constructed by HS [39]. He also shows that, within the class of (product) type spaces, the canonical product type space is (belief-)complete.

of the world to knowledge-types.[10] The given set of states of the world, however, may not have a product structure. Specifying a knowledge-type mapping is equivalent to specifying a knowledge operator. In Section 6 (Theorem 3), by extending HS's [39] hierarchical approach, we demonstrate a universal knowledge space in terms of hierarchies of knowledge-types. Each world in the resulting universal knowledge space consists of a state of nature and a profile of players' hierarchies of knowledge-types. We also characterize (in Theorem 4 in Section 6.2) our universal knowledge space in terms of coherent hierarchies of knowledge.

1.2 Technical Overview

HS [38] demonstrate that a universal standard partitional knowledge space generically does not exist, where recall that a standard partitional knowledge space allows any subset of underlying states of the world to be an object of knowledge. They show that, unlike σ-additive probabilistic beliefs, a non-trivial sequence of interactive knowledge can develop beyond any order (precisely, beyond any given ordinal number). The negative results are also obtained by Fagin [25], Fagin, Geanakoplos, Halpern, and Vardi (FGHV) [26], Fagin, Halpern, and Vardi (hereafter, FHV) [28], and HS [41], where they explicitly formalize notions of coherent hierarchies of knowledge. Moreover, Meier [54] shows, by invoking the result of HS [38] above, that there is no universal knowledge space even when knowledge is represented by a more general non-partitional structure. If there were a universal knowledge space in a class of such general knowledge spaces, then one could construct a universal partitional knowledge space from the given class, which is impossible.

How do our positive results reconcile with the negative results? What plays a crucial role in establishing a universal knowledge space is to specify a set algebra as objects of players' knowledge, i.e., a specification of the language that the players are allowed to use in their reasoning.

To see this point, let κ be an infinite cardinal number. We define a κ-complete algebra on underlying states of the world Ω to be a collection of subsets of Ω with the closure under complementation and under union and intersection of any sub-collection with cardinality less than κ. Thus, if a certain set is an object of knowledge, then so is its complement. Also, if each of a collection of events is an object of knowledge, then so are its union and intersection, provided that the collection has cardinality less than κ. The power set of Ω is always κ-complete. For example, a κ-complete algebra subsumes an algebra of sets if κ is the least infinite cardinal number. Likewise, a κ-complete algebra subsumes a σ-algebra if κ is the least uncountable cardinal number. With this definition in mind, we call a knowledge space a κ-knowledge space if its domain is a κ-complete algebra.

Specifying the domain of each knowledge space by a κ-complete algebra amounts to determining the language available to the players in reasoning about their interactive knowl-

[10] In Sections 5.2 and 5.3, we regard each player's knowledge-type mapping as a signal mapping that generates her knowledge in order to formally ask the (meta-)knowledge of the structure of a model.

edge regarding nature.[11] Namely, any κ-knowledge space can capture players' interactive knowledge of a form, player i knows that player j knows that ..., up to the level of κ. For example, any κ-knowledge space can capture any finite level of players' interactive knowledge when κ is the least infinite cardinal number. Likewise, any κ-knowledge space can capture any countable level of players' interactive knowledge when κ is the least uncountable cardinal number.

With this property in mind, we demonstrate (in Sections 3 and 6) that there is a universal κ-knowledge space by taking care of all the possible levels of interactive knowledge up to κ in a given class of κ-knowledge spaces. We do so for each class of knowledge spaces which respects given assumptions on players' knowledge.

Our construction has the following three implications. First, we circumvent the above mentioned non-existence results by explicitly specifying a domain of knowledge as a κ-complete algebra. On the one hand, the previously mentioned negative results imply that a sequence of interactive knowledge can generally develop beyond any depth of reasoning in a discontinuous way if any subset of states of the world is an object of knowledge. On the other hand, once we specify the language available to the players as a κ-complete algebra, any κ-knowledge space can generally take into consideration players' interactive knowledge up to the ordinality of κ.

Thus, we turn the previously mentioned negative results into the positive result in the following two ways. First, we enlarge a class of knowledge spaces by allowing the domain of a knowledge space to be a κ-complete algebra. Second, we find a universal κ-knowledge space by keeping track of all possible forms of reasoning up to the depth of κ attained in the given class of κ-knowledge spaces. Thus, unlike a universal (σ-additive) type space, our universal knowledge space usually has transfinite (precisely, κ) hierarchies of interactive knowledge incorporating all possible forms of interactive reasoning up to the depth of κ.

Observe the following analogy with σ-additive beliefs. The domain of each type space is a σ-algebra because a σ-additive probability measure may not necessarily be defined on the power set. That is, the domain specification is implicitly incorporated in type spaces. Put differently, the domain specification plays the following role. The domain of any σ-additive type space (σ-algebra) is the language available to the players in reasoning up to any countable form of interactive beliefs. The domain of any $\aleph_1$-knowledge space (where $\aleph_1$-complete algebra is a σ-algebra because $\aleph_1$ is the least uncountable cardinal) is the language available to the players in reasoning up to any countable form of interactive knowledge. While we keep track of any form of interactive knowledge up to the ordinality of $\aleph_1$ in establishing a universal knowledge space, the continuity property of σ-additive beliefs (or more precisely, the continuity of the operation "Δ") guarantees that the least

[11] It has been recognized as important in various strands of literature to explicitly endow the players with the language that they can (and cannot) use in their reasoning. Examples include: canonical representations of knowledge and belief, equivalence between syntactic and semantic models of knowledge and belief, and epistemic analyses of games. In relation to the first strand of literature, see, for instance, Aumann [5], Brandenburger [14], Brandenburger and Keisler [18], Fagin [25], FGHV [26], FHV [28], Gilboa [32], HS [39], Meier [55, 56, 57], Moss and Viglizzo [65, 66], Roy and Pacuit [71], Salonen [74], and Viglizzo [83].

infinite ordinality of interactive beliefs suffices to establish a universal σ-additive type space (see, for example, FGHV [26] and HS [39] for this point). That is, the least infinite depth of interactive beliefs can determine any subsequent countable order of players' interactive beliefs. To elaborate on this point further, suppose that players' beliefs are finitely additive. Meier [55] shows, in a similar way to HS [38], that a universal finitely additive belief space does not exist if all subsets are required to be measurable.[12] On the other hand, Meier [55] also shows that a universal finitely additive belief space exists once players' beliefs are defined on a κ-complete algebra.

The second implication of our construction is that existence hinges on the specification of a domain rather than on the assumptions on players' knowledge. Third, we assert that there is a universal partitional (non-partitional) κ-knowledge space if we explicitly specify domains of such partitional (non-partitional) κ-knowledge spaces.

More specifically, our existence results are related to the following two previous positive results. First, Meier [56] demonstrates the existence of a universal knowledge-belief space when players' knowledge operators operate only on given measurable subsets of the space on which players' type distributions are defined. Our framework nests Meier's [56] (on the part of players' knowledge) as a special class of $\aleph_1$-knowledge spaces (where $\aleph_1$ is the least uncountable cardinal) when his assumptions on players' knowledge and states of nature are met. Note that the class of partitional $\aleph_1$-knowledge spaces is a proper subclass of his $\aleph_1$-knowledge spaces. We also study how we can obtain tractable representations of players' knowledge within these classes. For example, in both cases, each player's knowledge is represented as a σ-sub-algebra.

Second, Aumann [5] constructs what he calls a canonical knowledge system (of a finitary epistemic $S5$ logic), where each state of the world is a "complete and coherent" set of formulas describing finite levels of players' interactive knowledge. We formally show (in Theorem 2 in Section 3.2) how Aumann's [5] canonical knowledge system and our universal $\aleph_0$-knowledge space (where $\aleph_0$ is the least infinite cardinal) are related in terms of HS's [39] logical construction. Indeed, we do so for any combination of assumptions on players' knowledge and for any domain (i.e., for any κ).

2 Knowledge Spaces

We denote by I a non-empty set of players. Let S be a set which we call the set of *states of nature*. An element of S is regarded as a specification of the exogenous parameters (e.g., strategies and payoff functions) that are relevant to the strategic interactions among the players. The set of states of nature S is endowed with a sub-collection $\mathcal{A}_S$ of $\mathcal{P}(S)$, where $\mathcal{P}(\cdot)$ is the power-set operation. Each element E of $\mathcal{A}_S$, called an event of nature, is an object about which players interactively reason.

[12] The related point is also made by FGHV [26, Example 4.5] in the context of belief hierarchies.

2.1 Technical Preliminaries

Each event $E \in \mathcal{A}_S$ plays a role of a "proposition" regarding states of nature S. Thus, in order to specify a language the players are allowed to use in making inferences about states of nature S and their interactive knowledge, we endow $\mathcal{A}_S$ with a "logical" (precisely, a set-algebraic) structure. To that end, we introduce the following four technical definitions.

First, we denote by κ an infinite cardinal number. Second, for an infinite cardinal κ, we introduce the notion of a κ-complete (Boolean) algebra (of sets). For an underlying set Ω, a subset $\mathcal{D}$ of $\mathcal{P}(\Omega)$ is said to be a *κ-complete algebra* if $\mathcal{D}$ is closed under complementation and is closed under arbitrary union and intersection of any sub-collection with cardinality less than κ.[13] For example, an $\aleph_0$-complete algebra is an algebra of sets, because $\aleph_0$ is the least infinite cardinal. In the similar vein, an $\aleph_1$-complete algebra is a σ-algebra, because $\aleph_1$ is the least uncountable cardinal.

Note that we follow the conventions that $\emptyset = \bigcup \emptyset$ and that, with Ω being an underlying set, $\Omega = \bigcap \emptyset$. Hence, the requirement that any κ-complete algebra $\mathcal{D}$ on Ω contains $\emptyset$ and Ω is subsumed by the requirement that $\mathcal{D}$ is closed under the empty union and intersection.

Third, we say that a subset $\mathcal{D}$ of $\mathcal{P}(\Omega)$ is said to be a *complete algebra* if $\mathcal{D}$ is closed under complementation and is closed under arbitrary union and intersection. In order to conveniently refer to κ-complete and complete algebras, we also introduce a symbol ∞ so that, by abuse of notation, a complete algebra is also referred to as an ∞-complete algebra.

Fourth, for any given infinite cardinal κ or $\kappa = \infty$, we denote by $\mathcal{A}_\kappa(\cdot)$ the operation taking the smallest κ-complete algebra containing a given collection. That is, for any given collection $\mathcal{D}$ of subsets of an underlying set Ω, we define $\mathcal{A}_\kappa(\mathcal{D})$ by

$$\mathcal{A}_\kappa(\mathcal{D}) := \bigcap \{\mathcal{A} \in \mathcal{P}(\mathcal{P}(\Omega)) \mid \mathcal{A} \text{ is a } \kappa\text{-complete algebra with } \mathcal{D} \subseteq \mathcal{A}\}.$$

With these four definitions in mind, for a given set of states of nature $(S, \mathcal{A}_S)$, we endow it with a κ-complete algebraic structure. Namely, we identify the given set of states of nature with $(S, \mathcal{A}_\kappa(\mathcal{A}_S))$.

This assumption means the following: (i) if E is an event of nature (i.e., an object of players' knowledge regarding states of nature S), then so is its complement E^c (we also denote the complement of E by $\neg E$); if E is an object of knowledge regarding S for each $E \in \mathcal{E}$ with $|\mathcal{E}| < \kappa$, then so are its union $\bigcup \mathcal{E} := \bigcup_{E \in \mathcal{E}} E$ and its intersection $\bigcap \mathcal{E} := \bigcap_{E \in \mathcal{E}} E$. Henceforth, we simply assume that $(S, \mathcal{A}_S)$ is a κ-complete algebra for a given κ, because our arguments hold by replacing $\mathcal{A}_S$ with $\mathcal{A}_\kappa(\mathcal{A}_S)$.

We have two further remarks regarding a κ-complete algebra $(S, \mathcal{A}_S)$. First, we usually assume that $\kappa > |I|$. This means that our "language" is fine enough to refer to statements regarding all the possible subsets of players. We will explicitly assume this assumption

[13] First, such $\mathcal{D}$ is also called a κ-complete field (of sets) or a κ-field (e.g., Meier [55]). Second, we say that $\mathcal{D}$ is closed under (non-empty) κ-intersection if it is closed under intersection of a (non-empty) sub-collection with cardinality less than κ. Likewise, we say that $\mathcal{D}$ is closed under (non-empty) κ-union if it is closed under union of a (non-empty) sub-collection with cardinality less than κ.

in Section 4.2, where we analyze the concepts of mutual and common knowledge among players.

Second, as mentioned in Meier [55, Remark 1], it is without loss of generality to restrict attention to κ-complete algebras for infinite regular cardinals κ or $\kappa = \infty$. If an infinite cardinal κ is not regular then any κ-complete algebra is indeed a κ^+-complete algebra, where κ^+ is the successor cardinal. Now, κ^+ is known to be regular (supposing the axiom of choice). Note also that $\aleph_0$ and $\aleph_1$ are indeed regular.

2.2 Knowledge Spaces in terms of Knowledge Operators

Now, we define a model of players' knowledge. We represent players' interactive knowledge regarding $(S, \mathcal{A}_S)$ on some "sample space" by knowledge operators.

Definition 1 (Knowledge Space)**.** *Let I be a non-empty set of players. Let $(S, \mathcal{A}_S)$ be a κ-complete algebra, where κ is an infinite (regular) cardinal or $\kappa = \infty$. A κ-knowledge space of I on $(S, \mathcal{A}_S)$ is a tuple $\overrightarrow{\Omega} := \langle(\Omega, \mathcal{D}), (K_i)_{i\in I}, \Theta\rangle$ with the following three properties.*

1. *$(\Omega, \mathcal{D})$ is a κ-complete algebra. Ω is the set of states of the world. Each element E of $\mathcal{D}$ is called an event (of the world).*

2. *$K_i : \mathcal{D} \to \mathcal{D}$ is player i's knowledge operator for each $i \in I$. For each $E \in \mathcal{D}$, the set $K_i(E)$ denotes the event that player i knows E (i.e., the set of states at which player i knows E). We say that a player $i \in I$ knows an event $E \in \mathcal{D}$ at a state $\omega \in \Omega$ if $\omega \in K_i(E)$.*

3. *$\Theta : \Omega \to S$ is a mapping such that $\Theta^{-1}(E) \in \mathcal{D}$ for any $E \in \mathcal{A}_S$. In other words, $\Theta : (\Omega, \mathcal{D}) \to (S, \mathcal{A}_S)$ is a κ-measurable mapping.*

We have three remarks regarding Definition 1. First, Condition (3) requires the inverse of Θ to be a well-defined mapping from $\mathcal{A}_S$ into $\mathcal{D}$. By this requirement, any set-algebraic ("logical") operations (taking intersections, unions, and complementation) in $\mathcal{A}_S$ are persevered (embedded) in the domain $\mathcal{D}$. In this regard, the mapping Θ can be regarded as a pair of mappings (Θ, Θ^{-1}) such (i) that Θ maps from Ω into S while Θ^{-1} maps $\mathcal{A}_S$ into $\mathcal{D}$ and (ii) that $\omega \in \Theta^{-1}(E)$ in $(\Omega, \mathcal{D})$ iff $\Theta(\omega) \in E$ in $(S, \mathcal{A}_S)$. Second, we do not impose any restriction on the cardinality of the sets S, $\mathcal{A}_S$, Ω, and $\mathcal{D}$. Third, we often call a κ-knowledge space $\overrightarrow{\Omega}$ of I on $(S, \mathcal{A}_S)$ to be a knowledge space $\overrightarrow{\Omega}$ by omitting κ, I, and $(S, \mathcal{A}_S)$.

While any subset of underlying states of the world is deemed to be an object of knowledge (i.e., $\mathcal{D} = \mathcal{P}(\Omega)$) in standard partitional models, our framework is more general. Especially, it intends to capture the following two cases.

First, it is often desirable to capture players' probabilistic beliefs in addition to their knowledge. To that end, we would like to have a model of knowledge whose domain is a certain set-algebra (such as a σ-algebra) in order to treat both knowledge and belief together. In contrast, suppose that players' probabilistic beliefs are represented on a σ-algebra of underlying states of the world while their knowledge is represented on the power set. Without additional assumptions on knowledge, however, the event that a player knows an E might not be measurable for a measurable event E.[14]

Second, in the literature giving logical foundations to state space models of knowledge, events are generated by some logical system, and thus the domain may only form a certain algebra of sets (depending on the given logical system).[15] In this regard, too, we would like to have a model of interactive knowledge amenable to such logical foundations.

Next, we define properties of knowledge to be analyzed.

Definition 2 (Propeties of Knowledge)**.** *Let* $\overrightarrow{\Omega} := \langle(\Omega, \mathcal{D}), (K_i)_{i\in I}, \Theta\rangle$ *be a* κ*-knowledge space. Fix* $i \in I$*.*

1. *The following properties of* K_i *are referred to as logical properties.*
   (a) *No-Contradiction Axiom:* $K_i(\emptyset) = \emptyset$.
   (b) *Consistency:* $K_i(E) \subseteq (\neg K_i)(E^c)$ *for any* $E \in \mathcal{D}$.
   (c) *Monotonicity:* $K_i(E) \subseteq K_i(F)$ *for any* $E, F \in \mathcal{D}$ *with* $E \subseteq F$.
   (d) *Necessitation:* $K_i(\Omega) = \Omega$.
   (e) *Non-empty* λ*-Conjunction:* $\bigcap_{E\in\mathcal{E}} K_i(E) \subseteq K_i(\bigcap \mathcal{E})$ *for any* $\mathcal{E}$ *with* $\emptyset \neq \mathcal{E} \subseteq \mathcal{D}$ *and* $|\mathcal{E}| < \lambda \leq \kappa$.[16]
2. *The following properties of* K_i *are referred to as introspective properties.*
   (a) *Truth Axiom:* $K_i(E) \subseteq E$ *(for any* $E \in \mathcal{D}$*).*
   (b) *Positive Introspection:* $K_i(E) \subseteq K_iK_i(E)$.
   (c) *Negative Introspection:* $(\neg K_i)(E) \subseteq K_i(\neg K_i)(E)$.
3. K_i *satisfies the Kripke property if, for each* $\omega \in \Omega$ *and* $E \in \mathcal{D}$*, the following holds:*

$$\omega \in K_i(E) \text{ if(f) } b_{K_i}(\omega) \subseteq E,$$

[14] For example, it is often assumed in standard partitional models that players' partitions are at most countable. See Maschler, Solan, and Zamir [52, Example 9.37] for an example where players' knowledge may not be measurable without such an assumption.

[15] Samet [78] is a model of knowledge whose primitive is a knowledge operator defined on an ($\aleph_0$-complete) algebra. Set-theoretical (semantic) models of knowledge where events are based on propositions include such papers as Aumann [5], Bacharach [6], Samet [75], and Shin [80]. In fact, it turns out later (in Section 3) that the domain of our universal knowledge space is generated by events corresponding to an "infinitary language" defined by nature and players' knowledge.

[16] We make two remarks. First, ∞-Conjunction means Arbitrary Conjunction. Second, for a given infinite (regular) cardinal κ or $\kappa = \infty$, we only consider λ-Conjunction with $\lambda \leq \kappa$. We could denote (κ, λ)-Conjunction if we need to emphasize that the domain $\mathcal{D}$ is a κ-complete algebra.

where $b_{K_i} : \Omega \to \mathcal{P}(\Omega)$ is defined as follows. For each $\omega \in \Omega$,

$$\begin{aligned} b_{K_i}(\omega) &:= \{\omega' \in \Omega \mid \omega' \in E \text{ for all } E \in \mathcal{D} \text{ with } \omega \in K_i(E)\} \\ &= \bigcap \{E \in \mathcal{D} \mid \omega \in K_i(E)\}. \end{aligned}$$

No-Contradiction Axiom means that a player cannot know any contradiction, which is represented as the empty set. Consistency means that, if a player knows an event E then she also consider E possible in the sense that she does not know its negation E^c. In other words, she cannot know E and E^c at the same time. This notion of possibility is considered to be the dual notion of knowledge (e.g., Hintikka [43]). Thus, we define the possibility operator $L_{K_i} : \mathcal{D} \to \mathcal{D}$ by $L_{K_i}(E) := (\neg K_i)(E^c)$. We say that a player i considers an event $E \in \mathcal{D}$ possible at a state $\omega \in \Omega$ if $\omega \in L_{K_i}(E)$. Now, Consistency means that knowledge implies possibility.

Monotonicity says that if a player knows some event then she knows any of its logical consequences. Necessitation means that a player knows any form of tautology such as $E \cup E^c$, where note that any such tautology is expressed as Ω.

Non-empty (λ-)Conjunction implies that a player knows any non-empty conjunction of events (with cardinality less than λ) if she knows each event. We refer to Non-empty $\aleph_0$-($\aleph_1$-)Conjunction as Non-empty Finite (Countable) Conjunction. Note that Non-empty (λ-)Conjunction and Necessitation is jointly equivalent to (λ-)Conjunction, since $\Omega = \bigcap \emptyset$ (i.e., Necessitation can be seen as the empty conjunction property).

Truth Axiom says that a player can only know what is true. Truth Axiom distinguishes knowledge from belief in the sense that belief can be false while knowledge has to be true. However, we generically refer to such (qualitative) belief as knowledge (without Truth Axiom). Positive Introspection states that if a player knows some event then she knows that she knows it. Negative Introspection states that if a player does not know some event then she knows that she does not know it.

Next, we examine what the Kripke property means. To that end, we define a notion of possibility between states. For states ω and ω' in Ω, we say that ω' is considered *possible* by i at ω if $\omega' \in b_{K_i}(\omega)$. In words, ω' is considered possible at ω iff for any event $E \in \mathcal{D}$ which player i knows at ω, the event E is true at ω'. Thus, $b_{K_i}(\omega)$ induces a binary relation (a possibility relation) in the sense that $b_{K_i}(\omega)$ comprises exactly of the set of states considered possible by i at ω. Note, however, that $b_{K_i}(\omega)$ may not necessarily be an event when $\mathcal{D}$ is not a complete algebra. Now, the Kripke property means that i knows an event E at a state ω iff E contains any state considered possible by i at ω.

We make three remarks on the definition of the possibility relation between states as well as the Kripke property. First, the logical and introspective properties of K_i can be stated as the corresponding properties of b_{K_i}. We will revisit some instances in Proposition 12 in Section 5.2. For example, under the Kripke property, K_i satisfies Truth Axiom, Positive Introspection, and Negative Introspection iff b_{K_i} yields a partition on Ω.

Second, the Kripke property implies the following form of conjunction property. For any subset $\mathcal{E}$ of $\mathcal{D}$ such that $\bigcap \mathcal{E} \in \mathcal{D}$ and $\omega \in K_i(E)$ for all $E \in \mathcal{E}$, we have $\omega \in K_i(\bigcap \mathcal{E})$. Thus, the Kripke property implies κ-Conjunction (including Necessitation) as well as Monotonicity. The converse, however, is not necessarily true unless $\mathcal{D}$ is a complete algebra.[17]

Third, it would be natural to ask how the possibility in terms of $\omega' \in b_{K_i}(\omega)$ relates to the possibility in the sense of $\omega \in L_{K_i}(\{\omega'\})$. For example, Morris [64] introduces a possibility correspondence from the possibility operator L_{K_i}. If every singleton set $\{\omega'\}$ is an event and if an i's knowledge satisfies Monotonicity, then b_{K_i} can be written as

$$b_{K_i}(\omega) = \{\omega' \in \Omega \mid \omega \in L_{K_i}(\{\omega'\})\}. \tag{1}$$

For completeness, the proof is in Remark A.1 in the Appendix.

In this section, so far, we have defined the notion of knowledge spaces (Definition 1) and the properties of knowledge operators (Definition 2). When we speak of a knowledge space, it resides in a given class of knowledge spaces satisfying given assumptions on players' knowledge. Note that we can assume different properties of knowledge on different players.[18] Note also that some property of knowledge implies another.[19]

2.3 Definition of a Universal Knowledge Space

Our main objective of the paper (Theorem 1 in Section 3) is to demonstrate the existence of a universal κ-knowledge space for any infinite (regular) cardinal κ and for any combination of assumptions on players' knowledge. To that end, we define a universal knowledge space (in a given class of knowledge spaces). It is a knowledge space to which every knowledge space (in the given class) is uniquely mapped in a knowledge-preserving manner. We start with formalizing the notion of a mapping that preserves states of nature and players' knowledge, i.e., the notion of a knowledge morphism between knowledge spaces.

Definition 3 (Knowledge Morphism)**.** *Suppose that* $\overrightarrow{\Omega} := \langle(\Omega, \mathcal{D}), (K_i)_{i\in I}, \Theta\rangle$ *and* $\overrightarrow{\Omega'} := \langle(\Omega', \mathcal{D}'), (K'_i)_{i\in I}, \Theta'\rangle$ *are knowledge spaces of a given class. A knowledge morphism* $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega'}$ *is a mapping* $\varphi : \Omega \to \Omega'$ *with the following three properties.*

[17] Samet [78] provides specific examples where a knowledge operator on an ($\aleph_0$-complete) algebra satisfies all the logical and introspective properties (i.e., "S5" knowledge, where Finite Conjunction is considered) but is not derived from a partition (i.e., fails to satisfy the Kripke property). Proposition A.1 in the Appendix provides a characterization of the Kripke property. This generalizes Samet's [78] condition under which "S5" knowledge is derived from a partition.

[18] For example, we can accommodate a case where players have multiple epistemic operators. Suppose that player i's knowledge operator (which satisfies Truth Axiom) is given by $K_{(0,i)}$ while her (non-probabilistic) belief operator (which may violate Truth Axiom) is given by $K_{(1,i)}$. We can carry out our analysis by regarding the set of players as $\{0, 1\} \times I$.

[19] For example, Truth Axiom implies No-Contradiction Axiom and Consistency. The combination of Consistency and Necessitation (that of Consistency and Finite Conjunction) also imply No-Contradiction Axiom. Regarding this point, Negative Introspection is somewhat strong in the following sense. Negative Introspection together with Truth Axiom imply Positive Introspection (see, for example, Aumann [5, p. 270]). Also, Negative Introspection, Monotonicity, and Truth Axiom imply κ-Conjunction including Necessitation (Corollary 2 in Section 4.1).

1. *For all $E' \in \mathcal{D}'$, $\varphi^{-1}(E') \in \mathcal{D}$.*

2. *For all $\omega \in \Omega$, $\Theta'(\varphi(\omega)) = \Theta(\omega)$.*

3. *For each $i \in I$ and $E' \in \mathcal{D}'$, $K_i(\varphi^{-1}(E')) = \varphi^{-1}(K'_i(E'))$.*

Condition (1) means that $\varphi : (\Omega, \mathcal{D}) \to (\Omega', \mathcal{D}')$ is κ-measurable. This is a version of the measurability condition imposed on a type morphism in the type space literature. This condition ensures that the inverse map $\varphi^{-1} : \mathcal{D}' \to \mathcal{D}$ embed set-algebraic operations equipped within $\mathcal{D}'$ into $\mathcal{D}$. It is to be noted that we do not require $\varphi(\mathcal{D}) \subseteq \mathcal{D}'$.

Condition (2) requires that the same state of nature prevail for two associated knowledge spaces. Condition (3) requires that players' knowledge be preserved from one space to another in the following sense: for any event E' in $\overrightarrow{\Omega'}$, player i knows E' at $\varphi(\omega)$ iff she knows $\varphi^{-1}(E')$ at ω.

Before we formally define the notion of a universal knowledge space, we examine the concept of knowledge morphism further. First, for any given knowledge space $\overrightarrow{\Omega}$, the identity map $\mathrm{id}_{\Omega} : \Omega \to \Omega$ is a knowledge morphism from $\overrightarrow{\Omega}$ into itself. We denote by $\mathrm{id}_{\overrightarrow{\Omega}} : \overrightarrow{\Omega} \to \overrightarrow{\Omega}$ the identity knowledge morphism.

Second, we call a knowledge morphism $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega'}$ a *knowledge isomorphism*, if there is a knowledge morphism $\psi : \overrightarrow{\Omega'} \to \overrightarrow{\Omega}$ such that $\psi \circ \varphi = \mathrm{id}_{\overrightarrow{\Omega}}$ and $\varphi \circ \psi = \mathrm{id}_{\overrightarrow{\Omega'}}$.

In other words, a knowledge morphism $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega'}$ is a knowledge isomorphism if φ is bijective and its inverse φ^{-1} is a knowledge morphism. Note that if φ is a knowledge isomorphism then its inverse φ^{-1} is unique. We say that knowledge spaces $\overrightarrow{\Omega}$ and $\overrightarrow{\Omega'}$ are isomorphic, if there is a knowledge isomorphism $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega'}$.

Third, we define the notion of a knowledge subspace in an analogous way to the notion of a belief subspace in the type space literature (Mertens and Zamir [59, Definition 2.15]).

Definition 4 (Knowledge Subspace)**.** *Let $\overrightarrow{\Omega} := \langle (\Omega, \mathcal{D}), (K_i)_{i \in I}, \Theta \rangle$ be a knowledge space. We call a knowledge space $\overrightarrow{\Omega'} = \langle (\Omega', \mathcal{D}'), (K'_i)_{i \in I}, \Theta' \rangle$ a knowledge subspace (of $\overrightarrow{\Omega}$) if $\Omega' \subseteq \Omega$ and if the inclusion map $i_{\Omega', \Omega} : \Omega' \to \Omega$ is a knowledge morphism. If a knowledge subspace $\overrightarrow{\Omega'}$ additionally satisfies that $K'_i(\Omega') = \Omega'$ for all $i \in I$, then we call $\overrightarrow{\Omega'}$ a knowledge closed subspace.*[20]

Now, we define a universal knowledge space. We formalize below the idea that a universal knowledge space "contains" all knowledge spaces in the sense that any knowledge space can be mapped uniquely to the universal space by a knowledge morphism.

Definition 5 (Universal Knowledge Space)**.** *Fix a class of (κ-)knowledge spaces (of I on $(S, \mathcal{A}_S)$). A knowledge space $\overrightarrow{\Omega^*}$ is said to be universal if, for any knowledge space $\overrightarrow{\Omega}$, there is a unique knowledge morphism $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega^*}$.*

[20]Note that we call such $\overrightarrow{\Omega'}$ a knowledge closed subspace even though the original knowledge space $\overrightarrow{\Omega}$ fails to satisfy Necessitation.

We make the following remarks. Fix an infinite cardinal κ (or $\kappa = \infty$), a non-empty set of players I, a κ-complete algebra of states of nature $(S, \mathcal{A}_S)$, and assumptions on players' knowledge. The following can be verified: (i) any composite of two knowledge morphisms is a knowledge morphism; (ii) composites of knowledge morphisms are associative; and (iii) the identity mapping is a knowledge morphism satisfying the identity law. This means that the collection of all κ-knowledge spaces of I on $(S, \mathcal{A}_S)$ forms a category, where each knowledge space $\overrightarrow{\Omega}$ is an *object* and a knowledge morphism is a *morphism*.[21]

In the language of category theory, a universal knowledge space is a *terminal (final)* object in the category of knowledge spaces.[22] As it is well known in category theory that a terminal object is unique up to isomorphism, a universal knowledge space is unique up to knowledge isomorphism.

In the following, we assume that S is not empty. For the degenerate case where $(S, \mathcal{A}_S) = (\emptyset, \{\emptyset\})$, the class of knowledge spaces of I on $(\emptyset, \{\emptyset\})$ consists of the trivial knowledge space $\overrightarrow{\emptyset} = \langle(\emptyset, \{\emptyset\}), (K_i)_{i\in I}, \Theta\rangle$, where $K_i = \mathrm{id}_{\{\emptyset\}}$ for all $i \in I$ and Θ is the empty function. This follows because there is no mapping from a non-empty set Ω into the empty set S. Thus, the category of knowledge spaces of I on $(\emptyset, \{\emptyset\})$ is the category consisting of a single object and its identity morphism, and the single object is terminal. In the language of category theory, the trivial knowledge space $\overrightarrow{\emptyset}$ is an initial knowledge space in a given category of knowledge spaces on any $(S, \mathcal{A}_S)$.

3 A Syntactic Approach to a Universal Knowledge Space

Throughout this section, fix a non-empty set of players I, an infinite regular cardinal κ, and a κ-complete algebra $(S, \mathcal{A}_S)$ of states of nature. We also fix assumptions on players' knowledge. Thus, any knowledge space refers to a κ-knowledge space of I on $(S, \mathcal{A}_S)$ in the given category.

3.1 Syntactic Construction of a Universal Knowledge Space

We demonstrate the existence of a universal (κ-)knowledge space by employing the "expressions-descriptions" approach (HS [39] and Meier [55, 56]).[23] We do so without imposing any restriction on a κ-complete algebra $(S, \mathcal{A}_S)$.[24]

[21] See also Remark A.2 in the Appendix for additional remarks on a class of knowledge spaces as a category.

[22] A knowledge space satisfying Definition 5 can be called a terminal knowledge space in line with the language of category theory (see also Brandenburger and Keisler [18, Section 11]). We, however, simply call such a space to be a universal space.

[23] See also Meier [57] and Moss and Viglizzo [65, 66] for related developments of this approach.

[24] This means that the constructions in Meier [55, 56] could be generalized. In Meier [55, 56], the following regularity ("separative") condition on $(S, \mathcal{A}_S)$ is imposed: for any distinct $s, s' \in S$, there is $E \in \mathcal{A}_S$ with

In the following, we take six steps in order to establish the existence of a universal knowledge space. The first step is to inductively define *expressions*. Expressions are syntactic formulas that express events defined solely in terms of nature and players' knowledge. Specifically, any event of nature $E \in \mathcal{A}_S$ is an object of interactive knowledge, so that any such E is an expression. Since objects of knowledge are closed under κ-union, κ-intersection, and complementation, we define the corresponding syntactic operations. Also, each player's knowledge is an object of interactive knowledge, and thus we define expressions denoting players' knowledge.[25]

Definition 6 (Expressions: Logical Formulas Expressing Nature and Interactive Knowledge)**.** *The set of all (κ-)expressions $\mathcal{L}(= \mathcal{L}^I_\kappa(\mathcal{A}_S))$ is the smallest set which satisfies the following.*

1. *Every $E \in \mathcal{A}_S$ is an expression.*
2. *If $\mathcal{E}$ is a set of expressions with $|\mathcal{E}| < \kappa$, then $(\bigwedge \mathcal{E})$ is an expression, with the following conventions. First, we let $S := \bigwedge \emptyset$. Second, we identify $\bigwedge \mathcal{E} := \bigcap \mathcal{E}$ if $\mathcal{E}$ is a subset of $\mathcal{A}_S$ with $|\mathcal{E}| < \kappa$.*
3. *If e is an expression then $(\neg e)$ is an expression, where we identify $(\neg E) := E^c$ for all $E \in \mathcal{A}_S$.*
4. *If e is an expression, then $(k_i(e))$ is an expression for each $i \in I$.*

For ease of notation, we often add or omit parentheses in denoting expressions. We also introduce other notations. First, if $\mathcal{E}$ is a set of expressions with $|\mathcal{E}| < \kappa$, then we let $(\bigvee \mathcal{E}) := \neg(\bigwedge\{(\neg e) \in \mathcal{L} \mid e \in \mathcal{E}\})$ with the convention that $\bigvee \emptyset := \emptyset$. Second, we interchangeably denote, for instance, $e_1 \wedge e_2 = \bigwedge\{e_1, e_2\}$ and $e_1 \vee e_2 = \bigvee\{e_1, e_2\}$.[26] Third, we interchangeably denote $\bigwedge_{j \in J} e_j = \bigwedge\{e_j \mid j \in J\}$ and $\bigvee_{j \in J} e_j = \bigvee\{e_j \mid j \in J\}$ when expressions are indexed by some index set J. Fourth, we introduce $(e \to f) := ((\neg e) \vee f)$ and $(e \leftrightarrow f) := ((e \to f) \wedge (f \to e))$.

We remark that the set $\mathcal{L}$ incorporates all the hierarchies of interactive knowledge regarding $(S, \mathcal{A}_S)$ up to the ordinality of κ. The following remark shows how the set of expressions $\mathcal{L}$ is inductively generated from the set of states of nature $(S, \mathcal{A}_S)$ in κ steps.

$s \in E$ and $s' \notin E$. In other words, $\{s\} = \bigcap\{E \in \mathcal{A}_S \mid s \in E\}$ for each $s \in S$, but it may be the case that $\{s\} \notin \mathcal{A}_S$ as $\mathcal{A}_S$ is not necessarily closed under arbitrary intersection.

[25] Thus, expressions are infinitary langauages. Fagin [25] and Heifetz [37] analyze their infinitary epistemic logics using such infinitary languages.

[26] Thus, for example, we simply do not distinguish $e_1 \vee e_2$ and $e_2 \vee e_1$. Similarly, since $\{e, e\} = \{e\}$, we simply identify $(e \wedge e)$ as e. These could be augmented by defining $(\bigwedge \mathcal{F})$ for an ordinal sequence of expressions $\mathcal{F}$ instead of a set of expressions.

Remark 1 (Restatement of Expressions $\mathcal{L}$)**.** We let $\mathcal{L}_0 := \mathcal{A}_S$. For any ordinal α with $0 < \alpha \leq \kappa$, we define

$$\mathcal{L}'_\alpha := \left(\bigcup_{\beta<\alpha} \mathcal{L}_\beta \right) \cup \bigcup_{i\in I} \{(k_i(e)) \mid e \in \bigcup_{\beta<\alpha} \mathcal{L}_\beta\}; \text{ and}$$

$$\mathcal{L}_\alpha := \mathcal{L}'_\alpha \cup \{(\neg e) \mid e \in \mathcal{L}'_\alpha\} \cup \left\{ \bigwedge \mathcal{F} \;\middle|\; \mathcal{F} \subseteq \mathcal{L}'_\alpha \text{ and } 0 < |\mathcal{F}| < \kappa \right\}.$$

Then, we have $\mathcal{L} = \mathcal{L}_\kappa$.

Intuitively, each expression $e \in \mathcal{L}_\alpha$ is an expression of "depth at most α." Thus, Remark 1 states that the set $\mathcal{L}$ consists exactly of expressions of "depth at most κ," i.e., logical formulas expressing interactive knowledge regarding $(S, \mathcal{A}_S)$ up to the ordinality of κ.

While expressions themselves are defined independently of any particular knowledge space, for any given knowledge space $\overrightarrow{\Omega}$, we can recursively identify each expression with an event in $\overrightarrow{\Omega}$ (i.e., an element of $\mathcal{D}$), by the measurability condition imposed on the knowledge space (i.e., $\Theta^{-1}(\mathcal{A}_S) \subseteq \mathcal{D}$).

Definition 7 (Expressions are Identified as Events)**.** *Fix a (κ-)knowledge space $\overrightarrow{\Omega}$. We inductively define the mapping $[\![\cdot]\!]_{\overrightarrow{\Omega}} : \mathcal{L} \to \mathcal{D}$, which we call the semantic interpretation function of $\overrightarrow{\Omega}$, as follows.*

1. $[\![E]\!]_{\overrightarrow{\Omega}} := \Theta^{-1}(E)$ *for every* $E \in \mathcal{A}_S$.
2. $[\![\bigwedge \mathcal{E}]\!]_{\overrightarrow{\Omega}} := \bigcap_{e\in\mathcal{E}} [\![e]\!]_{\overrightarrow{\Omega}}$ *for any* $\mathcal{E}$ *with* $\mathcal{E} \subseteq \mathcal{L}$ *and* $|\mathcal{E}| < \kappa$.
3. $[\![\neg e]\!]_{\overrightarrow{\Omega}} := \neg [\![e]\!]_{\overrightarrow{\Omega}} \left(= ([\![e]\!]_{\overrightarrow{\Omega}})^c\right)$ *for each expression* e.
4. $[\![k_i(e)]\!]_{\overrightarrow{\Omega}} := K_i([\![e]\!]_{\overrightarrow{\Omega}})$ *for each* $i \in I$ *and expression* e.

We call $[\![e]\!]_{\overrightarrow{\Omega}} \in \mathcal{D}$ *the denotation of* e *in* $\overrightarrow{\Omega}$.

Note that the semantic interpretation function of a given knowledge space is, by recursion, uniquely extended from Θ^{-1}. It gives the semantic meaning of an expression e in the very sense that $[\![e]\!]_{\overrightarrow{\Omega}}$ is the set of states of the world in which the expression e holds. Again, we often add or omit parentheses for ease of notation. Now, we establish that a knowledge morphism preserves semantics.

Lemma 1 (Knowledge Morphism Preserves Semantics/Meanings of Expressions)**.** *If $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega'}$ is a knowledge morphism between knowledge spaces $\overrightarrow{\Omega}$ and $\overrightarrow{\Omega'}$, then $[\![\cdot]\!]_{\overrightarrow{\Omega}} = \varphi^{-1}([\![\cdot]\!]_{\overrightarrow{\Omega'}})$.*

Before we go to the second step, we define the following semantic notions and see how a knowledge morphism preserves these notions. Our main purpose is to characterize these notions in a universal knowledge space.

Definition 8 (Semantic Properties). *1. We say that an expression $e \in \mathcal{L}$ is valid in a knowledge space $\overrightarrow{\Omega}$ (written $\models_{\overrightarrow{\Omega}} e$) if $[\![e]\!]_{\overrightarrow{\Omega}} = \Omega$. If e is valid in any knowledge space (of the given category), then we say that e is valid (written $\models e$) (in the given category).*

2. A set of expressions $\Phi(\subseteq \mathcal{L})$ is said to be satisfiable in $\overrightarrow{\Omega}$ if there is a state $\omega \in \Omega$ such that $\omega \in [\![f]\!]_{\overrightarrow{\Omega}}$ for all $f \in \Phi$. If there is a knowledge space $\overrightarrow{\Omega}$ such that Φ is satisfiable, then we say that Φ is satisfiable.

3. We say that $e \in \mathcal{L}$ is a semantic consequence of Φ in $\overrightarrow{\Omega}$ (written $\Phi \models_{\overrightarrow{\Omega}} e$) if, $\omega \in [\![e]\!]_{\overrightarrow{\Omega}}$ holds whenever $\omega \in [\![f]\!]_{\overrightarrow{\Omega}}$ for all $f \in \Phi$. We say that $e \in \mathcal{L}$ is a semantic consequence of Φ (written $\Phi \models e$) if $\Phi \models_{\overrightarrow{\Omega}} e$ for any knowledge space $\overrightarrow{\Omega}$.

We have four remarks regarding the notion of validity. First, irrespective of assumptions on players' knowledge, such expressions as S and $(e \vee (\neg e))$ are valid. Second, in a category of knowledge spaces where, for example, Truth Axiom is always assumed on player i, an expression of the form $(k_i(e) \to e)$ is always valid. On the other hand, in a category of knowledge spaces where Truth Axiom is not necessarily assumed on player i, an expression of the form $(k_i(e) \to e)$ is not necessarily valid. Third, it is generally possible that a given expression f happens to be valid in some knowledge space $\overrightarrow{\Omega}$ of a given category due to a particular representation of states (i.e., in a particular context) and players' knowledge while it is not a valid expression in other knowledge spaces of the same category.

Fourth, Lemma 1 implies that any valid expression e in $\overrightarrow{\Omega'}$ is also valid in $\overrightarrow{\Omega}$. This is because $[\![e]\!]_{\overrightarrow{\Omega}} = \varphi^{-1}([\![e]\!]_{\overrightarrow{\Omega'}}) = \varphi^{-1}(\Omega') = \Omega$. If Φ is satisfiable in $\overrightarrow{\Omega}$, then so is it in $\overrightarrow{\Omega'}$. Also, suppose that $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega'}$ is a surjective knowledge morphism. If e is a semantic consequence of Φ in $\overrightarrow{\Omega}$, then so is it in $\overrightarrow{\Omega'}$.

The second step is to define *descriptions* by the set of expressions that obtain at each state of the world in each given knowledge space, together with the corresponding state of nature. Since states of nature and expressions reside in different spaces, we define a description to be a subset of the disjoint union $S \sqcup \mathcal{L} (:= \{(0, s) \in \{0\} \times S \mid s \in S\} \cup \{(1, e) \in \{1\} \times \mathcal{L} \mid e \in \mathcal{L}\})$ as follows.[27]

Definition 9 (Description: Set of Expressions Prevailing at Some State of Some Knowledge Space). *For any given knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$, we define $D(\omega)$, which we call the description of ω, by*

$$\begin{aligned} D(\omega) &:= \{\Theta(\omega)\} \sqcup \{e \in \mathcal{L} \mid \omega \in [\![e]\!]_{\overrightarrow{\Omega}}\} \\ &:= \{(0, \Theta(\omega))\} \cup \{(1, e) \in \{1\} \times \mathcal{L} \mid \omega \in [\![e]\!]_{\overrightarrow{\Omega}}\}. \end{aligned}$$

[27] Throughout the paper, we keep denoting by $(0, s)$ and $(1, e)$ elements of $S \sqcup \mathcal{L}$.

Note that $(0, s) \in D(\omega)$ indicates which event of nature belongs to $D(\omega)$ in the following sense: for any $E \in \mathcal{A}_S$, we have $(1, E) \in D(\omega)$ iff $s \in E$. We, however, simply keep track of every expression e which is true at ω in the description $D(\omega)$.

Descriptions have two roles in establishing the existence of a universal knowledge space. First, we will construct a universal knowledge space in such a way that its underlying set Ω^* of states of the world is the set of all descriptions of states of the world ranged over all knowledge spaces (in the given category). Thus, we define Ω^* as follows:

$$\Omega^* := \{\omega^* \in \mathcal{P}(S \sqcup \mathcal{L}) \mid \omega^* = D(\omega) \text{ for some } \overrightarrow{\Omega} \text{ and } \omega \in \Omega\}. \tag{2}$$

By definition, the set Ω^* is not empty as long as there is a knowledge space $\overrightarrow{\Omega}$ with $\Omega \neq \emptyset$ in the given category. For any assumptions on players' knowledge, we can simply construct a knowledge space $\overrightarrow{\{s\}} := \langle (\{s\}, \mathcal{P}(\{s\})), (\mathrm{id}_{\mathcal{P}(\{s\})})_{i \in I}, i_{\{s\},S} \rangle$ where $s \in S$. It can be seen that each $K_i = \mathrm{id}_{\mathcal{P}(\{s\})}$ satisfies all the properties of knowledge. Thus, Ω^* is indeed not empty.

Note also that the set Ω^* depends on the choice of a class of knowledge spaces. Consider any two classes of knowledge spaces where the set of assumptions on players' knowledge in the first class is a subset of that in the second class. Denote by Ω^{1*} and Ω^{2*} the spaces constructed according to Equation (2). Then, we have $\Omega^{2*} \subseteq \Omega^{1*}$. This follows because if $\omega^* \in \Omega^{2*}$ then there are a knowledge space $\overrightarrow{\Omega}$ (in the second class) and a state $\omega \in \Omega$ such that $\omega^* = D(\omega)$. Since $\overrightarrow{\Omega}$ is also an object in the first class, we get $\omega^* = D(\omega) \in \Omega^{1*}$.

Second, we regard D as a mapping $D : \Omega \to \Omega^*$ for any given knowledge space $\overrightarrow{\Omega}$, and hence we call $D : \Omega \to \Omega^*$ the *description map*. We denote the description map by $D_{\overrightarrow{\Omega}} : \Omega \to \Omega^*$ when we stress the domain of D. The description map D will turn out to be a knowledge morphism between $\overrightarrow{\Omega}$ and our candidate universal knowledge space "$\overrightarrow{\Omega^*}$."

Before we go to the next step, we show that it follows from Lemma 1 that a knowledge morphism preserves the descriptions. To that end, observe the following. Fix knowledge spaces $\overrightarrow{\Omega}$ and $\overrightarrow{\Omega'}$. For any $(\omega, \omega') \in \Omega \times \Omega'$, we have $D_{\overrightarrow{\Omega}}(\omega) = D_{\overrightarrow{\Omega'}}(\omega')$ iff the following hold: (i) $\Theta(\omega) = \Theta'(\omega')$; and (ii) $\omega \in [\![e]\!]_{\overrightarrow{\Omega}}$ iff $\omega' \in [\![e]\!]_{\overrightarrow{\Omega'}}$ for all $e \in \mathcal{L}$. Thus, $D_{\overrightarrow{\Omega}}(\omega) = D_{\overrightarrow{\Omega'}}(\omega')$ means that the outside analysts would consider states ω and $\varphi(\omega)$ to be equivalent in terms of a prevailing state of nature and prevailing expressions, abstracting away from physical representations of $\overrightarrow{\Omega}$ and $\overrightarrow{\Omega'}$.[28]

Both conditions are met for any $(\omega, \varphi(\omega)) \in \Omega \times \Omega'$ such that $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega'}$ is a knowledge morphism. Thus, a knowledge morphism preserves the descriptions.

[28] This notion of equivalence (identicalness or indistingushability) is closely related to the following: Fagin [25, Section 4] in the context of epistemic logic and Mertens and Zamir [59] in terms of belief hierarchies in the universal type space literature. Also, as it turns out that the description map is a unique knowledge morphism into a universal knowledge space, this notion of equivalence corresponds to one notion of bisimulations called "behavioral equivalence" (Kurz [46]) in the literature of category theory, computer science, and logic. We discuss this point in Remark A.3 in the Appendix. For notions of bisimulations ("observational equivalence"), see, for instance, Jacobs and Rutten [44], Kurz [46], Rutten [73], and the references therein.

Corollary 1 (Knowledge Morphism Preserves Descriptions)**.** *Let $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega'}$ be a knowledge morphism between knowledge spaces $\overrightarrow{\Omega}$ and $\overrightarrow{\Omega'}$. Then, $D_{\overrightarrow{\Omega}} = D_{\overrightarrow{\Omega'}} \circ \varphi$.*

We make the following two remarks on Corollary 1. First, we say that a knowledge space $\overrightarrow{\Omega}$ is *non-redundant* (Mertens and Zamir [59, Definition 2.4]) if its description map D is injective. In other words, for any distinct ω and ω', either $\Theta(\omega) \neq \Theta(\omega')$ or they are separated by (a κ-complete sub-algebra) $\mathcal{D}_\kappa := \{[\![e]\!]_{\overrightarrow{\Omega}} \in \mathcal{D} \mid e \in \mathcal{L}\}$.[29] Note that, in Section 3.3, we will represent $\mathcal{D}_\kappa$ in terms of the primitives of the knowledge space $\overrightarrow{\Omega}$ alone (i.e., the notion of non-redundancy can be defined in terms of the primitives of a given knowledge space).

Second, Corollary 1 implies that if $\overrightarrow{\Omega'}$ is non-redundant then there is at most one knowledge morphism from a given space $\overrightarrow{\Omega}$ into $\overrightarrow{\Omega'}$.[30] If $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega'}$ and $\psi : \overrightarrow{\Omega} \to \overrightarrow{\Omega'}$ are knowledge morphisms then $D_{\overrightarrow{\Omega'}} \circ \varphi = D_{\overrightarrow{\Omega}} = D_{\overrightarrow{\Omega'}} \circ \psi$. Since $D_{\overrightarrow{\Omega'}}$ is injective, it follows that $\varphi = \psi$. We will show that the description map $D_{\overrightarrow{\Omega}} : \overrightarrow{\Omega} \to \overrightarrow{\Omega^*}$ is a unique knowledge morphism by demonstrating that $D_{\overrightarrow{\Omega^*}} : \overrightarrow{\Omega^*} \to \overrightarrow{\Omega^*}$ is the identity map.

The third step is to define the collection of events on Ω^* (i.e., the domain of our candidate universal knowledge space). Since each expression e corresponds to an object of knowledge, we define the set of descriptions $[e]$ that make e true (i.e., the set of descriptions that contain e) to be objects of players' knowledge on Ω^*.

Formally, for each $e \in \mathcal{L}$, we define the set of descriptions $[e]$ by $[e] := \{\omega^* \in \Omega^* \mid (1, e) \in \omega^*\}$. We then define the collection $\mathcal{D}^* := \{[e] \in \mathcal{P}(\Omega^*) \mid e \in \mathcal{L}\}$ on Ω^*. Now, we show that $\mathcal{D}^*$ is a legitimate domain (i.e., a κ-complete algebra).

Lemma 2 (Domain of Candidate Universal Space)**.** *$(\Omega^*, \mathcal{D}^*)$ is a κ-complete algebra. Moreover, for any knowledge space $\overrightarrow{\Omega}$, the description map $D : (\Omega, \mathcal{D}) \to (\Omega^*, \mathcal{D}^*)$ is a κ-measurable mapping such that $D^{-1}([e]) = [\![e]\!]_{\overrightarrow{\Omega}}$ for any $e \in \mathcal{L}$.*

The property that $D^{-1}_{\overrightarrow{\Omega}}([\cdot]) = [\![\cdot]\!]_{\overrightarrow{\Omega}}$ exhibits a duality between the semantic interpretation function and the description map in the following sense. By recursion, the semantic interpretation function $[\![\cdot]\!]_{\overrightarrow{\Omega}}$ is a unique map from the set of expressions $\mathcal{L}$ into the domain $\mathcal{D}$ of a given knowledge space. On the other hand, the description map $D_{\overrightarrow{\Omega}}$ is going to be a unique map from the underlying states Ω into the set of descriptions Ω^*.

Note also that the κ-complete sub-algebra $\mathcal{D}_\kappa$ is the one induced by $D_{\overrightarrow{\Omega}}$ in the sense that $\mathcal{D}_\kappa = \{D^{-1}_{\overrightarrow{\Omega}}([e]) \in \mathcal{D} \mid e \in \mathcal{L}\}$.

The fourth step is to introduce players' knowledge. We define each player's knowledge regarding $\mathcal{D}^*$ in a way that an agent i knows an event $[e]$ at a state ω^* iff ω^* contains $k_i(e)$

[29] In Fagin's [25] terminology, such a knowledge space is said to be non-flabby.

[30] An object having this property is called extensional in Kurz [46]. This is also related to a notion of "coinduction proof principle" in the literature of category theory, computer science, and logic (see, for example, Jacobs and Rutten [44], Kurz [46], Rutten [73], and the references therein) in the following sense. If $\overrightarrow{\Omega}$ is non-redundant, then in order for two states ω and ω' in Ω to be the same, it is enough to show that $D_{\overrightarrow{\Omega}}(\omega) = D_{\overrightarrow{\Omega}}(\omega')$ (i.e., they are behaviorally equivalent).

(i.e., $(1, k_i(e)) \in \omega^*$). We show that this is well defined: for any expressions e and f, if they are equivalent in the sense that $(1, e) \in \omega^*$ iff $(1, f) \in \omega^*$, then $k_i(e)$ and $k_i(f)$ are equivalent in the same sense.[31] This equivalence depends on the imposed assumptions on players' knowledge. For example, if Positive Introspection and Truth Axiom are imposed on player i, then $k_i(e)$ and $k_i k_i(e)$ are equivalent in the above sense. We will examine how assumptions on each player's knowledge are encoded within Ω^* itself.

Lemma 3 (Knowledege Operators of Candidate Universal Space)**.** *Fix $i \in I$. We define $K_i^* : \mathcal{D}^* \to \mathcal{D}^*$ by $K_i^*([e]) := [k_i(e)]$ for each $e \in \mathcal{L}$. Then, K_i^* is a well-defined knowledge operator which inherits the properties of knowledge imposed in the given category. Moreover, for any given knowledge space $\overrightarrow{\Omega}$, $D^{-1}(K_i^*([e])) = K_i(D^{-1}([e]))$ for all $[e] \in \mathcal{D}^*$.*

In the Appendix, we provide the following two results. First, we show in Lemma A.1 that each K_i^* can also inherit other potential properties satisfied in the given category.[32] Second, we show in Proposition A.2 that if there is a knowledge space $\overrightarrow{\Omega}$ which fails to satisfy a given property with respect to some event $[\![e]\!]_{\overrightarrow{\Omega}}$, then the knowledge operator K_i^* fails to satisfy that property with respect to $[e]$.

The fifth step is to construct the mapping $\Theta^* : \Omega^* \to S$ that associates with each state of the world $\omega^* \in \Omega^*$ a state of nature in the following way. For each $\omega^* \in \Omega^*$, we extract the unique state of nature s contained in ω^*.

Lemma 4 (Mapping of Candidate Universal Space)**.** *There is a κ-measurable mapping $\Theta^* : (\Omega^*, \mathcal{D}^*) \to (S, \mathcal{A}_S)$ with the following two properties: (i) $\Theta^*(D(\omega)) = \Theta(\omega)$ for any knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$; and (ii) $(\Theta^*)^{-1}(E) = [E] \in \mathcal{D}^*$ for all $E \in \mathcal{A}_S$.*

So far, we have established the following: (i) $\overrightarrow{\Omega^*} = \langle (\Omega^*, \mathcal{D}^*), (K_i^*)_{i \in I}, \Theta^* \rangle$ is a legitimate knowledge space in the given category of knowledge spaces; and (ii) for any knowledge space $\overrightarrow{\Omega}$, the description map $D : \Omega \to \Omega^*$ is a knowledge morphism.

Before we go on to the next (final) step, we examine how the knowledge space $\overrightarrow{\Omega^*}$ resolves the following form of self-reference: generally, each player's knowledge is defined on states while states are supposed to be complete descriptions of the world.[33] Recall that each player's knowledge at each state ω^* is built within the state ω^* itself in the sense that $\omega^* \in K_i^*([e])$ iff $(1, k_i(e)) \in \omega^*$. The following proposition shows how each player's knowledge is encoded in $\overrightarrow{\Omega^*}$.

[31] First, this equivalence is, in spirit, related to the inference rule of equivalence in a syntactic system (see, for example, Lismont and Mongin [50]) and the notion which Dekel, Lipman, and Rustichini (DLR) [21] calls event sufficiency. Second, we will discuss in Proposition 2 that if there are a knowledge space $\overrightarrow{\Omega}$ and a state $\omega \in \Omega$ such that $[\![e]\!]_{\overrightarrow{\Omega}} \neq [\![f]\!]_{\overrightarrow{\Omega}}$ then e and f are not equivalent (i.e., $[e] \neq [f]$). Thus, such identifications (equivalences) of expressions are minimal in $\overrightarrow{\Omega^*}$.

[32] For example, we will show in Section 7 that our arguments apply to richer settings such as dynamic knowledge spaces by using Lemma A.1.

[33] See, for example, Aumann [3, 4, 5], Bacharach [6, 7], Binmore and Brandenburger [11], Brandenburger and Dekel [16], BDG [17], Dekel and Gul [20], FGHV [26], Gilboa [32], Pires [70], Roy and Pacuit [71], Tan and Werlang [82].

Proposition 1 (How Knowledge is Encoded within States)**.** *Fix $i \in I$.*

1. *For each $\omega^* \in \Omega^*$ and $e \in \mathcal{L}$, either $(1, k_i(e)) \in \omega^*$ or $(1, (\neg k_i)(e)) \in \omega^*$.*
2. *For each $\omega^* \in \Omega^*$ and $e \in \mathcal{L}$, at least one of the following holds:*
$$(1, k_i(e)) \in \omega^*, (1, k_i(\neg e)) \in \omega^*, \text{ or } (1, (\neg k_i)(e) \wedge (\neg k_i)(\neg e)) \in \omega^*.$$
Moreover, i's knowledge satisfies Consistency iff exactly one of them holds for each $\omega^* \in \Omega^*$ and $e \in \mathcal{L}$.
3. *For each $\omega^* \in \Omega^*$ and $e \in \mathcal{L}$, exactly one of the following holds:*
$$(1, k_i(e)) \in \omega^*, (1, (\neg k_i)(e) \wedge k_i(\neg k_i)(\neg e)) \in \omega^*, \text{ or } (1, (\neg k_i)(e) \wedge (\neg k_i)(\neg k_i)(e)) \in \omega^*.$$
Moreover, i's knowledge satisfies Negative Introspection iff the third condition never occurs for any $\omega^* \in \Omega^*$ and $e \in \mathcal{L}$.

The first part of Proposition 1 states that, for any $e \in \mathcal{E}$, each state ω^* completely describes i's knowledge of e in the sense that ω^* contains exactly one of the above two expressions denoting (i) i knows e or (ii) i does not know e. The second and third parts characterize how the space Ω^* encodes properties of knowledge (specifically, Consistency and Negative Introspection).[34] We will characterize in Section 3.2 how states encode each property of knowledge.

The sixth step finally establishes that the description map D is a unique knowledge morphism. To that end, we show that the description map from $\overrightarrow{\Omega^*}$ into itself is the identity map on Ω^*.

Lemma 5 (Description Map of Candidate Universal Space)**.** *The description map $D : \Omega^* \to \Omega^*$ is the identity map on Ω^*.*

We consider the following five implications of Lemma 5. The first is that $[\cdot] = [\![\cdot]\!]_{\overrightarrow{\Omega^*}}$. That is, the semantics of e at ω^* is determined by whether $(1, e) \in \omega^*$. Indeed, we prove this property to obtain the result. Second, the knowledge space $\overrightarrow{\Omega^*}$ is non-redundant. The third is the uniqueness of a knowledge morphism $D_{\overrightarrow{\Omega}}$. While the uniqueness follows from the fact that $\overrightarrow{\Omega^*}$ is non-redundant, suppose that $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega^*}$ is a knowledge morphism. Then we have $D_{\overrightarrow{\Omega}}(\omega) = D_{\overrightarrow{\Omega^*}}(\varphi(\omega)) = \varphi(\omega)$. Thus, before we go on to the last two implications, we establish our main result, the existence of a universal knowledge space together with the subsequent remark on how the universal knowledge space Ω^* "contains" other knowledge spaces.

Theorem 1 ($\overrightarrow{\Omega^*}$ is Universal)**.** *The space $\overrightarrow{\Omega^*} = \langle (\Omega^*, \mathcal{D}^*), (K_i^*)_{i \in I}, \Theta^* \rangle$ is a universal knowledge space of I on $(S, \mathcal{A}_S)$ for each given category of knowledge spaces.*

[34] Proposition 1 is related to some of Gilboa's [32] consistency conditions for a state to be a complete description of the world.

Remark 2 (How Universal Space "Contains" Other Knowledge Spaces)**.** Let $\overrightarrow{\Omega}$ be a knowledge space. First, as discussed in Section 2.3, a universal knowledge space exists uniquely up to knowledge isomorphism. It is clear that $\overrightarrow{\Omega}$ is universal iff the description map $D_{\overrightarrow{\Omega}}$ is a knowledge isomorphism.

Second, if $\overrightarrow{\Omega}$ is non-redundant, then, by definition, it is embedded into $\overrightarrow{\Omega^*}$. Generally, there is a knowledge subspace $\overrightarrow{D(\Omega)}$ such that $D : \overrightarrow{\Omega} \to \overrightarrow{D(\Omega)}$ is a (surjective) knowledge morphism. Moreover, if $\overrightarrow{\Omega}$ satisfies Necessitation for every player, then $\overrightarrow{D(\Omega)}$ is a knowledge closed subspace.

The fourth implication of Lemma 5 is that, for any state ω of any particular knowledge space $\overrightarrow{\Omega}$, states $\omega \in \Omega$ and $D(\omega) \in \Omega^*$ are equivalent in the sense that the same state of nature $\Theta(\omega) = \Theta^*(D(\omega)) \in S$ prevails and the same set of expressions regarding nature and players' interactive knowledge obtain. This is because $D(\omega) = D_{\overrightarrow{\Omega^*}}(D(\omega))$. To restate, for any representation $\overrightarrow{\Omega}$ of players' interactive knowledge regarding $(S, \mathcal{A}_S)$ and for any realization $\omega \in \Omega$, the prevailing state of nature and the prevailing set of expressions at ω are encoded in the state $D(\omega)$ of the universal knowledge space $\overrightarrow{\Omega^*}$.

The fifth implication of Lemma 5 (indeed, $[\cdot] = [\![\cdot]\!]_{\overrightarrow{\Omega^*}}$) is the following result.

Proposition 2 (Informational Robustness of Universal Space)**.** *Let e be an expression, and let Φ be a set of expressions.*

1. *Φ is satisfiable iff Φ is satisfiable in $\overrightarrow{\Omega^*}$.*

2. *e is a semantic consequence of Φ in every knowledge space iff e is a semantic consequence of Φ in $\overrightarrow{\Omega^*}$.*

3. *e is valid in every knowledge space iff e is valid in $\overrightarrow{\Omega^*}$.*

The first part of Proposition 2 implies that the knowledge space $\overrightarrow{\Omega^*}$ exhausts all possible sets of satisfiable expressions (in some knowledge space $\overrightarrow{\Omega}$) within $\overrightarrow{\Omega^*}$. Put differently, if Φ is a set of expressions that hold at some state ω in some knowledge space $\overrightarrow{\Omega}$, then the expressions in Φ hold at $D(\omega)$ in $\overrightarrow{\Omega^*}$. This result reflects the insight by Moss and Viglizzo [65, 66] that their terminal object (for a "measure polynomial functor") in their category-theoretical version of the expression-description approach consists of all satisfied theories (descriptions) of all points in all objects.

The third part of Proposition 2 implies that if there are expressions e and f and a knowledge space $\overrightarrow{\Omega}$ such that $[\![e]\!]_{\overrightarrow{\Omega}} \neq [\![f]\!]_{\overrightarrow{\Omega}}$ (i.e., $(e \leftrightarrow f)$ is not valid in $\overrightarrow{\Omega}$) then $([e] =) [\![e]\!]_{\overrightarrow{\Omega^*}} \neq [\![f]\!]_{\overrightarrow{\Omega^*}} (= [f])$. For example, suppose that expressions e and $k_i(f)$ happen to satisfy $[\![e]\!]_{\overrightarrow{\Omega'}} = K_i([\![f]\!]_{\overrightarrow{\Omega'}})$ in a particular representation $\overrightarrow{\Omega'}$ (i.e., in a particular context). If there is another knowledge space $\overrightarrow{\Omega}$ which distinguishes these two expressions in the sense that $[\![e]\!]_{\overrightarrow{\Omega}} \neq K_i([\![f]\!]_{\overrightarrow{\Omega}})$, it follows in the universal knowledge space $\overrightarrow{\Omega^*}$ that $[\![e]\!]_{\overrightarrow{\Omega^*}} \neq K_i^*([\![f]\!]_{\overrightarrow{\Omega^*}})$.

Put differently, if two expressions satisfy $[\![e]\!]_{\overrightarrow{\Omega^*}} = K_i^*([\![f]\!]_{\overrightarrow{\Omega^*}})$, then it is always the case in any knowledge space that $[\![e]\!]_{\overrightarrow{\Omega}} = K_i([\![f]\!]_{\overrightarrow{\Omega}})$.

Generally, Proposition 2 means that such semantic notions as satisfiability, semantic consequence, and validness in $\overrightarrow{\Omega^*}$ are informationally robust in the sense that they do not depend on any particular representation $\overrightarrow{\Omega}$.

In relation to Theorem 1 and Proposition 2, the following sheds light on how the knowledge space $\overrightarrow{\Omega^*}$ exhausts nature and players' knowledge.

Proposition 3 (Universal Space Exhausts Interactive Knowledge)**.** *We let*

$$\Omega^{**} := \{(s, (\Psi_i)_{i\in I}) \in S \times \mathcal{P}(\mathcal{L})^I \mid \text{there are a knowledge space } \overrightarrow{\Omega} \text{ and } \omega \in \Omega \text{ such that } (s, (\Psi_i)_{i\in I}) = (\Theta(\omega), (\{e \in \mathcal{L} \mid \omega \in K_i([\![e]\!]_{\overrightarrow{\Omega}})\})_{i\in I})\}.$$

Then, the following mapping is a bijection:

$$\Omega^* \ni \omega^* \mapsto (\Theta^*(\omega^*), (\{e \in \mathcal{L} \mid \omega^* \in K_i^*([e])\})_{i\in I}) \in \Omega^{**}.$$

The "injection" part of Proposition 3 states that each state ω^* is in a one-to-one relation to a profile of the corresponding nature and each player's knowledge at ω^*. The "surjection" part simply follows from Lemmas 3 and 5.

To conclude this subsection, we pose the following two questions regarding the notion of common knowledge in the universal knowledge space. While the notion of common knowledge is formally defined in Section 4.2, we remark that the common knowledge operator $C : \mathcal{D} \to \mathcal{D}$ is defined in each knowledge space $\overrightarrow{\Omega}$. Thus, for each event $E \in \mathcal{D}$, the set $C(E)$ is the event that E is common knowledge among the players I. In the following, we assume that assumptions on players' knowledge include Monotonicity and Positive Introspection and that they are homogeneously given across players.

First, we remark that the universal knowledge space preserves the common knowledge among players. Proposition 10 in Section 4.2 establishes that, for any knowledge space $\overrightarrow{\Omega}$ (with the above assumptions), an event $D^{-1}([e]) \in \mathcal{D}$ is common knowledge at state $\omega \in \Omega$ iff $[e]$ is common knowledge at $D(\omega) \in \Omega^*$.

Second, at the interpretational level, an implicit assumption often made in state space models of knowledge is the "meta-knowledge" of the model itself (or the primitives of the model themselves).[35] Proposition 13 in Section 5.3 provides one test according to which we (i.e., the outside analysts) can say that the players commonly know the structure of the universal knowledge space. We defer the discussion because we need to formalize the knowledge and common knowledge of the "structure of a model of knowledge."

[35] See, for instance, Aumann [3, 4, 5], Bacharach [6, 7], Binmore and Brandenburger [11], Brandenburger and Dekel [16], BDG [17], Dekel and Gul [20], FGHV [26], Gilboa [32], Pires [70], Roy and Pacuit [71], Tan and Werlang [82].

3.2 Characterizing the Universal Knowledge Space by Coherent Sets of Expressions

In the last subsection, we have established the existence of a universal knowledge space by collecting all possible states that can realize in some knowledge space. Each state in the universal knowledge space consists of a set of expressions (together with a state of nature) satisfiable at some state of some knowledge space. Thus, for a general regular infinite cardinal κ (especially with $\kappa \geq \aleph_1$ where infinitary operations are allowed), states in the universal κ-knowledge space would generally be different from the collection of "maximally consistent" sets of expressions (together with a state of nature) in some syntax system (see also Moss and Viglizzo [65, 66]).[36]

Now, the question arises as to how (or whether) we can characterize each state ω^* and the set $\overrightarrow{\Omega^*}$ in a somewhat more explicit way. We aim to characterize the universal (κ-)knowledge space Ω^* in terms of the largest collection comprising of "coherent" and "complete" set of expressions.

To that end, recall that each state $\omega^* \in \Omega^*$ contains expressions that hold at ω^* (i.e., $(1, e) \in \omega^*$ iff $\omega^* \in [e]$) as well as the corresponding state of nature $s = \Theta^*(\omega^*)$. This fact suggests the following three ideas as to which expressions a given state ω^* contains. First, every state $\omega^* \in \Omega^*$ contains expressions that hold in any knowledge space (of the given class). That is, each state ω^* contains valid expressions. Second, for any expression e, if a state ω^* contains e then it does not contain $(\neg e)$. In other words, each ω^* is *coherent*. Third, on the other hand, if a state ω^* does not contain e then it contains $(\neg e)$. That is, each ω^* is *complete*. The terminologies of coherence and completeness in this context (i.e., coherence and completeness of each state in terms of the expressions $\mathcal{L}$) are from Aumann [5].

Here, we formally link the universal knowledge space obtained in the previous section with Aumann's [5] construction of what he calls a canonical knowledge system (of a finitary epistemic $S5$ logic) by generalizing Aumann's [5] idea in the following sense. We show that the universal knowledge space Ω^* is written as the largest set with the following two properties: (i) its element (state) is a complete and coherent set of expressions together with the corresponding state of nature; and (ii) its element (state) reflects the given properties of knowledge. Thus, we establish an alternative characterization of the universal knowledge space obtained in Section 3.1. This characterization holds irrespective of a given cardinality κ and assumptions on players' knowledge.

Theorem 2 (Largest Set of Complete and Coherent Expressions)**.** *The set Ω^* obtained in Section 3.1 is the largest set satisfying the following: (i) Ω^* satisfies all the conditions specified below; and (ii) for any set Ω satisfying all the conditions below, there is a knowledge space $\overrightarrow{\Omega}$ such that its description map $D_{\overrightarrow{\Omega}} : \overrightarrow{\Omega} \to \overrightarrow{\Omega^*}$ is an inclusion map (and thus $\Omega \subseteq \Omega^*$).*

[36] Indeed, the discrepancy between a semantic notion of satisfiability and a syntactic notion of maximal consistency would emerge even at the level of a propositional logic when infinitary operations are allowed (Karp [45]). See also Heifetz [37] and Meier [57] for this point.

1. *Each element $\omega \in \Omega$ is a subset of $S \sqcup \mathcal{L}$ with the following properties.*
   (a) *There is a unique $s \in S$ such that $(0, s) \in \omega$. Moreover, for such an $s \in S$, $(1, E) \in \omega$ for all $E \in \mathcal{A}_S$ with $s \in E$.*[37]
   (b) *Depending on the notion of knowledge, ω contains any instance of the following expressions.*
      i. *No-Contradiction Axiom: $(\emptyset \leftrightarrow k_i(\emptyset))$.*
      ii. *Consistency: $(k_i(e) \to (\neg k_i)(\neg e))$.*
      iii. *Non-empty λ-Conjunction: $((\bigwedge_{e \in \mathcal{E}} k_i(e)) \to k_i(\bigwedge \mathcal{E}))$ with $0 < |\mathcal{E}| < \lambda (\leq \kappa)$.*
      iv. *Necessitation: $k_i(S)$.*
      v. *Truth Axiom: $(k_i(e) \to e)$.*
      vi. *Positive Introspection: $(k_i(e) \to k_i k_i(e))$.*
      vii. *Negative Introspection: $((\neg k_i)(e) \to k_i(\neg k_i)(e))$.*
   (c) *If $(1, e) \in \omega$ and $(1, (e \to f)) \in \omega$ then $(1, f) \in \omega$.*
   (d) *For any $\mathcal{E}$ with $\mathcal{E} \subseteq \mathcal{L}$ and $|\mathcal{E}| < \kappa$, $(1, \bigwedge \mathcal{E}) \in \omega$ iff $(1, e) \in \omega$ for all $e \in \mathcal{E}$.*
   (e) *Coherency: For each $e \in \mathcal{L}$, if $(1, (\neg e)) \in \omega$ then $(1, e) \notin \omega$.*
   (f) *Completeness: For each $e \in \mathcal{L}$, if $(1, e) \notin \omega$ then $(1, (\neg e)) \in \omega$.*
2. *Ω satisfies the following conditions.*
   (a) *If expressions e and f are such that $(1, (e \leftrightarrow f)) \in \omega$ for all $\omega \in \Omega$, then $(1, (k_i(e) \leftrightarrow k_i(f))) \in \omega$ for all $\omega \in \Omega$.*
   (b) *Suppose that Monotonicity is imposed on i's knowledge. If expressions e and f are such that $(1, (e \to f)) \in \omega$ for all $\omega \in \Omega$, then $(1, (k_i(e) \to k_i(f))) \in \omega$ for all $\omega \in \Omega$.*
   (c) *Suppose that the Kripke property is imposed on i's knowledge. Then, $(1, k_i(e)) \in \omega$ for any $e \in \mathcal{L}$ and $\omega \in \Omega$ with the following condition: if $\omega' \in \Omega$ satisfies $(1, f) \in \omega'$ for all $f \in \mathcal{L}$ with $(1, k_i(f)) \in \omega$ then $(1, e) \in \omega'$.*

In Theorem 2, Condition (1a) states that each state of the world ω describes a corresponding state of nature s in a well-defined manner. Also, the state of the world ω contains those events of nature $E \in \mathcal{A}_S$ that are true at s (i.e., $s \in E$). Conditions (1c) through (1f) are logical requirements on how the world is described.

Condition (2a) requires that if two expressions e and f are equivalent in the sense that $(e \leftrightarrow f)$ is in any state of the world then expressions $k_i(e)$ and $k_i(f)$ are equivalent in the same sense. This condition allows us to define players' knowledge operators in a way such that if two expressions e and f correspond to the same event then the events associated

[37]By Condition (1e) below, Condition (1a) implies that, for a unique $s \in S$ with $(0, s) \in \omega$, we have $(1, E) \in \omega$ iff $s \in E$ for any $E \in \mathcal{A}_S$. Especially, we have $(1, S) \in \omega$ and $(1, \emptyset) \notin \omega$.

with the knowledge of e and f are the same. We have seen in Lemma 3 that $\overrightarrow{\Omega^*}$ satisfies this property.

Each condition in (1b) describes how each state of the world describes the corresponding property of players' knowledge. We have seen, in Proposition 1, a related characterization for Consistency and Negative Introspection. Monotonicity and the Kripke property are described, in (2b), and (2c), by conditions on the entire states of the world Ω. If these conditions are satisfied in Ω, then it means that assumptions on players' knowledge are encoded within Ω itself in the sense that we can induce players' knowledge operators which satisfy the given assumptions.

3.3 Comparison with the Previous Negative Results

We discuss how domain specifications together with an appropriate notion of knowledge morphism (which reflects domain specifications) have an essential role in establishing the existence of a universal knowledge space. The basic idea is that, for any given infinite regular cardinal κ, any κ-knowledge space can capture players' interactive knowledge of the ordinal depth up to κ. Theorem 1 establishes the existence of a universal κ-knowledge space within such class of κ-knowledge spaces.

To compare our existence result with the previous negative results, we look at the notion of a rank of a standard partitional knowledge space (HS [38]).[38] HS [38] demonstrate that there is no universal standard partitional knowledge space on the following two grounds. First, a knowledge morphism preserves the ranks. Second, there is a standard partitional knowledge space with arbitrarily high rank. These facts imply that, for any candidate universal standard partitional knowledge space, there exists a standard partitional knowledge space which has a higher rank and thus the candidate space must not be universal.

We extend their notion of a rank to that of a κ-rank of a κ-knowledge space, because not all subsets are expressible within the κ-complete algebra of a given κ-knowledge space. We see (i) that a knowledge morphism preserves the κ-ranks but (ii) that the κ-rank of any κ-knowledge space is at most κ. Below is the definition of the κ-rank of a κ-knowledge space.

Definition 10 (κ-Rank: Maximal Ordinality of Interactive Knowledge in Each Knowledge Space)**.** *Let $\overrightarrow{\Omega} := \langle (\Omega, \mathcal{D}), (K_i)_{i \in I}, \Theta \rangle$ be a κ-knowledge space of I on $(S, \mathcal{A}_S)$. The κ-rank of $\overrightarrow{\Omega}$ is defined as the least ordinal α such that $\mathcal{C}_\alpha = \mathcal{C}_{\alpha+1}$, where the sequence $(\mathcal{C}_\alpha)_\alpha$ is defined as follows:*

$$\mathcal{C}_\alpha := \begin{cases} \mathcal{A}_\kappa(\{\Theta^{-1}(E) \in \mathcal{D} \mid E \in \mathcal{A}_S\})(= \{\Theta^{-1}(E) \in \mathcal{D} \mid E \in \mathcal{A}_S\}) & \text{if } \alpha = 0 \\ \mathcal{A}_\kappa\left(\left(\bigcup_{\beta<\alpha} \mathcal{C}_\beta\right) \cup \bigcup_{i \in I} \{K_i(E) \in \mathcal{D} \mid E \in \bigcup_{\beta<\alpha} \mathcal{C}_\beta\}\right) & \text{if } \alpha > 0 \end{cases}.$$

[38] Fagin [25] considers a closely related concept, "distinguishing ordinals," in his modal logic framework.

With this definition in mind, we establish the following.

Proposition 4 (κ-Rank of a Universal κ-Knowledge Space). *1. If $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega'}$ is a knowledge morphism between κ-knowledge spaces $\overrightarrow{\Omega}$ and $\overrightarrow{\Omega'}$, then the κ-rank of $\overrightarrow{\Omega'}$ is at least as high as that of $\overrightarrow{\Omega}$.*

2. The κ-rank of any κ-knowledge space $\overrightarrow{\Omega}$ is at most κ.

The first part of Proposition 4 states that a knowledge morphism between κ-knowledge spaces preserves the κ-ranks. The second part, on the other hand, asserts the difference between κ-knowledge spaces and standard knowledge spaces whose domains are power sets. Namely, the κ-rank of any κ-knowledge space is at most κ. This result hinges on the fact that the set of expressions $\mathcal{L}(= \mathcal{L}^I_\kappa(\mathcal{A}_S))$ consists of expressions that involve player's interactive knowledge of the ordinality up to κ (Remark 1).[39]

Specifically, we define $\mathcal{D}_\alpha = \{[\![e]\!]_{\overrightarrow{\Omega}} \in \mathcal{D} \mid e \in \mathcal{L}_\alpha\}$ for each ordinal $\alpha \leq \kappa$, where $\mathcal{L}_\alpha$ is defined as in Remark 1. Then, we show in the proof that $\mathcal{D}_\alpha = \mathcal{C}_\alpha$ for each ordinal $\alpha \leq \kappa$. Then, it follows that $\mathcal{D}_\kappa = \mathcal{C}_\kappa = \mathcal{C}_{\kappa+1}$, i.e., the κ-rank of $\overrightarrow{\Omega}$ is at most κ. We also remark that the κ-complete sub-algebra $\{[\![e]\!]_{\overrightarrow{\Omega}} \in \mathcal{D} \mid e \in \mathcal{L}\}$ is equal to $\mathcal{C}_\kappa$. That is, the notion of non-redundancy can be examined through $\mathcal{C}_\kappa$.

We make two remarks regarding the second part of Proposition 4. First, for an infinite regular cardinal κ, HS's [38] non-existence argument does not apply to a given class of κ-knowledge spaces. Second, it is important to take care of all the κ levels of interactive knowledge in order to incorporate possible "discontinuity" of knowledge.[40]

By fixing the language that the players are allowed to use in reasoning about their interactive knowledge (within the ordinality of κ), a knowledge morphism (a description map) preserves interactive knowledge in a given (κ-)knowledge space to the universal (κ-)knowledge space. At the same time, such preservation concerns only to the extent that (κ-)expressions are preserved.

Finally, we discuss how our results reconcile with the existence of a universal partitional κ-knowledge space. First, fix any infinite regular cardinal κ. Since we can identify the conditions on each player's knowledge operator under which her knowledge is induced from a partition (namely, Truth Axiom, Negative Introspection, and the Kripke property), Theorem 1 demonstrates that there is a universal partitional κ-knowledge space. In partitional κ-knowledge spaces, the conjunction property is generally restricted to κ-Conjunction.

[39] In the hierarchical construction of a universal knowledge space in Section 6, we also define the space consisting of hierarchies of knowledge up to the ordinality of κ.

[40] HS [38, 39] attribute the non-existence of a universal (standard partitional) knowledge space to the "lack of continuity" of knowledge. On a related point, FHV [28], Fagin [25], FGHV [26], and HS [41] attribute the non-existence of the space of all coherent hierarchies of knowledge to the lack of "continuity" property of knowledge structures, as opposed to that of σ-additive probability measures. See also Meier [55, 56] for the discussion of the use of infinitary expressions.

On the contrary, consider the ∞-knowledge spaces of I on $(S, \mathcal{P}(S))$ (with $|S| \geq 2$ and $|I| \geq 2$).[41] In this case, the notion of an ∞-rank is equivalent to that defined by HS [38]. Thus, contrary to the case where κ is an infinite regular cardinal, there is no universal ∞-knowledge space on $(S, \mathcal{P}(S))$ satisfying all the logical and introspective properties (provided that $|S| \geq 2$ and $|I| \geq 2$). Moreover, this non-existence result shows that there is no universal ∞-knowledge space in a category of knowledge spaces on $(S, \mathcal{P}(S))$ which contains, as a subclass, the category of knowledge spaces satisfying all the logical and introspective properties. With respect to our previous discussion, the collection of "∞-expressions" (Definition 6 with $\kappa = \infty$) would be too large to be a set (in the realm of the standard set theory).

4 Knowledge and Common Knowledge in Knowledge Spaces

We study how we can capture players' knowledge and common knowledge in each knowledge space. First, in Section 4.1, we represent each player's knowledge by a set algebra if her knowledge satisfies Truth Axiom, Monotonicity, and Positive Introspection. The usefulness of this representation would lie in the fact that if each player's knowledge also satisfies Negative Introspection then her knowledge is represented as a sub-algebra without imposing an assumption on the cardinality of the sub-algebra.

Second, in Section 4.2, we introduce the notion of common knowledge regardless of assumptions on players' knowledge, where note that the discussion of the common knowledge pre-supposes notations in Section 4.1. The notion of common knowledge is used to ask the sense in which the players commonly know the structure of a universal knowledge space (Proposition 13 in Section 5.3).

This section aims to have representations of players' knowledge and common knowledge in the framework of knowledge spaces. In this regard, we generalize, in Appendix A.7.1, possibility correspondence models of knowledge under Monotonicity in order to claim that our framework of knowledge spaces admits a variety of knowledge representations. We also connect our framework to the literature of knowledge representation in psychology (Doignon and Falmagne [22, 23]) in Appendix A.7.4.

Note that, in this paper, we represent players' knowledge by (i) knowledge operators (Section 2.2), (ii) set algebras (Section 4.1), (iii) knowledge-type mappings (which are qualitative analogues of belief-type mappings, in Section 5.1), and generalized possibility correspondences (Appendix A.7.1). Under appropriate conditions, one framework is shown to be equivalent to another. This means that a universal knowledge space exists under various frameworks (with various combinations of properties on knowledge).

[41] Note that if $\mathcal{A}_S$ satisfies the separative property (see Footnote 24) then $\mathcal{A}_S = \mathcal{P}(S)$.

4.1 Set-algebraic Representation of Knowledge

Here, we provide a set-algebraic representation of knowledge. This embodies the intuition that the content of knowledge can be captured by a sub-collection of a given domain. To that end, we introduce the following two definitions. First, we say that an event E is *self-evident* to player i if $E \subseteq K_i(E)$.[42] In words, E is self-evident to i if she knows E whenever E obtains. We denote the collection of events which are self-evident to i, which we call i's *self-evident collection*, by $\mathcal{J}_{K_i} := \{E \in \mathcal{D} \mid E \subseteq K_i(E)\}$.[43]

Second, we say that a given sub-collection $\mathcal{J}_i$ of $\mathcal{D}$ satisfies the *maximality property* (w.r.t. $\mathcal{D}$) if $\emptyset \in \mathcal{J}_i$ and for any $E \in \mathcal{D}$, the largest element of $\mathcal{J}_i$ contained in E, $\max\{F \in \mathcal{D} \mid F \in \mathcal{J}_i \text{ and } F \subseteq E\}$, exists in $\mathcal{J}_i$.[44]

With these two definitions in mind, we establish the following two results. First, we show that properties of knowledge operators are equivalently expressed as set-algebraic properties of self-evident collections when Truth Axiom, Positive Introspection, and Monotonicity are imposed.

Specifically, given a knowledge operator K_i which satisfies Truth Axiom, Positive Introspection, and Monotonicity, the event $K_i(E)$ is the largest (in the sense of set inclusion) self-evident event to i which is contained in E. Other properties of K_i are translated into set-algebraic properties of $\mathcal{J}_{K_i}$. For example, Negative Introspection is translated as the closure under complementation. The self-evident collection $\mathcal{J}_{K_i}$, in turn, induces the knowledge operator through $K_i = K_{\mathcal{J}_{K_i}}(E) := \max\{F \in \mathcal{J}_{K_i} \mid F \subseteq E\}$.

Our second result is that, for any given sub-collection $\mathcal{D}'$ of $\mathcal{D}$, we can express the smallest sub-collection $\mathcal{J}(\mathcal{D}')$ containing $\mathcal{D}'$ and satisfying the maximality property. Thus, we can induce an informational content to a given sub-collection $\mathcal{D}'$ of events. Note that we will use this concept to define the notion of common knowledge in the next subsection.

Now, we go on to our first main result: a knowledge space admits representations of players' knowledge in terms of self-evident collections.

Proposition 5 (Equivalence between Knowledge Operators and Self-Evident Collections)**.** *Fix a κ-complete algebra $(\Omega, \mathcal{D})$.*

1. *For a given knowledge operator $K_i : \mathcal{D} \to \mathcal{D}$, the self-evident collection $\mathcal{J}_{K_i}$ satisfies the following.*

[42] Binmore and Brandenburger [11] call such an event to be a truism.

[43] We have a technical remark. If we express each player's knowledge by a self-evident collection as opposed to a knowledge operator, we specify it as a sub-collection of a given domain. Thus, two collections $\mathcal{J}_{K_i}$ on $\mathcal{D}$ and $\mathcal{J}'_{K'_i}$ on $\mathcal{D}'$ are considered to be different as long as the domains $\mathcal{D}$ and $\mathcal{D}'$ differ, even if $\mathcal{J}_{K_i}$ and $\mathcal{J}'_{K'_i}$ are extensionally equivalent (have the same elements).

[44] The terminology is due to Samet [78], where the maximality property is defined for a collection of events which forms an ($\aleph_0$-complete) sub-algebra of an ($\aleph_0$-complete) algebra $\mathcal{D}$. As Proposition 5 demonstrates, this corresponds to the knowledge (defined on an algebra) which satisfies Truth Axiom, Monotonicity, (Positive Introspection, Finite Conjunction), and Negative Introspection. In the sense that Negative Introspection is not necessarily imposed, the maximality property of a self-evident collection is generally not identical with the notion of "relative completeness" of a Boolean sub-algebra (Halmos [33, Section 4]).

(a) *If K_i satisfies Truth Axiom, Monotonicity, and Positive Introspection, then $\mathcal{J}_{K_i}$ satisfies the maximality property in the following sense. For any $E \in \mathcal{D}$,*

$$K_i(E) = \max\{F \in \mathcal{D} \mid F \in \mathcal{J}_{K_i} \text{ and } F \subseteq E\} \tag{3}$$
$$(= \{\omega \in \Omega \mid \text{there is } F \in \mathcal{J}_{K_i} \text{ with } \omega \in F \text{ and } F \subseteq E\}).$$

(b) *If K_i satisfies Non-empty λ-Conjunction (with $\lambda \leq \kappa$), then $\mathcal{J}_{K_i}$ is closed under non-empty λ-intersection.*

(c) *If K_i satisfies Necessitation, then $\Omega \in \mathcal{J}_{K_i}$.*

(d) *If K_i satisfies Truth Axiom and Negative Introspection, then $\mathcal{J}_{K_i}$ is closed under complementation.*

2. *Conversely, for a given sub-collection $\mathcal{J}_i$ of $\mathcal{D}$ which satisfies the maximality property, the operator $K_{\mathcal{J}_i} : \mathcal{D} \to \mathcal{D}$ satisfies the converse of the above assertion in the following sense, where $K_{\mathcal{J}_i}(E) := \max\{F \in \mathcal{J}_i \mid F \subseteq E\}$.*

   (a) *$K_{\mathcal{J}_i}$ satisfies Truth Axiom, Monotonicity, and Positive Introspection.*

   (b) *If $\mathcal{J}_i$ is closed under non-empty λ-intersection (with $\lambda \leq \kappa$), then $K_{\mathcal{J}_i}$ satisfies Non-empty λ-Conjunction.*

   (c) *If $\Omega \in \mathcal{J}_i$ then $K_{\mathcal{J}_i}$ satisfies Necessitation.*

   (d) *If $\mathcal{J}_i$ is closed under complementation then $K_{\mathcal{J}_i}$ satisfies Negative Introspection.*

3. *Starting with K_i, we have $K_{\mathcal{J}_{K_i}} = K_i$. Likewise, $\mathcal{J}_i$ induces $\mathcal{J}_i = \mathcal{J}_{K_{\mathcal{J}_i}}$.*

4. *If $\kappa = \infty$, then $\mathcal{J}_{K_i}$ ($\mathcal{J}_i$) satisfies the maximality property iff $\mathcal{J}_{K_i}$ ($\mathcal{J}_i$) is closed under arbitrary union. Generally, if $\mathcal{J}_{K_i}$ ($\mathcal{J}_i$) satisfies the maximality property, then $\mathcal{J}_{K_i}$ ($\mathcal{J}_i$) is closed under κ-union.*

An immediate corollary of Proposition 5 is that Negative Introspection, together with Truth Axiom and Monotonicity, imply all the other logical and introspective properties of knowledge postulated in Definition 2.

Corollary 2 (Negavtive Introspection Implies Conjunction)**.** *Let $\overrightarrow{\Omega}$ be a κ-knowledge space such that $K_i : \mathcal{D} \to \mathcal{D}$ satisfies Truth Axiom, Monotonicity, and Negative Introspection. Then, K_i also satisfies κ-Conjunction including Necessitation (as well as No-Contradiction Axiom, Consistency, and Positive Introspection).*

We make six additional remarks regarding Proposition 5. First, the maximality property ensures that $K_i(E) = \bigcup\{F \in \mathcal{D} \mid F \in \mathcal{J}_{K_i} \text{ and } F \subseteq E\} \in \mathcal{D}$ for each $E \in \mathcal{D}$, in spite of the fact that the self-evident collection (as well as the domain $\mathcal{D}$) may not necessarily be closed under arbitrary union.

Second, the self-evident collection satisfies $\mathcal{J}_{K_i} = \{K_i(E) \in \mathcal{D} \mid E \in \mathcal{D}\}$ as long as K_i satisfies Truth Axiom and Positive Introspection. Thus, it comprises solely of the largest self-evident event $K_i(E)$ contained in E for each $E \in \mathcal{D}$.

Third, all of Truth Axiom, Positive Introspection, and Monotonicity are essential in establishing the equivalence between a knowledge operator and a self-evident collection. If any one condition is absent, there is a simple example where $K_i \neq K_i'$ but $\mathcal{J}_{K_i} = \mathcal{J}_{K_i'}$. See Remark A.4 in the Appendix for concrete examples.

Fourth, Proposition 5 provides the sense in which knowledge takes a form of various set-algebras. For example, a self-evident collection in an ∞-knowledge space becomes a (sub-)topology ((sub-)topology closed under arbitrary intersection) under Truth Axiom, Monotonicity, Positive Introspection, and Finite (Arbitrary) Conjunction including Necessitation.[45]

Moreover, if knowledge satisfies Truth Axiom, Monotonicity, and Negative Introspection (again, recall Corollary 2) in a κ-knowledge space, then the self-evident collection becomes a κ-complete sub-algebra. To restate, for $\kappa \in \{\aleph_0, \aleph_1, \infty\}$, knowledge takes a form of a sub-algebra, σ-sub-algebra, and complete sub-algebra, respectively. Thus, a self-evident collection as a σ-sub-algebra could induce the conditional probability (belief). Since such σ-sub-algebra satisfies the maximality property, we would not need to impose an extra assumption on its cardinality.

At the same time, if a domain of knowledge is a complete algebra in the presence of probabilistic beliefs, then we would have to be careful about σ-measurability. This is because, knowledge, which takes a form of a complete sub-algebra, is usually richer than a σ-algebra.[46]

Fifth, a self-evident collection may not necessarily be closed under complementation in the absence of Negative Introspection, and hence we need to take care of this asymmetry when we interpret self-evident events. To see this point, observe that the possibility operator L_{K_i} satisfies $E^c \in \mathcal{J}_{K_i}$ iff $L_{K_i}(E) \subseteq E$ (i.e., E is true whenever i considers it possible). Put differently, when $E \in \mathcal{J}_{K_i}$ but $E^c \notin \mathcal{J}_{K_i}$, it is not the case that player i knows E is false (E^c is true) whenever E is false at ω. Colloquially, she cannot conclude that E is false just

[45] If an i's knowledge satisfies Truth Axiom, Monotonicity, Positive Introspection, and Arbitrary Conjunction (including Necessitation) in an ∞-knowledge space, then it will turn out (an an implication of Proposition 12 in Section 5.2) that $b_{K_i} : \Omega \to \mathcal{D}$ is a reflexive and transitive possibility correspondence. See Footnote 3 for the literature on non-partitional possibility correspondence models of knowledge.

[46] That is, a complete algebra may be "too large" for some σ-additive measure to be well defined. Also, assuming the Kripke property, an i's information set $b_{K_i}(\omega)$ may be null so that we have to take care of Bayesian updating on such a null set if i's posterior belief is derived from a prior distribution (see, for example, Brandenburger and Dekel [15] for the use of regular conditional probability measures or Nielsen [68] for enriching the underlying state space). On the other hand, modeling information by a σ-algebra may be problematic due to the fact that σ-algebras generated by partitions are not necessarily monotonic in the fineness of partitions (see Dubra and Echenique [24], Hérves-Beloso and Monteiro [42], Nielsen [68], Stinchcombe [81] and the references therein). We will shortly discuss how the maximality property can capture the preservation of informational contents.

by the fact that she does not observe E.[47]

If we start with the possibility operator, then the corresponding dual collection $\overline{\mathcal{J}}_{K_i} := \{F \in \mathcal{D} \mid F^c \in \mathcal{J}_{K_i}\}$ satisfies the minimality property in the following sense:

$$L_{K_i}(E) = \min\{F \in \overline{\mathcal{J}}_{K_i} \mid E \subseteq F\}(\in \overline{\mathcal{J}}_{K_i}).$$

That is, for any $E \in \mathcal{D}$, the dual collection $\overline{\mathcal{J}}_{K_i}$ always has the minimum event containing E. Since $\mathcal{J}_{K_i}$ is not necessarily closed under complementation, however, $\mathcal{J}_{K_i}$ does not necessarily satisfy the minimality property in the form of $L_{K_i}(E) = \min\{F \in \mathcal{J}_{K_i} \mid E \subseteq F\}$.[48] Without Negative Introspection of K_i, these two minimality properties do not necessarily coincide with each other.

Sixth, set-inclusion between self-evident collections naturally gives (global) "knowledgeability" relation. Namely, suppose that knowledge operators K_i and K_j satisfy Truth Axiom, Positive Introspection, and Monotonicity (equivalently, self-evident collections $\mathcal{J}_i$ and $\mathcal{J}_j$ satisfy the maximality property). Then, we have

$$\mathcal{J}_i \subseteq \mathcal{J}_j \textit{ iff } K_i(\cdot) \subseteq K_j(\cdot).$$

Generally, the second condition implies the first without any assumption on knowledge operators. This follows because $E \subseteq K_i(E) \subseteq K_j(E)$ for any $E \in \mathcal{J}_{K_i}$. If the self-evident collections represent players' knowledge (through the maximality property), then the first condition implies the second.

Now, we go on to our second result. For any given sub-collection $\mathcal{D}'$ of $\mathcal{D}$, we consider the smallest collection containing $\mathcal{D}'$, satisfying the maximality property and possibly various other logical and introspective properties. This observation enables us to give an informational content to any given sub-collection $\mathcal{D}'$, for example, in the following cases. First, for a given σ-algebra, we can consider the smallest σ-algebra containing the original one and satisfying the maximality property. Second, if a player's knowledge is given by her self-evident collection and she "learns" a certain set of events, then we can obtain the new smallest self-evident collection containing both collections. Third, for a given set of players, we can get the smallest self-evident collections containing the intersection of

[47] On the other hand, for any $E \in \mathcal{J}_{K_i}$ with $E^c \in \mathcal{J}_{K_i}$, player i knows whether E is true or not at any state (i.e., $\Omega = K_i(E) \cup K_i(E^c)$). In other words, if $\mathcal{J}_{K_i}$ forms a κ-complete sub-algebra of $\mathcal{D}$, then $\mathcal{J}_{K_i}$ comprises of an event E such that i knows whether E is true or not at any state. This formalizes the (common) sense in which knowledge is interpreted as a sub-algebra of a given domain (see, for example, Hérves-Beloso and Monteiro [42] and Stinchcombe [81]). Indeed, a self-evident event conicides with an event which Hérves-Beloso and Monteiro [42] call an informed set (with respect to a partition) in the following sense. If an agent's knowledge is represented by a partition (i.e., her knowledge satisfies the Kripke property, Truth Axiom, and Negative Introspection), then an event is self-evident to her iff it is an informed set with respect to her information partition (where technically we allow each partition cell $b_{K_i}(\omega)$ not to be an event).

[48] As an example, consider $\Omega = \{\omega_1, \omega_2\}$, $\mathcal{D} = \mathcal{P}(\Omega)$, and $\mathcal{J}_{K_i} = \{\emptyset, \{\omega_1\}\}$. We have $L_{K_i}(\{\omega_1\}) = \min\{F \in \overline{\mathcal{J}}_{K_i} \mid \{\omega_1\} \subseteq F\} = \Omega \neq \{\omega_1\} = \min\{F \in \mathcal{J}_{K_i} \mid E \subseteq F\}$. This happens because $\{\omega_1\}^c$ is not self-evident.

self-evident collections (i.e., the "publicly evident" (Milgrom [60]) collection $\bigcap_{i\in I} \mathcal{J}_i$). This is the "infimum" informational content associated with the group of players. We will, in fact, formalize the notion of common knowledge in Section 4.2 by slightly amending this observation. Fourth, we can also pool players' knowledge $\bigcup_{i\in I} \mathcal{J}_i$ and consider the smallest self-evident collection containing the pooling of thier knowledge. This corresponds to the notion of distributed knowledge (Halpern and Moses [34]).

Proposition 6 (Smallest Collection Satisfying Maximality)**.** *Let $\mathcal{D}'$ be a sub-collection of a κ-complete algebra $\mathcal{D}$ on Ω.*

1. *The smallest collection containing $\mathcal{D}'$ and satisfying the maximality property is*

$$\mathcal{J}(\mathcal{D}') := \{E \in \mathcal{D} \mid \textit{if } \omega \in E \textit{ then there is } F \in \mathcal{D}' \textit{ with } \omega \in F \subseteq E\}. \tag{4}$$

2. *The smallest collection containing $\mathcal{D}' \cup \{\Omega\}$ and satisfying the maximality property is $\mathcal{J}_{Nec}(\mathcal{D}') := \mathcal{J}(\mathcal{D}' \cup \{\Omega\}) = \mathcal{J}(\mathcal{D}') \cup \{\Omega\}$.*

3. *Fix $\lambda \leq \kappa$. If $\mathcal{D}'$ is closed under non-empty λ-intersection, so does $\mathcal{J}(\mathcal{D}')$. Generally, the smallest collection containing $\mathcal{D}'$, satisfying the maximality property, and closed under non-empty λ-intersection is*

$$\begin{aligned}\mathcal{J}_{\lambda\text{-}Con}(\mathcal{D}') &:= \mathcal{J}(\{E \in \mathcal{D} \mid E = \bigcap \mathcal{E} \textit{ for some } \mathcal{E} \subseteq \mathcal{D}' \textit{ with } 0 < |\mathcal{E}| < \lambda\}) \\ &= \{E \in \mathcal{D} \mid \textit{if } \omega \in E \textit{ then there is } \mathcal{E} \subseteq \mathcal{D}' \textit{ with } 0 < |\mathcal{E}| < \lambda \textit{ and } \omega \in \bigcap \mathcal{E} \subseteq E\}.\end{aligned}$$

Likewise, the smallest collection containing $\mathcal{D}'$, satisfying the maximality property and closed under λ-intersection is

$$\begin{aligned}\mathcal{J}_{\lambda\text{-}Con,\ Nec}(\mathcal{D}') &:= \mathcal{J}(\{E \in \mathcal{D} \mid E = \bigcap \mathcal{E} \textit{ for some } \mathcal{E} \subseteq \mathcal{D}' \textit{ with } |\mathcal{E}| < \lambda\}) \\ &= \{E \in \mathcal{D} \mid \textit{if } \omega \in E \textit{ then there is } \mathcal{E} \subseteq \mathcal{D}' \textit{ with } |\mathcal{E}| < \lambda \textit{ and } \omega \in \bigcap \mathcal{E} \subseteq E\}.\end{aligned}$$

4. *The collection $\mathcal{J}(\mathcal{D}')$ does not necessarily inherit the closure under complementation from $\mathcal{D}'$. The smallest collection $\mathcal{J}_{NI}(\mathcal{D}')$ containing $\mathcal{D}'$, satisfying the maximality property, and closed under complementation is given as follows.*[49] *We start with defining auxiliary collections. Let $\mathcal{C}_0 = \mathcal{D}'$. For a successor ordinal $\alpha = \beta + 1$, we let $\mathcal{C}_{\beta+1} := \mathcal{J}(\mathcal{C}_\beta) \cup \{E \in \mathcal{D} \mid E^c \in \mathcal{C}_\beta\}$. For a limit ordinal α, we let $\mathcal{C}_\alpha = \bigcup_{\beta<\alpha} \mathcal{C}_\beta$. Observe that $\mathcal{C}_\alpha \subseteq \mathcal{D}$ for all α. Letting α be the least ordinal with $\mathcal{C}_\alpha = \mathcal{C}_{\alpha+1}$, we define $\mathcal{J}_{NI}(\mathcal{D}') := \mathcal{C}_\alpha$.*

5. *Suppose that $\kappa = \infty$. Then, we have*

$$\begin{aligned}&\mathcal{J}(\mathcal{D}') = \bigcap \{\mathcal{J}' \in \mathcal{P}(\mathcal{D}) \mid \mathcal{D}' \subseteq \mathcal{J}' \textit{ and } \mathcal{J}' \textit{ is closed under arbitrary union}\}, \textit{ and} \\ &\mathcal{J}_{NI}(\mathcal{D}') = \mathcal{A}_\infty(\mathcal{D}').\end{aligned}$$

[49] Note that, by Proposition 5, $\mathcal{J}_{\mathrm{NI}}(\mathcal{D}')$ is also closed under κ-intersection.

We have the following four remarks on Proposition 6. First, we can write the knowledge operator associated with $\mathcal{J}(\mathcal{D}')$ through the maximality property as follows:

$$K_{\mathcal{J}(\mathcal{D}')}(E) = \{\omega \in \Omega \mid \text{there is } F' \in \mathcal{D}' \text{ with } \omega \in F' \subseteq E\} \text{ for each } E \in \mathcal{D}.$$

Second, Proposition 6 also implies that, with a κ-complete algebra $(\Omega, \mathcal{D})$ fixed, the family of all self-evident collections forms a complete lattice with respect to the set inclusion (i.e., the knowledgeability relation). Especially, for a given profile of self-evident collections $(\mathcal{J}_{K_i})_{i\in I}$, the infimum is given by $\mathcal{J}(\bigcap_{i\in I}\mathcal{J}_{K_i})$ while the supremum is given by $\mathcal{J}(\bigcup_{i\in I}\mathcal{J}_{K_i})$.

Third, it can be seen that the dual collection $\overline{\mathcal{J}(\mathcal{D}')}$ (i.e., $E \in \overline{\mathcal{J}(\mathcal{D}')}$ iff $E^c \in \mathcal{J}(\mathcal{D}')$) is given by

$$\overline{\mathcal{J}(\mathcal{D}')} = \{E \in \mathcal{D} \mid \text{if } E \cap F \neq \emptyset \text{ for any } (\omega, F) \in \Omega \times \mathcal{D}' \text{ with } \omega \in F, \text{ then } \omega \in E\}.$$

Fourth, we examine the sense in which $\mathcal{J}$ preserves collection of events. A part of the argument made by Dubra and Echenique [24] that σ-algebras are not necessarily adequate for modeling information is that the operation of taking the smallest σ-algebra does not necessarily preserve set inclusion with respect to partitions. Dubra and Echenique [24, Theorem A] and Hérves-Beloso and Monteiro [42, Propositions 1 and 5] study the operations that preserve an informational content (a partition) and establish Blackwell's theorem connecting the preferences over signals with information partitions. Here, we can see from Equation (4) that $\mathcal{J}$ preserves collections of events (including partitions) in the following sense.

Corollary 3 (Preservation of Information)**.** *Let $(\Omega, \mathcal{D})$ be a κ-complete algebra. Let $\mathcal{E}$ and $\mathcal{E}'$ be such that $\mathcal{E}, \mathcal{E}' \subseteq \mathcal{D}$. The following are equivalent.*

1. $\mathcal{J}(\mathcal{E}) \subseteq \mathcal{J}(\mathcal{E}')$.
2. *For any $E \in \mathcal{E}$ and $\omega \in E$, there is $E' \in \mathcal{E}'$ such that $\omega \in E' \subseteq E$.*

To conclude this subsection, we remark that Proposition A.3 in the Appendix characterizes knowledge morphisms in terms of self-evident collections. Thus, under the assumption that each player's knowledge satisfies Truth Axiom, Monotonicity, and Positive Introspection, players' self-evident collections can be given as a primitive of a knowledge space.

4.2 Common Knowledge

Here, we aim to formalize the notion of common knowledge among the group of players I by defining the common knowledge operator $C : \mathcal{D} \to \mathcal{D}$ in any given knowledge space.[50]

[50]First, as discussed in Footnote 4, we abuse the terminology, common knowledge (instead of common belief), even if individual players may violate Truth Axiom. Second, we can also introduce the common knowledge operator C_G among any subset of players $G \in \mathcal{P}(I)$, with the convention that $C_\emptyset := \mathrm{id}_{\mathcal{D}}$. Our analyses go through by replacing I with G. For example, recall the example in Footnote 18. There, player i's knowledge is represented by $K_{(0,i)}$ while her (non-probabilistic) belief is captured by $K_{(1,i)}$. In this case, we can consider the common knowledge among $\{0\} \times I$ and the common belief among $\{1\} \times I$.

Throughout this subsection, we simply assume that a given κ-knowledge space $\overrightarrow{\Omega}$ of I on $(S, \mathcal{A}_S)$ satisfies $|I| < \kappa$.

In order to define the notion of common knowledge for any domain and irrespective of assumptions on players' knowledge, we start with defining the following preliminary definition. An event $F \in \mathcal{D}$ is called a *common basis* (among I) if $F \subseteq K_I(E)$ for any $F \in \mathcal{D}$ with $F \subseteq E$, where $K_I(E) := \bigcap_{i \in I} K_i(E)$ is the event that every player in I knows E. Thus, an event F is a common basis if everybody knows any logical implication of F whenever F is true. We denote by $\mathcal{J}_I$ the collection of an event that forms a common basis among I. Clearly, if F is a common basis then F is publicly evident (i.e., $F \subseteq K_I(F)$). That is, $\mathcal{J}_I \subseteq \bigcap_{i \in I} \mathcal{J}_{K_i}$. Conversely, if every player's knowledge satisfies Monotonicity then these two notions coincide.

We take the self-referential approach to defining common knowledge by imposing the following three properties. First, the common knowledge of E implies the mutual knowledge of E: $C(E) \subseteq K_I(E)$. Second, if E is commonly known at ω, then there is a common basis F which is true at ω and which implies the very fact that E is common knowledge. Third, the event $C(E)$ is the largest event satisfying the above two properties.

Formally, we define the common knowledge operator $C : \mathcal{D} \to \mathcal{D}$ as follows.

$$C(E) := \max\{F \in \mathcal{J}(\mathcal{J}_I) \mid F \subseteq K_I(E)\} (\in \mathcal{J}(\mathcal{J}_I)) \text{ for each } E \in \mathcal{D}. \tag{5}$$

Note that it is $C(E) \in \mathcal{J}(\mathcal{J}_I)$ that ensures the second property (i.e., if $\omega \in C(E)$ then there is $F \in \mathcal{J}_I$ such that $\omega \in F \subseteq C(E)$). Note also that $C(E)$ is by definition a legitimate event.

In order to contrast common knowledge with mutual knowledge, we define the chain of mutual knowledge as follows. For a successor ordinal $\alpha = \beta + 1$, we define $K_I^\alpha := K_I \circ K_I^\beta$, starting with $K_I^1 := K_I$. For a limit ordinal $\alpha (> 0)$, we let $K_I^\alpha(\cdot) := \bigcap_{\beta : 1 \leq \beta < \alpha} K_I^\beta(\cdot)$. Generally, the mutual knowledge operator K_I^α is a well-defined operator for all $\alpha < \kappa$ in any κ-knowledge space.

From now on, we ask three strands of questions. The first is the relation between individual, mutual, and common knowledge. The second is how our definition of the common knowledge operator relates to those posited in the previous literature. The third is how a knowledge morphism preserves common knowledge. We start with examining the sense in which common knowledge inherits logical and introspective properties of individual knowledge.

Proposition 7 (Relation among Individual, Mutual, and Common Knowledge)**.** *Fix a κ-knowledge space $\overrightarrow{\Omega}$.*

1. *C always satisfies Positive Introspection.*
2. *If K_i satisfies No-Contradiction Axiom for some $i \in I$, then so does C. The same is true for Consistency and Truth Axiom.*
3. *If each K_i satisfies Monotonicity, then so does C.*

4. *Every K_i satisfies Necessitation iff C satisfies Necessitation.*

5. *If each K_i satisfies Monotonicity and Non-empty λ-Conjunction, then C satisfies Non-empty λ-Conjunction (as well as Monotonicity).*

6. *Negative Introspection of $(K_i)_{i \in I}$ does not necessarily imply that of C.*

7. *Let $\kappa = \infty$. Suppose that every K_i satisfies Truth Axiom, Positive Introspection, and Monotonicity. If each K_i also satisfies Negative Introspection, then C satisfies Negative Introspection.*

8. *Suppose that each K_i satisfies Positive Introspection and Monotonicity. Then, (i) C also satisfies Positive Introspection and Monotonicity; and (ii) for any knowledge operator $K : \mathcal{D} \to \mathcal{D}$ satisfying Positive Introspection and Monotonicity and satisfying $K(\cdot) \subseteq K_I(\cdot)$, we have $K(\cdot) \subseteq C(\cdot)$.*

9. *$C(\cdot) \subseteq K_I^{\alpha}(\cdot)$ for any ordinal $\alpha < \kappa$.*

The first eight statements of Proposition 7 demonstrate how the common knowledge operator does (and does not) inherit the properties of individual knowledge operators. The following are especially worth noting. First, Positive Introspection of C is endemic in its self-referential definition and the notion of common basis events. This holds regardless of assumptions on players' knowledge. In relation to this point, if an i' knowledge does not satisfy either Monotonicity or Positive Introspection, then i's knowledge is not necessarily the same as the common knowledge among $\{i\}$. See Remark A.5 in the Appendix for specific examples. They coincide with each other if both of Monotonicity and Positive Introspection are satisfied.

Second, while the common knowledge operator does not necessarily inherit Negative Introspection of individual knowledge operators, it does inherit Negative Introspection in an ∞-knowledge space satisfying Truth Axiom, Positive Introspection, and Monotonicity. This is because, in this case, the common knowledge is exactly the "individual" knowledge associated with the publicly evident events $\bigcap_{i \in I} \mathcal{J}_{K_i}$ and it inherits the closure under complementation, i.e., Negative Introspection.[51]

Proposition 7 (8) reveals the relation between individual and common knowledge in terms of knowledgeability. Put differently, it clarifies the sense in which common knowledge has been understood as the "infimum" of players' knowledge.[52] For example, as is well

[51] Publicly evident events (Milgrom [60]) are also termed as self-evident events (Aumann [5] and Rubinstein and Wolinsky [72]), common truisms (Binmore and Brandenburger [11]), public events (Geanakoplos [31]), belief closed events (Lismont and Mongin [50]), evident knowledge events (Monderer and Samet [61]), common information (Nielsen [68]), and so forth.

[52] In this respect, we can formally introduce the notion of distributed knowledge as a dual of common knowledge, i.e., the "supremum" of players' knowledge, under Truth Axiom, Monotonicity, and Positive Introspection. Namely, we define the distributed knowledge operator as the knowledge operator associated with the collection $\sup_{i \in I} \mathcal{J}_i = \mathcal{J}(\bigcup_{i \in I} \mathcal{J}_{K_i})$: $D_I(E) := \max\{F \in \mathcal{J}(\bigcup_{i \in I} \mathcal{J}_{K_i}) \mid F \subseteq E\}$. By construction,

known, common knowledge in a standard partitional model (Aumann [3]) is associated with the infimum partition (finest partition coarser than every player's partition). In this regard, Proposition 7 (7) also shows that the collection of publicly evident events $\bigcap_{i\in I}\mathcal{J}_{K_i}$ is exactly the infimum of players' knowledge in an ∞-knowledge space satisfying Truth Axiom, Positive Introspection, and Monotonicity. Indeed, for any κ-knowledge space satisfying Truth Axiom, Positive Introspection, and Monotonicity, the common knowledge is the infimum of players' knowledge because it is characterized by $\mathcal{J}(\bigcap_{i\in I}\mathcal{J}_{K_i})$. In other words, Proposition 7 (8) also holds for the combination of Truth Axiom, Positive Introspection, and Monotonicity.

Proposition 7 (9) means that the common knowledge of an event E implies the mutual knowledge of E up to any ordinal level $\alpha < \kappa$. Especially, if $\kappa \geq \aleph_1$ so that countable conjunction can legitimately be taken, the common knowledge of E implies the countable chain of mutual knowledge of E, which is often taken as an intuitive (or iterative) definition of common knowledge. That is, for a given κ-knowledge space with $\kappa \geq \aleph_1$, an event E is common knowledge, in an intuitive sense, within I at a state $\omega \in \Omega$ if, everybody (in I) knows E at ω, everybody knows that everybody knows E at ω, *ad infinitum*. We denote the intuitive notion of common knowledge of E among I by $K_I^{\infty}(E) := \bigcap_{n\in\mathbb{N}} K_I^n(E)$.[53] We will shortly ask when this intuitive notion coincides with the common knowledge operator defined by Equation (5).

The definition of common basis events plays an important role in establishing Proposition 7 (9) without imposing Monotonicity. McCarthy, Sato, Hayashi, and Igarashi [53] define the notion of what "any fool" knows in the sense that the knowledge of event by "any fool" implies the chain of mutual knowledge. In our definition, the notion of common knowledge can be interpreted as the individual knowledge associated with the common basis events.

Now, we go on to the second strand of questions. We start with showing that our definition of common knowledge nests the previous characterizations when Monotonicity is assumed. Specifically, we consider the characterization by Monderer and Samet [61] in terms of publicly evident collections (Equation (6) below) and the "fixed-point" characterization by Friedell [29] and Halpern and Moses [34] (Equation (7) below).

Proposition 8 (Characterizations of Common Knowledge under Monotonicity)**.** *Let* $\overrightarrow{\Omega}$ *be*

D_I satisfies Truth Axiom, Positive Introspection, and Monotonicity. This definition captures the original idea by Halpern and Moses [34] that an event E is distributed knowledge at a state ω if someone who were informed of everything that each player $i \in I$ knows at ω would know E at ω.

[53] First, we denote by $\mathbb{N}$ the set of positive integers. That is, we do not require C to be correct. Second, we use the superscript ∞ instead of the least infinite ordinal (ω). The use of ∞ here is different from that in the context of ∞-complete algebras. Third, another intuitive (and weaker) notion of common knowledge would be that, an event E is common knowledge among I at ω, if, for any finite sequence of players $(i_1, i_2, \ldots, i_n)$ in I with $n \in \mathbb{N}$, player i_1 knows that player i_2 knows that $\ldots$ player i_n knows E at ω (i.e., $\omega \in (K_{i_n} \circ \cdots \circ K_{i_2} \circ K_{i_1})(E)$). It can be seen that, under Monotonicity and (Non-empty) λ-Conjunction with $\lambda > |I|$, these two intuitive notions coincide.

a knowledge space such that every K_i satisfies Monotonicity. For each $E \in \mathcal{D}$,

$$C(E) = \{\omega \in \Omega \mid \text{there is } F \in \bigcap_{i \in I} \mathcal{J}_{K_i} \text{ with } \omega \in F \subseteq K_I(E)\} \tag{6}$$

$$= \max\{F \in \mathcal{D} \mid F \in \bigcap_{i \in I} \mathcal{J}_{K_i} \text{ and } F \subseteq K_I(E)\} \in \mathcal{D}; \text{ and}$$

$$C(E) = \max\{F \in \mathcal{D} \mid F = K_I(E) \cap K_I(F)\} (= \max\{F \in \mathcal{D} \mid F \subseteq K_I(E) \cap K_I(F)\}).$$

Suppose, in addition, that each K_i satisfies Finite Conjunction. For any $E \in \mathcal{D}$,

$$C(E) = \max\{F \in \mathcal{D} \mid F = K_I(E \cap F)\} (= \max\{F \in \mathcal{D} \mid F \subseteq K_I(E \cap F)\}). \tag{7}$$

Equation (6) follows because the common basis events coincide with the publicly evident events under Monotonicity. Thus, the last term in Equation (6) is well defined by the maximality property of $\mathcal{J}(\bigcap_{i \in I} \mathcal{J}_{K_i})$. In other words, C is the knowledge operator associated with $\mathcal{J}(\bigcap_{i \in I} \mathcal{J}_{K_i})$.

Monderer and Samet [61, Proposition] study the common knowledge of an event in a standard partitional knowledge space defined on the power set of underlying states of the world. On the other hand, the above characterization holds in any domain $\mathcal{D}$ and irrespective of assumptions on players' knowledge as long as Monotonicity is imposed. Likewise, the maximality property of $\mathcal{J}(\bigcap_{i \in I} \mathcal{J}_{K_i})$ ensures that Equation (7) be well defined.

Next, we ask when our notion of common knowledge coincides with the intuitive (iterative) notion. Imposing Monotonicity and Countable Conjunction on players' knowledge operators, the answer to the above question is to utilize a well-known (a variant of Tarski's) fixed point theorem stating that $K_I^{\infty}(E)$ is the greatest fixed point of the monotone and continuous set operator $f_E(F) = K_I(E \cap F)$ (see, for example, Halpern and Moses [34]). In other words, the intuitive definition ($C = K_I^{\infty}$), the fixed-point definition as well as the one by publicly evident events (Proposition 8), and our definition (Equation (5)) all agree as long as knowledge operators satisfy Monotonicity and Countable Conjunction.

Corollary 4 (Intuitive/Iterative Notion of Common Knowledge)**.** *Let $\overrightarrow{\Omega}$ be a κ-knowledge space where $\kappa \geq \aleph_1$ and knowledge operators satisfy Monotonicity and Countable Conjunction. Then, $C = K_I^{\infty}$.*

The reason that Corollary 4 may fail in the absence of Countable Conjunction is that K_I^{∞} may fail to satisfy Positive Introspection $K_I^{\infty}(E) \subseteq K_I^{\infty} K_I^{\infty}(E) (= K_I^{\omega+\omega}(E))$ by the very fact that the chain of mutual knowledge stops at the least infinite ordinal level.[54]

We consider an implication of Corollary 4 when each player's knowledge satisfies the Kripke property and Positive Introspection. In this case, observe first that the "everyone-knows" operator K_I has the Kripke property as we have $b_{K_I}(\cdot) = \bigcup_{i \in I} b_{K_i}(\cdot)$. By induction,

[54] While the first infinite ordinal number of chain of mutual knowledge is already hard to check in reality (Monderer and Samet [61]), the definition of common knowledge as a chain of mutual knowledge is in fact not strong enough to capture introspective properties of common knowledge (see, for example, Barwise [8] and Lismont and Mongin [50]).

each mutual knowledge operator K_I^n has the Kripke property. Indeed, it satisfies that $b_{K_I^n}(\cdot) = b_{K_I}^n(\cdot)$, where $b_{K_I}^1 := b_{K_I}$ and, for $n \geq 2$, $b_{K_I}^n(\omega) := \bigcup_{\omega' \in b_{K_I}(\omega)} b_{K_I}^{n-1}(\omega')$ for each $\omega \in \Omega$. Then, it can be seen that the common knowledge operator $C = K_I^\infty$ inherits the Kripke property, where $b_C(\cdot) = \bigcup_{n\in\mathbb{N}} b_{K_I}^n(\cdot)$ (i.e., $b_C(\cdot)$ is equal to the transitive closure of $b_{K_I}(\cdot)$).

We can connect our argument to the notion of reachability (Aumann [3]).[55] We say that ω' is *reachable* from ω if there are sequences $(\omega^j)_{j=1}^m$ of Ω and $(i^j)_{j=1}^{m-1}$ of I with $m \in \mathbb{N} \setminus \{1\}$ such that $\omega = \omega^1$, $\omega' = \omega^m$, and $\omega^{j+1} \in b_{K_{i^j}}(\omega^j)$ for all $j \in \{1, \ldots, m-1\}$. Then, we have the following.

Proposition 9 (Common Knowledge and the Kripke Property/Reachability)**.** *Let $\overrightarrow{\Omega}$ be a κ-knowledge space with $\kappa \geq \aleph_1$ such that each player's knowledge satisfies the Kripke property and Positive Introspection. Then, each K_I^n (with $n \in \mathbb{N}$) and C inherit the Kripke property. Moreover, the following hold.*

1. *For each $n \in \mathbb{N}$, $b_{K_I^n} = b_{K_I}^n$.*
2. *$b_C(\omega) = \bigcup_{n\in\mathbb{N}} b_{K_I}^n(\omega) = \{\omega' \in \Omega \mid \omega' \text{ is reachable from } \omega\}$ for each $\omega \in \Omega$.*

As the third strand of questions, we ask when a knowledge morphism $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega'}$ preserves common knowledge in the following sense: for any event $E' \in \mathcal{D}'$, the event E' is common knowledge at $\varphi(\omega)$ iff $\varphi^{-1}(E')$ is common knowledge at ω'. Clearly, if the common knowledge operator can be written in terms of composites of individual knowledge operators, common knowledge is preserved under a knowledge morphism. For example, in the category of κ-knowledge spaces (with $\kappa \geq \aleph_1$) satisfying Monotonicity and Countable Conjunction, $C = \bigcap_{n\in\mathbb{N}} K_I^n$ (recall Corollary 4) commutes with the set-algebraic operations (i.e., $\bigcap_{n\in\mathbb{N}} K^n(\varphi^{-1}(E')) = \varphi^{-1}(\bigcap_{n\in\mathbb{N}} K'^n(E'))$).[56]

Next, we examine how common knowledge is preserved between a given knowledge space and the universal knowledge space $\overrightarrow{\Omega^*}$ established in Section 3.1.[57]

Proposition 10 (Preservation of Common Knowledge in Universal Space)**.** *Suppose that assumptions on players' knowledge include Monotonicity and Positive Introspection and that they are homogeneously given across players in the given category of knowledge spaces. Then, for any knowledge space $\overrightarrow{\Omega}$, we have*

$$D_{\overrightarrow{\Omega}}^{-1}(C^*([e])) = C(D_{\overrightarrow{\Omega}}^{-1}([e])) \text{ for all } [e] \in \mathcal{D}^*,$$

where C and C^* are the common knowledge operators in $\overrightarrow{\Omega}$ and $\overrightarrow{\Omega^*}$, respectively.

[55] See also Halpern and Moses [34], Hérves-Beloso and Monteiro [42], and the references therein.

[56] On a related point, if $\kappa = \infty$ then Monotonicity guarantees that there is a limit ordinal α such that $K_I^\alpha = C$. Thus, common knowledge is preserved in such a case as well. See Remark A.6 in the Appendix.

[57] In Section 6, we also demonstrate the existence of a universal knowledge space in terms of hierarchies of knowledge. Proposition 10 also applies to the universal knowledge space established in Section 6.

We establish Proposition 10 in the following way. First, suppose that an event $[e]$ is commonly known at $D_{\overrightarrow{\Omega}}(\omega)$. Given that $\overrightarrow{\Omega}$ satisfies Monotonicity, it can be seen that $[\![e]\!]_{\overrightarrow{\Omega}} = D_{\overrightarrow{\Omega}}^{-1}([e])$ is commonly known at ω. For converse, it follows from the assumptions of the proposition that $C^* = K_i^*$ for all $i \in I$.

Generally, how does a knowledge morphism from one space to another preserve common knowledge? We provide a general characterization in the Appendix (Proposition A.4).

5 Knowledge and Common Knowledge of the Structure

Players' probabilistic beliefs are usually represented by the concept of types (Harsanyi [35]). Can we define a qualitative analogue of type mappings in order to capture players' knowledge?

In Section 5.1, we represent each player's knowledge at a state as a "knowledge-type" and show that this knowledge-type approach is equivalent to the knowledge operator approach for any given assumptions on players' knowledge. In Section 6, we establish a universal knowledge space in terms of hierarchies of knowledge-types (Theorem 3), and thus we formally connect Harsanyi's idea of representing players' probabilistic beliefs by belief-types to that of representing players' knowledge by knowledge-types.

The main goal of this section, however, is to formalize a notion of the structure of the model of knowledge by using players' knowledge-type mappings. We regard each player's knowledge-type mapping with a "signal" and ask the sense in which she knows hew own signal. Thus, in Section 5.2, we formally define what it means by a signal and examine the sense in which each player knows her knowledge-type mapping in terms of her introspective properties. In Section 5.3, we apply this formalization to the universal knowledge space established in Section 3. By using this formalization, we (the outside analysts) can say that the players know (indeed, commonly know) the structure of the universal knowledge space.

5.1 Representing Knowledge by Knowledge-Types

Throughout the subsection, fix a κ-complete algebra $(\Omega, \mathcal{D})$, where κ is an infinite (regular) cardinal or $\kappa = \infty$. Each knowledge-type is a mapping $\mu : \mathcal{D} \to \{0,1\}$ (i.e., $\mu \in \{0,1\}^{\mathcal{D}}$), where the knowledge of an event $E \in \mathcal{D}$ is captured by $\mu(E) = 1$. We represent each player's knowledge by a mapping, which we call a knowledge-type mapping, $t_i : \Omega \to \{0,1\}^{\mathcal{D}}$ with the following interpretation: player i knows an event E at state ω if $t_i(\omega)(E) = 1$ with $t_i(\omega)$ being her knowledge-type at ω.

Specifically, we define a knowledge-type mapping t_i in the following two steps. The first is to define the set of knowledge-types (a subset of $\{0,1\}^{\mathcal{D}}$) that reflects given assumptions on knowledge. Put differently, just as $\Delta(\cdot)$ stands for the set of σ-additive probability measures over a measurable space, we aim to formalize the set of legitimate binary "measures" $M(\cdot)$

which represents knowledge on a κ-complete algebra. The second is to define a knowledge-type mapping t_i as a mapping from $(\Omega, \mathcal{D})$ into $(M(\Omega, \mathcal{D}), \mathcal{M}(\Omega, \mathcal{D}))$ which satisfies the notion of knowledge in question, where $\mathcal{M}(\Omega, \mathcal{D})$ is a κ-complete algebra on $M(\Omega, \mathcal{D})$.

Our first task is to define $M(\Omega)(= M(\Omega, \mathcal{D}))$ as a given subset of $\{0,1\}^{\mathcal{D}}$ based on a concept of knowledge we adopt, as well as the κ-complete algebra $\mathcal{M}(\Omega, \mathcal{D})$. Recall that some properties of knowledge are referred to as logical properties (i.e., No-Contradiction Axiom, Consistency, Monotonicity, Non-empty λ-Conjunction with $\lambda \leq \kappa$, and Necessitation). The others are referred to as the introspective properties (Truth Axiom, Positive Introspection, and Negative Introspection) and the Kripke property.

We translate the logical properties of knowledge in terms of knowledge-types.

Definition 11 (Logical Properties of Knowledge-Types). *Fix* $\mu \in \{0,1\}^{\mathcal{D}}$.

1. *No-Contradiction Axiom:* $\mu(\emptyset) = 0$.
2. *Consistency:* $\mu(E) \leq 1 - \mu(E^c)$ *for all* $E \in \mathcal{D}$.
3. *Monotonicity:* $\mu(E) \leq \mu(F)$ *for all* $E, F \in \mathcal{D}$ *with* $E \subseteq F$.
4. *Non-empty* λ*-Conjunction:* $\min_{E \in \mathcal{E}} \mu(E) \leq \mu(\bigcap \mathcal{E})$ *for all* $\mathcal{E} \subseteq \mathcal{D}$ *with* $0 < |\mathcal{E}| < \lambda \leq \kappa$.
5. *Necessitation:* $\mu(\Omega) = 1$.

The logical properties of μ are completely analogous to the corresponding logical properties of knowledge operators. Unlike (σ-)additive probabilities, however, it is possible that players do not know an event E nor its negation E^c at a particular state, and thus it is possible that $\mu(E) + \mu(E^c) = 0$.

Going back to the properties of knowledge operators, recall that properties of knowledge operators are related with each other. For example, Truth Axiom, which is referred to as an introspective property, implies the logical properties of No-Contradiction Axiom and Consistency. Thus, for a given concept of knowledge, fix a set of properties of knowledge to be assumed and consider all the implied logical properties. For example, if Truth Axiom, Monotonicity, and Negative Introspection are assumed, then we adopt all the logical properties as shown in Corollary 2 in Section 4.1, where λ-Conjunction becomes κ-Conjunction.

We define $M(\Omega, \mathcal{D})$ to be the set of mappings $\mu \in \{0,1\}^{\mathcal{D}}$ which satisfy all the implied logical properties in question. Thus, note that the set $M(\Omega, \mathcal{D})$ depends on properties of each player's knowledge we adopt. That is, while we suppress players' identities, $M(\Omega, \mathcal{D})$ depends on the identity of a player i if different assumptions on knowledge are assumed across players.

Next, we introduce a κ-complete algebra $\mathcal{M}(\mathcal{D})(= \mathcal{M}(\Omega, \mathcal{D}))$ on $M(\Omega)$. In an analogous way to the probabilistic type space approach (e.g., HS [39]), we define $\mathcal{M}(\mathcal{D})$ to be the κ-complete algebra generated by the sets of the form $\mathcal{M}_E := \{\mu \in M(\Omega) \mid \mu(E) = 1\}$ for all $E \in \mathcal{D}$. Abusing the notation, we sometimes write $\mathcal{M}(E) := \mathcal{M}_E$ when it is clear that

E is not equal to the domain $\mathcal{D}$. Also, we denote by $\mathcal{M}^{(\Omega,\mathcal{D})}(E)$ if we stress the underlying κ-complete algebra $(\Omega, \mathcal{D})$. Thus, the κ-complete algebra that we introduce on $M(\Omega)$ is given by

$$\mathcal{M}(\mathcal{D}) := \mathcal{A}_\kappa(\{\mathcal{M}_E \in \mathcal{P}(M(\Omega)) \mid E \in \mathcal{D}\}).$$

The second task is to define a *knowledge-type mapping* as a κ-measurable mapping $t_i : (\Omega, \mathcal{D}) \to (M(\Omega), \mathcal{M}(\mathcal{D}))$ which satisfies all the given properties of knowledge. This is done in the following three steps.

First, the above κ-measurablity condition of t_i means that the set $t_i^{-1}(\mathcal{M}_E)$ (i.e., the set of states at which player i knows an E) is an event. Formally, for all $E \in \mathcal{D}$,

$$t_i^{-1}(\mathcal{M}_E) = t_i^{-1}(\{\mu \in M(\Omega) \mid \mu(E) = 1\}) \in \mathcal{D}.$$

Second, we specify the logical properties of a knowledge-type mapping. A knowledge-type mapping t_i satisfies a given logical property iff $t_i(\omega)$ satisfies it for all $\omega \in \Omega$. Thus, for example, we say that t_i satisfies *No-Contradiction Axiom* if $t_i(\omega)$ satisfies it (i.e., $t_i(\omega)(\emptyset) = 0$) for all $\omega \in \Omega$. The other logical properties (Consistency, Monotonicity, Non-empty λ-Conjunction, and Necessitation) are defined in the same way.

Third, we define below the introspective and Kripke properties of a knowledge-type mapping. Then, a knowledge-type mapping is formally defined as a κ-measurable mapping $t_i : (\Omega, \mathcal{D}) \to (M(\Omega), \mathcal{M}(\mathcal{D}))$ which satisfies the required properties of knowledge.

Definition 12 (Introspective Properties of Knowledge-Types)**.** *Fix a κ-measurable mapping $t_i : (\Omega, \mathcal{D}) \to (M(\Omega), \mathcal{M}(\mathcal{D}))$. We define the introspective and Kripke properties of t_i as follows.*

1. *Truth Axiom: for any $\omega \in \Omega$ and $E \in \mathcal{D}$, $t_i(\omega)(E) = 1$ implies $\omega \in E$.*

2. *Positive Introspection: for any $\omega \in \Omega$ and $E \in \mathcal{D}$, $t_i(\omega)(E) = 1$ implies $t_i(\omega)(\{\omega' \in \Omega \mid t_i(\omega')(E) = 1\}) = 1$ (i.e., $t_i(\omega)(t_i^{-1}(\mathcal{M}_E)) = 1$).*

3. *Negative Introspection: for any $\omega \in \Omega$ and $E \in \mathcal{D}$, $t_i(\omega)(E) = 0$ implies $t_i(\omega)(\{\omega' \in \Omega \mid t_i(\omega')(E) = 0\}) = 1$ (i.e., $t_i(\omega)(\neg t_i^{-1}(\mathcal{M}_E)) = 1$).*

4. *The Kripke property: for any $\omega \in \Omega$ and $E \in \mathcal{D}$, $b_{t_i}(\omega)(:= \bigcap\{F \in \mathcal{D} \mid t_i(\omega)(F) = 1\}) \subseteq E$ implies $t_i(\omega)(E) = 1$.*

We call a κ-measurable mapping $t_i : (\Omega, \mathcal{D}) \to (M(\Omega), \mathcal{M}(\mathcal{D}))$ to be a knowledge-type mapping if it satisfies all the required properties of knowledge in question.

In order to conclude this subsection, we establish that this knowledge-type approach is equivalent to the knowledge operator approach. Fix a notion of knowledge (i.e., a category of knowledge spaces). Suppose that each player's knowledge is assigned by her knowledge operator $K_i : \mathcal{D} \to \mathcal{D}$. Let $M(\Omega, \mathcal{D})$ be the set of i's knowledge-types which satisfy the logical properties of knowledge derived from i's knowledge operators of the given category (note again that we suppress players' identities on $M(\Omega)$).

Now, let $t_{K_i} : \Omega \to M(\Omega)$ be such that for each $\omega \in \Omega$ and $E \in \mathcal{D}$,

$$t_{K_i}(\omega)(E) = \begin{cases} 1 & \text{if } \omega \in K_i(E) \\ 0 & \text{otherwise} \end{cases}.$$

The mapping t_{K_i} clearly inherits all the properties of K_i, and hence it is well defined as a map $t_{K_i} : \Omega \to M(\Omega)$.[58] We also have $t_{K_i}^{-1}(\mathcal{M}_E) = K_i(E) \in \mathcal{D}$, so that $t_{K_i} : (\Omega, \mathcal{D}) \to (M(\Omega), \mathcal{M}(\mathcal{D}))$ is κ-measurable. Hence, t_{K_i} is a knowledge-type mapping.

Conversely, suppose that a knowledge-type mapping $t_i : (\Omega, \mathcal{D}) \to (M(\Omega), \mathcal{M}(\mathcal{D}))$ is given. Then, we define the knowledge operator $K_{t_i} : \mathcal{D} \to \mathcal{D}$ as follows:

$$K_{t_i}(E) := \{\omega \in \Omega \mid t_i(\omega)(E) = 1\} (= t_i^{-1}(\mathcal{M}_E)) \in \mathcal{D} \text{ for each } E \in \mathcal{D}.$$

It is the κ-measurability of t_i that ensures that K_{t_i} send events to events. It can be easily seen that K_{t_i} inherits all the properties imposed on t_i. Again, see Remark A.7.

Finally, it is clear that $K_{t_{K_i}} = K_i$ and $t_{K_{t_i}} = t_i$. In fact, given K_i, we have $\omega \in K_{t_{K_i}}(E)$ iff $t_{K_i}(\omega)(E) = 1$ iff $\omega \in K_i(E)$. Given t_i, we have $t_{K_{t_i}}(\omega)(E) = 1$ iff $\omega \in K_{t_i}(E)$ iff $t_i(\omega)(E) = 1$. The above equalities suggest that a knowledge space is equivalently defined by $\langle (\Omega, \mathcal{D}), (t_i)_{i \in I}, \Theta \rangle$, where each $t_i : (\Omega, \mathcal{D}) \to (M(\Omega), \mathcal{M}(\mathcal{D}))$ is i's knowledge-type mapping.[59]

To conclude this subsection, we have two remarks. First, we can define the possibility-type mapping $\bar{t}_i$ by $\bar{t}_i(\omega)(E) := 1 - t_i(\omega)(E^c)$ in light of the possibility operator induced from a knowledge operator. It is clear that specifying possibility-types is equivalent to specifying knowledge-types. Using the notion of possibility-types, for example, we can re-write Consistency as $t_i(\omega)(\cdot) \leq \bar{t}_i(\omega)(\cdot)$ for all $\omega \in \Omega$.

Second, specifying an i's knowledge-type at each state is equivalent to specifying the collection of events that i knows at that state.[60] For example, for a given knowledge-type mapping t_i, the collection of events that i knows at ω is given by $(t_i(\omega))^{-1}(\{1\}) = \{E \in \mathcal{D} \mid t_i(\omega)(E) = 1\}$.

5.2 Meta-Knowledge of Knowledge-Type Mappings

In a given knowledge space, the objects of knowledge are events, i.e., a subset of states of the world. How can we extend the notion of the knowledge of an event to that of a mapping such as players' strategies defined on the states of the world?

[58] For completeness, we provide a proof in Remark A.7 in the Appendix.

[59] Formally, the category of knowledge spaces in terms of knowledge operators is equivalent (in the category theoretical sense) to that of knowledge spaces in terms of knowledge-type mappings. See Remark A.8 in the Appendix for the proof.

[60] This latter specification corresponds to a neighborhood (Montague-Scott) system in the context of modal logic.

In this subsection, we first define a notion of a signal mapping. It is simply a mapping defined on underlying states of the world. We formalize the idea that a player knows a signal mapping. Examples of signal mappings include action/decision, strategies, random variables, and (knowledge-)type mappings. Second, we define a notion of informativeness derived from a signal mapping. Third, we apply these notions to players' type mappings in order to study the meta-knowledge of the structure of a model.[61] Especially, we examine what it means by the fact that a player knows her knowledge-type mapping.

Throughout the subsection, $(\Omega, \mathcal{D})$ refers to a κ-complete algebra, where κ is an infinite (regular) cardinal or $\kappa = \infty$. We start with defining a notion of a signal mapping. Let X be a set and let $\mathcal{D}_X$ be a subset of $\mathcal{P}(X)$. A *signal* mapping is a mapping $x : (\Omega, \mathcal{D}) \to (X, \mathcal{D}_X)$ satisfying $x^{-1}(\mathcal{D}_X) \subseteq \mathcal{D}$. Examples include strategies, action/decision functions, random variables, and so on. Now, we define a notion that an agent knows a signal mapping.

Definition 13 (Knowledge of Signal)**.** *Let $\overrightarrow{\Omega}$ be a knowledge space, and let $x : (\Omega, \mathcal{D}) \to (X, \mathcal{D}_X)$ be a signal mapping. An agent i is said to know the signal mapping $x : (\Omega, \mathcal{D}) \to (X, \mathcal{D}_X)$ at ω (or, she knows the mapping x at ω with respect to $\mathcal{D}_X$) if $x^{-1}(E_X) \subseteq K_i(x^{-1}(E_X))$ for any $E_X \in \mathcal{D}_X$ with $x(\omega) \in E_X$. If i knows the signal mapping x at any state (equivalently, if $x^{-1}(\mathcal{D}_X) \subseteq \mathcal{J}_{K_i}$), then we say that i knows x.*

We have the following four remarks. First, at an interpretational level, when we examine whether an agent i knows a signal mapping $x : (\Omega, \mathcal{D}) \to (X, \mathcal{D}_X)$, we pre-suppose that the mapping x is observable to i with respect to $\mathcal{D}_X$.

Second, in words, an agent i knows a signal mapping $x : (\Omega, \mathcal{D}) \to (X, \mathcal{D}_X)$ at ω when any event $x^{-1}(E_X)$ containing ω is self-evident to i. If $\mathcal{D}_X$ contains the singleton $\{x(\omega)\}$ and if the agent i knows the signal mapping x at ω, then $[x(\omega)] := x^{-1}(\{x(\omega)\}) = \{\omega' \in \Omega \mid x(\omega') = x(\omega)\}$ is self-evident to her.[62]

Third, the knowledge of the signal mapping x is considered to be the "measurability" of $x : (\Omega, \mathcal{J}_{K_i}) \to (X, \mathcal{D}_X)$.[63] Indeed, recall that $\mathcal{J}_{K_i}$ forms a (κ-complete) sub-algebra of $\mathcal{D}$ if i's knowledge satisfies Truth Axiom, Monotonicity, and Negative Introspection.

Fourth, we can extend the notion of a signal mapping to common knowledge. We say that $x : (\Omega, \mathcal{D}) \to (X, \mathcal{D}_X)$ is *common knowledge* at ω (or, x is *common knowledge* at ω with respect to $\mathcal{D}_X$) if, for any $E_X \in \mathcal{D}_X$ with $x(\omega) \in E_X$, we have $x^{-1}(E_X) \subseteq C(x^{-1}(E_X))$. Next, we say that x is *common knowledge* if it is common knowledge at any state (i.e., $x^{-1}(E_X) \subseteq C(x^{-1}(E_X))$ for all $E_X \in \mathcal{D}_X$).

[61] See Footnote 35 for the literature posing the meta-knowledge of the structure itself. Bacharach [6, 7] formalizes the event that an agent has an information partition by regarding it as a signal mapping in a related manner.

[62] In the literature (e.g., the one on a characterization of a solution concept of a game in state space models of knowledge), the knowledge of a mapping $x : (\Omega, \mathcal{D}) \to X$ at ω is often defined by the requirement that $[x(\omega)]$ is self-evident (e.g., BDG [17] and Geanakoplos [31]). This corresponds to the case where $\mathcal{D}_X = \{\{x(\omega)\} \in \mathcal{P}(X) \mid \omega \in \Omega\}$ (provided that $x^{-1}(\mathcal{D}_X) \subseteq \mathcal{D}$). If, for example, x is an action function, then $\mathcal{D}_X = \{\{x(\omega)\} \mid \omega \in \Omega\}$ is associated with actions that could have been taken at each state.

[63] For example, Aumann [4] defines the knowledge of a strategy by its measurability with respect to a partition.

We proceed with defining a concept of informativeness derived from a signal mapping.

Definition 14 (Informativeness according to Signal)**.** *Fix states ω and ω' in Ω. We say that ω is at least as informative as ω' according to a signal $x : (\Omega, \mathcal{D}) \to (X, \mathcal{D}_X)$ if*

$$\{E_X \in \mathcal{D}_X \mid \omega' \in x^{-1}(E_X)\} \subseteq \{E_X \in \mathcal{D}_X \mid \omega \in x^{-1}(E_X)\}.$$

Likewise, we say that states ω and ω' are equally informative according to $x : (\Omega, \mathcal{D}) \to (X, \mathcal{D}_X)$ if

$$\{E_X \in \mathcal{D}_X \mid \omega' \in x^{-1}(E_X)\} = \{E_X \in \mathcal{D}_X \mid \omega \in x^{-1}(E_X)\}.$$

The ideas behind Definition 14 are (i) that the informational content of a signal mapping $x : (\Omega, \mathcal{D}) \to (X, \mathcal{D}_X)$ at ω is expressed as the collection of sets $\{E_X \in \mathcal{D}_X \mid x(\omega) \in E_X\}$ and (ii) that informational contents are ranked by the implication in the form of set inclusion.[64] The notion of informativeness is clearly reflexive and transitive.

For the rest of this subsection, we apply these notions to knowledge-type mappings. Thus, we firstly examine the fact that a player knows her knowledge-type mapping in terms of introspection. Second, we study the relation between informativeness and possibility in order to characterize the sense in which a player knows her knowledge-type mapping in terms of informativeness.

The first aim is thus to characterize what it means by the fact that player i knows her knowledge-type mapping, in terms of introspection.

Proposition 11 (Knowledge of Type and Introspection)**.** *Let $\overrightarrow{\Omega} := \langle (\Omega, \mathcal{D}), (t_i)_{i \in I}, \Theta \rangle$ be a knowledge space. Fix $i \in I$.*

1. *Player i knows her knowledge-type mapping t_i with respect to $\{\mathcal{M}(E) \mid E \in \mathcal{D}\}$ iff $K_{t_i}(\cdot) \subseteq K_{t_i} K_{t_i}(\cdot)$.*

2. *Player i knows t_i with respect to $\{\neg\mathcal{M}(E) \mid E \in \mathcal{D}\}$ iff $(\neg K_{t_i})(\cdot) \subseteq K_{t_i}(\neg K_{t_i})(\cdot)$.*

We make the following three remarks. First, in the literature of (standard) non-partitional models of knowledge, the lack of Negative Introspection is interpreted as the fact that an agent does not know her own possibility correspondence.[65] Without imposing Negative Introspection, an agent does not know her own knowledge-type mapping with respect to $\{\mathcal{M}(E) \mid E \in \mathcal{D}\} \cup \{\neg\mathcal{M}(E) \mid E \in \mathcal{D}\}$. Rather, she takes her own information at face value in the sense that she only knows her knowledge-type mapping with respect to her own knowledge (i.e., $\{\mathcal{M}(E) \mid E \in \mathcal{D}\}$).

Second, it is worth pointing out that Positive Introspection and Negative Introspection in Proposition 11 pertain to every event including $E = K_j(F)$ for some $F \in \mathcal{D}$. This means

[64] The notion of informativeness is closely related to that of information studied by Bonanno [13]. Lipman [49] and Mukerji [67] also study the idea that informational contents are ranked by the implication in the form of set inclusion.

[65] See, for instance, BDG [17], Dekel and Gul [20], and Geanakoplos [31].

that if player i knows her knowledge-type mapping t_i with respect to $\{\mathcal{M}(E) \mid E \in \mathcal{D}\}$ then she also knows such knowledge-type mapping as $t_{K_{t_i}K_{t_j}}$ with respect to $\{\mathcal{M}(E) \mid E \in \mathcal{D}\}$, where $t_{K_{t_i}K_{t_j}}$ is the knowledge-type mapping associated with the operator $K_{t_i}K_{t_j}$ (i.e., $t_{K_{t_i}K_{t_j}}(\omega)(E) = 1$ iff $\omega \in K_{t_i}K_{t_j}(E)$). On the one hand, the fact that each i knows $t_{K_{t_i}K_{t_j}}$ could possibly be a justification for why we (the outside analysts) can assume that "i knows j's knowledge operator in i's mind." On the other hand, it is implicitly assumed that player i figures out what K_{t_j} is. We'll revisit (in Proposition 13 in Section 5.3) the question whether we (the outside analysts) can assume that each player knows each other's knowledge-type mapping in relation to a universal knowledge space.

Third, we can also characterize Positive Introspection and Negative Introspection in terms of player i's possibility correspondence $b_i : (\Omega, \mathcal{D}) \to (\mathcal{P}(\Omega), \{\{F \in \mathcal{D} \mid F \subseteq E\} \mid E \in \mathcal{D}\})$ (where b_i is induced from either t_i or K_i) if i's knowledge satisfies the Kripke property (thus, b_i satisfies the regularity condition that $b_i^{-1}(\{F \in \mathcal{D} \mid F \subseteq E\}) \in \mathcal{D}$ for all $E \in \mathcal{D}$). Then, i's knowledge satisfies Positive Introspection iff i knows her possibility correspondence b_i with respect to $\{\{F \in \mathcal{D} \mid F \subseteq E\} \mid E \in \mathcal{D}\}$. Likewise, i's knowledge satisfies Negative Introspection iff i knows b_i with respect to $\{\{F \in \mathcal{D} \mid E^c \cap F \neq \emptyset\} \mid E \in \mathcal{D}\}$.

We go on to the second aim. Namely, we apply the notion of informativeness to i's knowledge-type mapping $t_i : (\Omega, \mathcal{D}) \to (M(\Omega), \{\mathcal{M}(E) \mid E \in \mathcal{D}\})$. Observe the following. For states ω and ω' in Ω, ω is at least as informative as ω' to i (precisely, according to $t_i : (\Omega, \mathcal{D}) \to (M(\Omega), \{\mathcal{M}(E) \mid E \in \mathcal{D}\})$) iff $t_i(\omega')(E) \leq t_i(\omega)(E)$ for all $E \in \mathcal{D}$. Likewise, states ω and ω' are equally informative according to i iff $t_i(\omega) = t_i(\omega')$.

We define the set of states that are at least as informative to i as ω:

$$(\uparrow t_i(\omega)) = \{\omega' \in \Omega \mid t_i(\omega)(E) \leq t_i(\omega')(E) \text{ for all } E \in \mathcal{D}\}.$$

We also define, for each $\omega \in \Omega$,

$$\begin{aligned}(\downarrow t_i(\omega)) &= \{\omega' \in \Omega \mid t_i(\omega')(E) \leq t_i(\omega)(E) \text{ for all } E \in \mathcal{D}\} \text{ and} \\ [t_i(\omega)] &= \{\omega' \in \Omega \mid t_i(\omega) = t_i(\omega')\} (= (\uparrow t_i(\omega)) \cap (\downarrow t_i(\omega))).\end{aligned}$$

If $\omega' \in [t_i(\omega)]$ then ω and ω' are indistinguishable to player i in the sense that her knowledge-types (and thus the collections of events that she knows) are exactly the same at these states. Put differently, the equal informativeness is translated into the indistinguishability. Thus, the collection $\{[t_i(\omega)] \mid \omega \in \Omega\}$ forms a partition of Ω generated by the knowledge-type mapping t_i. Note that $(\uparrow t_i(\omega))$, $(\downarrow t_i(\omega))$, and $[t_i(\omega)]$ may not necessarily be events.[66]

As a remark, we mention that if ω' is at least as informative to i as ω (i.e., $\omega' \in (\uparrow t_i(\omega))$), then we have

$$b_{t_i}(\omega') = \bigcap \{E \in \mathcal{D} \mid t_i(\omega')(E) = 1\} \subseteq \bigcap \{E \in \mathcal{D} \mid t_i(\omega)(E) = 1\} = b_{t_i}(\omega).$$

[66] Suppose that $\mathcal{D}$ is a complete algebra. Since $\{\mu\} = \bigcap \{\mathcal{M}_E \in \mathcal{M}(\mathcal{D}) \mid \mu \in \mathcal{M}_E \text{ and } E \in \mathcal{D}\} \in \mathcal{M}(\mathcal{D})$, it follows that $\mathcal{M}(\mathcal{D}) = \mathcal{P}(M(\Omega, \mathcal{D}))$. Now, it is always the case that $(\uparrow t_i(\omega)) = t_i^{-1}(\{\mu \in M(\Omega) \mid t_i(\omega)(E) \leq \mu(E) \text{ for all } E \in \mathcal{D}\}) \in \mathcal{D}$, $(\downarrow t_i(\omega)) = t_i^{-1}(\{\mu \in M(\Omega) \mid \mu(E) \leq t_i(\omega)(E) \text{ for all } E \in \mathcal{D}\}) \in \mathcal{D}$, and $[t_i(\omega)] = t_i^{-1}(\{t_i(\omega)\}) \in \mathcal{D}$ for all $\omega \in \Omega$.

That is, the informativeness relation implies the set-inclusion between possibility sets. Note that if t_i satisfies the Kripke property, then the converse also holds: $b_{t_i}(\omega') \subseteq b_{t_i}(\omega)$ implies $\omega' \in (\uparrow t_i(\omega))$. This is simply because, if $t_i(\omega)(E) = 1$ then $b_{t_i}(\omega') \subseteq b_{t_i}(\omega) \subseteq E$ and thus $t_i(\omega')(E) = 1$. In other words, if the Kripke property is met, then the possibility set can be alternatively used to define the notion of informativeness.

Now, we examine the sense in which a player knows her knowledge-type mapping by studying how introspective properties imply the relations between informativeness and possibility.

Proposition 12 (Relation between Possibility and Informativeness)**.** *Let $\overrightarrow{\Omega}$ be a knowledge space.*

1. *t_i satisfies Truth Axiom iff $(\omega \in [t_i(\omega)] \subseteq)(\uparrow t_i(\omega)) \subseteq b_{t_i}(\omega)$ (for all $\omega \in \Omega$).*
2. *If t_i satisfies Positive Introspection, then $b_{t_i}(\omega) \subseteq (\uparrow t_i(\omega))$. If t_i satisfies the Kripke property, the converse is also true.*
3. *If t_i satisfies Negative Introspection, then $b_{t_i}(\omega) \subseteq (\downarrow t_i(\omega))$. If t_i satisfies the Kripke property, the converse is also true.*
4. *If t_i satisfies Truth Axiom, (Positive Introspection), and Negative Introspection, then $(\uparrow t_i(\omega)) = (\downarrow t_i(\omega)) = [t_i(\omega)] = b_{t_i}(\omega)$. If t_i satisfies the Kripke property, the converse is also true.*

First, when player i's knowledge-type mapping satisfies Truth Axiom, informativeness implies possibility. Conversely, when t_i satisfies Positive Introspection, possibility implies informativeness. Hence, when player i's knowledge-type mapping satisfies Truth Axiom and Positive Introspection, the notions of informativeness and possibility coincide. A simple corollary of this argument is that, as with standard possibility correspondence models, if t_i satisfies Truth Axiom and Positive Introspection as well as the Kripke property, then b_{t_i} is reflexive and transitive.[67]

Second, when player i's knowledge-type mapping satisfies Truth Axiom, Positive Introspection, and Negative Introspection, then either notion of informativeness or possibility induces the same informational partition $\{b_{t_i}(\omega) \mid \omega \in \Omega\} = \{[t_i(\omega)] \mid \omega \in \Omega\}$ of Ω with the following property: what she knows at ω coincides with that at ω' for any $\omega' \in [t_i(\omega)] = b_{t_i}(\omega)$.

Finally, we mention that, under Truth Axiom and Positive Introspection, we can define the notion of possibility between states in terms of a self-evident collection as $b_{t_i}(\omega) = \bigcap\{E \in \mathcal{J}_{t_i} \mid \omega \in E\}$, where $\mathcal{J}_{t_i} := \{E \in \mathcal{D} \mid t_i(\omega)(E) = 1 \text{ for any } \omega \in E\}$. This follows because $\mathcal{J}_{t_i} = \{K_{t_i}(E) \mid E \in \mathcal{D}\}$ under Truth Axiom and Positive Introspection. If $\kappa = \infty$ and if player i's knowledge satisfies Arbitrary Conjunction and Monotonicity, then

[67]See the references cited in Footnote 3 for such reflexive and transitive (non-partitional) possibility correspondence models.

$b_{t_i}(\omega)$ is indeed the minimal self-evident event containing ω. Furthermore, the notion of common knowledge is characterized by $\bigcap_{i\in I}\mathcal{J}_{t_i}$ in this special case. Thus the possibility correspondence associated with common knowledge is simply given by $\bigcap\{E \in \bigcap_{i\in I}\mathcal{J}_{t_i} \mid \omega \in E\}$ in this special case.

5.3 How Do the Players Know the Structure of a Universal Knowledge Space?

As we have discussed at the end of Section 3.1, we ask the sense in which the players know (or commonly know) the structure of the model itself.[68] Consider the universal knowledge space $\overrightarrow{\Omega}^*$ established in Section 3.1. We apply the notion of the knowledge of signal mappings to the knowledge-type mappings associated with each player's knowledge operator K_i^* and the common knowledge operator C^*.[69]

Proposition 13 (Do Players Know the Model of Knowledge?)**.** *Suppose that, in the given category of knowledge spaces, assumptions on players' knowledge include Monotonicity and Positive Introspection and that they are homogeneously given across players. Consider the knowledge-type mappings* $t_{K_i^*} : (\Omega, \mathcal{D}) \to (M(\Omega^*, \mathcal{D}^*), \{\mathcal{M}([e]) \mid [e] \in \mathcal{D}^*\})$ *and* $t_{C^*} : (\Omega, \mathcal{D}) \to (M(\Omega^*, \mathcal{D}^*), \{\mathcal{M}([e]) \mid [e] \in \mathcal{D}^*\})$*. Each mapping is commonly known among* I*.*

We establish Proposition 13 in the following way. First, if assumptions on players' knowledge are homogeneous, then every player's knowledge-type mapping is identical in the universal knowledge space $\overrightarrow{\Omega}^*$. Under Positive Introspection and Monotonicity, every player's knowledge-type mapping is also identical with the type mapping associated with the common knowledge operator. Second, Positive Introspection enables each player to know their own type mappings (see Proposition 11). But since their type mappings are homogeneous, it follows that every player know every player's type mapping. Indeed, players' knowledge-type mappings (and the knowledge-type mapping associated with their common knowledge) are commonly known.

The last step in the above reasoning, however, is worth scrutinizing. First, in concluding that the knowledge-type mappings are commonly known, we (i.e., the outside analysts) use the exogenous assumption that each player's knowledge is homogeneously given. Thus, the above proposition renders the sense in which we (i.e., the analysts) can conclude that the type-mappings are commonly known among the players within the model. Second, in order for us (i.e., the outside analysts) to ask i's knowledge about j's knowledge-type mapping, it is an important underlying assumption that assumptions on players' knowledge are homogeneously given. This is because j's knowledge-type mapping

[68]For the literature, see Footnote 35.

[69]In Section 6, we establish a universal knowledge space in terms of hierarchies of knowledge-types. There, the knowledge-type mappings are a primitive of the knowledge space, and thus we can directly apply our discussion (instead of defining the knowledge-type mappings $t_{K_i^*}$ from K_i^*).

$t_{K_j^*} : (\Omega, \mathcal{D}) \to (M(\Omega^*, \mathcal{D}^*), \{\mathcal{M}([e]) \mid [e] \in \mathcal{D}^*\})$ would be unobservable to player i without such an exogenous assumption.

Also, note that Proposition 13 does not necessarily claim that players' knowledge-type mappings are commonly known with respect to $\{\mathcal{M}([e]) \mid [e] \in \mathcal{D}^*\} \cup \{\neg\mathcal{M}([e]) \mid [e] \in \mathcal{D}^*\}$ unless otherwise each player's knowledge satisfies Negative Introspection. Indeed, if some player fails to satisfy Negative Introspection, she does not know her knowledge-type mapping with respect to $\{\mathcal{M}([e]) \mid [e] \in \mathcal{D}^*\} \cup \{\neg\mathcal{M}([e]) \mid [e] \in \mathcal{D}^*\}$.

Finally, we make the following two remarks. First, it is clear (from Lemma 3 and Proposition 7) that if every player's knowledge is assumed to satisfy Necessitation then the common knowledge operator satisfies Necessitation (i.e., $\Omega^* = C(\Omega^*)$). Thus, any valid expression is commonly known among the players.

Second, consider a category of knowledge spaces which satisfies the assumptions imposed in Proposition 13. Take any knowledge space $\overrightarrow{\Omega}$ in the given category. Then, as we have seen in Remark 2, $\overrightarrow{D(\Omega)}$ is a knowledge subspace of $\overrightarrow{\Omega^*}$. The statements of Proposition 13 hold with respect to the knowledge subspace $\overrightarrow{D(\Omega)}$. That is, for any knowledge space $\overrightarrow{\Omega}$, players' type mappings are commonly known in $\overrightarrow{D(\Omega)}$. In contrast, players' type mappings may not be necessarily commonly known in $\overrightarrow{\Omega}$. This is because players' knowledge may be different across players in the given space $\overrightarrow{\Omega}$. See Proposition A.5 of the Appendix.

6 A Hierarchical Approach to a Universal Knowledge Space

Here, we provide an alternative construction of a universal knowledge space by representing players' interactive knowledge as hierarchies of knowledge-types instead of infinitary languages. In this section, each player's knowledge in a given knowledge space is given by her knowledge-type mapping.

Specifically, we follow HS's [39] hierarchical approach to establishing a universal type space.[70] Before we go on to the construction, we make the following two preliminary remarks. First, we define a product κ-complete algebra as follows.

Remark 3 (Product κ-Complete Algebra)**.** Let $(\Omega_j, \mathcal{D}_j)$ be a κ-complete algebra for each $j \in J$, where J is a non-empty index set. We define the product κ-complete algebra $\prod_{j\in J} \mathcal{D}_j$ on the product space $\prod_{j \in J} \Omega_j$ as

$$\prod_{j\in J} \mathcal{D}_j := \mathcal{A}_\kappa \left(\bigcup_{j \in J} \{\pi_j^{-1}(E) \mid E \in \mathcal{D}_j\} \right),$$

[70]First, other hierarchical representations of knowledge include Fagin [25], FGHV [26], FHV [28], and HS [40, 41] in related contexts. Second, Viglizzo [83] studies HS's [39] hierarchical approach as an alternative construction of Moss and Viglizzo's [65, 66] terminal (final) coalgebra for a measure polynomial functor.

where $\pi_j : \prod_{j\in J} \Omega_j \to \Omega_j$ is the projection for each $j \in J$. As usual, if $\mathcal{D} = \mathcal{D}_j$ for all $j \in J$, then we denote $\prod_{j\in J} \mathcal{D}_j$ by $\mathcal{D}^J$. We also define finite products such as $\mathcal{D}_1 \times \mathcal{D}_2 = \prod_{j\in\{1,2\}} \mathcal{D}_j$ in the usual manner.

Second, we express knowledge morphisms in terms of knowledge-type mappings, which is analogous to the one in the literature on the (belief-)type spaces.

Remark 4 (Knowledge Morphism in terms of Knowledge-Tye Mapping)**.** Let $\overrightarrow{\Omega}$ and $\overrightarrow{\Omega'}$ be knowledge spaces. Let $\varphi : \Omega \to \Omega'$ be a mapping. Condition (3) in Definition 3 can be written as follows.

$$t'_i(\varphi(\omega))(E') = t_i(\omega)(\varphi^{-1}(E')) \text{ for all } \omega \in \Omega \text{ and } E' \in \mathcal{D}'.$$

Throughout the section, fix an infinite regular cardinal κ, a κ-complete algebra of states of nature $(S, \mathcal{A}_S)$, and a non-empty set of players I. We also fix assumptions on players' knowledge.

6.1 A Hierarchical Construction of a Universal Knowledge Space

We proceed with constructing a universal knowledge space in terms of hierarchies in four steps. The first step is to define the hierarchies space, which encompasses all the hierarchies of players' interactive knowledge regarding $(S, \mathcal{A}_S)$ up the ordinality of κ.

Definition 15 (Hierarchies Space)**.** *We define the sequence of κ-complete algebras $(H^\alpha, \mathcal{H}^\alpha)$ for ordinals $\alpha \leq \kappa$ as follows.*

1. *For $\alpha = 0$, we let $(H^0, \mathcal{H}^0) := (S, \mathcal{A}_S)$.*
2. *For any ordinal α with $0 < \alpha < \kappa$, we let*
$$(H^\alpha, \mathcal{H}^\alpha) := \left(S \times \prod_{\beta<\alpha} M(H^\beta, \mathcal{H}^\beta)^I, \mathcal{A}_S \times \prod_{\beta<\alpha} \mathcal{M}(H^\beta, \mathcal{H}^\beta)^I \right).$$
3. *For $\alpha = \kappa$, we define $(H^\kappa, \mathcal{H}^\kappa)$ by $H^\kappa := S \times \prod_{\alpha<\kappa} M(H^\alpha, \mathcal{H}^\alpha)^I$ and*
$$\mathcal{H}^\kappa := \{(\pi^{\kappa,\alpha})^{-1}(E^\alpha) \mid E^\alpha \in \mathcal{H}^\alpha \text{ for some } \alpha < \kappa\},$$
where $\pi^{\alpha,\beta} : H^\alpha \to H^\beta$ is the projection for all ordinals (α, β) with $0 \leq \beta \leq \alpha \leq \kappa$.

When $\alpha = \gamma$, we omit the superscript κ. Thus, we denote $(H, \mathcal{H}) := (H^\kappa, \mathcal{H}^\kappa)$. We remark that $(H, \mathcal{H})$ is indeed a κ-complete algebra (see Remark A.9 in the Appendix).

We often identify the hierarchies space H with the product of the set of states of nature S and each player's hierarchies space. To that end, for each $i \in I$ and α with $1 \leq \alpha \leq \kappa$, we let $H_i^\alpha := \prod_{\beta<\alpha} M(H^\beta, \mathcal{H}^\beta)$. Then, for any ordinal $\alpha \geq 1$, we can identify

$$H^\alpha = S \times \prod_{i\in I} H_i^\alpha. \tag{8}$$

Again, we omit the superscript κ when $\alpha = \kappa$. For all ordinals (α, β) with $1 \leq \beta \leq \alpha \leq \kappa$, we denote by $\pi_i^{\alpha,\beta} : H_i^\alpha \to H_i^\beta$ the projection.

In the second step, we define descriptions in terms of hierarchies. For any given knowledge space $\overrightarrow{\Omega}$, we define the (hierarchical) description map $h : \Omega \to H$.

Definition 16 (Hierarchical Description)**.** *Given a knowledge space $\overrightarrow{\Omega} = \langle(\Omega, \mathcal{D}), (t_i)_{i \in I}, \Theta\rangle$, we define the (hierarchical) description map $h : \Omega \to H$ as follows.*

1. *For $\alpha = 0$, we define $h^0 : \Omega \to H^0$ by $h^0 := \Theta$.*
2. *For a successor ordinal $\alpha = \beta + 1$, we define $h^\alpha : \Omega \to H^\alpha$ by*
$$h^\alpha(\omega) := (h^\beta(\omega), t(\omega) \circ (h^\beta)^{-1}) := (h^\beta(\omega), (t_i(\omega) \circ (h^\beta)^{-1})_{i \in I}) \text{ for all } \omega \in \Omega.$$
3. *For any limit ordinal α with $(0 <)\alpha \leq \kappa$, we let $h^\alpha : \Omega \to H^\alpha$ be the unique mapping which satisfies $h^\beta = \pi^{\alpha,\beta} \circ h^\alpha$ for all $\beta < \alpha$.*

We omit the superscript κ when $\alpha = \kappa$, i.e., we let $h = h^\kappa$. We call each $h(\omega)$ to be the (hierarchical) description of ω.

In light of Equation (8), for each $\alpha \geq 1$, we can identify each $h^\alpha \in H^\alpha$ in Definition 16 with $h^\alpha = (h^0, (h_i^\alpha)_{i \in I})$ as follows. Let $h_i^1(\omega) := t_i(\omega) \circ (h^0)^{-1}$ for all $\omega \in \Omega$. For a successor ordinal $\alpha = \beta + 1$ with $\beta \geq 1$, we define $h_i^\alpha : \Omega \to H_i^\alpha$ by $h_i^\alpha(\omega) := (h_i^\beta(\omega), t_i(\omega) \circ (h^\beta)^{-1})$ for all $\omega \in \Omega$. For a limit ordinal α, we define $h_i^\alpha : \Omega \to H_i^\alpha$ by the unique mapping which satisfies $h_i^\beta = \pi_i^{\alpha,\beta} \circ h_i^\alpha$ for all $\beta < \alpha$.

Now, we define an underlying set Ω^* of a candidate universal knowledge space as the set of all descriptions of states of the world ranged over all knowledge spaces of I on $(S, \mathcal{A}_S)$. That is,
$$\Omega^* := \{\omega^* \in H \mid \omega^* = h(\omega) \text{ for some } \overrightarrow{\Omega} \text{ and } \omega \in \Omega\}. \tag{9}$$
It is clear that Ω^* is not empty as long as there is a knowledge space $\overrightarrow{\Omega}$ with $\Omega \neq \emptyset$ in the given category of knowledge spaces. As we have already seen in Section 3.1, we indeed have $\Omega^* \neq \emptyset$.

Once Ω^* is defined as a subset of H, we can induce a κ-complete algebra $\mathcal{D}^*$ on Ω^* from $\mathcal{H}$ by
$$\mathcal{D}^* = \{(\pi^\alpha|_{\Omega^*})^{-1}(E^\alpha) \in \mathcal{P}(\Omega^*) \mid E^\alpha \in \mathcal{H}^\alpha \text{ for some } \alpha < \kappa\}, \tag{10}$$
where $\pi^\alpha|_{\Omega^*} : \Omega^* \to H^\alpha$ is the restriction of $\pi^\alpha : H \to H^\alpha$ on Ω^* for each $\alpha < \kappa$. Note that $\mathcal{D}^* = \{E \cap \Omega^* \in \mathcal{P}(\Omega^*) \mid E \in \mathcal{H}\}$, since $(\pi^\alpha|_{\Omega^*})^{-1}(\cdot) = (\pi^\alpha)^{-1}(\cdot) \cap \Omega^*$.

From now on, for any knowledge space $\overrightarrow{\Omega}$, we identify the description map h as $h : \Omega \to \Omega^*$. We denote the description map by $h_{\overrightarrow{\Omega}} : \Omega \to \Omega^*$ when we stress its domain. We establish that the description map $h : (\Omega, \mathcal{D}) \to (\Omega^*, \mathcal{D}^*)$ is κ-measurable and that a knowledge morphism preserves hierarchical descriptions as in Corollary 1. Thus, for any two states $(\omega, \varphi(\omega)) \in \Omega \times \Omega'$ where $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega'}$ is a knowledge morphism, they induce the same hierarchy of knowledge.

Lemma 6 (Preservation of Descriptions through Knowledge Morphism)**.** *Let $\overrightarrow{\Omega}$ and $\overrightarrow{\Omega'}$ be knowledge spaces.*

1. *The description map $h_{\overrightarrow{\Omega}} : (\Omega, \mathcal{D}) \to (\Omega^*, \mathcal{D}^*)$ is κ-measurable.*

2. *If $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega'}$ is a knowledge morphism, then $h_{\overrightarrow{\Omega}} = h_{\overrightarrow{\Omega'}} \circ \varphi$.*

We make the following remark. As in Section 3.1, we say that a knowledge space $\overrightarrow{\Omega}$ is *non-redundant* if its description map h is injective. If a knowledge space $\overrightarrow{\Omega}$ is non-redundant, then $(\Theta, (t_i)_{i\in I}) : \Omega \to S \times M(\Omega)^I$ is injective. This follows because if $(\Theta, (t_i)_{i\in I})(\omega) = (\Theta, (t_i)_{i\in I})(\omega')$ for some $\omega, \omega' \in \Omega$ then $h(\omega) = h(\omega')$ and thus $\omega = \omega'$.

In the third step, we define the mapping $\Theta^* : \Omega^* \to S$ and players' knowledge-type mappings. We define $\Theta^* : \Omega^* \to S$ by the projection $\Theta^* = \pi^0|_{\Omega^*}$. By construction, $\Theta^* : (\Omega^*, \mathcal{D}^*) \to (S, \mathcal{A}_S)$ is κ-measurable. Moreover, for any knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$, we have

$$\Theta(\omega) = \pi^0(h(\omega)) = \Theta^*(h(\omega)).$$

Hence, Θ^* preserves states of nature. Next, we define players' type mappings as follows.

Lemma 7 (Players' Knowledge in Candidate Universal Space)**.** *For each $i \in I$ and $\omega^* \in \Omega^*$, we define*

$$t_i^*(\omega^*)((\pi^\alpha|_{\Omega^*})^{-1}(E^\alpha)) := (\omega^*)_i^{\alpha+1}(E^\alpha) \text{ for each } \alpha < \kappa \text{ and } E^\alpha \in \mathcal{H}^\alpha, \tag{11}$$

where note that $(\omega^*)_i^{\alpha+1} \in M(H^\alpha, \mathcal{H}^\alpha)$.[71] *Then, we have the following.*

1. *$t_i^* : (\Omega^*, \mathcal{D}^*) \to (M(\Omega^*, \mathcal{D}^*), \mathcal{M}(\Omega^*, \mathcal{D}^*))$ is a well-defined κ-measurable mapping.*

2. *t_i^* inherits all the properties of knowledge imposed in the given category of knowledge spaces.*

3. *Fix a knowledge space $\overrightarrow{\Omega}$. Then, for any $\omega \in \Omega$ and $E^* \in \mathcal{D}^*$,*

$$t_i^*(h(\omega))(E^*) = t_i(\omega)(h^{-1}(E^*)). \tag{12}$$

So far, we have established the following: (i) the space $\overrightarrow{\Omega^*} = \langle(\Omega^*, \mathcal{D}^*), (t_i^*)_{i\in I}, \Theta^*\rangle$ is a legitimate knowledge space with $\Omega^* \neq \emptyset$; and (ii) for any given knowledge space $\overrightarrow{\Omega} = \langle(\Omega, \mathcal{D}), (t_i)_{i\in I}, \Theta\rangle$, the description map $h : \Omega \to \Omega^*$ is a knowledge morphism. In the knowledge space $\overrightarrow{\Omega^*}$, players' knowledge at state ω^* is encoded within the state ω^* through Equation (11).

In the fourth and final step, we show that the description map is a unique knowledge morphism from a given knowledge space into the knowledge space $\overrightarrow{\Omega^*}$. As in Lemma 5 in Section 3.1, we show that the description map from $\overrightarrow{\Omega^*}$ into itself is the identity map.

[71] If we denote $\omega^* = (s, (\mu_i)_{i\in I}) \in \Omega^*$ and $\mu_i = (\mu_i^\alpha)_{0\le\alpha<\kappa}$ with $\mu_i^\alpha \in M(H^\alpha, \mathcal{H}^\alpha)$, then we have $(\omega^*)_i^{\alpha+1} = \mu_i^\alpha$.

Lemma 8 (Description Map of Candidate Universal Space)**.** *The description map* $h^* : \overrightarrow{\Omega^*} \to \Omega^*$ *is the identity map on* Ω^**.*

Thus, we establish that the knowledge space $\overrightarrow{\Omega^*} = \langle(\Omega^*, \mathcal{D}^*), (t_i^*)_{i\in I}, \Theta^*\rangle$ is universal. Also, the mapping $(\Theta^*, (t_i^*)_{i\in I}) : \Omega^* \to S \times M(\Omega^*, \mathcal{D}^*)^I$ is injective. Thus, the universal knowledge space $\overrightarrow{\Omega^*}$ is in a bijective relation to a subset of $S \times M(\Omega^*, \mathcal{D}^*)^I$ that respects given introspective (as well as Kripke) properties in the following sense.

Theorem 3 (Existence of a Universal Knowledge Space)**.** *The knowledge space* $\overrightarrow{\Omega^*} = \langle(\Omega^*, \mathcal{D}^*), (t_i^*)_{i\in I}, \Theta^*\rangle$ *is universal in each given category of knowledge spaces of* I *on* $(S, \mathcal{A}_S)$*. The mapping* $(\Theta^*, (t_i^*)_{i\in I}) : \Omega^* \to \Omega^{**}$ *is bijective, where*

$$\Omega^{**} := \{(s, (\mu_i)_{i\in I}) \in S \times M(\Omega^*, \mathcal{D}^*)^I \mid \textit{there are a knowledge space } \overrightarrow{\Omega} \textit{ and } \omega \in \Omega \\ \textit{such that } (s, (\mu_i)_{i\in I}) = (\Theta(\omega), (t_i(\omega) \circ h^{-1})_{i\in I})\}.$$

We note that the discussions in Remark 2 in Section 3.1 apply by replacing the appropriate notations. In the Appendix, we establish the following fact by using category theory (specifically, the theory of coalgebra).

Remark 5 (Property of Universal Space with No Introspection)**.** Suppose that no introspective property nor the Kripke property is imposed in the given class of knowledge spaces. Then, $(\Theta^*, (t_i^*)_{i\in I}) : \Omega^* \to S \times M(\Omega^*, \mathcal{D}^*)^I$ is bijective (indeed, a knowledge isomorphism).

We make the following two remarks. First, a mathematical intuition behind Remark 5 is that a knowledge space $\overrightarrow{\Omega} = \langle(\Omega, \mathcal{D}), (t_i)_{i\in I}, \Theta\rangle$ can be identified as a pair of the κ-complete algebra $(\Omega, \mathcal{D})$ and the mapping $(\Theta, (t_i)_{i\in I}) : (\Omega, \mathcal{D}) \to (F(\Omega), \mathcal{F}(\mathcal{D}))$, where $(F(\Omega), \mathcal{F}(\mathcal{D}))$ is a κ-complete algebra satisfying $F(\Omega) = S \times M(\Omega, \mathcal{D})^I$. In the language of category theory, the pair $\langle(\Omega, \mathcal{D}), (\Theta, (t_i)_{i\in I})\rangle$ is called a coalgebra over the endofunctor F. Thus, the category of knowledge(-type) spaces is seen as the full subcategory of F-coalgebras such that each t_i satisfies the given introspective and/or Kripke properties, if F is taken to be $S \times M(\Omega, \mathcal{D})^I$.[72] If we do not assume any introspective or the Kripke property, then a knowledge space can be seen as a coalgebra. Now, as is well known in the theory of coalgebras (namely, the Lambek lemma [47]), a terminal (final) coalgebra Ω^* is isomorphic to $F(\Omega^*) = S \times M(\Omega^*)^I$.

Second, we briefly discuss the role of the domain specification. To make the exposition simplest, we impose no assumption on players' knowledge. Suppose that the domain of a knowledge space is always the power set. If such a class of knowledge spaces has a universal knowledge space, then we would have to have a bijection between Ω^* and $S \times \mathcal{P}(\mathcal{P}(\Omega^*))^I$, where note that $M(\Omega^*)$ is in a bijective relation to $\mathcal{P}(\mathcal{P}(\Omega^*))$. This is clearly impossible. This argument is closely related to Brandenburger's result on the non-existenece of a (belief-)complete possibility structure (Brandenburger [14]).

[72] First, see, for example, Kurz [46], Jacobs and Rutten [44], and Rutten [73] for a general treatment of coalgebras. Second, Moss and Viglizzo [65, 66] take a coalgebraic approach to a universal type space. In the type space approach, introspective properties can be encoded by the requirement that each player's type distribution be defined on the product of states of nature and the other players' types.

6.2 Coherent Hierarchies of Knowledge

In the previous subsection, we have established the existence of a universal knowledge space in terms of hierarchies of knowledge-types. The natural question, then, is to examine how our construction relates to the previous literature on universal type spaces, in which the notion of coherent hierarchies plays an important role.

Below, we define a coherent subset of the hierarchies space H and show that the universal knowledge space $\overrightarrow{\Omega^*}$ (established in Theorem 3) is the largest coherent subset of H. Note that the notion of coherency is defined in the context of the knowledge hierarchies H. We start with the definition of a coherent subset of H.

Definition 17 (Coherent Hierarchies of Knowledge)**.** *1. A subset Ω of H is said to be coherent if $(\Omega, \mathcal{D})$ satisfies the following conditions, where $\mathcal{D}$ is a κ-complete algebra on Ω defined by*

$$\mathcal{D} := \{(\pi^\alpha|_\Omega)^{-1}(E^\alpha) \in \mathcal{P}(\Omega) \mid E^\alpha \in \mathcal{H}^\alpha \text{ for some } \alpha < \kappa\}.$$

Take any $(s, (\mu_i)_{i \in I}) \in \Omega$, where $\mu_i = (\mu_i^\alpha)_{0 \leq \alpha < \kappa}$ with $\mu_i^\alpha \in M(H^\alpha, \mathcal{H}^\alpha)$.

(a) For any ordinals (α, β) with $0 \leq \beta \leq \alpha < \kappa$, if $(\pi^\alpha|_\Omega)^{-1}(E^\alpha) = (\pi^\beta|_\Omega)^{-1}(F^\beta)$ for some $E^\alpha \in \mathcal{H}^\alpha$ and $F^\beta \in \mathcal{H}^\beta$, then $\mu_i^\alpha(E^\alpha) = \mu_i^\beta(F^\beta)$ for all $i \in I$.

(b) Suppose that Truth Axiom is imposed on player i. If $\mu_i^\alpha(E^\alpha) = 1$ for some $E^\alpha \in \mathcal{H}^\alpha$, then $(s, (\mu_i)_{i \in I}) \in (\pi^\alpha|_\Omega)^{-1}(E^\alpha)$.

(c) Suppose that Positive Introspection is imposed on player i. If $\mu_i^\alpha(E^\alpha) = 1$ for some $E^\alpha \in \mathcal{H}^\alpha$, then $\mu_i^\alpha(\{(s', (\mu'_j)_{j \in I}) \in H \mid (\mu'_i)^\alpha \in \mathcal{M}(E^\alpha)\}) = 1$.

(d) Suppose that Negative Introspection is imposed on player i. If $\mu_i^\alpha(E^\alpha) = 0$ for some $E^\alpha \in \mathcal{H}^\alpha$, then $\mu_i^\alpha(\{(s', (\mu'_j)_{j \in I}) \in H \mid (\mu'_i)^\alpha \in \neg\mathcal{M}(E^\alpha)\}) = 1$.

(e) Suppose that the Kripke property is imposed on player i. If $\bigcap\{E \in \mathcal{D} \mid E = (\pi^\alpha|_\Omega)^{-1}(E^\alpha)$ and $\mu_i^\alpha(E^\alpha) = 1$ for some $\alpha < \kappa$ and $E^\alpha \in \mathcal{H}^\alpha\} \subseteq (\pi^\beta|_\Omega)^{-1}(F^\beta)$ for some $\beta < \kappa$ and $F^\beta \in \mathcal{H}^\beta$, then $\mu_i^\beta(F^\beta) = 1$.

2. If Ω is a coherent subset of H, then we define the induced knowledge space $\overrightarrow{\Omega} = \langle(\Omega, \mathcal{D}), (t_i)_{i \in I}, \Theta\rangle$ as follows.

(a) We define $\Theta : (\Omega, \mathcal{D}) \to (S, \mathcal{A}_S)$ by the projection $\pi^0|_\Omega$.

(b) We define $t_i : (\Omega, \mathcal{D}) \to (M(\Omega, \mathcal{D}), \mathcal{M}(\Omega, \mathcal{D}))$ by

$$t_i(s, (\mu_j)_{j \in I})((\pi^\alpha|_\Omega)^{-1}(E^\alpha)) := \mu_i^\alpha(E^\alpha) \text{ for each } E^\alpha \in \mathcal{H}^\alpha.$$

For any subset Ω of H, we can induce a κ-complete algebra $\mathcal{D}$ from $\mathcal{H}$. Also, the mapping Θ is naturally defined by the projection through Condition (2a). Now, in order for Ω to be a coherent subset of H, Condition (1a) requires the set Ω to induce well-defined knowledge-type mappings within Ω through Condition (2b).[73] This condition requires different levels of players' knowledge not to contradict with each other. The other conditions stipulated in (1) ensures that the knowledge-type mappings respect the required properties of knowledge.

Now, we establish our main result of this subsection that $\overrightarrow{\Omega^*}$ is a largest coherent subset of H.

Theorem 4 (Universal Knowledge Space is Largest Coherent Space)**.** *The universal knowledge space Ω^* established in Section 6.1 is the largest coherent subset of H in the following sense: (i) Ω^* is coherent; and (ii) for any coherent subset Ω of H, the description map $h : \overrightarrow{\Omega} \to \overrightarrow{\Omega^*}$ is an inclusion map (so that $\Omega \subseteq \Omega^*$).*

7 Applications to Richer Settings

In this section, we discuss applicability of our framework. Specifically, since our framework can accommodate a case where players have multiple epistemic operators, we can consider the following extensions. First, we can add time dynamics so that each player's knowledge at each time is described by her knowledge operator at that time. Second, we can add other epistemic operators such as non-probabilistic belief and unawareness.[74]

In Section 7.1, we consider dynamic knowledge-belief spaces, where players have qualitative beliefs as well as knowledge in a dynamic setting. Next, in Section 7.2, we consider knowledge-unawareness spaces in the context of state space models. Henceforth in this section, fix a κ-complete algebra of states of nature $(S, \mathcal{A}_S)$ and a non-empty set of players I, where κ is an infinite regular cardinal.

7.1 Universal Dynamic Knowlwdge-(Non-probabilistic-)Belief Spaces

Epistemic analyses of dynamic games usually call for players' knowledge and belief. We demonstrate that there exists a universal dynamic knowledge-belief space, where players' beliefs are defined by their qualitative (i.e., non-probabilistic) belief operators. The class of dynamic knowledge-belief spaces that we study is based on Battigali and Bonanno [9].

We assume, for simplicity, that time runs through $\mathbb{N}$. Each player i's knowledge at time $t \in \mathbb{N}$ is represented by a knowledge operator $K_{i,t} : \mathcal{D} \to \mathcal{D}$, while player i's (non-probabilistic) belief at time $t \in \mathbb{N}$ is captured by a belief operator on the same domain: $B_{i,t} : \mathcal{D} \to \mathcal{D}$. Specifically, we define a dynamic knowledge-belief space of I on $(S, \mathcal{A}_S)$ as follows.

[73] The corresponding condition that induces players' knowledge in a well-defined manner in the syntactic approach is Condition (2a) in Theorem 2 of Section 3.2.

[74] Although we entirely omit it from our discussion, counterfactual/hypothetical reasoning would be another interesting example.

Definition 18 (Dynamic Knowledge-Belief Space)**.** *A dynamic (κ-)knowledge-belief space of I on $(S, \mathcal{A}_S)$ is a tuple $\overrightarrow{\Omega} = \langle(\Omega, \mathcal{D}), (K_{i,t})_{(i,t)\in I\times\mathbb{N}}, (B_{i,t})_{(i,t)\in I\times\mathbb{N}}, \Theta\rangle$ with the following properties:*

1. *$(\Omega, \mathcal{D})$ is a κ-complete algebra and the mapping $\Theta : (\Omega, \mathcal{D}) \to (S, \mathcal{A}_S)$ is κ-measurable.*
2. *Each knowledge operator $K_{i,t} : \mathcal{D} \to \mathcal{D}$ satisfies Truth Axiom, Positive Introspection, Monotonicity, κ-Conjunction, Necessitation, and Negative Introspection.*
3. *Each belief operator $B_{i,t} : \mathcal{D} \to \mathcal{D}$ satisfies Consistency, Positive Introspection, Monotonicity, κ-Conjunction, Necessitation, and Negative Introspection.*
4. *Knowledge and belief operators satisfy the following joint conditions.*
   (a) *$K_{i,t}(E) \subseteq B_{i,t}(E)$ for any $E \in \mathcal{D}$.*
   (b) *$B_{i,t}(E) \subseteq K_{i,t}B_{i,t}(E)$ for any $E \in \mathcal{D}$.*
   (c) *$B_{i,t}(E) = B_{i,t}B_{i,t+1}(E)$ for each $E \in \mathcal{D}$.*

Conditions (4) specify relations between knowledge and belief. The first condition (4a) means that knowledge implies belief at each time. The second (4b) states that each player knows her own belief at each time, where note that $(\neg B_{i,t})(\cdot) \subseteq K_{i,t}(\neg B_{i,t})(\cdot)$ is satisfied, given Truth Axiom and Negative Introspection of knowledge. The third condition (4c) captures the idea of belief persistence by Battigali and Bonanno [9]: player i believes E at time t iff she believes at t that she (will) believe E at $t+1$. We say that player i's knowledge satisfies perfect recall if $K_{i,t}(\cdot) \subseteq K_{i,t+1}(\cdot)$ for all $t \in \mathbb{N}$. A dynamic knowledge-belief space *with perfect recall* is a dynamic knowledge-belief space such that each player's knowledge satisfies perfect recall.

The point is that a dynamic knowledge-belief space is mathematically a knowledge space of $I \times \mathbb{N} \times \{0, 1\}$, where "player" $(i, t, 0)$'s knowledge operator is i's knowledge operator at time t while "player" $(i, t, 1)$'s knowledge operator is i's (non-probabilistic) belief operator at time t, with the specified conditions.

A knowledge-belief morphism can be regarded as a knowledge morphism of this extended knowledge space. That is, a (dynamic) knowledge-belief morphism from $\overrightarrow{\Omega}$ to $\overrightarrow{\Omega'}$ is a κ-measurable mapping $\varphi : (\Omega, \mathcal{D}) \to (\Omega', \mathcal{D}')$ with the following properties: (i) $\Theta' \circ \varphi = \Theta$; and (ii) for all $i \in I$ and $t \in \mathbb{N}$, $K_{i,t}(\varphi^{-1}(\cdot)) = \varphi^{-1}(K'_{i,t}(\cdot))$ and $B_{i,t}(\varphi^{-1}(\cdot)) = \varphi^{-1}(B'_{i,t}(\cdot))$.

A universal dynamic knowledge-belief space is a universal knowledge space in this category of extended knowledge spaces. That is, a dynamic knowledge-belief space $\overrightarrow{\Omega^*}$ of I on $(S, \mathcal{A}_S)$ is said to be *universal* if, for any dynamic knowledge-belief space $\overrightarrow{\Omega}$ of I on $(S, \mathcal{A}_S)$ there is a unique (dynamic) knowledge-belief morphism $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega^*}$.

Our previous arguments apply to the existence of a universal dynamic knowledge-belief space.

Theorem 5 (Existence of Universal Dynamic Knowledge-Belief Space)**.** *There exists a universal dynamic knowledge-belief space (with/without perfect recall) $\overrightarrow{\Omega^*}$ of I on $(S, \mathcal{A}_S)$.*

7.2 Universal Knowledge-Unawareness Spaces

In our framework, we can accommodate a wide variety of assumptions on players' knowledge. This raises questions as to how our framework can accommodate notions of unawareness. While studies of unawareness in the framework goes beyond the scope of our present paper, we illustrate the idea that our arguments can be applied to the existence of a universal space involving a notion of unawareness. We do so within the (simple) framework of standard state space models (see, for example, Chen, Ely, and Luo [19] and DLR [21]).[75]

Specifically, we represent each player i's knowledge and unawareness by her knowledge operator $K_i : \mathcal{D} \to \mathcal{D}$ and unawareness operator $U_i : \mathcal{D} \to \mathcal{D}$, respectively, where $\mathcal{D}$ is a κ-complete algebra on an underlying set Ω. Player i is *unaware* of an event $E \in \mathcal{D}$ at a state $\omega \in \Omega$ if $\omega \in U_i(E)$. On the other hand, player i is *aware* of an event $E \in \mathcal{D}$ at a state $\omega \in \Omega$ if $\omega \in A_i(E) := (\neg U_i)(E)$.

It is to be noted that any set $E \in \mathcal{P}(\Omega) \setminus \mathcal{D}$ is simply not an object of players' unawareness. Hence, the fact that a player is unaware of a certain event should be distinguished from the fact that some event is not an object of her unawareness.

While we can establish the existence of a universal knowledge-unawareness space where a notion of unawareness satisfies various other properties, here we restrict attention to the following three axioms on players' unawareness taken from DLR [21].[76]

The first axiom is Plausibility: $U_i(\cdot) \subseteq (\neg K_i)(\cdot) \cap (\neg K_i)^2(\cdot)$. It states that if a player is unaware of an event E then she does not know E and she does not know that she does not know E. The second is KU Introspection: $K_i U_i(\cdot) = \emptyset$. It requires that any player cannot know any event of which she is unaware. The third is AU Introspection: $U_i(\cdot) \subseteq U_i U_i(\cdot)$. It says that if a player is unaware of an event then she is unaware of being unaware of that event.

Definition 19 (Knowledge-Unawareness Space)**.** *A knowledge-unawareness space of I on $(S, \mathcal{A}_S)$ is a tuple $\overrightarrow{\Omega} := \langle (\Omega, \mathcal{D}), (K_i)_{i \in I}, (U_i)_{i \in I}, \Theta \rangle$ with the following properties.*

1. *$\langle (\Omega, \mathcal{D}), (K_i)_{i \in I}, \Theta \rangle$ is a κ-knowledge space of I on $(S, \mathcal{A}_S)$.*
2. *Each unawareness operator $U_i : \mathcal{D} \to \mathcal{D}$ satisfies (some of) Plausibility, KU Introspection, and AU Introspection.*

We say that a knowledge-unawareness space $\overrightarrow{\Omega}$ captures a non-trivial form of unawareness ($\overrightarrow{\Omega}$ is non-trivial, for short) if there is $(i, E) \in I \times \mathcal{D}$ such that $U_i(E) \neq \emptyset$.

[75] First, as Modica and Rustichini [62] and DLR [21] demonstrate, standard state space models have limitations in representing unawareness. Second, it would be an interesting question to ask whether or how our results in this paper apply to models of unawareness that have richer structures than standard state space models (see, Schipper [79] and the references therein for the richer structures). Third, the second chapter of this dissertation (Fukuda [30]) studies how a standard state space model can (and cannot) represent notions of unawareness within our present framework.

[76] For other properties of unawareness, see Schipper [79] and the references therein.

Note that we can regard a knowledge-unawareness space as a knowledge space by identifying unawareness operators with "knowledge" operators. With keeping this in mind, we define a knowledge-unawareness morphism and a universal knowledge-unawareness space. Also, note that a given knowledge space induces a knowledge-unawareness space, where an unawareness operator is, for example, defined by $U_i(\cdot) = (\neg K_i)(\cdot) \cap (\neg K_i)^2(\cdot)$.

A κ-measurable mapping $\varphi : (\Omega, \mathcal{D}) \to (\Omega', \mathcal{D}')$ between knowledge-unawareness spaces $\overrightarrow{\Omega}$ and $\overrightarrow{\Omega'}$ is called a knowledge-unawareness morphism if (i) $\Theta' \circ \varphi = \Theta$ and if (ii) $K_i(\varphi^{-1}(\cdot)) = \varphi^{-1}(K_i'(\cdot))$ and $U_i(\varphi^{-1}(\cdot)) = \varphi^{-1}(U_i'(\cdot))$ for all $i \in I$.

We say that a knowledge-unawareness space $\overrightarrow{\Omega^*}$ of I on $(S, \mathcal{A}_S)$ is *universal* if, for any knowledge-unawareness space Ω of I on $(S, \mathcal{A}_S)$ there is a unique knowledge-unawareness morphism $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega^*}$.

Now, we show that there exists a universal knowledge-unawareness space Ω^* of I on $(S, \mathcal{A}_S)$. We also present a mild sufficient condition for guaranteeing its non-triviality.[77]

Theorem 6 (Existence of Universal Knowledge-Unawareness Space)**.** *There exists a universal knowledge-unawareness space $\overrightarrow{\Omega^*}$ of I on $(S, \mathcal{A}_S)$. Moreover, suppose that there is a knowledge-unawareness space $\overrightarrow{\Omega}$ in a given category of knowledge-unawareness spaces such that $U_i([\![e]\!]_{\overrightarrow{\Omega}}) \neq \emptyset$ for some $i \in I$ and $e \in \mathcal{L}$. Then, $\overrightarrow{\Omega^*}$ is non-trivial.*

References

[1] Ahn, D. S. "Hierarchies of Ambiguous Beliefs," J. Econ. Theory 136 (2007), 286-301.

[2] Armbruster, W. and W. Böge. "Bayesian Game Theory." in O. Moeschlin and D. Pallaschke (Eds.), *Game Theory and Related Topics*, North-Holland, 1979, 17-28.

[3] Aumann, R. J. "Agreeing to Disagree." Ann. Statist. 4 (1976), 1236-1239.

[4] Aumann, R. J. "Correlated Equilibrium as an Expression of Bayesian Rationality." Econometrica 55 (1987), 1-18.

[5] Aumann, R. J. "Interactive Epistemology I, II." Int. J. Game Theory 28 (1999), 263-300, 301-314.

[6] Bacharach, M. "Some Extensions of a Claim of Aumann in an Axiomatic Model of Knowledge." J. Econ. Theory 37 (1985), 167-190.

[7] Bacharach, M. "When Do We Have Information Partitions?" in M. Bacharach and M. Dempster. (Eds.) *Mathematical Models in Economics.* Oxford University Press, 1990, 1-23.

[77] In light of this result, here we do not study conditions on the knowledge and unawareness operators which guarantee a non-trivial form of unawareness. The second chapter of this dissertation (Fukuda [30]) studies this question.

[8] Barwise, J. “On the Model Theory of Common Knowledge.” in J. Barwise. *The Situation in Logic*. CSLI, 1989, 201-220.

[9] Battigalli, P. and G. Bonanno. “The Logic of Belief Persistence.” Econ. Philos. 13 (1997), 39-59.

[10] Battigali, P. and M. Siniscalchi. “Hierarchies of Conditional Beliefs and Interactive Epistemology in Dynamic Games.” J. Econ. Theory 88 (1999), 188-230.

[11] Binmore, K. and A. Brandenburger. “Common Knowledge and Game Theory.” in K. Binmore *Essays on the Foundations of Game Theory*. Basil Blackwell, 1990, 105-150.

[12] Böge, W. and T. Eisele. “On Solutions of Bayesian Games.” Int. J. Game Theory 8 (1979), 193-215.

[13] Bonanno, G. “Information, Knowledge and Belief.” Bull. Econ. Res. 54 (2002), 47-67.

[14] Brandenburger, A. “On the Existence of a “Complete” Possibility Structure.” in Basili M., Dimitri N., Gilboa I. (Eds.) *Cognitive Processes and Economic Behavior*. Routledge, 2003, 30-34.

[15] Brandenburger, A. and E. Dekel. “Common Knowledge with Probability 1.” J. Math. Econ. 16 (1987), 237-245.

[16] Brandenburger, A. and E. Dekel. “Hierarchies of Beliefs and Common Knowledge.” J. Econ. Theory 59 (1993), 189-198.

[17] Brandenburger, A., E. Dekel, and J. Geanakoplos. “Correlated Equilibrium with Generalized Information Structures.” Games Econ. Behav. 4 (1992), 182-201.

[18] Brandenburger, A. and H. J. Keisler. “An Impossibility Theorem on Beliefs in Games.” Studia Logica. 84 (2006), 211-240.

[19] Chen, Y-C., J. C. Ely, and X. Luo. “Note on Unawareness: Negative Introspection versus AU Introspection (and KU Introspection).” Int. J. Game Theory 41 (2012), 325-329.

[20] Dekel, E., and F. Gul, “Rationality and Knowledge in Game Theory,” in Kreps, D. M., and K. F. Wallis (Eds.), *Advances in Economics and Econometrics: Theory and Applications*, volume 1, Seventh World Congress, Cambridge University Press, 1997.

[21] Dekel, E., B. L. Lipman, and A. Rustichini. “Standard State-Space Models Preclude Unawareness,” Econometrica 66 (1998), 159-173.

[22] Doignon, J.-P. and J.-C. Falmagne. “Spaces for the Assessment of Knowledge.” Int. J. Man Mach. Stud. 23 (1985), 175-196.

[23] Doignon, J.-P. and J.-C. Falmagne. Knowledge Spaces. Springer-Verlag, Berlin Heidelberg, 1999.

[24] Dubra, J. and F. Echenique. "Information is not about Measurability." Math. Soc. Sci. 47 (2004), 177-185.

[25] Fagin, R. "A Quantitative Analysis of Modal Logic." J. Symb. Logic 59 (1994), 209-252.

[26] Fagin, R., J. Geanakoplos, J. Y. Halpern, and M. Y. Vardi. "The Hierarchical Approach to Modeling Knowledge and Common Knowledge." Int. J. Game Theory 28 (1999), 331-365.

[27] Fagin, R. and J. Y. Halpern "Belief, Awareness, and Limited Reasoning." Art. Intell. 34 (1988), 39-76.

[28] Fagin, R., J. Y. Halpern, and M. Y. Vardi. "A Model-theoretic Analysis of Knowledge." J. ACM 38 (1991), 382-428.

[29] Friedell, M. F. "On the Structure of Shared Awareness." Behav. Sci. 14 (1969), 28-39.

[30] Fukuda, S. "Representing Unawareness on State Spaces." 2017.

[31] Geanakoplos, J. D. "Games Theory without Partitions, and Applications to Speculation and Consensus," Cowles Foundation Discussion Paper No. 914, Yale University, 1989.

[32] Gilboa, I. Information and Meta-Information. Proceedings of the 2nd Conference on Theoretical Aspects of Reasoning about Knowledge, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 1988, pp. 227-243.

[33] Halmos, P. R. "Algebraic Logic I. Monadic Boolean Algebras. Compositio Math. 12 (1955), 217-249. In P. R. Halmos. *Algebraic Logic*. Chelsea, 1962.

[34] Halpern, J. Y. and Y. Moses. "Knowledge and Common knowledge in a Distributed Environment." J. ACM 37 (1990), 549-587.

[35] Harsanyi, J. C. "Games with Incomplete Information Played by "Bayesian" Players, I-III." Manag. Sci., 14 (1967-1968), pp. 159-182, 320-334, 486-502.

[36] Heifetz, A. "The Bayesian Formulation of Incomplete Information—the Non-Compact Case." Int. J. Game Theory 21 (1993), 329-338.

[37] Heifetz, A. Infinitary $S5$-Epistemic Logic. Math. Log. Quart. 43 (1997) 333-342.

[38] Heifetz, A. and D. Samet. "Knowledge Spaces with Arbitrarily High Rank." Games Econ. Behav. 22 (1998), 260-273.

[39] Heifetz, A. and D. Samet. "Typology-Free Typology of Beliefs." J. Econ. Theory 82 (1998), 324-341.

[40] Heifetz, A. and D. Samet. "Coherent Beliefs are not always Types." J. Math. Econ. 32 (1999), 475-488.

[41] Heifetz, A. and D. Samet. "Hierarchies of Knowledge: An Unbounded Stairway." Math. Soc. Sci. 22 (1999), 157-170.

[42] Hérves-Beloso, C. and P. K. Monteiro. "Information and σ-algebras." Econ. Theory. 54 (2013), 405-418.

[43] Hintikka, J. *Knowledge and Belief: An Introduction to the Logic of the Two Notions.* Cornell University Press, 1962.

[44] Jacobs, B. and J. Rutten. "An Introduction to (co)algebra and (co)induction. in D. Sangiorgi and J. Rutten. (Eds.) *Advanced Topics in Bisimulation and Coinduction.* Cambridge University Press, 2012, 38-99.

[45] Karp, C. R. *Language with Expressions of Infinite Length.* North Holland, 1964.

[46] Kurz, A. *Logics for Coalgebras and Applications to Computer Science.* Dissertation, Ludwig-Maximilians-Universiät München, 2000.

[47] Lambek, J. "A Fixpoint Theorem for Complete Categories." Math. Zeitschr. 103 (1968) 151-161.

[48] Lewis, D. K. *Convention: A Philosophical Study.* Cambridge: Harvard University Press, 1969.

[49] Lipman, B. L. "Information Processing and Bounded Rationality: A Survey." Can. J. Econ. 28 (1995), 42-67.

[50] Lismont, L. and P. Mongin. "On the Logic of Common Belief and Common Knowledge." Theory Dec. 37 (1994), 75-106.

[51] Mariotti, T., M. Meier, and M. Piccione. "Hierarchies of Beliefs for Compact Possibility Models." J. Math. Econ. 41 (2005), 303-324.

[52] Maschler, M., E. Solan, and S. Zamir. Translated by Z. Hellman. *Game Theory.* Cambridge University Press, 2013.

[53] McCarthy, J., M. Sato, T. Hayashi, and S. Igarashi. (1978) "On the Model Theory of Knowledge." Stanford Artificial Intelligence Laboratory AIM-312, Computer Science Department STAN-CS-78-657.

[54] Meier, M. "On the Nonexistence of Universal Information Structures." J. Econ. Theory 122 (2005), 132-139.

[55] Meier, M. "Finitely Additive Beliefs and Universal Type Spaces." Ann. Probab. 34 (2006), 386-422.

[56] Meier, M. "Universal Knowledge-Belief Structures." Games Econ. Behav. 62 (2008), 53-66.

[57] Meier, M. "An Infinitely Probability Logic for Type Spaces." Israel J. Math. 192 (2012), 1-58.

[58] Mertens, J. F., S. Sorin, and S. Zamir. "Repeated Games: Part A: Background Material." CORE Discussion Paper No. 9420, Université Catholique de Louvain, 1994. Published as Mertens, J. F., S. Sorin, and S. Zamir. *Repeated Games*. Cambridge University Press, 2015.

[59] Mertens, J. F., and S. Zamir. "Formulation of Bayesian Analysis for Games with Incomplete Information." Int. J. Game Theory 14 (1985), 1-29.

[60] Milgrom, P. "An Axiomatic Characterization of Common Knowledge." Econometrica 49 (1981), 219-222.

[61] Monderer, D. and D. Samet. "Approximating Common Knowledge with Common Beliefs," Games Econ. Behav. 1 (1989), 1-29.

[62] Modica S. and A. Rustichini. "Awareness and Partitional Information Structures," Theory Dec. 37 (1994), 107-124.

[63] Modica S. and A. Rustichini. "Unawareness and Partitional Information Structures," Games Econ. Behav. 27 (1999), 265-298.

[64] Morris, S. "The Logic of Belief and Belief Change: A Decision Theoretic Approach." J. Econ. Theory. 69 (1996), 1-23.

[65] Moss, L. and I. D. Viglizzo. "Harsanyi Type Spaces and Final Coalgebras Constructed from Satisfied Theories." Electron. Notes Theor. Comput. Sci. 106 (2004), 279-295.

[66] Moss, L. and I. D. Viglizzo. "Final Coalgebras for Functors on Measurable Spaces." Inform. Comput. 204 (2006), 610-636.

[67] Mukerji, S. "Understanding the Nonadditive Probability Decision Model." Econ. Theory 9 (1997), 23-46.

[68] Nielsen, L. T. "Common Knowledge, Communication, and Convergence of Beliefs." Math. Soc. Sci. 8 (1984), 1-14.

[69] Pintér, M. "Type Space on a Purely Measurable Parameter Space." Econ. Theory 26 (2005), 129-139.

[70] Pires, C. P. "Do I know Ω? An Axiomatic Model of Awareness and Knowledge." A Manuscript, 1994.

[71] Roy, O. and E. Pacuit. "Substantive Assumptions in Interaction: A Logical Perspective." Synthese 190 (2013), 891-908.

[72] Rubinstein, A., and A. Wolinsky. "On the Logic of "Agreeing to Disagree" Type Results." J. Econ. Theory 51 (1990), 184-193.

[73] Rutten, J. J. M. M. "Universal Coalgebra: a Theory of Systems." Theor. Comput. Sci. 249 (2000), 3-80.

[74] Salonen, H. "Common Theories." Math. Soc. Sci. 58 (2009), 279-289.

[75] Samet, D. "Ignoring Ignorance and Agreeing to Disagree." J. Econ. Theory 52 (1990), 190-207.

[76] Samet, D. "Agreeing to Disagree in Infinite Information Structures." Int. J. Game Theory 21 (1992), 213-218.

[77] Samet, D. "Agreeing to Disagree: The Non-Probabilistic Case." Games Econ. Behav. 69 (2010), 169-174.

[78] Samet, D. "S5 Knowledge without Partitions." Synthese. 172 (2010), 145-155.

[79] Schipper, B. C. "Awareness." in H. van Ditmarsch, J.Y. Halpern, W. van der Hoek and B. Kooi (Eds.) *Handbook of Epistemic Logic*. College Publications, London, 2015, 77-146.

[80] Shin, H. S. "Logical Structure of Common Knowledge." J. Econ. Theory 60 (1993), 1-13.

[81] Stinchcombe, M. B. "Bayesian Information Topologies." J. Math. Econ. 19 (1990), 233-253.

[82] Tan, T. C.-C. and S. R. D. C. Werlang. "On Aumann's Notion of Common Knowledge: An Alternative Approach." R. Bras. Econ. 46 (1992), 151-166.

[83] Viglizzo, I. D. "Final Sequences and Final Coalgebras for Measurable Spaces." in J. L. Fiadeiro, N. Harman, M. Roggenbach, J. Rutten (Eds.) *Algebra and Coalgebra in Computer Science*. Springer-Verlag Berlin Heidelberg, 2005, 395-407.

A Appendix

A.1 Section 2

Remark A.1 (Proof of Equation (1))**.** Since $\omega' \notin \{\omega'\}^c$, if $\omega' \in b_{K_i}(\omega)$ then $\omega \in (\neg K_i)(\{\omega'\}^c) = L_{K_i}(\{\omega'\})$. Conversely, suppose to the contrary that there is $E \in \mathcal{D}$ with $\omega' \notin E$ (i.e., $E \subseteq \{\omega'\}^c$) and $\omega \in K_i(E)$. By Monotonicity, we have $\omega \in K_i(\{\omega'\}^c)$, a contradiction.

Proposition A.1 (Criteria for the Kripke Property)**.** *Let $\overrightarrow{\Omega}$ be a knowledge space such that K_i satisfies Monotonicity. The Kripke property of K_i is characterized as follows. If $(\omega, F) \in \Omega \times \mathcal{D}$ satisfies that $E \cap F \neq \emptyset$ for all $E \in \mathcal{D}$ with $\omega \in K_i(E)$, then (ω, F) satisfies $b_{K_i}(\omega) \cap F \neq \emptyset$.*

Proof of Proposition A.1. Assume the Kripke property. Take any $(\omega, F) \in \Omega \times \mathcal{D}$ such that $E \cap F \neq \emptyset$ for all $E \in \mathcal{D}$ with $\omega \in K_i(E)$. Now, if it were the case that $b_{K_i}(\omega) \cap F = \emptyset$, then we would have $b_{K_i}(\omega) \subseteq F^c$ and thus $F^c \cap F \neq \emptyset$, a contradiction.

Conversely, suppose to the contrary that there is $E' \in \mathcal{D}$ such that $b_{K_i}(\omega) \subseteq E'$ and $\omega \notin K_i(E')$. Since $b_{K_i}(\omega) \cap (E')^c = \emptyset$, there is $E \in \mathcal{D}$ such that $\omega \in K_i(E)$ and $E \cap (E')^c = \emptyset$. Then, we have $E \subseteq E'$. By Monotonicity, $\omega \in K_i(E')$, which is a contradiction. $\square$

Remark A.2 (Class of Knowlede Spaces as a Category)**.** First, the category of knowledge spaces is large (but locally small), whenever $S \neq \emptyset$. Indeed, we can introduce players' knowledge for any set Ω. Second, let $\mathbf{K}$ and $\mathbf{K}'$ be two categories of knowledge spaces where the assumptions on players' knowledge in $\mathbf{K}'$ are also assumed in $\mathbf{K}$. Then, $\mathbf{K}$ is a full subcategory of $\mathbf{K}'$: (i) any $\mathbf{K}$-object is also a $\mathbf{K}'$-object; (ii) any $\mathbf{K}$-morphism is also a $\mathbf{K}'$-morphism with the same identity and composite morphisms; and (iii) for any $\mathbf{K}$-objects $\overrightarrow{\Omega}$ and $\overrightarrow{\Omega}'$ and for any $\mathbf{K}'$-morphism $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega}'$, the mapping φ is also a $\mathbf{K}$-morphism.

A.2 Section 3

A.2.1 Section 3.1

First, Figure 1 illustrates the interrelations among the definitions and lemmas for the construction of a universal knowledge space in established Theorem 1.

Proof of Remark 1. We first show that $\mathcal{L}_\kappa \subseteq \mathcal{L}$. Clearly, $\mathcal{L}_0 \subseteq \mathcal{L}_\kappa$. If $\mathcal{L}_\beta \subseteq \mathcal{L}$ for all $\beta < \alpha$, then it is clear that $\mathcal{L}_\alpha \subseteq \mathcal{L}$. Hence, $\mathcal{L}_\kappa \subseteq \mathcal{L}$.

Conversely, it can be seen that if $e \in \mathcal{L}_\kappa$ then there is $\alpha < \kappa$ such that $e \in \mathcal{L}_\alpha$. Now, we establish that $\mathcal{L} \subseteq \mathcal{L}_\kappa$. First, it is immediate that $\mathcal{A}_S \subseteq \mathcal{L}_\kappa$. Second, if $e \in \mathcal{L}_\kappa$ then $e \in \mathcal{L}_\alpha$ for some $\alpha < \kappa$ and thus $(\neg e) \in \mathcal{L}_{\alpha+1} \subseteq \mathcal{L}_\kappa$. Third, let $\mathcal{F}$ be such that $\mathcal{F} \subseteq \mathcal{L}_\kappa$ and $0 < |\mathcal{F}| < \kappa$. Then, there is $\gamma < \kappa$ such that $e \in \mathcal{L}_\gamma \subseteq \mathcal{L}_\kappa$ for all $e \in \mathcal{F}$, and thus $\bigwedge \mathcal{F} \in \mathcal{L}_\kappa$. Hence, $\mathcal{L} \subseteq \mathcal{L}_\kappa$. $\square$

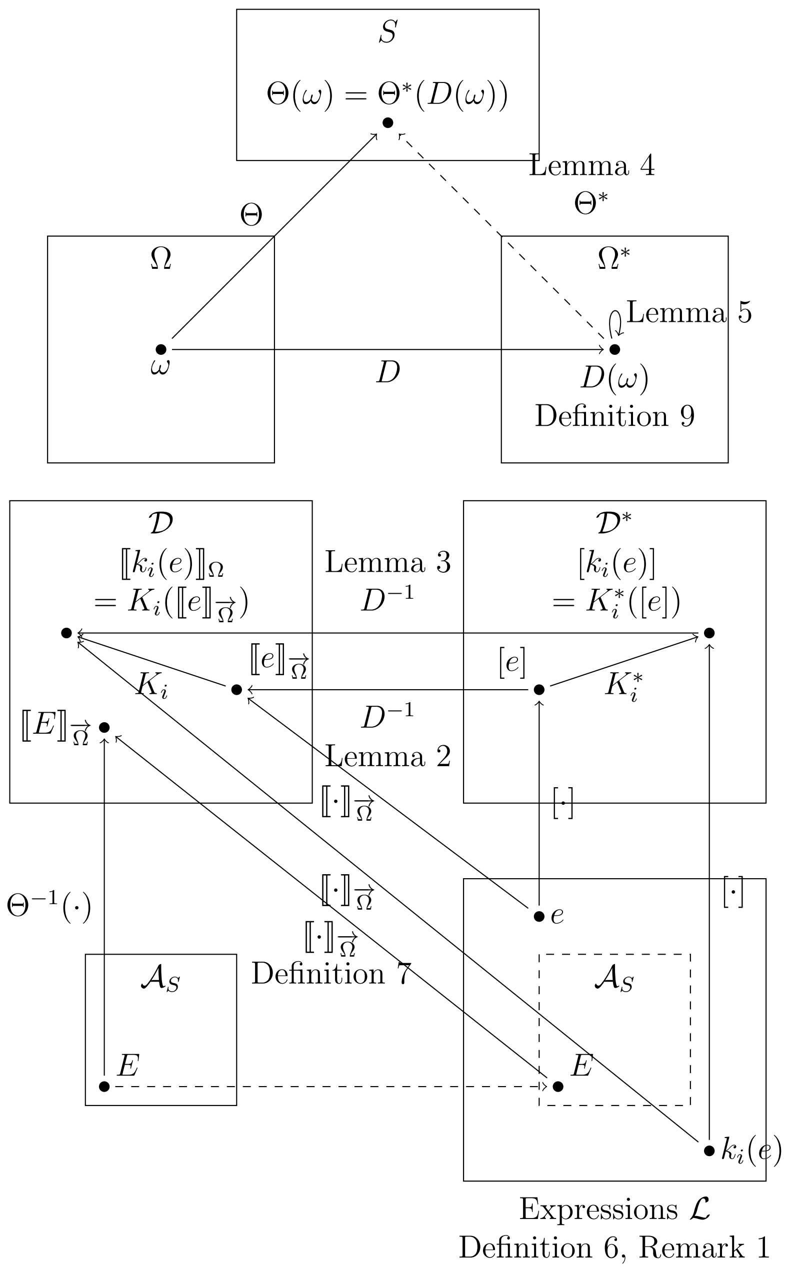


Figure 1: Interrelations among the Definitions and Lemmas for Theorem 1.

Proof of Lemma 1. In a similar way to the proof of HS [39, Proposition 4.1], Meier [55, Proposition 2], and Meier [56, Proposition 1], we show the statement by induction on the formulation of the expressions.

First, for each $E \in \mathcal{A}_S$, we have: $\omega \in [\![E]\!]_{\overrightarrow{\Omega}} := \Theta^{-1}(E)$ iff $\Theta(\omega) \in E$ iff $\Theta'(\varphi(\omega)) \in E$ iff

$\varphi(\omega) \in (\Theta')^{-1}(E) =: [\![E]\!]_{\overrightarrow{\Omega'}}$, where $\Theta(\omega) = \Theta'(\varphi(\omega))$ follows from the assumption that φ is a knowledge morphism.

Second, let $\mathcal{E}$ be a set of expressions with $|\mathcal{E}| < \kappa$. If $\mathcal{E} = \emptyset$, then observe that $\bigwedge \emptyset := S \in \mathcal{A}_S$. Hence, we let $\mathcal{E} \neq \emptyset$. Given the induction hypothesis, we have

$$\omega \in \left[\!\!\left[\bigwedge \mathcal{E}\right]\!\!\right]_{\overrightarrow{\Omega}} := \bigcap_{e \in \mathcal{E}} [\![e]\!]_{\overrightarrow{\Omega}} \text{ iff } \omega \in [\![e]\!]_{\overrightarrow{\Omega}} \text{ for all } e \in \mathcal{E}$$
$$\text{iff } \varphi(\omega) \in [\![e]\!]_{\overrightarrow{\Omega'}} \text{ for all } e \in \mathcal{E} \text{ iff } \varphi(\omega) \in \bigcap_{e \in \mathcal{E}} [\![e]\!]_{\overrightarrow{\Omega'}} =: \left[\!\!\left[\bigwedge \mathcal{E}\right]\!\!\right]_{\overrightarrow{\Omega'}}.$$

Third, assuming the induction hypothesis, we get

$$\omega \in [\![\neg e]\!]_{\overrightarrow{\Omega}} := \neg [\![e]\!]_{\overrightarrow{\Omega}} \text{ iff } \omega \notin [\![e]\!]_{\overrightarrow{\Omega}} \text{ iff } \varphi(\omega) \notin [\![e]\!]_{\overrightarrow{\Omega'}} \text{ iff } \varphi(\omega) \in \neg [\![e]\!]_{\overrightarrow{\Omega'}} =: [\![\neg e]\!]_{\overrightarrow{\Omega'}}.$$

Fourth, assuming the induction hypothesis, we have

$$\omega \in [\![k_i(e)]\!]_{\overrightarrow{\Omega}} := K_i([\![e]\!]_{\overrightarrow{\Omega}}) = K_i(\varphi^{-1}([\![e]\!]_{\overrightarrow{\Omega'}})) = \varphi^{-1}(K_i'([\![e]\!]_{\overrightarrow{\Omega'}}))$$
$$\text{iff } \varphi(\omega) \in K_i'([\![e]\!]_{\overrightarrow{\Omega'}}) =: [\![k_i(e)]\!]_{\overrightarrow{\Omega'}}.$$

The induction is complete. □

Proof of Corollary 1. Fix $\omega \in \Omega$. First, since φ is a knowledge morphism, we have $\Theta(\omega) = \Theta'(\varphi(\omega))$. Second, it follows from Lemma 1 that $\omega \in [\![e]\!]_{\overrightarrow{\Omega}}$ iff $\varphi(\omega) \in [\![e]\!]_{\overrightarrow{\Omega'}}$ for all $e \in \mathcal{L}$. It follows from these two arguments that $D_{\overrightarrow{\Omega}}(\omega) = D_{\overrightarrow{\Omega'}}(\varphi(\omega))$. □

Remark A.3 (Remark on Footnote 28)**.** Two states, which possibly reside in different knowledge spaces, are identified when the descriptions are identical. We remark that this notion is related to behavioral equivalence (Kurz [46]). Let $\overrightarrow{\Omega}$ and $\overrightarrow{\Omega'}$ be knowledge spaces in a given category. A pair of states $(\omega, \omega') \in \Omega \times \Omega'$ is said to be *behaviorally equivalent* if there are a knowledge space $\overrightarrow{\Omega''}$ and a pair of knowledge morphisms $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega''}$ and $\varphi' : \overrightarrow{\Omega'} \to \overrightarrow{\Omega''}$ such that $\varphi(\omega) = \varphi'(\omega')$.

We show that $(\omega, \omega') \in \Omega \times \Omega'$ is behaviorally equivalent iff $D_{\overrightarrow{\Omega}}(\omega) = D_{\overrightarrow{\Omega'}}(\omega')$. This means that, in order to show that two states are identical in terms of players' hierarchies of knowledge, it is enough to show that they are behaviorally equivalent. The proof goes as follows. If $(\omega, \omega') \in \Omega \times \Omega'$ is behaviorally equivalent then we have

$$D_{\overrightarrow{\Omega}}(\omega) = D_{\overrightarrow{\Omega''}}(\varphi(\omega)) = D_{\overrightarrow{\Omega''}}(\varphi'(\omega')) = D_{\overrightarrow{\Omega'}}(\omega'),$$

where the first and third equalities follow from Corollary 1. The converse trivially holds once we show that the description map is a knowledge morphism.

Proof of Lemma 2. We divide the proof into the following three steps. In the first step, we prove the following correspondence between syntactic and semantic operations.

1. $[(\neg e)] = \neg[e]$ for any $e \in \mathcal{L}$.

2. $[S] = \Omega^*$ and $[\emptyset] = \emptyset$. In other words, $[\bigwedge \emptyset] = \Omega^*$ and $[\bigvee \emptyset] = \emptyset$.

3. Let $\mathcal{E}$ be a set of expressions with $(0 <)|\mathcal{E}| < \kappa$. Then, $\left[\bigwedge_{e\in\mathcal{E}} e\right] = \bigcap_{e\in\mathcal{E}}[e]$ and $\left[\bigvee_{e\in\mathcal{E}} e\right] = \bigcup_{e\in\mathcal{E}}[e]$.

1. Fix $e \in \mathcal{L}$. We have the following:

$$\begin{aligned}\omega^* \in [(\neg e)] \text{ iff } (1,(\neg e)) \in \omega^* = D(\omega) \text{ iff } \omega \in [\![\neg e]\!]_{\overrightarrow{\Omega}} = \neg[\![e]\!]_{\overrightarrow{\Omega}} \\ \text{iff } \omega \notin [\![e]\!]_{\overrightarrow{\Omega}} \text{ iff } (1,e) \notin D(\omega) = \omega^* \text{ iff } \omega^* \notin [e] \text{ iff } \omega^* \in (\neg[e]),\end{aligned}$$

where a knowledge space $\overrightarrow{\Omega}$ and a state $\omega \in \Omega$ satisfy $\omega^* = D(\omega)$. Thus, we obtain $[(\neg e)] = \neg[e]$.

2. First, we show that $[S] = \Omega^*$. It is sufficient to prove that $\Omega^* \subseteq [S]$. For any $\omega^* \in \Omega^*$, there are a knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$ such that $\omega^* = D(\omega)$. Now, noting that $[\![S]\!]_{\overrightarrow{\Omega}} = \Theta^{-1}(S) = \Omega$, we have $\omega \in \Omega = [\![S]\!]_{\overrightarrow{\Omega}}$. By definition, we get $(1,S) \in D(\omega) = \omega^*$ and thus $\omega^* \in [S]$. Hence, $\Omega^* \subseteq [S]$. Now, we also have $[\emptyset] = [\neg S] = \neg[S] = \emptyset$.

3. Let $\mathcal{E}$ be a set of expressions with $(0 <)|\mathcal{E}| < \kappa$. It is enough to show that $[\bigwedge \mathcal{E}] = \bigcap_{e\in\mathcal{E}}[e]$. If $\omega^* \in [\bigwedge \mathcal{E}]$, then $(1, \bigwedge \mathcal{E}) \in \omega^*$. There are a knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$ such that $D(\omega) = \omega^*$. Thus, we have $(1, \bigwedge \mathcal{E}) \in \omega^* = D(\omega)$. By Definitions 7 and 9, we have $\omega \in [\![\bigwedge \mathcal{E}]\!]_{\overrightarrow{\Omega}} = \bigcap_{e\in\mathcal{E}}[\![e]\!]_{\overrightarrow{\Omega}}$. Since $\omega \in [\![e]\!]_{\overrightarrow{\Omega}}$ iff $(1,e) \in D(\omega) = \omega^*$ iff $\omega^* \in [e]$, it follows that $\omega^* \in \bigcap_{e\in\mathcal{E}}[e]$.

   Conversely, suppose $\omega^* \in \bigcap_{e\in\mathcal{E}}[e]$. There are a knowledge space $\overrightarrow{\Omega'}$ and ω' such that $D(\omega') = \omega^*$. Again, since $\omega^* \in [e]$ iff $(1,e) \in D(\omega') = \omega^*$ iff $\omega' \in [\![e]\!]_{\overrightarrow{\Omega'}}$, we have $\omega' \in \bigcap_{e\in\mathcal{E}}[\![e]\!]_{\overrightarrow{\Omega'}} = [\![\bigwedge \mathcal{E}]\!]_{\overrightarrow{\Omega'}}$. It follows from Definition 9 that $(1, \bigwedge \mathcal{E}) \in D(\omega') = \omega^*$, i.e., $\omega^* \in [\bigwedge \mathcal{E}]$.

In the second step, we show that $\mathcal{D}^*$ is a κ-complete algebra on Ω^*. It follows from the first step that $\emptyset = [\emptyset] \in \mathcal{D}^*$ and $\Omega^* = [S] \in \mathcal{D}^*$, where note that $\emptyset$ and S are expressions. In other words, $\mathcal{D}^*$ is closed under the empty intersection and union.

Next, we show that $\mathcal{D}^*$ is closed under complementation. If $[e] \in \mathcal{D}^*$, then it follows from the first step that $\neg[e] = [(\neg e)]$. Since $(\neg e) \in \mathcal{L}$, we have $\neg[e] = [(\neg e)] \in \mathcal{D}^*$.

Next, we show that $\mathcal{D}^*$ is closed under non-empty κ-intersection (and κ-union). It is enough to take a subset $\mathcal{E}$ of $\mathcal{L}$ with $(0 <)|\mathcal{E}| < \kappa$. Then, it follows from the first step that $\bigcap_{e\in\mathcal{E}}[e] = [\bigwedge \mathcal{E}]$ (and $\bigcup_{e\in\mathcal{E}}[e] = [\bigvee \mathcal{E}]$). Since $\bigwedge \mathcal{E} \in \mathcal{L}$ (and $\bigvee \mathcal{E} \in \mathcal{L}$), we have $\bigcap_{e\in\mathcal{E}}[e] = [\bigwedge \mathcal{E}] \in \mathcal{D}^*$ (and $\bigcup_{e\in\mathcal{E}}[e] = [\bigvee \mathcal{E}] \in \mathcal{D}^*$).

In the third step, we show that, for any given knowledge space $\overrightarrow{\Omega}$, the description map $D : \Omega \to \Omega^*$ satisfies that $D^{-1}([e]) = [\![e]\!]_{\overrightarrow{\Omega}}$ for all $[e] \in \mathcal{D}^*$. Fix $[e] \in \mathcal{D}^*$. We have $\omega \in D^{-1}([e])$ iff $D(\omega) \in [e]$ iff $(1,e) \in D(\omega)$ iff $\omega \in [\![e]\!]_{\overrightarrow{\Omega}}$. □

Next, we establish Lemma 3 in a general way so that we can also demonstrate the preservation of other potential properties on knowledge. To that end, we consider the following two lemmas.

Lemma A.1 (Preservation of Varieties of Properties of Knowledge)**.** *We express a property of players' knowledge by using operators* $f_{\overrightarrow{\Omega}} : \mathcal{D} \to \mathcal{D}$ *and* $g_{\overrightarrow{\Omega}} : \mathcal{D} \to \mathcal{D}$ *for each* κ*-knowledge space* $\overrightarrow{\Omega}$*. To that end, suppose that, for any given pair of knowledge spaces* $(\overrightarrow{\Omega}, \overrightarrow{\Omega'})$ *and knowledge morphism* $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega'}$*, the operations* $f_{\overrightarrow{\Omega}}$ *and* $f_{\overrightarrow{\Omega'}}$ *(likewise,* $g_{\overrightarrow{\Omega}}$ *and* $g_{\overrightarrow{\Omega'}}$*) satisfy* $\varphi^{-1}(f_{\overrightarrow{\Omega'}}(E')) = f_{\overrightarrow{\Omega}}(\varphi^{-1}(E'))$ *(likewise,* $\varphi^{-1}(g_{\overrightarrow{\Omega'}}(E')) = g_{\overrightarrow{\Omega}}(\varphi^{-1}(E'))$*) for all* $E' \in \mathcal{D}'$*.*

For example, $f_{\overrightarrow{\Omega}}$ *and* $g_{\overrightarrow{\Omega}}$ *are operators generated by composing knowledge operators* $(K_i)_{i \in I}$ *and set-algebraic as well as constant and identity operations. Specific examples are:* $f_{\overrightarrow{\Omega}}(\cdot) = K_i(\cdot)$*,* $f_{\overrightarrow{\Omega}}(\cdot) = K_i(\neg K_i)(\cdot)$*,* $f_{\overrightarrow{\Omega}}(\cdot) = \bigcap_{i \in I} K_i(\cdot)$*,* $f_{\overrightarrow{\Omega}}(\cdot) = \mathrm{id}_{\mathcal{D}}(\cdot)$*,* $f_{\overrightarrow{\Omega}}(\cdot) = K_i K_j(\cdot)$*,* $f_{\overrightarrow{\Omega}}(\cdot) = K_i(\emptyset)$*,* $f_{\overrightarrow{\Omega}}(\cdot) = \Omega$*,* $f_{\overrightarrow{\Omega}}(\cdot) = \emptyset$*, and so on.*

Now, we have the following.

1. *Suppose that* $\omega \in g_{\overrightarrow{\Omega}}(E)$ *for all* $E \in \mathcal{D}$ *with* $\omega \in f_{\overrightarrow{\Omega}}(E)$*. Then, for any* $\omega \in \Omega$ *and* $E' \in \mathcal{D}$*, if* $\varphi(\omega) \in f_{\overrightarrow{\Omega}}(E')$ *then* $\varphi(\omega) \in g_{\overrightarrow{\Omega}}(E')$*.*
2. (a) *If* $f_{\overrightarrow{\Omega}}(\cdot) \subseteq g_{\overrightarrow{\Omega}}(\cdot)$ *then* $f_{\overrightarrow{\Omega^*}}(\cdot) \subseteq g_{\overrightarrow{\Omega^*}}(\cdot)$*.*[A.1]
   (b) *Suppose that* $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega'}$ *is a surjective knowledge morphism. If* $f_{\overrightarrow{\Omega}}(\cdot) \subseteq g_{\overrightarrow{\Omega}}(\cdot)$ *then* $f_{\overrightarrow{\Omega'}}(\cdot) \subseteq g_{\overrightarrow{\Omega'}}(\cdot)$*.*
3. *Suppose that* $f_{\overrightarrow{\Omega}}(E) \subseteq f_{\overrightarrow{\Omega}}(F)$ *for all* $E, F \in \mathcal{D}$ *with* $E \subseteq F$*. Then,* $\varphi(\omega) \in f_{\overrightarrow{\Omega'}}(E')$ *implies* $\varphi(\omega) \in f_{\overrightarrow{\Omega'}}(F')$ *for all* $E', F' \in \mathcal{D}'$ *with* $E' \subseteq F'$ *and* $\omega \in \Omega$*.*
4. *Suppose that* $f_{\overrightarrow{\Omega}}(E) \subseteq f_{\overrightarrow{\Omega}}(F)$ *for all* $E, F \in \mathcal{D}$ *with* $E \subseteq F$*.*
   (a) $f_{\overrightarrow{\Omega^*}}([e]) \subseteq f_{\overrightarrow{\Omega^*}}([f])$ *for all* $[e], [f] \in \mathcal{D}^*$ *with* $[e] \subseteq [f]$*.*
   (b) *Suppose that* $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega'}$ *is a surjective knowledge morphism. Then,* $f_{\overrightarrow{\Omega'}}(E') \subseteq f_{\overrightarrow{\Omega'}}(F')$ *for all* $E', F' \in \mathcal{D}'$ *with* $E' \subseteq F'$*.*
5. *Suppose that* $\bigcap_{E \in \mathcal{E}} f_{\overrightarrow{\Omega}}(E) \subseteq f_{\overrightarrow{\Omega}}(\bigcap \mathcal{E})$ *for all* $\mathcal{E} \subseteq \mathcal{D}$ *with* $(0 <)|\mathcal{E}| < \kappa$*. Then,* $\varphi(\omega) \in \bigcap_{E' \in \mathcal{E}'} f_{\overrightarrow{\Omega'}}(E')$ *implies* $\varphi(\omega) \in f_{\overrightarrow{\Omega'}}(\bigcap \mathcal{E}')$ *for all* $\omega \in \Omega$ *and* $\mathcal{E}' \subseteq \mathcal{D}'$ *with* $(0 <)|\mathcal{E}'| < \kappa$*.*
6. *Suppose that* $\bigcap_{E \in \mathcal{E}} f_{\overrightarrow{\Omega}}(E) \subseteq f_{\overrightarrow{\Omega}}(\bigcap \mathcal{E})$ *for all* $\mathcal{E} \subseteq \mathcal{D}$ *with* $(0 <)|\mathcal{E}| < \kappa$*.*
   (a) $\bigcap_{[e] \in \mathcal{E}^*} f_{\overrightarrow{\Omega^*}}([e]) \subseteq f_{\overrightarrow{\Omega^*}}(\bigcap \mathcal{E}^*)$ *for all* $\mathcal{E}^* \subseteq \mathcal{D}^*$ *with* $(0 <)|\mathcal{E}^*| < \kappa$*.*
   (b) *Suppose that* $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega'}$ *is a surjective knowledge morphism. Then,* $\bigcap_{E' \in \mathcal{E}'} f_{\overrightarrow{\Omega'}}(E') \subseteq f_{\overrightarrow{\Omega'}}(\bigcap \mathcal{E}')$ *for all* $\mathcal{E}' \subseteq \mathcal{D}'$ *with* $(0 <)|\mathcal{E}'| < \kappa$*.*

[A.1] To be precise, we abuse the notation $f_{\overrightarrow{\Omega^*}}(\cdot)$ to denote the given property expressed by the operation f with respect to players' knowledge in $(\Omega^*, \mathcal{D}^*)$.

Proof of Lemma A.1. 1. Fix $\omega \in \Omega$ and $E' \in \mathcal{D}'$. Suppose that $\varphi(\omega) \in f_{\overrightarrow{\Omega'}}(E')$. Then, $\omega \in \varphi^{-1}(f_{\overrightarrow{\Omega'}}(E')) = f_{\overrightarrow{\Omega}}(\varphi^{-1}(E'))$. It follows from the assumption that $\omega \in g_{\overrightarrow{\Omega}}(\varphi^{-1}(E')) = \varphi^{-1}(g_{\overrightarrow{\Omega'}}(E'))$, i.e., $\varphi(\omega) \in g_{\overrightarrow{\Omega'}}(E')$.

2. (a) Fix $[e] \in \mathcal{D}^*$. If $\omega^* \in f_{\overrightarrow{\Omega^*}}([e])$ then there are a knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$ with $D(\omega) = \omega^* \in f_{\overrightarrow{\Omega^*}}([e])$. Now, it follows from the previous argument that $\omega^* = D(\omega) \in g_{\overrightarrow{\Omega^*}}([e])$.
   (b) Fix $E' \in \mathcal{D}'$. If $\omega' \in f_{\overrightarrow{\Omega'}}(E')$ then there is $\omega \in \Omega$ with $\varphi(\omega) = \omega' \in f_{\overrightarrow{\Omega'}}(E')$. Now, it follows from the previous argument that $\omega' = \varphi(\omega) \in g_{\overrightarrow{\Omega'}}(E')$.

3. The proof is similar to that of (1). Fix $\omega \in \Omega$ and $E', F' \in \mathcal{D}'$ with $E' \subseteq F'$. Suppose that $\varphi(\omega) \in f_{\overrightarrow{\Omega'}}(E')$. Then, $\omega \in \varphi^{-1}(f_{\overrightarrow{\Omega'}}(E')) = f_{\overrightarrow{\Omega}}(\varphi^{-1}(E'))$. Since $\varphi^{-1}(E') \subseteq \varphi^{-1}(F')$, it follows from the assumption that $\omega \in f_{\overrightarrow{\Omega}}(\varphi^{-1}(F')) = \varphi^{-1}(f_{\overrightarrow{\Omega'}}(F'))$, i.e., $\varphi(\omega) \in f_{\overrightarrow{\Omega'}}(F')$.

4. The proof is similar to that of (2).
   (a) Fix $[e], [f] \in \mathcal{D}^*$ with $[e] \subseteq [f]$. If $\omega^* \in f_{\overrightarrow{\Omega^*}}([e])$ then there are a knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$ with $D(\omega) = \omega^* \in f_{\overrightarrow{\Omega^*}}([e])$. Now, it follows from the previous argument that $\omega^* = D(\omega) \in f_{\overrightarrow{\Omega^*}}([f])$.
   (b) Fix $E', F' \in \mathcal{D}'$ with $E' \subseteq E'$. If $\omega' \in f_{\overrightarrow{\Omega'}}(E')$ then there is $\omega \in \Omega$ with $\varphi(\omega) = \omega' \in f_{\overrightarrow{\Omega'}}(E')$. Now, it follows from the previous argument that $\omega' = \varphi(\omega) \in f_{\overrightarrow{\Omega'}}(F')$.

5. The proof is similar to that of (1). Fix $\omega \in \Omega$ and $\mathcal{E}' \subseteq \mathcal{D}'$ with $|\mathcal{E}'| < \kappa$. Suppose that $\varphi(\omega) \in \bigcap_{E' \in \mathcal{E}'} f_{\overrightarrow{\Omega'}}(E')$. Then, $\omega \in \varphi^{-1}(f_{\overrightarrow{\Omega'}}(E')) = f_{\overrightarrow{\Omega}}(\varphi^{-1}(E'))$ for all $E' \in \mathcal{E}'$, i.e., $\omega \in \bigcap_{E' \in \mathcal{E}'} f_{\overrightarrow{\Omega}}(\varphi^{-1}(E'))$. Now, it follows from the assumption that $\omega \in f_{\overrightarrow{\Omega}}(\bigcap_{E' \in \mathcal{E}'} \varphi^{-1}(E')) = f_{\overrightarrow{\Omega}}(\varphi^{-1}(\bigcap \mathcal{E}')) = \varphi^{-1}(f_{\overrightarrow{\Omega'}}(\bigcap \mathcal{E}'))$, i.e., $\varphi(\omega) \in f_{\overrightarrow{\Omega'}}(\bigcap \mathcal{E}')$.

6. The proof is similar to that of (2).
   (a) Fix $\mathcal{E}^* \subseteq \mathcal{D}^*$ with $|\mathcal{E}^*| < \kappa$. If $\omega^* \in \bigcap_{[e] \in \mathcal{E}^*} f_{\overrightarrow{\Omega^*}}([e])$ then there are a knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$ with $D(\omega) = \omega^* \in \bigcap_{[e] \in \mathcal{E}^*} f_{\overrightarrow{\Omega^*}}([e])$. Now, it follows from the previous argument that $\omega^* = D(\omega) \in f_{\overrightarrow{\Omega^*}}(\bigcap \mathcal{E}^*)$.
   (b) Suppose that φ is surjective. Fix $\mathcal{E}' \subseteq \mathcal{D}'$ with $|\mathcal{E}'| < \kappa$. If $\omega' \in \bigcap_{E' \in \mathcal{E}'} f_{\overrightarrow{\Omega'}}(E')$ then there is $\omega \in \Omega$ with $\varphi(\omega) = \omega' \in \bigcap_{E' \in \mathcal{E}'} f_{\overrightarrow{\Omega'}}(E')$. Now, it follows from the previous argument that $\omega' = \varphi(\omega) \in f_{\overrightarrow{\Omega'}}(\bigcap \mathcal{E}')$.

$\square$

We also establish the following lemma in order to assert the preservation of the Kripke property.

Lemma A.2 (Preservation of the Kripke Property)**.** *Let $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega'}$ be a knowledge morphism between knowledge spaces $\overrightarrow{\Omega}$ and $\overrightarrow{\Omega'}$. For each $\omega \in \Omega$, we have $\varphi(b_{K_i}(\omega)) \subseteq b'_{K'_i}(\varphi(\omega))$, where recall that*

$$b_{K_i}(\omega) = \bigcap\{E \in \mathcal{D} \mid \omega \in K_i(E)\} \text{ and } b'_{K'_i}(\varphi(\omega)) = \bigcap\{E' \in \mathcal{D}' \mid \varphi(\omega) \in K'_i(E')\}.$$

Proof of Lemma A.2. Fix $\omega \in \Omega$. If $\varphi(\tilde{\omega}) \notin b'_{K'_i}(\varphi(\omega))$, then there is $E' \in \mathcal{D}'$ such that $\varphi(\omega) \in K'_i(E')$ and $\varphi(\tilde{\omega}) \notin K'_i(E')$. Since φ is a knowledge morphism, $\omega \in K_i(\varphi^{-1}(E'))$ and $\tilde{\omega} \notin K_i(\varphi^{-1}(E'))$. This implies that $\tilde{\omega} \notin b_{K_i}(\omega)$. This yields the desired expression.

In an alternative way, we can directly assert the desired expression as follows.

$$\begin{aligned} b'_{K'_i}(\varphi(\omega)) = \bigcap\{E' \in \mathcal{D}' \mid \omega \in K_i(\varphi^{-1}(E'))\} &\supseteq \bigcap_{E' \in \mathcal{D}' : \omega \in K_i(\varphi^{-1}(E'))} \varphi(\varphi^{-1}(E')) \\ &\supseteq \varphi(\bigcap_{E' \in \mathcal{D}' : \omega \in K_i(\varphi^{-1}(E'))} \varphi^{-1}(E')) \supseteq \varphi(b_{t_i}(\omega)). \end{aligned}$$

□

Now, we establish Lemma 3.

Proof of Lemma 3. Fix $i \in I$. We prove the statement in the following three steps. First, we show that the operator $K_i^* : \mathcal{D}^* \to \mathcal{D}^*$ is well defined, i.e., $K_i^*([e]) = K_i^*([f])$ for any $e, f \in \mathcal{D}$ with $[e] = [f]$.

Let $e, f \in \mathcal{D}$ be such that $[e] = [f]$. If $\omega^* \in K_i^*([e]) = [k_i(e)]$, then there are a knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$ such that $D(\omega) \in [k_i(e)]$, i.e., $(1, k_i(e)) \in D(\omega)$. Thus, we have $\omega \in [\![k_i(e)]\!]_{\overrightarrow{\Omega}} = K_i([\![e]\!]_{\overrightarrow{\Omega}}) = K_i(D^{-1}([e]))$. Since $[e] = [f]$, we obtain $\omega \in [\![k_i(f)]\!]_{\overrightarrow{\Omega}}$. That is, we have $\omega^* = D(\omega) \in [k_i(f)] = K_i^*([f])$. By changing the role of e and f, we conclude that $K_i^*([e]) = K_i^*([f])$.

Second, for any $[e] \in \mathcal{D}^*$, we have

$$K_i(D^{-1}([e])) = K_i([\![e]\!]_{\overrightarrow{\Omega}}) = [\![k_i(e)]\!]_{\overrightarrow{\Omega}} = D^{-1}([k_i(e)]) = D^{-1}(K_i^*([e])).$$

Third, we show that each K_i^* inherits all the properties imposed in the given category of knowledge spaces.

1. For No-Contradiction Axiom, apply Lemma A.1 (2a) by taking $(f_{\overrightarrow{\Omega}}, g_{\overrightarrow{\Omega}}) = (K_i(\emptyset), \emptyset)$.
2. For Consistency, apply Lemma A.1 (2a) by taking $(f_{\overrightarrow{\Omega}}, g_{\overrightarrow{\Omega}}) = (K_i(\cdot) \cap (\neg K_i)(\cdot), \emptyset)$.
3. For Truth Axiom, apply Lemma A.1 (2a) by taking $(f_{\overrightarrow{\Omega}}, g_{\overrightarrow{\Omega}}) = (K_i(\cdot), \mathrm{id}_{\mathcal{D}}(\cdot))$.
4. For Necessitation, apply Lemma A.1 (2a) by taking $(f_{\overrightarrow{\Omega}}, g_{\overrightarrow{\Omega}}) = (\Omega, K_i(\Omega))$.
5. For Positive Introspection, apply Lemma A.1 (2a) by taking $(f_{\overrightarrow{\Omega}}, g_{\overrightarrow{\Omega}}) = (K_i(\cdot), K_i K_i(\cdot))$.

6. For Negative Introspection, apply Lemma A.1 (2a) by taking $(f_{\overrightarrow{\Omega}}, g_{\overrightarrow{\Omega}}) = (K_i(\cdot), K_i(\neg K_i)(\cdot))$.
7. For Monotonicity, apply Lemma A.1 (4a) by taking $f_{\overrightarrow{\Omega}} = K_i(\cdot)$.
8. For Non-empty λ-Conjunction, apply Lemma A.1 (6a) by taking $f_{\overrightarrow{\Omega}}(\cdot) = K_i(\cdot)$.
9. Finally, consider the Kripke property. Note that, by Lemma A.2, we have $D(b_{K_i}(\omega)) \subseteq b^*_{K^*_i}(D(\omega))$ for each $\omega \in \Omega$. Suppose that $b^*_{K^*_i}(\omega^*) \subseteq [e]$. There are a knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$ such that $\omega^* = D(\omega)$. Then, we have

$$b_{K_i}(\omega) \subseteq D^{-1}(D(b_{K_i}(\omega))) \subseteq D^{-1}(b^*_{K^*_i}(\omega^*)) \subseteq D^{-1}([e]).$$

By the Kripke property of $\overrightarrow{\Omega}$, it follows that $\omega \in K_i(D^{-1}([e])) = D^{-1}(K^*_i([e]))$, i.e., $\omega^* = D(\omega) \in K^*_i([e])$.

$\square$

We provide the following two results. The first is a lemma regarding the preservation of properties of knowledge under surjective mappings.

Lemma A.3 (Surjection Preserves Properties of Knowledge)**.** *Let $\overrightarrow{\Omega}$ and $\overrightarrow{\Omega}'$ be knowledge spaces which may reside in different classes of knowledge spaces. Let $\varphi : \Omega \to \Omega'$ be a surjective mapping such that $K_i(\varphi^{-1}(E')) = \varphi^{-1}(K'_i(E'))$ for all $E' \in \mathcal{D}'$. Then, the knowledge operator K'_i inherits all the properties of K_i.*

Proof of Lemma A.3. The proof is similar to the argument in the proof of Lemma 3 (i.e., applications of Lemmas A.1 and A.2), and hence omitted. $\square$

Second, we provide a sufficient condition for K^*_i to violate some property of knowledge if it is not assumed in the given category of knowledge spaces.

Proposition A.2 (When Does Candidate Universal Space Violate Properties of Knowledge?)**.**

1. *Suppose that a property of knowledge is expressed by $f_{\overrightarrow{\Omega}}(\cdot) \subseteq g_{\overrightarrow{\Omega}}(\cdot)$ for each knowledge space $\overrightarrow{\Omega}$. Suppose that there is a knowledge space $\overrightarrow{\Omega}$ which violates this property with respect to some $e \in \mathcal{L}$. That is, $f_{\overrightarrow{\Omega}}([\![e]\!]_{\overrightarrow{\Omega}}) \not\subseteq g_{\overrightarrow{\Omega}}([\![e]\!]_{\overrightarrow{\Omega}})$. Then, we have $f_{\overrightarrow{\Omega^*}}([e]) \not\subseteq g_{\overrightarrow{\Omega^*}}([e])$.*
2. *Suppose that a property of knowledge is expressed by the monotonicity of $f_{\overrightarrow{\Omega}}(\cdot)$ for each knowledge space $\overrightarrow{\Omega}$. Suppose that there is a knowledge space $\overrightarrow{\Omega}$ which violates this property with respect to some expressions $e, f \in \mathcal{L}$ with $[\![e]\!]_{\overrightarrow{\Omega}} \subseteq [\![f]\!]_{\overrightarrow{\Omega}}$. That is, $f_{\overrightarrow{\Omega}}([\![e]\!]_{\overrightarrow{\Omega}}) \not\subseteq f_{\overrightarrow{\Omega}}([\![f]\!]_{\overrightarrow{\Omega}})$. Then, we have $f_{\overrightarrow{\Omega^*}}([e]) \not\subseteq f_{\overrightarrow{\Omega^*}}([f])$.*

3. *Suppose that a property of knowledge is expressed by the non-empty λ-conjunction of $f_{\overrightarrow{\Omega}}(\cdot)$ for each knowledge space $\overrightarrow{\Omega}$. Suppose that there is a knowledge space $\overrightarrow{\Omega}$ which violates this property with respect to some set of expressions $\mathcal{E}(\subseteq \mathcal{L})$ with $0 < |\mathcal{E}| < \lambda(\leq \kappa)$. That is, $\bigcap_{e\in\mathcal{E}} f_{\overrightarrow{\Omega}}([\![e]\!]_{\overrightarrow{\Omega}}) \not\subseteq f_{\overrightarrow{\Omega}}(\bigcap_{e\in\mathcal{E}}[\![e]\!]_{\overrightarrow{\Omega}})$. Then, we have $\bigcap_{e\in\mathcal{E}} f_{\overrightarrow{\Omega^*}}([e]) \not\subseteq f_{\overrightarrow{\Omega^*}}(\bigcap_{e\in\mathcal{E}}[e])$.*

Proof of Proposition A.2. 1. Suppose that there are a knowledge space $\overrightarrow{\Omega}$ and an expression $e \in \mathcal{L}$ such that $f_{\overrightarrow{\Omega}}([\![e]\!]_{\overrightarrow{\Omega}}) \not\subseteq g_{\overrightarrow{\Omega}}([\![e]\!]_{\overrightarrow{\Omega}})$. Then, there is $\omega \in \Omega$ such that $\omega \in f_{\overrightarrow{\Omega}}([\![e]\!]_{\overrightarrow{\Omega}})$ and $\omega \notin g_{\overrightarrow{\Omega}}([\![e]\!]_{\overrightarrow{\Omega}})$. Since $f_{\overrightarrow{\Omega}}([\![e]\!]_{\overrightarrow{\Omega}}) = D^{-1}(f_{\overrightarrow{\Omega^*}}([e]))$ and $g_{\overrightarrow{\Omega}}([\![e]\!]_{\overrightarrow{\Omega}}) = D^{-1}(g_{\overrightarrow{\Omega^*}}([e]))$, we obtain $D(\omega) \in f_{\overrightarrow{\Omega^*}}([e])$ and $D(\omega) \notin g_{\overrightarrow{\Omega^*}}([e])$.

2. Suppose that there are a knowledge space $\overrightarrow{\Omega}$ and expressions $e, f \in \mathcal{L}$ such that $f_{\overrightarrow{\Omega}}([\![e]\!]_{\overrightarrow{\Omega}}) \not\subseteq f_{\overrightarrow{\Omega}}([\![f]\!]_{\overrightarrow{\Omega}})$. Then, there is $\omega \in \Omega$ such that $\omega \in f_{\overrightarrow{\Omega}}([\![e]\!]_{\overrightarrow{\Omega}})$ and $\omega \notin f_{\overrightarrow{\Omega}}([\![f]\!]_{\overrightarrow{\Omega}})$. Now, we obtain $D(\omega) \in f_{\overrightarrow{\Omega^*}}([e])$ and $D(\omega) \notin f_{\overrightarrow{\Omega^*}}([f])$.

3. Suppose that there are a knowledge space $\overrightarrow{\Omega}$ and a set of expressions $\mathcal{E}(\subseteq \mathcal{L})$ such that $\bigcap_{e\in\mathcal{E}} f_{\overrightarrow{\Omega}}([\![e]\!]_{\overrightarrow{\Omega}}) \not\subseteq f_{\overrightarrow{\Omega}}(\bigcap_{e\in\mathcal{E}}[\![e]\!]_{\overrightarrow{\Omega}})$ and $0 < |\mathcal{E}| < \lambda(\leq \kappa)$. Then, there is $\omega \in \Omega$ such that $\omega \in \bigcap_{e\in\mathcal{E}} f_{\overrightarrow{\Omega}}([\![e]\!]_{\overrightarrow{\Omega}}) = D^{-1}(\bigcap_{e\in\mathcal{E}} f_{\overrightarrow{\Omega^*}}([e]))$ and

$$\omega \notin f_{\overrightarrow{\Omega}}(\bigcap_{e\in\mathcal{E}}[\![e]\!]_{\overrightarrow{\Omega}}) = f_{\overrightarrow{\Omega}}([\![\bigwedge \mathcal{E}]\!]_{\overrightarrow{\Omega}}) = D^{-1}(f_{\overrightarrow{\Omega^*}}([\bigwedge \mathcal{E}])) = D^{-1}(f_{\overrightarrow{\Omega^*}}(\bigcap_{e\in\mathcal{E}}[e])).$$

Now, we obtain $D(\omega) \in \bigcap_{e\in\mathcal{E}} f_{\overrightarrow{\Omega^*}}([e])$ and $D(\omega) \notin f_{\overrightarrow{\Omega^*}}(\bigcap_{e\in\mathcal{E}}[e])$.

□

Proof of Lemma 4. For any $\omega^* \in \Omega^*$, there are a knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$ such that $\omega^* = D(\omega)$. Choose such $\overrightarrow{\Omega}$ and ω and define $\Theta^*(\omega^*) := \Theta(\omega)$, where $(0, \Theta(\omega)) \in D(\omega)$.

It can be easily seen that $\Theta^* : \Omega^* \to S$ is well defined (i.e., Θ^* does not depend on a particular choice of $\overrightarrow{\Omega}$ and ω with $\omega^* = D(\omega)$). Indeed, suppose that $\omega^* = D_{\overrightarrow{\Omega}}(\omega) = D_{\overrightarrow{\Omega'}}(\omega')$ for some $\omega \in \Omega$ and $\omega' \in \Omega'$, where $\overrightarrow{\Omega}$ and $\overrightarrow{\Omega'}$ are knowledge spaces. Then, $(0, \Theta(\omega)) = (0, \Theta'(\omega'))$, i.e., $\Theta(\omega) = \Theta'(\omega')$.

Finally, take $E \in \mathcal{A}_S$. We show $(\Theta^*)^{-1}(E) = [E] \in \mathcal{D}^*$ as follows:

$$\begin{aligned}\omega^* \in (\Theta^*)^{-1}(E) &\text{ iff } \Theta^*(\omega^*) = \Theta^*(D(\omega)) \in E \text{ iff } \Theta(\omega) \in E\\ &\text{ iff } \omega \in \Theta^{-1}(E) = [\![E]\!]_{\overrightarrow{\Omega}} \text{ iff } E \in D(\omega) = \omega^* \text{ iff } \omega^* \in [E],\end{aligned}$$

where $\overrightarrow{\Omega}$ and $\omega \in \Omega$ satisfy $\omega^* = D(\omega)$. □

In order to prove Proposition 1, consider the following lemma.

Lemma A.4 (Logical Properties of Each State)**.** *Fix $\omega^* \in \Omega^*$.*

1. *For each $e \in \mathcal{L}$, $(1, e) \notin \omega^*$ iff $(1, (\neg e)) \in \omega^*$.*

2. *For any $\mathcal{E}$ with $\mathcal{E} \subseteq \mathcal{L}$ and $|\mathcal{E}| < \kappa$, we have $(1, \bigwedge \mathcal{E}) \in \omega^*$ iff $(1, e) \in \omega^*$ for all $e \in \mathcal{E}$.*

Proof of Lemma A.4. Fix $\omega^* \in \Omega^*$.

1. Suppose that $(1, e) \in \omega^*$, i.e., $\omega^* \in [e]$. Now, if it were the case that $(1, (\neg e)) \in \omega^*$, then we also have $\omega^* \in [\neg e] = \neg[e]$, a contradiction. Conversely, suppose that $(1, e) \notin \omega^*$. Then, $\omega^* \notin [e]$, i.e., $\omega^* \in \neg[e] = [\neg e]$. Thus, we have $(1, (\neg e)) \in \omega^*$.

2. Let $\mathcal{E}$ be a subset of $\mathcal{L}$ with $|\mathcal{E}| < \kappa$. If $(1, e) \in \omega^*$ (i.e., $\omega^* \in [e]$) for all $e \in \mathcal{E}$, then we obtain $\omega^* \in \bigcap_{e \in \mathcal{E}} [e] = [\bigwedge \mathcal{E}]$. Thus, we have $(1, \bigwedge \mathcal{E}) \in \omega^*$. Conversely, suppose that $(1, \bigwedge \mathcal{E}) \in \omega^*$, i.e., $\omega^* \in [\bigwedge \mathcal{E}] = \bigcap_{e \in \mathcal{E}} [e]$. It is clear that $\omega^* \in [e]$, i.e., $(1, e) \in \omega^*$, for all $e \in \mathcal{E}$.

□

Proof of Proposition 1. Fix $i \in I$.

1. The assertion clearly follows from Lemma A.4.

2. Suppose that $(1, k_i(e)) \notin \omega^*$ and $(1, k_i(\neg e)) \notin \omega^*$. Then, we have $\omega^* \in [(\neg k_i)(e)] \cap [(\neg k_i)(\neg e)] = [(\neg k_i)(e) \wedge (\neg k_i)(\neg e)]$. This implies $(1, (\neg k_i)(e) \wedge (\neg k_i)(\neg e)) \in \omega^*$.

   Next, it is clear that if $(1, k_i(e)) \in \omega^*$ then $(1, (\neg k_i)(e) \wedge (\neg k_i)(\neg e)) \notin \omega^*$. Likewise, $(1, k_i(\neg e)) \in \omega^*$ then $(1, (\neg k_i)(e) \wedge (\neg k_i)(\neg e)) \notin \omega^*$. Also, if $(1, (\neg k_i)(e) \wedge (\neg k_i)(\neg e)) \in \omega^*$ then $(1, k_i(e)) \notin \omega^*$ and $(1, k_i(\neg e)) \notin \omega^*$. Without Consistency, however, it might be possible that $(1, k_i(e)) \in \omega^*$ and $(1, k_i(\neg e)) \in \omega^*$.

   Next, assume that i's knowledge satisfies Consistency. Suppose to the contrary that $(1, k_i(e)) \in \omega^*$ and $(1, k_i(\neg e)) \in \omega^*$. Then, we get $\omega \in K_i^*([e]) \cap K_i^*(\neg[e]) = \emptyset$, a contradiction.

   Conversely, assume that exactly one of the three conditions holds. If $\omega^* \in K_i^*([e])$ then $(1, k_i(e)) \in \omega^*$. Then, we get $(1, k_i(\neg e)) \notin \omega^*$, i.e., $\omega^* \in (\neg K_i^*)([\neg e]) = (\neg K_i^*)([e]^c)$, establishing Consistency.

3. The first assertion clearly follows from Lemma A.4. The second condition follows because we have

$$(1, (\neg k_i)(e) \wedge (\neg k_i)(\neg k_i)(e)) \in \omega^* \text{ iff } \omega^* \in (\neg K_i^*)([e]) \cap (\neg K_i^*)(\neg K_i^*)([e]).$$

□

Proof of Lemma 5. First, we show that there is a unique $s \in S$ such that $\omega^* \cap \{0\} \times S = D(\omega^*) \cap \{0\} \times S = \{(0, s)\}$. Second, in a similar way to HS [39, Lemma 4.6], Meier [55, Lemma 6], and Meier [56, Lemma 4], we show that $[\![e]\!]_{\overrightarrow{\Omega^*}} = [e]$ for every $e \in \mathcal{L}$ by induction on the formation of expressions. Once these assertions are established, we have

$\omega^* = \{s\} \sqcup \{e \in \mathcal{L} \mid (1,e) \in \omega^*\} = \{s\} \sqcup \{e \in \mathcal{L} \mid \omega^* \in [e]\} = \{s\} \sqcup \{e \in \mathcal{L} \mid \omega^* \in [\![e]\!]_{\overrightarrow{\Omega^*}}\} = D(\omega^*)$ for any $\omega^* \in \Omega^*$, where s is a unique element associated with ω^* and $D(\omega^*)$.

In the first step, assume that $(0,s) \in \omega^*$ and $(0,s') \in D(\omega^*)$. There are a knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$ such that $s = \Theta(\omega)$ and $s' = \Theta^*(\omega^*) = \Theta^*(D(\omega))$. Now, since $\Theta(\omega) = \Theta^*(D(\omega))$, we obtain $s = s'$. Note that the argument does not depend on a particular choice of knowledge spaces.

In the second step, we show by induction that $[\![e]\!]_{\overrightarrow{\Omega^*}} = [e]$ for every $e \in \mathcal{L}$. For any $E \in \mathcal{A}_S$, we have the following:

$$\begin{aligned}\omega^* \in [\![E]\!]_{\overrightarrow{\Omega^*}} (:= (\Theta^*)^{-1}(E)) \text{ iff } \Theta^*(\omega^*) \in E \text{ iff } \Theta^*(D(\omega)) = \Theta(\omega) \in E \\ \text{iff } \omega \in \Theta^{-1}(E) = [\![E]\!]_{\overrightarrow{\Omega}} \text{ iff } (1,E) \in D(\omega) \text{ iff } \omega^* = D(\omega) \in [E],\end{aligned}$$

where $\overrightarrow{\Omega}$ and $\omega \in \Omega$ satisfy $\omega^* = D(\omega)$.

Next, let $\mathcal{E}$ be a non-empty set of expressions with $|\mathcal{E}| < \kappa$. Assume the induction hypothesis that $[\![e]\!]_{\overrightarrow{\Omega^*}} = [e]$ for all $e \in \mathcal{E}$. Then, we have $[\![\bigwedge \mathcal{E}]\!]_{\overrightarrow{\Omega^*}} = \bigcap_{e \in \mathcal{E}} [\![e]\!]_{\overrightarrow{\Omega^*}} = \bigcap_{e \in \mathcal{E}} [e] = [\bigwedge \mathcal{E}]$, where the last equality follows from the proof of Lemma 2.

Next, assume the induction hypothesis that $[\![e]\!]_{\overrightarrow{\Omega^*}} = [e]$. Then, by definition, we have $[k_i(e)] = K_i^*([e]) = K_i^*([\![e]\!]_{\overrightarrow{\Omega^*}}) = [\![k_i(e)]\!]_{\overrightarrow{\Omega^*}}$. It can also be seen that $[(\neg e)] = \neg[e] = \neg[\![e]\!]_{\overrightarrow{\Omega^*}} = [\![\neg e]\!]_{\overrightarrow{\Omega^*}}$. The proof is complete. □

Proof of Theorem 1. Recall that we have shown that $\overrightarrow{\Omega^*} = \langle (\Omega^*, \mathcal{D}^*), (K_i^*)_{i \in I}, \Theta^* \rangle$ is a knowledge space of I on $(S, \mathcal{A}_S)$ of the given category (Lemmas 2, 3, and 4) and that for any knowledge space $\overrightarrow{\Omega} = \langle (\Omega, \mathcal{D}), (K_i)_{i \in I}, \Theta \rangle$ of I on $(S, \mathcal{A}_S)$, the description map $D_{\overrightarrow{\Omega}} : \Omega \to \Omega^*$ is a knowledge morphism (Lemmas 2, 3, and 4).

Thus, we need only to show that $D_{\overrightarrow{\Omega}}$ is a unique knowledge morphism from $\overrightarrow{\Omega}$ into $\overrightarrow{\Omega^*}$. Suppose that $\varphi : \Omega \to \Omega^*$ is a knowledge morphism. Then, it follows from Corollary 1 that $D_{\overrightarrow{\Omega}}(\omega) = D_{\overrightarrow{\Omega^*}}(\varphi(\omega))$ for any $\omega \in \Omega$. On the other hand, it follows from Lemma 5 that the description map $D_{\overrightarrow{\Omega^*}} : \Omega^* \to \Omega^*$ is the identity on Ω^*. Thus, we get $D_{\overrightarrow{\Omega}} = \varphi$. The proof is complete. □

Proof of Remark 2. Here, we only prove the second statement as the first one is clear. Fix a knowledge space $\overrightarrow{\Omega}$. We let $\Omega' := D(\Omega)$. We denote by $i_{\Omega'}$ the inclusion map from Ω' into Ω^*. We define the knowledge space $\overrightarrow{\Omega'}$ as follows. First, we let $\mathcal{D}' := \{(i_{\Omega'})^{-1}([e]) \mid [e] \in \mathcal{D}^*\}$. By construction, $(\Omega', \mathcal{D}')$ is a κ-complete algebra. Second, we let $\Theta' = \Theta^*|_{\Omega'} (= \Theta^* \circ i_{\Omega'})$. By construction, we have $(\Theta')^{-1}(E) = (i_{\Omega'})^{-1}((\Theta^*)^{-1}(E)) \in \mathcal{D}'$ for all $E \in \mathcal{A}_S$. Third, we define $K_i'((i_{\Omega'})^{-1}([e])) := (i_{\Omega'})^{-1}(K_i^*([e]))$ for all $[e] \in \mathcal{D}^*$. We show that this is well defined. Suppose that $(i_{\Omega'})^{-1}([e]) = (i_{\Omega'})^{-1}([f])$ for some $e, f \in \mathcal{L}$. If $\omega^* \in (i_{\Omega'})^{-1}(K_i^*([e])) = K_i^*([e]) \cap D(\Omega)$, then there is $\omega \in \Omega$ such that $\omega \in D^{-1}(K_i^*([e])) = K_i(D^{-1}([e])) = K_i(D^{-1}(i_{\Omega'})^{-1}([e])) = K_i(D^{-1}(i_{\Omega'})^{-1}([f]))$. Then, it can be easily seen that $\omega^* \in (i_{\Omega'})^{-1}(K_i^*([f]))$. By changing the role of e and f, we obtain $(i_{\Omega'})^{-1}(K_i^*([e])) = (i_{\Omega'})^{-1}(K_i^*([f]))$. Thus, K_i' is well defined.

Now, we consider $D' : \Omega \to \Omega'$. By construction, it is surjective. Also, we obtain

$$\begin{aligned}(D')^{-1}K_i'((i_{\Omega'})^{-1}([e])) &= (D')^{-1}(i_{\Omega'})^{-1}K_i^*([e]) = D^{-1}(K_i^*([e])) \\ &= K_i(D^{-1}([e])) = K_i((D')^{-1}((i_{\Omega'})^{-1}([e]))).\end{aligned}$$

It follows (i) that $\overrightarrow{\Omega'}$ inherits the properties of knowledge imposed in $\overrightarrow{\Omega}$ (recall Lemma A.3) and (ii) that $D' : \Omega \to \Omega'$ is a knowledge morphism.

By construction, $\overrightarrow{\Omega'}$ is a knowledge subspace of $\overrightarrow{\Omega^*}$, because the inclusion map $i_{\Omega'}$ is a knowledge morphism. □

Proof of Proposition 2.

1. It is sufficient to show that if Φ is satisfiable then it is satisfiable in $\overrightarrow{\Omega^*}$. Suppose that Φ is satisfiable, i.e., there are a knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$ such that $[\![f]\!]_{\overrightarrow{\Omega}} = D^{-1}([f])$ for all $f \in \Phi$. Then, we have $D(\omega) \in [f] = [\![f]\!]_{\overrightarrow{\Omega^*}}$ for all $f \in \Phi$.

2. It is enough to show that $\Phi \models_{\overrightarrow{\Omega^*}} e$ implies $\Phi \models e$. Let $\overrightarrow{\Omega}$ be any knowledge space. If $\omega \in [\![f]\!]_{\overrightarrow{\Omega}} = D^{-1}([f])$ for all $f \in \Phi$, then again $D(\omega) \in [f] = [\![f]\!]_{\overrightarrow{\Omega^*}}$ for all $f \in \Phi$. By assumption, we have $D(\omega) \in [\![e]\!]_{\overrightarrow{\Omega^*}} = [e]$, i.e., $\omega \in D^{-1}([e]) = [\![e]\!]_{\overrightarrow{\Omega}}$. Thus, $\Phi \models e$.

3. This assertion can be seen as a special case of the last assertion. Indeed, suppose that $\Omega^* = [\![e]\!]_{\overrightarrow{\Omega^*}} = [e]$. For any $\overrightarrow{\Omega}$ and $\omega \in \Omega$, we have $D(\omega) \in [e]$ and thus $\omega \in D^{-1}([e]) = [\![e]\!]_{\overrightarrow{\Omega}}$. Hence, $\Omega = [\![e]\!]_{\overrightarrow{\Omega}}$.

□

Proof of Proposition 3. We divide the proof into the two steps.

Step 1. We show that the mapping defined in the statement of the proposition is injective. Suppose that $\Theta^*(\omega^*) = \Theta^*(\tilde{\omega}^*)$ and

$$\{e \in \mathcal{L} \mid \omega^* \in K_i^*([e])\} = \{e \in \mathcal{L} \mid \tilde{\omega}^* \in K_i^*([e])\} \text{ for each } i \in I,$$

where note that $[\cdot] = [\![\cdot]\!]_{\overrightarrow{\Omega^*}}$.

First, we have $(0, \Theta^*(\omega^*)) \in \omega^*$ and $(0, \Theta^*(\tilde{\omega}^*)) \in \tilde{\omega}^*$. Thus, ω^* and $\tilde{\omega}^*$ contain the same unique element of S.

Second, we show by induction that ω^* and $\tilde{\omega}^*$ contain the same set of expressions. For any $E \in \mathcal{A}_S$ with $(1, E) \in \omega^*$ (i.e., $\omega^* \in [E] = (\Theta^*)^{-1}(E)$), we have $\Theta^*(\tilde{\omega}^*) = \Theta^*(\omega^*) \in E$ and thus $(1, E) \in \tilde{\omega}^*$. The converse is also true.

Next, we have $(1, (\neg e)) \in \omega^*$ iff $(1, e) \notin \omega^*$ iff $(1, e) \notin \tilde{\omega}^*$ iff $(1, (\neg e)) \in \tilde{\omega}^*$.

Let $\mathcal{E}$ be a subset of $\mathcal{L}$ with $(0 <)|\mathcal{E}| < \kappa$. Then, $(1, \bigwedge \mathcal{E}) \in \omega^*$ iff $(1, e) \in \omega^*$ for all $e \in \mathcal{E}$ iff $(1, e) \in \tilde{\omega}^*$ for all $e \in \mathcal{E}$ iff $(1, \bigwedge \mathcal{E}) \in \tilde{\omega}^*$.

Fix $i \in I$. We have $(1, k_i(e)) \in \omega^*$ iff $\omega^* \in K_i^*([e])$ iff $\tilde{\omega}^* \in K_i^*([e])$ iff $(1, k_i(e)) \in \tilde{\omega}^*$. Thus, the induction is complete. We have $\omega^* = \tilde{\omega}^*$.

Step 2. Next, we show that the mapping defined in the statement of the proposition is surjective. Let $\overrightarrow{\Omega}$ be a knowledge space, and let $\omega \in \Omega$ be a state. Take any $(\Theta(\omega), (\{e \in \mathcal{L} \mid \omega \in K_i([\![e]\!]_{\overrightarrow{\Omega}})\})_{i \in I})$. Then, it follows from follows from Lemmas 3 and 5 that

$$\begin{aligned}(\Theta(\omega), (\{e \in \mathcal{L} \mid \omega \in K_i([\![e]\!]_{\overrightarrow{\Omega}})\})_{i \in I}) &= (\Theta^*(D(\omega)), (\{e \in \mathcal{L} \mid D(\omega) \in K_i^*([\![e]\!]_{\overrightarrow{\Omega^*}})\})_{i \in I}) \\ &= (\Theta^*(D(\omega)), (\{e \in \mathcal{L} \mid D(\omega) \in K_i^*([e])\})_{i \in I}).\end{aligned}$$

The proof is complete. □

A.2.2 Section 3.2 (Proof of Theorem 2)

Step 1. The proof consists of the following two steps. In the first step, we show that Ω^* satisfies all the properties mentioned in Theorem 2.

1. Fix $\omega^* \in \Omega^*$. Since Conditions (d) to (f) follow from Lemma A.4, we establish (a)-(c).
   (a) There is a unique $s = \Theta^*(\omega^*) \in S$ such that $(0, s) \in \omega^*$. For any $E \in \mathcal{A}_S$ with $s \in E$, we have $\omega^* \in (\Theta^*)^{-1}(E) = [E]$, i.e., $(1, E) \in \omega^*$.
   (b) It is enough to show that each of the following expressions is valid in $\overrightarrow{\Omega^*}$.
      i. Suppose that i's knowledge satisfies No-Contradiction Axiom. Then, it can be easily seen that $[\![\emptyset \leftrightarrow k_i(\emptyset)]\!]_{\overrightarrow{\Omega^*}} = (\neg K_i^*)(\emptyset) = \Omega^*$, and thus $(\emptyset \leftrightarrow k_i(\emptyset))$ is valid in $\overrightarrow{\Omega^*}$.
      ii. If i's knowledge satisfies Consistency, then we have
$$[\![k_i(e) \to (\neg k_i)(\neg e)]\!]_{\overrightarrow{\Omega^*}} = (\neg K_i^*([\![e]\!]_{\overrightarrow{\Omega^*}})) \cup (\neg K_i^*)(\neg [\![e]\!]_{\overrightarrow{\Omega^*}}) = \Omega^*.$$
      iii. Suppose that i's knowledge satisfies Non-empty λ-Conjunction. Take any $\mathcal{E}$ such that $\mathcal{E} \subseteq \mathcal{L}$ and $0 < |\mathcal{E}| < \kappa$. Then, we have
$$\left[\!\!\left[(\bigwedge_{e \in \mathcal{E}} k_i(e)) \to k_i(\bigwedge \mathcal{E})\right]\!\!\right]_{\overrightarrow{\Omega^*}} = \left(\neg \bigcap_{e \in \mathcal{E}} K_i^*([\![e]\!]_{\overrightarrow{\Omega^*}})\right) \cup K_i^*(\bigcap_{e \in \mathcal{E}} [\![e]\!]_{\overrightarrow{\Omega^*}}) = \Omega^*.$$
      iv. If Necessitation is imposed on player i, then $k_i(S)$ is valid because $[\![k_i(S)]\!]_{\overrightarrow{\Omega^*}} = K_i^*([\![S]\!]_{\overrightarrow{\Omega^*}}) = K_i^*(\Omega^*) = \Omega^*$.
      v. If i's knowledge satisfies Truth Axiom, then we have
$$[\![k_i(e) \to e]\!]_{\overrightarrow{\Omega^*}} = (\neg K_i^*)([\![e]\!]_{\overrightarrow{\Omega^*}}) \cup [\![e]\!]_{\overrightarrow{\Omega^*}} = \Omega^*.$$
      vi. If i's knowledge satisfies Positive Introspection, then we have
$$[\![k_i(e) \to k_i k_i(e)]\!]_{\overrightarrow{\Omega^*}} = (\neg K_i^*)([\![e]\!]_{\overrightarrow{\Omega^*}}) \cup K_i^* K_i^*([\![e]\!]_{\overrightarrow{\Omega^*}}) = \Omega.$$

vii. If i's knowledge satisfies Negative Introspection, then we obtain

$$[\![(\neg k_i)(e) \to k_i(\neg k_i)(e)]\!]_{\overrightarrow{\Omega^*}} = K_i^*([\![e]\!]_{\overrightarrow{\Omega^*}}) \cup K_i^*(\neg K_i^*)([\![e]\!]_{\overrightarrow{\Omega^*}}) = \Omega^*.$$

(c) Suppose that $(1, e) \in \omega^*$ and $(1, (e \to f)) \in \omega^*$. Then, we get $\omega^* \in [e]$ and $\omega^* \in [e \to f] = [(\neg e) \vee f] = \neg[e] \cup [f]$. Thus, we must have $\omega^* \in [f]$, i.e., $(1, f) \in \omega^*$.

2. We show that Ω^* satisfies the following three conditions.

(a) If $(e \leftrightarrow f)$ is valid, then we have $[\![e]\!]_{\overrightarrow{\Omega}} = [\![f]\!]_{\overrightarrow{\Omega}}$ for any $\overrightarrow{\Omega}$. Then, we have $K_i([\![e]\!]_{\overrightarrow{\Omega}}) = K_i([\![f]\!]_{\overrightarrow{\Omega}})$, i.e., $[\![k_i(e)]\!]_{\overrightarrow{\Omega}} = [\![k_i(f)]\!]_{\overrightarrow{\Omega}}$. Now, it is easy to see that $[\![k_i(e) \leftrightarrow k_i(f)]\!]_{\overrightarrow{\Omega}} = \Omega$. Since $\overrightarrow{\Omega}$ is arbitrary, $(k_i(e) \leftrightarrow k_i(f))$ is valid.

(b) Suppose that i's knowledge satisfies Monotonicity. The proof is similar to the above one. If $(e \to f)$ is valid, then we have $[\![e]\!]_{\overrightarrow{\Omega}} \subseteq [\![f]\!]_{\overrightarrow{\Omega}}$ for any $\overrightarrow{\Omega}$. Then, we have $K_i([\![e]\!]_{\overrightarrow{\Omega}}) \subseteq K_i([\![f]\!]_{\overrightarrow{\Omega}})$, i.e., $[\![k_i(e)]\!]_{\overrightarrow{\Omega}} \subseteq [\![k_i(f)]\!]_{\overrightarrow{\Omega}}$. Now, it is easy to see that $[\![k_i(e) \to k_i(f)]\!]_{\overrightarrow{\Omega}} = \Omega$. Since $\overrightarrow{\Omega}$ is arbitrary, $(k_i(e) \to k_i(f))$ is valid.

(c) Suppose that $\{f \in \mathcal{L} \mid (1, k_i(f)) \in \omega^*\} \models_{\overrightarrow{\Omega^*}} e$. Since $\omega^* \in K_i^*([\![f]\!]_{\overrightarrow{\Omega^*}})$ iff $(1, k_i(f)) \in \omega^*$, we obtain $\bigcap\{[\![f]\!] \in \mathcal{D}^* \mid \omega^* \in K_i^*([\![f]\!]_{\overrightarrow{\Omega^*}})\} \subseteq [\![e]\!]_{\overrightarrow{\Omega^*}}$. By the Kripke property of $\overrightarrow{\Omega^*}$, we obtain $\omega^* \in K_i^*([\![e]\!]_{\overrightarrow{\Omega^*}}) = [k_i(e)]$, i.e., $(1, k_i(e)) \in \omega^*$.

In sum, Ω^* satisfies all the properties mentioned in Theorem 2.

Step 2. In the second step, we show that $\Omega \subseteq \Omega^*$ for any set Ω satisfying conditions specified in Theorem 2. To that end, we introduce a knowledge structure on Ω and show that the description map $D : \overrightarrow{\Omega} \to \overrightarrow{\Omega^*}$ is an inclusion map.

Step 2.1. We start with defining a κ-complete algebra on Ω. We let $\mathcal{D} := \{[e]_{\overrightarrow{\Omega}} \in \mathcal{P}(\Omega) \mid e \in \mathcal{L}\}$, where $[e]_{\overrightarrow{\Omega}} := \{\omega \in \Omega \mid (1, e) \in \omega\}$. Note the following. First, $[\cdot]_{\overrightarrow{\Omega}} \in \mathcal{D}$ is different from $[\cdot] \in \mathcal{D}^*$. Second, we use the notation $[\cdot]_{\overrightarrow{\Omega}}$ rather than $[\cdot]_\Omega$ although we have not introduced a knowledge structure to Ω.

We show that $(\Omega, \mathcal{D})$ is a κ-complete algebra. First, we show that $[\emptyset]_{\overrightarrow{\Omega}} = \emptyset$. Suppose to the contrary that $\omega \in [\emptyset]_{\overrightarrow{\Omega}}$. Then, by definition, $(1, \emptyset) \in \omega$, which is impossible. Hence, $\emptyset = [\emptyset]_{\overrightarrow{\Omega}} \in \mathcal{D}$. Second, it is immediate that $[S]_{\overrightarrow{\Omega}} = \Omega$ and thus $\Omega = [S]_{\overrightarrow{\Omega}} \in \mathcal{D}$.

Third, we show that $[\neg e]_{\overrightarrow{\Omega}} = \neg[e]_{\overrightarrow{\Omega}}$. Indeed, if $\omega \in [\neg e]_{\overrightarrow{\Omega}}$ then $(1, (\neg e)) \in \omega$ and thus $(1, e) \notin \omega$. Hence, $\omega \in \neg[e]_{\overrightarrow{\Omega}}$. The converse is similarly established. Hence, $\mathcal{D}$ is closed under complementation.

Fourth, we show that $[\bigwedge \mathcal{E}]_{\overrightarrow{\Omega}} = \bigcap_{e \in \mathcal{E}} [e]_{\overrightarrow{\Omega}}$ for any $\mathcal{E}$ with $\mathcal{E} \subseteq \mathcal{L}$ and $(0 <)|\mathcal{E}| < \kappa$. If $\omega \in \bigcap_{e \in \mathcal{E}} [e]_{\overrightarrow{\Omega}}$, then $(1, e) \in \omega$ for all $e \in \mathcal{E}$. Hence, we obtain $(1, \bigwedge \mathcal{E}) \in \omega$, i.e., $\omega \in [\bigwedge \mathcal{E}]_{\overrightarrow{\Omega}}$. Conversely, suppose that $\omega \in [\bigwedge \mathcal{E}]_{\overrightarrow{\Omega}}$, i.e., $(1, \bigwedge \mathcal{E}) \in \omega$. Then, we have $(1, e) \in \omega$ (i.e., $\omega \in [e]_{\overrightarrow{\Omega}}$) for all $e \in \mathcal{E}$, that is, $\omega \in \bigcap_{e \in \mathcal{E}} [e]_{\overrightarrow{\Omega}}$. This completes the proof of the fact that

$(\Omega, \mathcal{D})$ is a κ-complete algebra.

Step 2.2. Next, we define the mapping $\Theta : \Omega \to S$ which associates, with each state of the world ω, the unique state of nature $s \in S$ with $(0, s) \in \omega$. Then, the mapping $\Theta : \Omega \to S$ is a well-defined κ-measurable map such that $(\Theta)^{-1}(E) = [E]_{\overrightarrow{\Omega}}$ for each $E \in \mathcal{A}_S$.

Indeed, take $E \in \mathcal{A}_S$. If $\omega \in [E]_{\overrightarrow{\Omega}}$, then $(1, E) \in \omega$. Hence, $\Theta(\omega) \in E$, i.e., $\omega \in \Theta^{-1}(E)$. Conversely, if $\omega \in \Theta^{-1}(E)$ then $\Theta(\omega) \in E$, and thus $(1, E) \in \omega$. Hence, $\omega \in [E]_{\overrightarrow{\Omega}}$.

Step 2.3. Next, we define players' knowledge operators on $(\Omega, \mathcal{D})$. Fix $i \in I$. We let $K_i : \mathcal{D} \to \mathcal{D}$ be such that $K_i([e]_{\overrightarrow{\Omega}}) := [k_i(e)]_{\overrightarrow{\Omega}}$ for each $[e] \in \mathcal{D}$.

We first show that each K_i is well defined. Suppose that $[e]_{\overrightarrow{\Omega}} = [f]_{\overrightarrow{\Omega}}$. Then, it is clear that $[(e \leftrightarrow f)]_{\overrightarrow{\Omega}} = \Omega$. This implies that $[(k_i(e) \leftrightarrow k_i(f))]_{\overrightarrow{\Omega}} = \Omega$. Thus, we obtain $[k_i(e)] = [k_i(f)]$.

Second, we show below that K_i inherits the properties of knowledge imposed in the given category.

1. Suppose that i's knowledge satisfies No-Contradiction Axiom in the given category. If $\omega \in K_i([\emptyset]_{\overrightarrow{\Omega}}) = [k_i(\emptyset)]_{\overrightarrow{\Omega}}$, then $(1, k_i(\emptyset)) \in \omega$. Since $(\emptyset \leftrightarrow k_i(\emptyset)) \in \omega$, it follows that $(1, \emptyset) \in \omega$, which is impossible.

2. Consider Consistency of i's knowledge. If $\omega \in K_i([e]_{\overrightarrow{\Omega}}) = [k_i(e)]_{\overrightarrow{\Omega}}$ then $(1, k_i(e)) \in \omega$. Now, since $(1, (k_i(e) \to (\neg k_i)(\neg e))) \in \omega$, it follows that $(1, (\neg k_i)(\neg e))) \in \omega$, i.e., $\omega \in [(\neg k_i)(\neg e)]_{\overrightarrow{\Omega}} = (\neg K_i)(\neg [e]_{\overrightarrow{\Omega}})$.

3. Consider Necessitation of i's knowledge. We have $\Omega = [k_i(S)]_{\overrightarrow{\Omega}} = K_i([S]_{\overrightarrow{\Omega}}) = K_i(\Omega)$.

4. Consider Monotonicity of i's knowledge. Take $[e]_{\overrightarrow{\Omega}}, [f]_{\overrightarrow{\Omega}} \in \mathcal{D}$ with $[e]_{\overrightarrow{\Omega}} \subseteq [f]_{\overrightarrow{\Omega}}$. Then, observe that $[e \to f]_{\overrightarrow{\Omega}} = \Omega$. Now, we obtain $[k_i(e) \to k_i(f)]_{\overrightarrow{\Omega}} = \Omega$, i.e., $[k_i(e)]_{\overrightarrow{\Omega}} \subseteq [k_i(f)]_{\overrightarrow{\Omega}}$. Thus, we have $K_i([e]_{\overrightarrow{\Omega}}) \subseteq K_i([f]_{\overrightarrow{\Omega}})$.

5. Suppose that i's knowledge satisfies Non-empty λ-Conjunction in the given category. Let $\mathcal{E}$ be such that $\mathcal{E} \subseteq \mathcal{L}$ and $0 < |\mathcal{E}| < \kappa$. If $\omega \in \bigcap_{e \in \mathcal{E}} K_i([e]_{\overrightarrow{\Omega}}) = [\bigwedge_{e \in \mathcal{E}} k_i(e)]_{\overrightarrow{\Omega}}$ then $(1, \bigwedge_{e \in \mathcal{E}} k_i(e)) \in \omega$. Now, since $(1, (\bigwedge_{e \in \mathcal{E}} k_i(e) \to k_i(\bigwedge \mathcal{E})) \in \omega$, it follows that $(1, k_i(\bigwedge \mathcal{E})) \in \omega$, i.e., $\omega \in [k_i(\bigwedge \mathcal{E})]_{\overrightarrow{\Omega}} = K_i(\bigcap_{e \in \mathcal{E}} [e]_{\overrightarrow{\Omega}})$.

6. Consider Truth Axiom of i's knowledge. If $\omega \in K_i([e]_{\overrightarrow{\Omega}}) = [k_i(e)]_{\overrightarrow{\Omega}}$ then $(1, k_i(e)) \in \omega$. Now, since $(1, (k_i(e) \to e) \in \omega$, it follows that $(1, e) \in \omega$, i.e., $\omega \in [e]_{\overrightarrow{\Omega}}$.

7. Suppose that i's knowledge satisfies Positive Introspection in the given category. If $\omega \in K_i([e]_{\overrightarrow{\Omega}}) = [k_i(e)]_{\overrightarrow{\Omega}}$ then $(1, k_i(e)) \in \omega$. Now, since $(1, (k_i(e) \to k_i k_i(e))) \in \omega$, it follows that $(1, k_i k_i(e)) \in \omega$, i.e., $\omega \in [k_i k_i(e)]_{\overrightarrow{\Omega}} = K_i K_i([e]_{\overrightarrow{\Omega}})$.

8. Consider Negative Introspection of i's knowledge. If $\omega \in (\neg K_i)([e]_{\overrightarrow{\Omega}}) = [(\neg k_i)(e)]_{\overrightarrow{\Omega}}$ then $(1, (\neg k_i)(e)) \in \omega$. Now, since $(1, ((\neg k_i)(e) \to k_i(\neg k_i)(e))) \in \omega$, it follows that $(1, k_i(\neg k_i)(e)) \in \omega$, i.e., $\omega \in [k_i(\neg k_i)(e)]_{\overrightarrow{\Omega}} = K_i(\neg K_i)([e]_{\overrightarrow{\Omega}})$.

9. Suppose that the Kripke property is imposed on i in the given category. It follows from the assumption that $\omega \in K_i([e]_{\overrightarrow{\Omega}})$ for any $(\omega, [e]_{\overrightarrow{\Omega}}) \in \Omega \times \mathcal{D}$ such that $\omega \in \bigcap\{[f]_{\overrightarrow{\Omega}} \in \mathcal{D} \mid \omega \in K_i([f]_{\overrightarrow{\Omega}})\}$.

Step 2.4. The above arguments establish that $\overrightarrow{\Omega} := \langle(\Omega, \mathcal{D}), (K_i)_{i\in I}, \Theta\rangle$ is a knowledge space of the given category. Finally, we show that the description map $D : \overrightarrow{\Omega} \to \overrightarrow{\Omega^*}$ is an inclusion map so that $\Omega \subseteq \Omega^*$.

To that end, we first establish by induction that $[\cdot]_{\overrightarrow{\Omega}} : \mathcal{L} \to \mathcal{D}$, viewed as a mapping, coincides with the semantic interpretation function $[\![\cdot]\!]_{\overrightarrow{\Omega}}$. Fix $E \in \mathcal{A}_S$. We have $\omega \in [\![E]\!]_{\overrightarrow{\Omega}} = \Theta^{-1}(E)$ iff $\Theta(\omega) \in E$ iff $\omega \in [E]_{\overrightarrow{\Omega}}$.

Next, suppose that $[\![e]\!]_{\overrightarrow{\Omega}} = [e]_{\overrightarrow{\Omega}}$ for some e. Then,

$$\omega \in [\![\neg e]\!]_{\overrightarrow{\Omega}} = \neg[\![e]\!]_{\overrightarrow{\Omega}} \text{ iff } \omega \notin [\![e]\!]_{\overrightarrow{\Omega}} \text{ iff } \omega \notin [e]_{\overrightarrow{\Omega}} \text{ iff } \omega \in [\neg e]_{\overrightarrow{\Omega}}.$$

Next, suppose that $[\![e]\!]_{\overrightarrow{\Omega}} = [e]_{\overrightarrow{\Omega}}$ for all $e \in \mathcal{E}$ with $\mathcal{E} \subseteq \mathcal{L}$ and $(0 <)|\mathcal{E}| < \kappa$. Then,

$$\omega \in \left[\!\!\left[\bigwedge \mathcal{E}\right]\!\!\right]_{\overrightarrow{\Omega}} = \bigcap_{e\in\mathcal{E}} [\![e]\!]_{\overrightarrow{\Omega}} \text{ iff } \omega \in \bigcap_{e\in\mathcal{E}} [e]_{\overrightarrow{\Omega}} = \left[\bigwedge \mathcal{E}\right]_{\overrightarrow{\Omega}}.$$

Next, assuming the induction hypothesis, we have $[\![k_i(e)]\!]_{\overrightarrow{\Omega}} = K_i([\![e]\!]_{\overrightarrow{\Omega}}) = K_i([e]_{\overrightarrow{\Omega}}) = [k_i(e)]_{\overrightarrow{\Omega}}$. The induction is complete.

Now, we show that $D(\omega) = \omega$ for all $\omega \in \Omega$. First, we consider expressions. If $(1, e) \in D(\omega)$ then $D(\omega) \in [e]$, and thus $\omega \in D^{-1}([e]) = [\![e]\!]_{\overrightarrow{\Omega}} = [e]_{\overrightarrow{\Omega}}$. Then, $(1, e) \in \omega$. Conversely, if $(1, e) \in \omega$ then $\omega \in [e]_{\overrightarrow{\Omega}} = [\![e]\!]_{\overrightarrow{\Omega}} = D^{-1}([e])$. Thus, $D(\omega) \in [e]$ and hence $(1, e) \in D(\omega)$.

Second, if $(0, s) \in \omega$ then $s = \Theta(\omega) = \Theta^*(D(\omega))$, and thus $(0, s) \in D(\omega)$. Conversely, if $(0, s) \in D(\omega)$ then $s = \Theta^*(D(\omega)) = \Theta(\omega)$, and thus $(0, s) \in \omega$. This completes the proof of the statement that D is an inclusion map, and the proof of Theorem 2 is now complete. $\square$

A.2.3 Section 3.3

Proof of Proposition 4. Part 1. Let $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega'}$ be a knowledge morphism between κ-knowledge spaces $\overrightarrow{\Omega}$ and $\overrightarrow{\Omega'}$. In a similar way to HS [38], we show by induction that $\mathcal{C}_\alpha = \varphi^{-1}(\mathcal{C}'_\alpha)$ for all α.

Let $\alpha = 0$. For any $E \in \mathcal{A}_S$, we have $(\Theta)^{-1}(E) = \varphi^{-1}((\Theta')^{-1}(E))$. Thus, $\mathcal{C}_0 = \varphi^{-1}(\mathcal{C}'_0)$. Now, suppose that $\mathcal{C}_\beta = \varphi^{-1}(\mathcal{C}'_\beta)$ for all $\beta < \alpha$. Then, $\mathcal{C}_\alpha = \varphi^{-1}(\mathcal{C}'_\alpha)$ follows because the set-algebraic operations and the knowledge operators commute with the inverse φ^{-1}. Indeed, we have

$$\begin{aligned}
\mathcal{C}_\alpha &= \mathcal{A}_\kappa\left(\{\varphi^{-1}(E') \in \mathcal{D} \mid E' \in \bigcup_{\beta<\alpha}\mathcal{C}'_\beta\} \cup \bigcup_{i\in I}\{K_i(\varphi^{-1}(E')) \in \mathcal{D} \mid E' \in \bigcup_{\beta<\alpha}\mathcal{C}'_\beta\}\right) \\
&= \mathcal{A}_\kappa\left(\{\varphi^{-1}(E') \in \mathcal{D} \mid E' \in \bigcup_{\beta<\alpha}\mathcal{C}'_\beta\} \cup \bigcup_{i\in I}\{\varphi^{-1}(K'_i(E')) \in \mathcal{D} \mid E' \in \bigcup_{\beta<\alpha}\mathcal{C}'_\beta\}\right) \\
&= \varphi^{-1}\left(\mathcal{A}_\kappa\left(\{E' \in \mathcal{D} \mid E' \in \bigcup_{\beta<\alpha}\mathcal{C}'_\beta\} \cup \bigcup_{i\in I}\{K'_i(E') \in \mathcal{D} \mid E' \in \bigcup_{\beta<\alpha}\mathcal{C}'_\beta\}\right)\right) = \varphi^{-1}(\mathcal{C}'_\alpha).
\end{aligned}$$

Thus, if $\mathcal{C}'_\alpha = \mathcal{C}'_{\alpha+1}$, then $\mathcal{C}_\alpha = \mathcal{C}_{\alpha+1}$. In other words, if the κ-rank of $\overrightarrow{\Omega'}$ is α then that of $\overrightarrow{\Omega}$ is at most α.

Part 2. Fix a κ-knowledge space $\overrightarrow{\Omega}$. We show that $\mathcal{D}_\alpha = \mathcal{C}_\alpha$ for all $\alpha \leq \kappa$, where $\mathcal{D}_\alpha$ is defined in the main text. For $\alpha = 0$, it is clear that $\mathcal{D}_0 = \{\Theta^{-1}(E) \in \mathcal{D} \mid E \in \mathcal{A}_S\} = \mathcal{C}_0$. Next, suppose that $\mathcal{D}_\beta = \mathcal{C}_\beta$ for all $\beta < \alpha$. Then, it can be seen that

$$\mathcal{D}_\alpha = \mathcal{A}_\kappa\left(\left(\bigcup_{\beta<\alpha}\mathcal{D}_\beta\right) \cup \bigcup_{i\in I}\{K_i(\llbracket e \rrbracket) \in \mathcal{D} \mid \llbracket e \rrbracket \in \bigcup_{\beta<\alpha}\mathcal{D}_\beta\}\right) = \mathcal{C}_\alpha.$$

Now, we have $\mathcal{C}_\kappa = \mathcal{D}_\kappa = \{\llbracket e \rrbracket_{\overrightarrow{\Omega}} \in \mathcal{D} \mid e \in \mathcal{L}\}$. Then, we obtain $\mathcal{C}_\kappa = \mathcal{C}_{\kappa+1}$, and the κ-rank of $\overrightarrow{\Omega}$ is at most κ. □

A.3 Section 4

A.3.1 Section 4.1

Proof of Proposition 5. Part 1. First, since $\emptyset \subseteq K_i(\emptyset)$, it is immediate that $\emptyset \in \mathcal{J}_{K_i}$.

Second, we show that $K_i(E) = \max\{F \in \mathcal{D} \mid F \in \mathcal{J}_{K_i} \text{ and } F \subseteq E\} \in \mathcal{J}_{K_i}$ for each $E \in \mathcal{D}$ if K_i satisfies Truth Axiom, Monotonicity and Positive Introspection. Once this equality is established, $\mathcal{J}_{K_i}$ clearly satisfies the maximality property.

Fix $E \in \mathcal{D}$. For any $F \in \mathcal{J}_{K_i}$ with $F \subseteq E$, we have $F \subseteq K_i(F) \subseteq K_i(E)$, where the second set inclusion follows from Monotonicity. On the other hand, Positive Introspection and Truth Axiom imply $K_i(E) \in \{F \in \mathcal{J}_{K_i} \mid F \subseteq E\}$, yielding the desired equality $K_i(E) = \max\{F \in \mathcal{J}_{K_i} \mid F \subseteq E\}$.

Third, assume Non-empty λ-Conjunction of K_i. Let $\mathcal{E}$ be a non-empty subset of $\mathcal{J}_{K_i}$ with $|\mathcal{E}| < \lambda$. We have $\bigcap \mathcal{E} \subseteq \bigcap_{E\in\mathcal{E}} K_i(E) \subseteq K_i(\bigcap \mathcal{E})$, implying $\bigcap \mathcal{E} \in \mathcal{J}_{K_i}$.

Fourth, it is clear that Necessitation implies $\Omega \in \mathcal{J}_{K_i}$, since $K_i(\Omega) = \Omega$.

Fifth, we show that Negative Introspection, together with Truth Axiom, imply that $\mathcal{J}_{K_i}$ is closed under complementation. If $E \in \mathcal{J}_{K_i}$ then $E = K_i(E)$, which implies $E^c = (\neg K_i)(E)$. It follows from Negative Introspection that $E^c = (\neg K_i)(E) \subseteq K_i(\neg K_i)(E) =$

$K_i(E^c)$, from which we obtain $E^c \in \mathcal{J}_{K_i}$.

Part 2. Let $\mathcal{J}_i$ be a sub-collection of $\mathcal{D}$ satisfying the maximality property. First, we show that $K_{\mathcal{J}_i}$ satisfies Truth Axiom, Positive Introspection, and Monotonicity. We obtain Truth Axiom of $K_{\mathcal{J}_i}$ because $K_{\mathcal{J}_i}(E) = \max\{F \in \mathcal{J}_i \mid F \subseteq E\} \subseteq E$ for any $E \in \mathcal{D}$. Next, since $K_{\mathcal{J}_i}(E) \in \mathcal{J}_i$, we have $K_{\mathcal{J}_i}(E) \subseteq \max\{F \in \mathcal{J}_i \mid F \subseteq K_{\mathcal{J}_i}(E)\} = K_{\mathcal{J}_i}K_{\mathcal{J}_i}(E)$, establishing Positive Introspection. Finally, if $E \subseteq F$ then $K_{\mathcal{J}_i}(E)(\subseteq E) \subseteq F$. Since $K_{\mathcal{J}_i}(E) \in \{F' \in \mathcal{J}_i \mid F' \subseteq F\}$, we get $K_{\mathcal{J}_i}(E) \subseteq \max\{F' \in \mathcal{J}_i \mid F' \subseteq F\} = K_{\mathcal{J}_i}(F)$, i.e., Monotonicity is established.

Second, assume that $\mathcal{J}_i$ is closed under non-empty λ-intersection. Let $\mathcal{E}$ be a non-empty subset of $\mathcal{D}$ with $|\mathcal{E}| < \lambda$. Note that by the previous argument, $K_{\mathcal{J}_i}$ satisfies Truth Axiom and Monotonicity. It follows from Truth Axiom that $\bigcap_{E\in\mathcal{E}} K_{\mathcal{J}_i}(E) \subseteq \bigcap \mathcal{E}$. Now, $\mathcal{J}_i$ being closed under non-empty λ-intersection, we get $\bigcap_{E\in\mathcal{E}} K_{\mathcal{J}_i}(E) \in \mathcal{J}_i$, and thus $\bigcap_{E\in\mathcal{E}} K_{\mathcal{J}_i}(E) \subseteq K_{\mathcal{J}_i}(\bigcap_{E\in\mathcal{E}} K_{\mathcal{J}_i}(E)) \subseteq K_{\mathcal{J}_i}(\bigcap \mathcal{E})$, where the second set inclusion follows from Monotonicity. Hence, Non-empty λ-Conjunction is established.

Third, if $\Omega \in \mathcal{J}_i$, then $\Omega = \max\{E \in \mathcal{J}_i \mid E \subseteq \Omega\} = K_{\mathcal{J}_i}(\Omega)$, establishing Necessitation.

Fourth, assume that $\mathcal{J}_i$ is closed under complementation. Since $K_{\mathcal{J}_i}(E) \in \mathcal{J}_i$, we have $(\neg K_{\mathcal{J}_i})(E) \in \mathcal{J}_i$ and thus $(\neg K_{\mathcal{J}_i})(E) \subseteq \max\{F \in \mathcal{J}_i \mid F \subseteq (\neg K_{\mathcal{J}_i})(E)\} = K_{\mathcal{J}_i}(\neg K_{\mathcal{J}_i})(E)$.

Part 3. First, $K_i = K_{\mathcal{J}_{K_i}}$ simply follows from Equation (3). Next, we show that $\mathcal{J}_i = \mathcal{J}_{K_{\mathcal{J}_i}}$. If $E \in \mathcal{J}_i$, then we obtain $E \subseteq \max\{F \in \mathcal{J}_i \mid F \subseteq E\} = K_{\mathcal{J}_i}(E)$ and thus $E \in \mathcal{J}_{K_{\mathcal{J}_i}}$, establishing $\mathcal{J}_i \subseteq \mathcal{J}_{K_{\mathcal{J}_i}}$. Conversely, if $E \in \mathcal{J}_{K_{\mathcal{J}_i}}$, then we have $E = K_{\mathcal{J}_i}(E) = \max\{F \in \mathcal{J}_i \mid F \subseteq E\} \in \mathcal{J}_i$ by the maximality property of $\mathcal{J}_i$. Thus, we have $\mathcal{J}_{K_{\mathcal{J}_i}} \subseteq \mathcal{J}_i$.

Part 4. Let κ be an infinite cardinal or $\kappa = \infty$. We show that the maximality property implies that $\mathcal{J}_{K_i}$ is closed under κ-union, where ∞-union refers to arbitrary union. Recall that the maximality property, by definition, implies $\emptyset \in \mathcal{J}_{K_i}$. Thus, $\mathcal{J}_{K_i}$ is closed under the empty union. Now, let $\mathcal{E}$ be any non-empty subset of $\mathcal{J}_{K_i}$ with $|\mathcal{E}| < \kappa$. For any $E \in \mathcal{E}$, we have $E \subseteq K_i(E) \subseteq K_i(\bigcup \mathcal{E})$, and thus $\bigcup \mathcal{E} \subseteq K_i(\bigcup \mathcal{E})$, establishing $\bigcup \mathcal{E} \in \mathcal{J}_{K_i}$. Indeed, we have shown that, for any $\mathcal{E} \subseteq \mathcal{J}_{K_i}$, if $\bigcup \mathcal{E} \in \mathcal{D}$ then $\bigcup \mathcal{E} \in \mathcal{J}_{K_i}$.

We establish the converse when $\kappa = \infty$. If $\mathcal{J}_{K_i}$ is closed under arbitrary union, then, for each $E \in \mathcal{D}$, we have

$$\max\{F \in \mathcal{D} \mid F \in \mathcal{J}_{K_i} \text{ and } F \subseteq E\} = \bigcup \{F \in \mathcal{D} \mid F \in \mathcal{J}_{K_i} \text{ and } F \subseteq E\} \in \mathcal{J}_{K_i}.$$

Thus, $\mathcal{J}_{K_i}$ satisfies the maximality property. □

Proof of Corollary 2. Note that Truth Axiom and Negative Introspection imply Positive Introspection (again, see, for example, Aumann [5, p. 270]). By Proposition 5, the knowledge operator K_i satisfies κ-Conjunction (including Necessitation) iff $\mathcal{J}_{K_i}$ is closed under κ-intersection. Now, it follows again from Proposition 5 that $\mathcal{J}_{K_i}$ is closed under κ-union and complementation. Thus, it is closed under κ-intersection. □

Remark A.4 (Remark on Proposition 5: Counterexamples without Truth Axiom, Monotonicity, or Positive Introspection)**.** Throughout the following three examples, let $\Omega = \{\omega_1, \omega_2\}$ and $\mathcal{D} = \mathcal{P}(\Omega)$. First, we define K_i and K_i' as follows: $K_i(\emptyset) = \emptyset$, $K_i(E) = \{\omega_1\}$ for any $E \in \mathcal{P}(\Omega) \setminus \{\emptyset\}$; and $K_i'(\emptyset) = K_i'(\{\omega_2\}) = \emptyset$, $K_i'(\{\omega_1\}) = K_i'(\Omega) = \{\omega_1\}$. The operator K_i' violates Truth Axiom. It is clear that $K_i \neq K_i'$ but $\mathcal{J}_{K_i} = \mathcal{J}_{K_i'}$. As an additional remark, while any operator K_i satisfying Positive Introspection and Monotonicity also trivially satisfies the equality $K_i(E) = \max\{F \in \mathcal{D} \mid F \in \mathcal{J}_{K_i} \text{ and } F \subseteq K_i(E)\}$, there may be multiple operators which have the same self-evident collections due to the lack of Truth Axiom.

Second, let K_i and K_i' be defined as follows: $K_i(\{\omega_1\}) = K_i(\Omega) = \{\omega_1\}$, $K_i(E) = \emptyset$ for any $E \in \{\emptyset, \{\omega_2\}\}$; and $K_i'(\{\omega_1\}) = \{\omega_1\}$ and $K_i'(E) = \emptyset$ for any $E \in \mathcal{D} \setminus \{\{\omega_1\}\}$. The operator K_i' violates Monotonicity. It is clear that $K_i \neq K_i'$ but $\mathcal{J}_{K_i} = \mathcal{J}_{K_i'}$.

Third, we define K_i and K_i' as follows: $K_i(E) = \emptyset$ for any $E \in \mathcal{D}$; and $K_i'(\Omega) = \{\omega_1\}$ and $K_i'(E) = \emptyset$ for any $E \in \mathcal{D} \setminus \{\Omega\}$. The operator K_i' violates Positive Introspection. It is clear that $K_i \neq K_i'$ but $\mathcal{J}_{K_i} = \mathcal{J}_{K_i'}$.

Proof of Proposition 6.

1. First, it is clear that $\mathcal{D}' \subseteq \mathcal{J}(\mathcal{D}')$. Second, we show that $\mathcal{J}(\mathcal{D}')$ satisfies the maximality property. It is clear that $\emptyset \in \mathcal{J}(\mathcal{D}')$. Fix $E \in \mathcal{D}$. It follows from the definition that
$$\begin{aligned}&\{F \in \mathcal{D} \mid F \in \mathcal{J}(\mathcal{D}') \text{ and } F \subseteq E\}\\ &= \{F \in \mathcal{D} \mid \text{if } \omega \in F \text{ then there is } F' \in \mathcal{D}' \text{ with } \omega \in F' \subseteq F \subseteq E\}.\end{aligned}$$
Then, it can be seen that
$$\begin{aligned}&\max\{F \in \mathcal{D} \mid F \in \mathcal{J}(\mathcal{D}') \text{ and } F \subseteq E\}\\ &= \{\omega \in \Omega \mid \text{there is } F' \in \mathcal{D}' \text{ with } \omega \in F' \subseteq E\} \in \mathcal{J}(\mathcal{D}').\end{aligned}$$
Third, take any collection $\mathcal{J}'$ containing $\mathcal{D}'$ and satisfying the maximality property. Take any $E \in \mathcal{J}(\mathcal{D}')$. Then, for any $\omega \in E$, there is $F \in \mathcal{D}'$ with $\omega \in F \subseteq E$. Since $\mathcal{D}' \subseteq \mathcal{J}'$, we have $F \in \mathcal{J}'$. Now, we obtain
$$E = \{\omega \in \Omega \mid \text{there is } F \in \mathcal{J}' \text{ such that } \omega \in F \subseteq E\} = \max\{F \in \mathcal{J}' \mid F \subseteq E\} \in \mathcal{J}'.$$
Thus, we get $\mathcal{J}(\mathcal{D}') \subseteq \mathcal{J}'$.

2. First, we have $\mathcal{J}(\mathcal{D}') \cup \{\Omega\} \subseteq \mathcal{J}(\mathcal{D}' \cup \{\Omega\})$. Second, it can be easily seen that $\mathcal{J}(\mathcal{D}') \cup \{\Omega\}$ satisfies the maximality property and contains $\mathcal{D}' \cup \{\Omega\}$, leading to $\mathcal{J}(\mathcal{D}' \cup \{\Omega\}) \subseteq \mathcal{J}(\mathcal{D}') \cup \{\Omega\}$.

3. It is enough to prove the second assertion. By construction, $\mathcal{J}_{\lambda\text{-Con}}$ contains $\mathcal{D}$, is closed under non-empty λ-intersection, and satisfies the maximality property. For any collection $\mathcal{J}'$ containing $\mathcal{D}'$ and satisfying the maximality property and non-empty λ-Conjunction, we have
$$\{E \in \mathcal{D} \mid E = \bigcap \mathcal{E} \text{ for some } \mathcal{E} \subseteq \mathcal{D}' \text{ with } 0 < |\mathcal{E}| < \lambda\} \subseteq \mathcal{J}',$$
establishing $\mathcal{J}_{\lambda\text{-Con}}(\mathcal{D}') \subseteq \mathcal{J}'$.

4. The following counterexample shows that $\mathcal{J}(\mathcal{D}')$ is not necessarily closed under complementation even if $\mathcal{D}'$ is. Let $(\Omega, \mathcal{D}) = (\{\omega_1, \omega_2, \omega_3\}, \mathcal{P}(\Omega))$. We consider $\mathcal{D}' = \{\emptyset, \{\omega_1\}, \{\omega_3\}, \{w_1, \omega_2\}, \{\omega_2, \omega_3\}, \Omega\}$. Then, we have $\mathcal{J}(\mathcal{D}') = \mathcal{P}(\Omega) \setminus \{\{\omega_2\}\}$. While $\mathcal{D}'$ is closed under complementation, $\mathcal{J}(\mathcal{D}')$ is not.

   Let α be an ordinal which satisfies the condition specified in the statement of the proposition. We show that $\mathcal{C}_\alpha$ is the desired collection. First, it is clear that $\mathcal{D}' \subseteq \mathcal{C}_\alpha$. Second, if $E \in \mathcal{C}_\alpha$, then $E^c \in \mathcal{C}_{\alpha+1} = \mathcal{C}_\alpha$. Thus, $\mathcal{C}_\alpha$ is closed under complementation.

   Third, we show that $\mathcal{C}_\alpha$ satisfies the maximality property. For any $E \in \mathcal{D}$, we have

   $$\{\omega \in \Omega \mid \text{if } \omega \in E \text{ then there is } F \in \mathcal{C}_\alpha \text{ such that } \omega \in F \subseteq E\} \in \mathcal{J}(\mathcal{C}_\alpha) = \mathcal{C}_\alpha,$$

   where the last equality follows because $\mathcal{C}_\alpha \subseteq \mathcal{J}(\mathcal{C}_\alpha) \subseteq \mathcal{C}_{\alpha+1} = \mathcal{C}_\alpha$. Note that the above set is the largest event in $\mathcal{C}_\alpha$ which is contained in E. Thus, $\mathcal{C}_\alpha$ satisfies the maximality property.

   Fourth, let $\mathcal{J}'$ be a collection with the following properties: (i) $\mathcal{D}' \subseteq \mathcal{J}'$; (ii) $\mathcal{J}'$ is closed under complementation; and (iii) $\mathcal{J}'$ satisfies the maximality property. We show that $\mathcal{C}_\alpha \subseteq \mathcal{J}'$. By assumption, $\mathcal{C}_0 \subseteq \mathcal{J}'$. If $\mathcal{C}_\beta \subseteq \mathcal{J}'$, then it is clear that $\mathcal{C}_{\beta+1} \subseteq \mathcal{J}'$. If $\mathcal{C}_\beta \subseteq \mathcal{J}'$ for all $\beta < \gamma$, then it is clear that $\bigcup_{\beta<\gamma} \mathcal{C}_\beta \subseteq \mathcal{J}'$. Hence, $\mathcal{C}_\alpha \subseteq \mathcal{J}'$.

5. It follows from Proposition 5 that the maximality property is characterized as the closure under arbitrary union. Then, the assertions clearly hold.

□

Proof of Corollary 3. The first condition clearly implies the second. Conversely, if $F \in \mathcal{J}(\mathcal{E})$ and $\omega \in F$ then it follows from Equation (4) that there is $E \in \mathcal{E}$ with $\omega \in E \subseteq F$. Now, by the second condition, there is $E' \in \mathcal{E}'$ such that $\omega \in E' \subseteq E \subseteq F$. Since $\omega \in F$ is arbitrary, we have $F \in \mathcal{J}(\mathcal{E}')$. □

Finally, since a knowledge space can be equivalently defined by players' self-evident collections instead of their knowledge operators, we express the condition for a mapping to preserve players' knowledge (i.e., Condition (2) in Definition 3) solely in terms of players' self-evident collections when players' knowledge satisfy Truth Axiom, Positive Introspection, and Monotonicity.

Proposition A.3 (Characterzing Knowledge Morphisms by Self-Evident Collections)**.** *Let $\overrightarrow{\Omega}$ and $\overrightarrow{\Omega'}$ be knowledge spaces such that player i's knowledge operator satisfies Truth Axiom, Positive Introspection, and Monotonicity. Let $\varphi : \Omega \to \Omega'$ be a mapping. Condition (2) in Definition 3 can be equivalently stated as the following two conditions for player i.*

1. *If $E' \in \mathcal{J}'_{K'_i}$, then $\varphi^{-1}(E') \in \mathcal{J}_{K_i}$.*

2. *For any $F \in \mathcal{J}_{K_i}$ and $E' \in \mathcal{D}'$ with $\varphi(F) \subseteq E'$, there is $F' \in \mathcal{J}'_{K'_i}$ such that $F' \subseteq E'$ and $\varphi(F) \subseteq F'$.*

By slightly abusing the notion of continuity in topology, we can regard the first condition as the regularity condition that a knowledge morphism "continuously" maps each player's knowledge from one space to another. On the other hand, if a self-evident event $E \in \mathcal{J}_{K_i}$ satisfies $\varphi(E) \in \mathcal{D}'$, then it is indeed self-evident, i.e., $\varphi(E) \in \mathcal{J}'_{K'_i}$. A knowledge morphism, however, does not necessarily map an event to an event, and this latter condition captures the sense in which a knowledge morphism preserves self-evident events with the presence of domain restrictions.

Proof of Proposition A.3. Suppose Condition (2) in Definition 3. First, if $E' \in \mathcal{J}'_{K'_i}$, then $E' \subseteq K'_i(E')$. Now, we have $\varphi^{-1}(E') \subseteq \varphi^{-1}(K'_i(E')) = K_i(\varphi^{-1}(E'))$, implying that $\varphi^{-1}(E') \in \mathcal{J}_{K_i}$. Second, let $F \in \mathcal{J}_{K_i}$ and $E' \in \mathcal{D}'$ with $\varphi(F) \subseteq E'$. We have $F \subseteq \varphi^{-1}(\varphi(F)) \subseteq \varphi^{-1}(E')$. Then, $F \subseteq K_i(F) \subseteq K_i(\varphi^{-1}(E')) = \varphi^{-1}(K'_i(E'))$. Let $F' := K'_i(E') \in \mathcal{J}'_{K'_i}$. We have $F' \subseteq E'$ and $\varphi(F) \subseteq \varphi(\varphi^{-1}(F')) \subseteq F'$. Conversely, suppose the two conditions in the statement. Take any $E' \in \mathcal{D}'$. We have $K'_i(E') \in \mathcal{J}'_{K'_i}$. It follows from the first condition that $\varphi^{-1}(K'_i(E')) \in \mathcal{J}_{K_i}$, and hence $\varphi^{-1}(K'_i(E')) \subseteq K_i(\varphi^{-1}(K'_i(E'))) \subseteq K_i(\varphi^{-1}(E'))$. Next, we have $\varphi^{-1}(E') \in \mathcal{D}$, $F := K_i(\varphi^{-1}(E')) \in \mathcal{J}_{K_i}$, and $\varphi(K_i(\varphi^{-1}(E'))) \subseteq \varphi(\varphi^{-1}(E')) \subseteq E'$. Now, the second condition implies that there exists $F' \in \mathcal{J}_{K'_i}$ such that $\varphi(K_i(\varphi^{-1}(E'))) = \varphi(F) \subseteq F' \subseteq K'_i(E')$. Thus, we obtain $K_i(\varphi^{-1}(E')) \subseteq \varphi^{-1}(K'_i(E'))$. The proof is complete. □

A.3.2 Section 4.2

Proof of Proposition 7.

1. Fix $E \in \mathcal{D}$. If $\omega \in C(E)$, then there is $F \in \mathcal{J}_I$ such that $\omega \in F \subseteq C(E)$. Since F is a common basis, we have $F \subseteq K_I(C(E))$. This implies that $\omega \in F \subseteq C(C(E))$, establishing Positive Introspection of C.

2. First, suppose that there is $i \in I$ such that K_i satisfies No-Contradiction Axiom. Then, $\emptyset \subseteq C(\emptyset) \subseteq K_I(\emptyset) \subseteq K_i(\emptyset) = \emptyset$. Second, if K_i satisfies Consistency for some $i \in I$, then $C(E) \cap C(E^c) \subseteq K_I(E) \cap K_I(E^c) \subseteq K_i(E) \cap K_i(E^c) = \emptyset$. Third, if K_i satisfies Truth Axiom for some $i \in I$, then $C(E) \subseteq K_I(E) \subseteq K_i(E) \subseteq E$.

3. Suppose that each K_i satisfies Monotonicity. Fix $E, F \in \mathcal{D}$ with $E \subseteq F$. If $\omega \in C(E)$ then there is $E' \in \mathcal{J}_I$ with $\omega \in E' \subseteq K_I(E) \subseteq K_I(F)$. Then, we get $\omega \in C(F)$, establishing Monotonicity of C.

4. Assume that each K_i satisfies Necessitation. Then, we have $\Omega \in \mathcal{J}_I$ and $K_I(\Omega) = \Omega$, leading to $C(\Omega) = \Omega$. Conversely, if C satisfies Necessitation then $\Omega \subseteq C(\Omega) \subseteq K_I(\Omega) \subseteq K_i(\Omega) \subseteq \Omega$ for all $i \in I$.

5. Assume that each K_i satisfies Non-empty λ-Conjunction as well as Monotonicity. Given Monotonicity, we have $\mathcal{J}_I = \bigcap_{i \in I} \mathcal{J}_{K_i}$. Now, since each $\mathcal{J}_{K_i}$ is closed under non-empty λ-intersection, so is $\mathcal{J}_I$. Then, it follows from Proposition 6 that $\mathcal{J}(\mathcal{J}_I)$ is also closed under non-empty λ-intersection. Let $\mathcal{E}$ be a non-empty set of events

with $0 < |\mathcal{E}| < \lambda$ such that $\omega \in \bigcap_{E \in \mathcal{E}} C(E)$. For each $E \in \mathcal{E}$, there is $F_E \in \mathcal{J}_I$ such that $\omega \in F_E \subseteq K_I(E)$. Then, we obtain $\omega \in \bigcap_{E \in \mathcal{E}} F_E \subseteq \bigcap_{E \in \mathcal{E}} \bigcap_{i \in I} K_i(E) = \bigcap_{i \in I} \bigcap_{E \in \mathcal{E}} K_i(E) \subseteq K_I(\bigcap \mathcal{E})$. Since $\bigcap_{E \in \mathcal{E}} F_E \in \mathcal{J}_I$, it follows that $\omega \in C(\bigcap \mathcal{E})$. Hence, C satisfies Non-empty λ-Conjunction.

Next, we provide a counterexample when Monotonicity is violated. Let $(\Omega, \mathcal{D}) = (\{\omega_1, \omega_2, \omega_3\}, \mathcal{P}(\Omega))$. Let K_i be defined as in the table below for each $i \in I = \{1, 2\}$.

E	$K_1(E)$	$K_2(E)$	$K_I(E)$	$C(E)$	$CC(E)$	$(\neg C)(E)$	$C(\neg C)(E)$
$\emptyset$	Ω	Ω	Ω	Ω	Ω	$\emptyset$	Ω
$\{\omega_1\}$	$\{\omega_1, \omega_2\}$	$\{\omega_1, \omega_2\}$	$\{\omega_1, \omega_2\}$	$\{\omega_1, \omega_2\}$	$\{\omega_1, \omega_2\}$	$\{\omega_3\}$	$\emptyset$
$\{\omega_2\}$	$\{\omega_2, \omega_3\}$	$\{\omega_2, \omega_3\}$	$\{\omega_2, \omega_3\}$	$\{\omega_2, \omega_3\}$	Ω	$\{\omega_1\}$	$\{\omega_1, \omega_2\}$
$\{\omega_3\}$	$\{\omega_2, \omega_3\}$	$\{\omega_1, \omega_3\}$	$\{\omega_3\}$	$\emptyset$	Ω	Ω	Ω
$\{\omega_1, \omega_2\}$	$\{\omega_1, \omega_2\}$	$\{\omega_1, \omega_2\}$	$\{\omega_1, \omega_2\}$	$\{\omega_1, \omega_2\}$	$\{\omega_1, \omega_2\}$	$\{\omega_3\}$	$\emptyset$
$\{\omega_1, \omega_3\}$	$\{\omega_1, \omega_2\}$	$\{\omega_1, \omega_3\}$	$\{\omega_1\}$	$\{\omega_1\}$	$\{\omega_1, \omega_2\}$	$\{\omega_2, \omega_3\}$	Ω
$\{\omega_2, \omega_3\}$	Ω	Ω	Ω	Ω	Ω	$\emptyset$	Ω
Ω	Ω	Ω	Ω	Ω	Ω	$\emptyset$	Ω

Clearly, we have $\mathcal{J}_1 = \mathcal{D} \setminus \{\{\omega_1, \omega_3\}\}$ and $\mathcal{J}_2 = \mathcal{D}$. On the other hand, we have $\mathcal{J}_I = \mathcal{D} \setminus \{\{\omega_3\}, \{\omega_1, \omega_3\}\}$. Then, the common knowledge operator C is depicted as in the table. C violates (Non-empty) Finite Conjunction because $C(\{\omega_1, \omega_3\}) \cap C(\{\omega_2, \omega_3\}) = \{\omega_1\} \not\subseteq \emptyset = C(\{\omega_3\})$. Incidentally, C violates all the properties except for Positive Introspection and Necessitation.

6. The above example is indeed an example in which C violates Negative Introspection while individual knowledge operators satisfy Negative Introspection (see the table below).

E	$K_1(E)$	$K_2(E)$	$(\neg K_1)(E)$	$(\neg K_2)(E)$	$K_1(\neg K_1)(E)$	$K_2(\neg K_2)(E)$
$\emptyset$	Ω	Ω	$\emptyset$	$\emptyset$	Ω	Ω
$\{\omega_1\}$	$\{\omega_1, \omega_2\}$	$\{\omega_1, \omega_2\}$	$\{\omega_3\}$	$\{\omega_3\}$	$\{\omega_2, \omega_3\}$	$\{\omega_1, \omega_3\}$
$\{\omega_2\}$	$\{\omega_2, \omega_3\}$	$\{\omega_2, \omega_3\}$	$\{\omega_1\}$	$\{\omega_1\}$	$\{\omega_1, \omega_2\}$	$\{\omega_1, \omega_2\}$
$\{\omega_3\}$	$\{\omega_2, \omega_3\}$	$\{\omega_1, \omega_3\}$	$\{\omega_1\}$	$\{\omega_2\}$	$\{\omega_1, \omega_2\}$	$\{\omega_2, \omega_3\}$
$\{\omega_1, \omega_2\}$	$\{\omega_1, \omega_2\}$	$\{\omega_1, \omega_2\}$	$\{\omega_3\}$	$\{\omega_3\}$	$\{\omega_2, \omega_3\}$	$\{\omega_1, \omega_3\}$
$\{\omega_1, \omega_3\}$	$\{\omega_1, \omega_2\}$	$\{\omega_1, \omega_3\}$	$\{\omega_3\}$	$\{\omega_2\}$	$\{\omega_2, \omega_3\}$	$\{\omega_2, \omega_3\}$
$\{\omega_2, \omega_3\}$	Ω	Ω	$\emptyset$	$\emptyset$	Ω	Ω
Ω	Ω	Ω	$\emptyset$	$\emptyset$	Ω	Ω

7. Assume that $(\Omega, \mathcal{D})$ is a complete algebra and that each player's knowledge satisfies Truth Axiom, Positive Introspection, and Monotonicity. First, it is clear that $\mathcal{J}(\mathcal{J}_I) = \bigcap_{i \in I} \mathcal{J}_{K_i}$, as the maximality property is characterized by the closure under arbitrary union (Proposition 5). In this case, $\bigcap_{i \in I} \mathcal{J}_{K_i}$ also inherits Negative Introspection (as well as λ-Conjunction) as they are characterized by the closure under complementation (λ-intersection).

8. The first assertion follows from the previous arguments. Next, fix $E \in \mathcal{D}$. Observe that we have

$$K(E) = \{\omega \in \Omega \mid \text{there is } F \in \mathcal{J}_K \text{ such that } \omega \in F \subseteq K(E)\}.$$

Observe also that if $F \in \mathcal{J}_K$ then $F \subseteq K(F) \subseteq K_I(F)$. Thus, $\mathcal{J}_K \subseteq \bigcap_{i \in I} \mathcal{J}_{K_i} = \mathcal{J}_I$, where the last equality follows from Monotonicity. Then, we can see that

$$K(E) \subseteq \{\omega \in \Omega \mid \text{there is } F \in \mathcal{J}_I \text{ such that } \omega \in F \subseteq K_I(E)\} = C(E).$$

9. Note that $K_I^\alpha(E)$ is an event of a given κ-knowledge space for any $E \in \mathcal{D}$ and for any $\alpha < \kappa$. By definition, $C(\cdot) \subseteq K_I^1(\cdot)$. Suppose $C(\cdot) \subseteq K_I^\beta(\cdot)$. Now, if $\omega \in C(E)$, then there is $F \in \mathcal{J}_I$ such that $\omega \in F \subseteq C(E) \subseteq K_I^\beta(E)$. Since F is a common basis, we have $\omega \in F \subseteq K_I^{\beta+1}(E)$. Next, if $C(E) \subseteq K_I^\beta(E)$ for all $\beta < \alpha$, then it is clear that $C(E) \subseteq \bigcap_{\beta: 1 \leq \beta < \alpha} K_I^\beta(E) = K_I^\alpha(E)$.

$\square$

Remark A.5 (Common Knowledge among $\{i\}$ need not be i's Knowledge without Monotonicity or Positive Introspection). Let $(\Omega, \mathcal{D}) = (\{\omega_1, \omega_2\}, \mathcal{P}(\Omega))$. Consider the knowledge operators K_1 and K_2 as depicted in the following table. While K_1 violates Positive Introspection, K_2 does Monotonicity. In either case, $K_i \neq C_{\{i\}}$.

E	$K_1(E)$	$K_1K_1(E)$	$C_{\{1\}}(E)$	E	$K_2(E)$	$K_2K_2(E)$	$C_{\{2\}}(E)$
$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\{\omega_1\}$	$\{\omega_1\}$	$\emptyset$
$\{\omega_1\}$	$\emptyset$	$\emptyset$	$\emptyset$	$\{\omega_1\}$	$\{\omega_1\}$	$\{\omega_1\}$	$\emptyset$
$\{\omega_2\}$	$\{\omega_1\}$	$\emptyset$	$\emptyset$	$\{\omega_2\}$	$\{\omega_1\}$	$\{\omega_1\}$	$\emptyset$
Ω	$\{\omega_1\}$	$\emptyset$	$\emptyset$	Ω	$\emptyset$	$\{\omega_1\}$	$\emptyset$

Proof of Proposition 8. Given that each K_i satisfies Finite Conjunction, the operator K_I satisfies Finite Conjunction as well. That is, $K_I(E \cap F) = K_I(E) \cap K_I(F)$ for any $E, F \in \mathcal{D}$. Now, we show that if each K_i satisfies Monotonicity then

$$C(E) = \max\{F \in \mathcal{D} \mid F \subseteq K_I(E) \cap K_I(F)\}. \tag{A.1}$$

First, by definition, $C(E) \subseteq K_I(E)$. Second, if $\omega \in C(E)$, then there is $F \in \mathcal{J}_I$ such that $\omega \in F \subseteq C(E)$. Now, since F is a common basis, we have $\omega \in F \subseteq K_I(C(E))$, establishing $C(E) \subseteq K_I(C(E))$.

Conversely, take any $F \in \mathcal{D}$ with $F \subseteq K_I(E) \cap K_I(F)$. Then, $F \in \bigcap_{i \in I} \mathcal{J}_{K_i} = \mathcal{J}_I$ (recall that the equality follows from Monotonicity) and $F \subseteq K_I(E)$. Thus, we have $F \subseteq C(E)$. Hence, we obtain the desired expression (A.1).

Finally, consider

$$\max\{F \in \mathcal{D} \mid F \subseteq K_I(E) \cap K_I(F)\} = \max\{F \in \mathcal{D} \mid F = K_I(E) \cap K_I(F)\}.$$

This is a variant of Tarski's fixed point theorem, stating that the greatest element F satisfying $F \subseteq f_E(F) := K_I(E) \cap K_I(F)$ of a monotone operator $f_E(\cdot)$, given that it exists, is indeed the greatest fixed point. Indeed, since $F \subseteq f_E(F)$, we have $f_E(F) \subseteq f_E(f_E(F))$. Since F is the greatest such point, we have $f_E(F) \subseteq F$, i.e., $F = f_E(F) = K_I(E) \cap K_I(F)$. □

Proof of Corollary 4. Here we provide an alternative proof instead of the reasoning in the main text. Fix $E \in \mathcal{D}$. First, it follows from Proposition 7 (9) that $C(E) \subseteq K_I^\infty(E)$ (without Monotonicity or Countable Conjunction). Conversely, observe first that $K^{n+1}(E) \subseteq K_i(K^n(E))$ for all $n \in \mathbb{N}$ and $i \in I$. Since each K_i satisfies Countable Conjunction, we have $\bigcap_{n\in\mathbb{N}} K_i(K^n(E)) = K_i(\bigcap_{n\in\mathbb{N}} K^n(E))$, which shows that $K_I^\infty(E) \in \mathcal{J}_i$ for each $i \in I$, that is, $K_I^\infty(E) \in \bigcap_{i\in I} \mathcal{J}_i = \mathcal{J}_I$, where the last equality follows from Monotonicity of $(K_i)_{i\in I}$. Since $K_I^\infty(E) \subseteq K_I(E)$, we obtain $K_I^\infty(E) \subseteq C(E)$. □

Proof of Proposition 9. The assertions follow from the reasoning in the main text, and thus the proof is omitted. □

Remark A.6 (Preservation of Common Knowledge in ∞-Knowledge Spaces)**.** Consider any ∞-knowledge space satisfying Monotonicity. Fix $E \in \mathcal{D}$. Consider a sequence $(K_I^\alpha(E))_\alpha$ where α is a limit ordinal. Since it is decreasing, there is a least ordinal such that $K_I^\alpha(E) = K_I^\beta(E)$ for any limit ordinal β with $\beta \geq \alpha$. Let α_E be such a limit ordinal for each $E \in \mathcal{D}$. Now, we have a limit ordinal α such that $\alpha \geq \alpha_E$ for all $E \in \mathcal{D}$. Then, we have $K_I^\alpha(E) \subseteq K_I(E)$ and $K_I^\alpha(E) \subseteq K^{\alpha+1}(E)$, from which we obtain $K_I^\alpha(E) \subseteq C(E)$. Indeed, the equality holds by Proposition 7 (9). This means that common knowledge is automatically preserved in ∞-knowledge spaces satisfying Monotonicity.

Generally, the following proposition characterizes the preservation of common knowledge.

Proposition A.4 (Preservation of Common Knowledge)**.** *Let $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega'}$ be a knowledge morphism between knowledge spaces $\overrightarrow{\Omega}$ and $\overrightarrow{\Omega'}$.*

1. *The following are equivalent.*
   (a) *$\varphi^{-1}(C'(E')) \subseteq C(\varphi^{-1}(E'))$ for all $E' \in \mathcal{D}'$.*
   (b) *For any $F' \in \mathcal{J}'_I$ and $E' \in \mathcal{D}'$ with $F' \subseteq K'_I(E')$, there is $F \in \mathcal{J}(\mathcal{J}_I)$ such that $\varphi^{-1}(F') \subseteq F$ and $F \subseteq K_I(\varphi^{-1}(E'))$.*
2. *The following are equivalent.*
   (a) *$C(\varphi^{-1}(E')) \subseteq \varphi^{-1}(C'(E'))$ for all $E' \in \mathcal{D}'$.*
   (b) *For any $F \in \mathcal{J}_I$ and $E' \in \mathcal{D}'$ with $\varphi(F) \subseteq K'_I(E')$, there is $F' \in \mathcal{J}(\mathcal{J}'_I)(= \mathcal{J}_{\mathcal{D}'}(\mathcal{J}'_I))$ such that $\varphi(F) \subseteq F'$ and $F' \subseteq K_I(E')$.*

In Proposition A.4, if the knowledge space $\overrightarrow{\Omega}$ satisfies Monotonicity then Condition (1b) is always satisfied. This is because we can take $F = \varphi^{-1}(F')$.

Proof of Proposition A.4. 1. Suppose that $\varphi^{-1}(C'(\cdot)) \subseteq C(\varphi^{-1}(\cdot))$. Take $F' \in \mathcal{J}'_I$ and $E' \in \mathcal{D}'$ with $F' \subseteq K'_I(E')$. Then, $F' \subseteq C'(E') \subseteq K'_I(E')$. We obtain $\varphi^{-1}(F') \subseteq \varphi^{-1}(C'(E')) \subseteq F \subseteq K_I(\varphi^{-1}(E'))$, where $F := C(\varphi^{-1}(E')) \in \mathcal{J}(\mathcal{J}_I)$.

Conversely, suppose (1b). Take any $E' \in \mathcal{D}'$ and assume that $\omega \in \varphi^{-1}(C'(E'))$, i.e., $\varphi(\omega) \in C'(E')$. Then, there is $F' \in \mathcal{J}'_I$ such that $\varphi(\omega) \in F' \subseteq K'_I(E')$. By (1b), there is $F \in \mathcal{J}(\mathcal{J}_I)$ such that $\omega \in \varphi^{-1}(F') \subseteq F \subseteq \varphi^{-1}(K'_I(E')) = K_I(\varphi^{-1}(E'))$. Now, it follows that $\omega \in C(\varphi^{-1}(E'))$.

2. Suppose that $C(\varphi^{-1}(\cdot)) \subseteq \varphi^{-1}(C'(\cdot))$. Let $F \in \mathcal{J}_I$ and $E' \in \mathcal{D}'$ with $\varphi(F) \subseteq E'$. Then, we have $F \subseteq \varphi^{-1}(\varphi(F)) \subseteq \varphi^{-1}(E')$. Since $F \subseteq K_I(\varphi^{-1}(E'))$, we have $F \subseteq C(\varphi^{-1}(E')) \subseteq \varphi^{-1}(C'(E'))$. Let $F' := C'(E') \in \mathcal{J}'(\mathcal{J}'_I)$. We have $F' \subseteq K'_I(E')$ and $\varphi(F) \subseteq \varphi(\varphi^{-1}(F')) \subseteq F'$.

   Conversely, suppose (2b). Take any $E' \in \mathcal{D}'$ and suppose that $\omega \in C(\varphi^{-1}(E'))$. Then, there is $F \in \mathcal{J}_I$ such that $\omega \in F \subseteq K_I(\varphi^{-1}(E')) = \varphi^{-1}(K'_I(E'))$. Thus, $\varphi(\omega) \in \varphi(F) \subseteq \varphi\varphi^{-1}(K'_I(E')) \subseteq K'_I(E')$. Now, there is $F' \in \mathcal{J}(\mathcal{J}'_I)$ such that $\varphi(\omega) \in \varphi(F) \subseteq F' \subseteq K'_I(E')$. Then, we obtain $\varphi(\omega) \in C'(E')$, i.e., $\omega \in \varphi^{-1}(C'(E'))$. □

Proof of Proposition 10. Since Monotonicity is assumed, it follows from Proposition A.4 that we need only to prove that $D^{-1}(C^*([e])) \supseteq C(D^{-1}([e]))$ for all $[e] \in \mathcal{D}^*$. Now, it follows from the assumptions of the statement that $C^* = K_i^*$ for all $i \in I$. Then, we have $C(D^{-1}([e])) \subseteq K_i(D^{-1}([e])) = D^{-1}(C^*([e]))$ for all $[e] \in \mathcal{D}^*$. □

A.4 Section 5

A.4.1 Section 5.1

Remark A.7 (Equivalence between Knowledge Operaors and Knowledge-Type Mappings)**.** In the first step, we show that t_{K_i} inherits the properties of K_i. Fix $\omega \in \Omega$.

1. t_{K_i} clearly inherits No-Contradiction Axiom: $t_{K_i}(\omega)(\emptyset) = 0$.

2. Consider Consistency. If $t_{K_i}(\omega)(E) = 1$ then $\omega \in K_i(E) \subseteq (\neg K_i)(E^c)$, and thus $t_{K_i}(\omega)(E^c) = 0$, establishing $t_{K_i}(\omega)(E) \leq 1 - t_{K_i}(\omega)(E^c)$.

3. Consider Monotonicity. Let $E, F \in \mathcal{D}$ be such that $E \subseteq F$. If $t_{K_i}(\omega)(E) = 1$ then $\omega \in K_i(E) \subseteq K_i(F)$ and thus $t_{K_i}(\omega)(F) = 1$.

4. Necessitation is immediate: $t_{K_i}(\omega)(\Omega) = 1$.

5. Consider Non-empty λ-Conjunction. Let $\mathcal{E}$ be such that $\mathcal{E} \subseteq \mathcal{D}$ and $0 < |\mathcal{E}| < \kappa$. Suppose that $t_{K_i}(\omega)(E) = 1$ for all $E \in \mathcal{E}$ (i.e., $\min_{E \in \mathcal{E}} t_{K_i}(\omega)(E) = 1$). Then, since $\omega \in \bigcap_{E \in \mathcal{E}} K_i(E) \subseteq K_i(\bigcap \mathcal{E})$, we obtain $t_{K_i}(\omega)(\bigcap \mathcal{E}) = 1$.

6. Consider Truth Axiom. If $t_{K_i}(\omega)(E) = 1$ then $\omega \in K_i(E) \subseteq E$.

7. Consider Positive Introspection. If $t_{K_i}(\omega)(E) = 1$ then $\omega \in K_i(E) \subseteq K_iK_i(E)$. Thus, $t_{K_i}(\omega)(\{\omega' \in \Omega \mid t_{K_i}(\omega')(E) = 1\}) = t_{K_i}(\omega)(K_i(E)) = 1$.

8. Consider Negative Introspection. If $t_{K_i}(\omega)(E) = 0$ then $\omega \in (\neg K_i)(E) \subseteq K_i(\neg K_i)(E)$. Thus, $t_{K_i}(\omega)(\{\omega' \in \Omega \mid t_{K_i}(\omega')(E) = 0\}) = t_{K_i}(\omega)((\neg K_i)(E)) = 1$.

9. The Kripke property follows because $b_{t_{K_i}} = b_{K_i}$.

In the second step, we show that K_{t_i} inherits the properties of t_i. Fix $\omega \in \Omega$.

1. K_{t_i} inherits No-Contradiction Axiom: $K_{t_i}(E) = \{\omega \in \Omega \mid t_i(\omega)(\emptyset) = 1\} = \emptyset$.

2. Consider Consistency. If $\omega \in K_{t_i}(E)$ then $t_i(\omega)(E) = 1 \leq 1 - t_i(\omega)(E^c)$, and thus $t_i(\omega)(E^c) = 0$. Then, $\omega \in (\neg K_{t_i})(E^c)$.

3. Consider Monotonicity. Let $E, F \in \mathcal{D}$ be such that $E \subseteq F$. If $\omega \in K_{t_i}(E)$ then $1 = t_i(\omega)(E) \leq t_i(\omega)(F)$. Thus, $\omega \in K_{t_i}(F)$.

4. Consider Necessitation. We have $K_{t_i}(\Omega) = \{\omega \in \Omega \mid t_i(\omega)(\Omega) = 1\} = \Omega$.

5. Consider Non-empty λ-Conjunction. Let $\mathcal{E}$ be such that $\mathcal{E} \subseteq \mathcal{D}$ and $0 < |\mathcal{E}| < \kappa$. Suppose that $\omega \in \bigcap_{E \in \mathcal{E}} K_i(E)$. Then, $t_i(\omega)(E) = 1$ for all $E \in \mathcal{E}$, and thus $1 = \min_{E \in \mathcal{E}} t_i(\omega)(E) \leq t_i(\omega)(\bigcap \mathcal{E})$. Thus, $\omega \in K_{t_i}(\bigcap \mathcal{E})$

6. Consider Truth Axiom. If $\omega \in K_{t_i}(E)$ then $t_i(\omega)(E) = 1$ and thus $\omega \in E$.

7. Consider Positive Introspection. If $\omega \in K_{t_i}(E)$ then $t_i(\omega)(E) = 1$. We then have $t_i(\omega)(\{\omega' \in \Omega \mid t_i(\omega')(E) = 1\}) = t_i(\omega)(K_{t_i}(E)) = 1$, leading to $\omega \in K_{t_i}K_{t_i}(E)$.

8. Consider Negative Introspection. If $\omega \in (\neg K_{t_i})(E)$ then $t_i(\omega)(E) = 0$. We then have $t_i(\omega)(\{\omega' \in \Omega \mid t_i(\omega')(E) = 0\}) = t_i(\omega)((\neg K_{t_i})(E)) = 1$, leading to $\omega \in K_{t_i}(\neg K_{t_i})(E)$.

9. The Kripke property follows because $b_{K_{t_i}} = b_{t_i}$.

Remark A.8 (Category Theoretical Equivalence of Definitions of Knowledge Spaces by Knowledge Operators and Knowledge-Type Mappings)**.** We denote by $\mathbf{K}_{oprt}$ and $\mathbf{K}_{type}$ the category of knowledge spaces given in terms of knowledge operators and knowledge-types, respectively. We define the following functor $T^{type}_{oprt} : \mathbf{K}_{oprt} \to \mathbf{K}_{type}$.

For any knowledge space $\overrightarrow{\Omega} = \langle (\Omega, \mathcal{D}), (K_i)_{i \in I}, \Theta \rangle$, we let $T^{type}_{oprt}(\overrightarrow{\Omega}) = \langle (\Omega, \mathcal{D}), (t_{K_i})_{i \in I}, \Theta \rangle$. For a knowledge morphism $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega'}$, we let $T^{type}_{oprt}(\varphi) = \varphi$ as a (κ-measurable) mapping $\varphi : (\Omega, \mathcal{D}) \to (\Omega', \mathcal{D}')$.

Likewise, for any given knowledge space $\overrightarrow{\Omega} = \langle (\Omega, \mathcal{D}), (t_i)_{i \in I}, \Theta \rangle$, we let $T^{oprt}_{type}(\overrightarrow{\Omega}) = \langle (\Omega, \mathcal{D}), (K_{t_i})_{i \in I}, \Theta \rangle$. For a knowledge morphism $\varphi : \overrightarrow{\Omega} \to \overrightarrow{\Omega'}$, we let $T^{oprt}_{type}(\varphi) = \varphi$.

Clearly, we have $T^{oprt}_{type} \circ T^{type}_{oprt} = I_{oprt}$ and $T^{type}_{oprt} \circ T^{oprt}_{type} = I_{type}$, where I_{oprt} and I_{type} are the constant functors. In this way, we can establish the (category-theoretical) equivalence between the categories of knowledge spaces given in terms of knowledge operators and knowledge-types. We can establish the equivalence for the other combinations of primitives in a similar way.

A.4.2 Section 5.2

Proof of Proposition 11.

1. By definition, i knows t_i with respect to $\{\mathcal{M}(E) \mid E \in \mathcal{D}\}$ iff $\omega \in t_i^{-1}(\mathcal{M}(E))$ implies $t_i(\omega)(t_i^{-1}(\mathcal{M}(E))) = 1$ for each $E \in \mathcal{D}$.

2. Again, by definition, i knows t_i with respect to $\{\neg\mathcal{M}(E) \mid E \in \mathcal{D}\}$ iff $\omega \in \neg t_i^{-1}(\mathcal{M}(E))$ implies $t_i(\omega)(\neg t_i^{-1}(\mathcal{M}(E))) = 1$ for each $E \in \mathcal{D}$.

□

Proof of Proposition 12.

1. It follows from Truth Axiom that $\omega' \in b_{t_i}(\omega')$, because it implies that $\omega' \in E$ for all $E \in \mathcal{D}$ with $t_i(\omega')(E) = 1$. Hence, we have $\omega' \in b_{t_i}(\omega') \subseteq b_{t_i}(\omega)$ for all $\omega' \in (\uparrow t_i(\omega))$. The converse is immediate because $\omega \in [t_i(\omega)] \subseteq (\uparrow t_i(\omega)) \subseteq b_{t_i}(\omega)$. This means that for any $E \in \mathcal{D}$ with $t_i(\omega)(E) = 1$, $\omega \in b_{t_i}(\omega) \subseteq E$.

2. Suppose that $\omega' \in b_{t_i}(\omega)$. Now, suppose to the contrary that $\omega' \notin (\uparrow t_i(\omega))$, i.e., there is $F \in \mathcal{D}$ such that $t_i(\omega')(F) = 0 < 1 = t_i(\omega)(F)$. Then, by Positive Introspection, we have $t_i(\omega)(t_i^{-1}(\mathcal{M}_F)) = 1$, and thus $\omega' \in t_i^{-1}(\mathcal{M}_F)$, i.e., $t_i(\omega')(F) = 1$, a contradiction.

   Next, suppose that t_i satisfies the Kripke property. Assume that $b_{t_i}(\omega) \subseteq (\uparrow t_i(\omega))$. Suppose that $t_i(\omega)(E) = 1$. Then, $b_{t_i}(\omega) \subseteq E$. Now, in order to establish Positive Introspection, it is enough to show that $b_i(\omega') \subseteq E$ for all $\omega' \in b_{t_i}(\omega)$. Take any $\omega' \in b_{t_i}(\omega)$. Since $\omega' \in (\uparrow t_i(\omega))$ and $t_i(\omega)(b_{t_i}(\omega)) = 1$ holds, we have $t_i(\omega')(b_{t_i}(\omega)) = 1$. Thus, $b_{t_i}(\omega') \subseteq b_{t_i}(\omega) \subseteq E$.

3. The proof is analogous to the last assertion. Suppose that $\omega' \in b_{t_i}(\omega)$. Now, suppose to the contrary that $\omega' \notin (\downarrow t_i(\omega))$, i.e., there is $F \in \mathcal{D}$ such that $t_i(\omega)(F) = 0 < 1 = t_i(\omega')(F)$. Then, by Negative Introspection, we have $t_i(\omega)(\neg t_i^{-1}(\mathcal{M}_F)) = 1$, and thus $\omega' \in \neg t_i^{-1}(\mathcal{M}_F)$, i.e., $t_i(\omega')(F) = 0$, a contradiction.

   Next, suppose that t_i satisfies the Kripke property. Assume that $b_{t_i}(\omega) \subseteq (\downarrow t_i(\omega))$. Suppose that $t_i(\omega)(E) = 0$. Then, $b_{t_i}(\omega) \cap E^c \neq \emptyset$. Now, in order to establish Negative Introspection, it is enough to show that $b_i(\omega') \cap E^c \neq \emptyset$ for all $\omega' \in b_{t_i}(\omega)$. Take any $\omega' \in b_{t_i}(\omega)$. Since $\omega' \in (\downarrow t_i(\omega))$ and $t_i(\omega)(E) = 0$ holds, we have $t_i(\omega')(E) = 0$, i.e., $b_i(\omega') \cap E^c \neq \emptyset$.

4. By the previous assertions, we have $[t_i(\omega)] = (\uparrow t_i(\omega)) = b_{t_i}(\omega) \subseteq (\downarrow t_i(\omega))$. Thus, we show $(\downarrow t_i(\omega)) \subseteq [t_i(\omega)]$. Suppose not, i.e., there is $\omega' \in \Omega$ with $\omega' \in (\downarrow t_i(\omega))$ but $\omega' \notin (\uparrow t_i(\omega))$. Thus, $t_i(\omega')(F) = 0 < 1 = t_i(\omega)(F)$ for some $F \in \mathcal{D}$. Then, Negative Introspection implies that $t_i(\omega')((\neg t_i^{-1}(\mathcal{M}_F)) = 1$, so that $t_i(\omega)((\neg t_i^{-1}(\mathcal{M}_F)) = 1$.

Truth Axiom then implies $\omega \in \neg t_i^{-1}(\mathcal{M}_F)$, i.e., $t_i(\omega)(F) = 0$, which is a contradiction to $t_i(\omega)(F) = 1$. Finally, when t_i satisfies the Kripke property, it is immediate that the converse is true.

□

A.4.3 Section 5.3

Proof of Proposition 13. First, since each player's knowledge satisfies Positive Introspection, it follows from Proposition 11 that each player i knows her own type-mapping $t_{K_i^*} : \Omega \to M(\Omega^*, \mathcal{D}^*)$ with respect to $\{\mathcal{M}([e]) \mid [e] \in \mathcal{D}^*\}$. Second, it follows from the assumptions of the statement that $C^* = K_i^*$ for all $i \in I$. □

The fact that players' knowledge operators are identical plays an important role in the proof of Proposition 13. The following proposition clarifies this point.

Proposition A.5 (Knowledge of Opponent's Type Implies Knowledgeability)**.** *Let $\overrightarrow{\Omega} := \langle (\Omega, \mathcal{D}), (t_i)_{i \in I}, \Theta \rangle$ be a knowledge space. Assume the same assumptions on knowledge of players i and j. Moreover, we impose Truth Axiom, Monotonicity, and Positive Introspection on them. Then, i knows j's knowledge-type mapping t_j with respect to $\{\mathcal{M}(E) \mid E \in \mathcal{D}\}$ iff $K_{t_j}(E) \subseteq K_{t_i}(E)$ for all $E \in \mathcal{D}$ (i.e., i is at least as knowledgeable as j).*

Proof of Proposition A.5. If i knows j's knowledge-type mapping t_j, then we have $K_{t_j}(E) \subseteq K_{t_i} K_{t_j}(E) \subseteq K_{t_i}(E)$ for all $E \in \mathcal{D}$, where the second set inclusion follows from Truth Axiom of j and Monotonicity of i. Conversely, suppose that $K_{t_j}(E) \subseteq K_{t_i}(E)$ for all $E \in \mathcal{D}$. By Positive Introspection of j, we obtain $K_{t_j}(E) \subseteq K_{t_j} K_{t_j}(E) \subseteq K_{t_i} K_{t_j}(E)$ for all $E \in \mathcal{D}$. □

Proposition A.5 implies the following. Under the above assumptions, players' type mappings may not be necessarily commonly known in a given knowledge space $\overrightarrow{\Omega}$ when players' knowledge is different.

A.5 Section 6

A.5.1 Section 6.1

Proof of Remark 4. First, it follows from the condition of the statement that

$$\begin{aligned} K_{t_i}(\varphi^{-1}(E')) &= \{\omega \in \Omega \mid t_i(\omega)(\varphi^{-1}(E')) = 1\} = \{\omega \in \Omega \mid t_i'(\varphi(\omega))(E') = 1\} \\ &= \{\omega \in \Omega \mid \varphi(\omega) \in K'_{t_i'}(E')\} = \varphi^{-1}(K'_{t_i'}(E')). \end{aligned}$$

Conversely, it follows frm Condition (3) in Definition 3 that

$$\begin{aligned} t_{K_i}(\omega)(\varphi^{-1}(E')) = 1 &\text{ iff } \omega \in K_i(\varphi^{-1}(E')) = \varphi^{-1}(K_i'(E')) \\ &\text{ iff } \varphi(\omega) \in K_i'(E') \text{ iff } t'_{K_i'}(\varphi(\omega))(E') = 1. \end{aligned}$$

□

Remark A.9 ($(H, \mathcal{H})$ is a κ-Complete Algebra)**.** We show that $(H, \mathcal{H})$ is a κ-complete algebra. First, $H \in \mathcal{H}$ is obvious. Second, $\mathcal{H}$ is clearly closed under complementation. Indeed, if $(\pi^\alpha)^{-1}(E^\alpha) \in \mathcal{H}$ (with $\alpha < \kappa$) then $\neg(\pi^\alpha)^{-1}(E^\alpha) = (\pi^\alpha)^{-1}(\neg E^\alpha) \in \mathcal{H}$. Third, since κ is regular, it is also closed under κ-union/intersection. Take any subset A of ordinals $\{\alpha \mid \alpha < \kappa\}$ whose cardinal $|A|$ is less than κ. Then, its supremum γ has cardinal less than κ, given that κ is regular. Now, each $(\pi^\alpha)^{-1}(E^\alpha)$ is projected on H^γ. Then, since H^γ is closed under κ-intersection, the proof is complete.

Proof of Lemma 6. Part 1. We show by induction that $h^{-1}((\pi^\alpha|_{\Omega^*})^{-1}(E^\alpha)) = (h^\alpha)^{-1}(E^\alpha) \in \mathcal{D}$ for all $\alpha < \kappa$ and $E^\alpha \in \mathcal{H}^\alpha$. Let $\alpha = 0$. We have $(h^0)^{-1}(E^0) = \Theta^{-1}(E^0) \in \mathcal{D}$ for any $E^0 \in \mathcal{A}_S$. For a successor ordinal $\alpha = \beta + 1$, it is sufficient to show that $(t_i \circ (h^\beta)^{-1})^{-1}(\mathcal{M}^{(H^\beta, \mathcal{H}^\beta)}(E^\beta)) \in \mathcal{D}$ for each $E^\beta \in \mathcal{H}^\beta$ and $i \in I$. Fix $E^\beta \in \mathcal{H}^\beta$ and $i \in I$. Then, we have

$$
\begin{aligned}
(t_i \circ (h^\beta)^{-1})^{-1}(\mathcal{M}^{(H^\beta, \mathcal{H}^\beta)}(E^\beta)) &= \{\omega \in \Omega \mid t_i(\omega)((h^\beta)^{-1}(E^\beta)) = 1\} \\
&= t_i^{-1}(\mathcal{M}^{(\Omega, \mathcal{D})}((h^\beta)^{-1}(E^\beta))) \in \mathcal{D}.
\end{aligned}
$$

For a limit ordinal α, if $h^{-1}((\pi^\beta|_{\Omega^*})^{-1}(E^\beta)) = (h^\beta)^{-1}(E^\beta) \in \mathcal{D}$ for all $\beta < \alpha$, then it is clear that $h^{-1}((\pi^\alpha|_{\Omega^*})^{-1}(E^\alpha)) = (h^\alpha)^{-1}(E^\alpha) \in \mathcal{D}$.

Part 2. For notational ease, we denote $h = h_{\vec{\Omega}}$ and $h' = h_{\vec{\Omega}'}$. We show by induction that, for each ordinal $\alpha < \kappa$, $h^\alpha(\omega) = (h')^\alpha(\varphi(\omega))$ for each $\omega \in \Omega$. Let $\alpha = 0$. Since φ is a knowledge morphism, we have $h^0(\omega) = \Theta(\omega) = \Theta'(\varphi(\omega)) = (h')^0(\varphi(\omega))$ for each $\omega \in \Omega$. Let $\alpha = \beta + 1$ be a successor ordinal. Since we have

$$
\begin{aligned}
h^{\beta+1}(\omega) &= \left(h^\beta(\omega), t(\omega) \circ (h^\beta)^{-1}\right) \text{ and} \\
(h')^{\beta+1}(\varphi(\omega)) &= \left(h'^\beta(\varphi(\omega)), t'(\varphi(\omega)) \circ (h'^k)^{-1}\right),
\end{aligned}
$$

it suffices to show that $t(\omega) \circ (h^\beta)^{-1} = t'(\varphi(\omega)) \circ (h'^\beta)^{-1}$. Now, since φ is a knowledge morphism, we have, for each $i \in I$,

$$
\begin{aligned}
t'_i(\varphi(\omega)) \left((h'^\beta)^{-1}(\cdot)\right) &= t_i(\omega) \left(\varphi^{-1}\left((h'^\beta)^{-1}(\cdot)\right)\right) \\
&= t_i(\omega) \left(\left(h'^\beta \circ \varphi\right)^{-1}(\cdot)\right) = t_i(\omega) \left((h^\beta)^{-1}(\cdot)\right).
\end{aligned}
$$

Let α be a limit ordinal. Fix $\omega \in \Omega$. By the definitions of h^α and $(h')^\alpha$, it is immediate that $h^\alpha(\omega) = (h')^\alpha(\varphi(\omega))$ if $h^\beta(\omega) = (h')^\beta(\varphi(\omega))$ for all $\beta < \alpha$. The induction is complete. $\square$

Proof of Lemma 7. Fix $i \in I$. We prove the results in the following five steps. In the first step, we start with showing that t_i^* is a well-defined mapping on Ω^*. Fix $\omega^* \in \Omega^*$. We show that, for any ordinals (α, β) with $0 \leq \beta \leq \alpha < \kappa$, if $(\pi^\alpha|_{\Omega^*})^{-1}(E^\alpha) = (\pi^\beta|_{\Omega^*})^{-1}(F^\beta)$ for some $E^\alpha \in \mathcal{H}^\alpha$ and $E^\beta \in \mathcal{H}^\beta$, then $(\omega^*)_i^{\alpha+1}(E^\alpha) = (\omega^*)_i^{\beta+1}(F^\beta)$.

Observe first that, for any a knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$ such that $\omega^* = h(\omega)$, we have

$$t_i^*(\omega^*) \circ (\pi^\alpha|_{\Omega^*})^{-1} = (h(\omega))_i^{\alpha+1} = t_i(\omega) \circ (h^\alpha)^{-1} = t_i(\omega) \circ (\pi^\alpha|_{\Omega^*} \circ h)^{-1}. \tag{A.2}$$

Thus, we have

$$(\omega^*)_i^{\alpha+1}(E^\alpha) = t_i(\omega)((\pi^\alpha|_{\Omega^*})^{-1}(E^\alpha)) = t_i(\omega)((\pi^\beta|_{\Omega^*})^{-1}(F^\beta)) = (\omega^*)_i^{\beta+1}(F^\beta).$$

This equation holds irrespective of a choice of knowledge spaces.

In the second step, we establish Equation (12). Indeed, it follows from Equation (A.2) that, for each $\alpha < \kappa$ and $E^\alpha \in \mathcal{H}^\alpha$,

$$t_i^*(h(\omega))((\pi^\alpha|_{\Omega^*})^{-1}(E^\alpha)) = t_i(\omega)(h^{-1}((\pi^\alpha|_{\Omega^*})^{-1}(E^\alpha))).$$

In the third step, we show that t_i^* inherits all the logical properties of knowledge so that it is a mapping from Ω^* into $M(\Omega^*, \mathcal{D}^*)$. Indeed, the statement clearly holds because the inverse image h^{-1} commutes with set-algebraic operations.

1. Suppose that i's knowledge satisfies No-Contradiction Axiom. For any $\omega^* \in \Omega^*$, there are a knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$ such that $\omega^* = h(\omega)$, and thus $t_i^*(\omega^*)(\emptyset) = t_i(\omega)(h^{-1}(\emptyset)) = 0$.

2. Suppose that i's knowledge satisfies Consistency. Suppose that there are $\omega^* \in \Omega^*$ and $E^* \in \mathcal{D}^*$ such that $t_i^*(\omega^*)(E^*) = 1$. Then, there are a knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$ such that $\omega^* = h(\omega)$ and $t_i(\omega)(h^{-1}(E^*)) = 1$. By Consistency of t_i, we have $0 = t_i(\omega)(h^{-1}(\neg E^*)) = t_i(\omega)(\neg h^{-1}(E^*))$.

3. Suppose that i's knowledge satisfies Necessitation. For any $\omega^* \in \Omega^*$, there are a knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$ such that $\omega^* = h(\omega)$ and $1 = t_i(\omega)(\Omega) = t_i(\omega)(h^{-1}(\Omega^*)) = t_i^*(\omega^*)(\Omega^*)$.

4. Suppose that i's knowledge satisfies Monotonicity. Take any $\omega^* \in \Omega^*$ and $E^*, F^* \in \mathcal{D}^*$ with $E^* \subseteq F^*$. Now, there are a knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$ such that $\omega^* = h(\omega)$ and
$$t_i^*(\omega^*)(E^*) = t_i(\omega)(h^{-1}(E^*)) \leq t_i(\omega)(h^{-1}(F^*)) = t_i^*(\omega^*)(F^*),$$
where the weak inequality follows from $h^{-1}(E^*) \subseteq h^{-1}(F^*)$ and Monotonicity of t_i.

5. Suppose that i's knowledge satisfies Non-empty λ-Conjunction. Take any non-empty subset $\mathcal{F} \subseteq \mathcal{D}^*$ with $|\mathcal{F}| < \lambda$. Suppose that $t_i^*(\omega^*)(F^*) = 1$ for all $F^* \in \mathcal{F}$. Now, there are a knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$ such that $\omega^* = h(\omega)$ and $t_i(\omega)(h^{-1}(F^*)) = 1$ for all $F^* \in \mathcal{F}$. Since t_i satisfies Non-empty λ-Conjunction, we have $1 = t_i(\omega)(\bigcap_{F^* \in \mathcal{F}} h^{-1}(F^*)) = t_i(\omega)(h^{-1}(\bigcap \mathcal{F}))$, establishing $t_i^*(\omega^*)(\bigcap \mathcal{F}) = 1$.

In the fourth step, we show that $t_i^* : (\Omega^*, \mathcal{D}^*) \to (M(\Omega^*, \mathcal{D}^*), \mathcal{M}(\Omega^*, \mathcal{D}^*))$ is κ-measurable. Let $E^* = (\pi^\alpha|_{\Omega^*})^{-1}(E^\alpha)$ with $E^\alpha \in \mathcal{H}^\alpha$ and $\alpha < \kappa$. Then, we have

$$(t_i^*)^{-1}(\mathcal{M}(E^*)) = \{\omega^* \in \Omega^* \mid (\omega^*)_i^{\alpha+1}(E^\alpha) = 1\} = \{\omega^* \in \Omega^* \mid (\omega^*)_i^{\alpha+1} \in \mathcal{M}(E^\alpha)\}.$$

In the fifth step, we show that t_i^* inherits the introspective properties of knowledge.

1. Consider Truth Axiom. Fix $E^* \in \mathcal{D}^*$. If $t_i^*(\omega^*)(E^*) = 1$, then there exist a knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$ such that $\omega^* = h(\omega)$ and $1 = t_i^*(\omega^*)(E^*) = t_i(\omega)(h^{-1}(E^*))$. Now, Truth Axiom of t_i implies $\omega \in h^{-1}(E^*)$, and thus $\omega^* = h(\omega) \in E^*$.

2. Next, consider Positive Introspection. Fix $E^* \in \mathcal{D}^*$ and $\omega^* \in \Omega^*$ such that $t_i^*(\omega^*)(E^*) = 1$. Then, there are a knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$ such that $\omega^* = h(\omega)$, $1 = t_i^*(\omega^*)(E^*) = t_i(\omega)(h^{-1}(E^*))$, and $t_i^*(\omega^*)((t_i^*)^{-1}(\mathcal{M}_{E^*})) = t_i(\omega)(h^{-1}((t_i^*)^{-1}(\mathcal{M}_{E^*})))$. Now, Positive Introspection of t_i implies that $1 = t_i(\omega)(t_i^{-1}(\mathcal{M}_{h^{-1}(E^*)}))$. Next, we show that $t_i^{-1}(\mathcal{M}_{h^{-1}(E^*)}) = h^{-1}((t_i^*)^{-1}(\mathcal{M}_{E^*}))$:

$$\begin{aligned} \omega \in t_i^{-1}(\mathcal{M}_{h^{-1}(E^*)}) &\text{ iff } t_i(\omega)(h^{-1}(E^*)) = 1 \text{ iff } t_i^*(\omega^*)(E^*) = 1 \\ &\text{ iff } h(\omega) = \omega^* \in (t_i^*)^{-1}(\mathcal{M}_{E^*}) \text{ iff } \omega \in h^{-1}((t_i^*)^{-1}(\mathcal{M}_{E^*})). \end{aligned}$$

   Then, we obtain

$$1 = t_i(\omega)(t_i^{-1}(\mathcal{M}_{h^{-1}(E^*)})) = t_i(\omega)(h^{-1}((t_i^*)^{-1}(\mathcal{M}_{E^*}))) = t_i^*(\omega^*)((t_i^*)^{-1}(\mathcal{M}_{E^*})).$$

3. We show that t_i^* inherits Negative Introspection. Fix $E^* \in \mathcal{D}^*$ and $\omega^* \in \Omega^*$ with $t_i^*(\omega^*)(E^*) = 0$. Then, there are a knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$ such that $\omega^* = h(\omega)$ and $0 = t_i^*(\omega^*)(E^*) = t_i(\omega)(h^{-1}(E^*))$. Now, Negative Introspection of t_i implies that $1 = t_i(\omega)(\neg t_i^{-1}(\mathcal{M}_{h^{-1}(E^*)}))$. Then, it follows from the previous argument that $t_i^{-1}(\mathcal{M}_{h^{-1}(E^*)}) = h^{-1}((t_i^*)^{-1}(\mathcal{M}_{E^*}))$, and hence we obtain

$$1 = t_i(\omega)(\neg t_i^{-1}(\mathcal{M}_{h^{-1}(E^*)})) = t_i(\omega)(h^{-1}(\neg(t_i^*)^{-1}(\mathcal{M}_{E^*}))) = t_i^*(\omega^*)(\neg(t_i^*)^{-1}(\mathcal{M}_{E^*})).$$

4. We show that t_i^* inherits the Kripke property. It follows from Lemma A.2 that $h(b_{t_i}(\omega)) \subseteq b^*_{t_i^*}(h(\omega))$ for each $\omega \in \Omega$. Now, if $b^*_{t_i^*}(\omega^*) \subseteq E^*$, then there are a knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$ such that $\omega^* = h(\omega)$, and thus

$$b_{t_i}(\omega) \subseteq h^{-1}(h(b_{t_i}(\omega))) \subseteq h^{-1}(b^*_{t_i^*}(\omega^*)) \subseteq h^{-1}(E^*).$$

   By the Kripke property of $\overrightarrow{\Omega}$, it follows that $1 = t_i(\omega)(h^{-1}(E^*)) = t_i^*(\omega^*)(E^*)$.

□

Proof of Lemma 8. We show by induction that $(h^*)^\alpha : \Omega^* \to H^\alpha$ is the projection $\pi^\alpha|_{\Omega^*}$ for all ordinal $\alpha < \kappa$, where note that $(h^*)^\alpha = \pi^\alpha \circ h^*$. For $\alpha = 0$, we have $(h^*)^0 = \pi^0|_{\Omega^*}$. Let $\alpha = \beta + 1$ be a successor ordinal. Then, for each ω^*, we have

$$\begin{aligned}(h^*)^{\beta+1}(\omega^*) &= ((h^*)^\beta(\omega^*), t^*(\omega^*) \circ ((h^*)^\beta)^{-1}) = (\pi^\beta|_{\Omega^*}(\omega^*), t^*(\omega^*) \circ (\pi^\beta|_{\Omega^*})^{-1}) \\ &= (\pi^\beta|_{\Omega^*}(\omega^*), (\omega^*)^{\beta+1}) = \pi^{\beta+1}|_{\Omega^*}(\omega^*),\end{aligned}$$

where $t^*(\omega^*) \circ (\pi^\beta|_{\Omega^*})^{-1} = (\omega^*)^{\beta+1}$ follows from Equation (A.2). Then, we get $(h^*)^{\beta+1} = \pi^{\beta+1}|_{\Omega^*}$. For a limit ordinal α, the statement clearly holds by construction. The induction is complete. $\square$

Proof of Theorem 3. First, we establish that $\overrightarrow{\Omega^*}$ is universal. To that end, as is discussed in the main text, it is sufficient to show the uniqueness of a knowledge morphism $h : \overrightarrow{\Omega} \to \overrightarrow{\Omega^*}$. The uniqueness, however, follows from Lemma 8 as in the proof of Theorem 1.

Second, Lemma 8 also implies that that $(\Theta^*, (t_i^*)_{i\in I}) : \Omega^* \to \Omega^{**}$ is injective. Thus, we show that $(\Theta^*, (t_i^*)_{i\in I}) : \Omega^* \to \Omega^{**}$ is surjective. For any $(s, (\mu_i)_{i\in I}) \in \Omega^{**}$, there are a knowledge space $\overrightarrow{\Omega}$ and $\omega \in \Omega$ such that

$$(s, (\mu_i)_{i\in I}) = (\Theta(\omega), (t_i(\omega) \circ h^{-1})_{i\in I}) = (\Theta^*(h(\omega)), (t_i^*(h(\omega)))_{i\in I}).$$

$\square$

A.5.2 Proof of Remark 5 (Section 6.1)

We prove the statement in Remark 5 by applying Lambek's lemma (Lambek [47]) in category theory. To that end, we represent a knowledge space in terms of (i) an underlying structure $(\Omega, \mathcal{D})$ and (ii) a tuple of mappings $(\Theta, (t_i)_{i\in I}) : \Omega \to S \times M(\Omega)^I$ in a category theoretical manner.

We first define underlying structures of knowledge spaces. We let $\mathbf{B}_\kappa$ be the category of κ-complete (Boolean) algebra (of sets). That is, each $\mathbf{B}_\kappa$-object is a κ-complete algebra $(\Omega, \mathcal{D})$, and each $\mathbf{B}_\kappa$-morphism is a κ-measurable mapping $\varphi : (\Omega, \mathcal{D}) \to (\Omega', \mathcal{D}')$.

Next, we define $(\Theta, (t_i)_{i\in I}) : (\Omega, \mathcal{D}) \to (F(\Omega, \mathcal{D}), \mathcal{F}(\Omega, \mathcal{D}))$ as a κ-measurable mapping as follows. Let $(\Omega, \mathcal{D})$ be a $\mathbf{B}_\kappa$-object. We define $F(\Omega, \mathcal{D}) := S \times M(\Omega, \mathcal{D})^I$ and $\mathcal{F}(\Omega, \mathcal{D}) := \mathcal{A}_S \times \mathcal{M}(\Omega, \mathcal{D})^I$. We often denote $(F(\Omega), \mathcal{F}(\mathcal{D})) = (F(\Omega, \mathcal{D}), \mathcal{F}(\Omega, \mathcal{D}))$.

Note that we can rewrite $\mathcal{F}(\Omega, \mathcal{D})$ as follows. To that end, we define the projections $\pi_S : F(\Omega, \mathcal{D}) \to S$ and $\pi_i : F(\Omega, \mathcal{D}) \to M(\Omega, \mathcal{D})$ for each $i \in I$. Then, we have

$$\mathcal{F}(\Omega, \mathcal{D}) = \mathcal{A}_\kappa \left(\{\pi_S^{-1}(E_S) \mid E_S \in \mathcal{A}_S\} \cup \bigcup_{i\in I} \{\pi_i^{-1}(\mathcal{M}_E) \mid E \in \mathcal{D}\} \right).$$

For each $\mathbf{B}_\kappa$-morphism $\varphi : (\Omega, \mathcal{D}) \to (\Omega', \mathcal{D}')$, we define a mapping $F(\varphi)$ defined on $F(\Omega, \mathcal{D})$ by

$$F(\varphi)(s, (\mu_i)_{i\in I}) := (s, (\mu_i \circ \varphi^{-1})_{i\in I}),$$

where note that $(\mu_i \circ \varphi^{-1})(E') := \mu_i(\varphi^{-1}(E'))$ for each $E' \in \mathcal{D}'$.

We show that $F : \mathbf{B}_\kappa \to \mathbf{B}_\kappa$ defined above is an endofunctor in the following three steps. In the first step, we show that $F(\varphi)$ is a mapping from $F(\Omega, \mathcal{D})$ into $F(\Omega', \mathcal{D}')$ for a given $\mathbf{B}_\kappa$-morphism $\varphi : (\Omega, \mathcal{D}) \to (\Omega', \mathcal{D}')$. This is, however, clear because the inverse image preserves the logical properties.

In the second step, we show that $F(\varphi) : (F(\Omega), \mathcal{F}(\mathcal{D})) \to (F(\Omega'), \mathcal{F}(\mathcal{D}'))$ is a $\mathbf{B}_\kappa$-morphism (a κ-measurable mapping) for each κ-measurable mapping $\varphi : (\Omega, \mathcal{D}) \to (\Omega', \mathcal{D}')$. It is enough to show that the inverse $(F(\varphi))^{-1}$ maps generators of $\mathcal{F}(\mathcal{D}')$ into $\mathcal{F}(\mathcal{D})$. To that end, consider first $\mathcal{A}_S$. For each $(\pi'_S)^{-1}(E_S)$ with $E_S \in \mathcal{A}_S$, we have

$$(F(\varphi))^{-1}((\pi'_S)^{-1}(E_S)) = \pi_S^{-1}(E_S) \in \mathcal{F}(\mathcal{D}).$$

Next, for each $(\pi'_i)^{-1}(\{\mu' \in M(\Omega') \mid \mu'(E') = 1\})$ with $i \in I$ and $E' \in \mathcal{D}'$, we have

$$(F(\varphi))^{-1}((\pi'_i)^{-1}(\{\mu' \in M(\Omega') \mid \mu'(E') = 1\})) = (\pi_i)^{-1}(\{\mu \in M(\Omega) \mid \mu(\varphi^{-1}(E')) = 1\}) \in \mathcal{F}(\mathcal{D}).$$

This establishes that $F(\varphi)$ is κ-measurable.

In the third step, we start with showing that $F(\psi \circ \varphi) = F(\psi) \circ F(\varphi)$. Indeed, we have

$$\begin{aligned} F(\psi \circ \varphi)(s, (\mu_i)_{i\in I}) &= (s, (\mu_i \circ (\psi \circ \varphi)^{-1})_{i\in I}) = (s, (\mu_i(\varphi^{-1}(\psi^{-1}(\cdot))))_{i\in I}) \\ &= F(\psi) \circ F(\varphi). \end{aligned}$$

Next, it is easily seen that $F(\mathrm{id}_\Omega) = \mathrm{id}_{F(\Omega)}$. Thus, $F : \mathbf{B}_\kappa \to \mathbf{B}_\kappa$ is an endofunctor.

Recall that a knowledge space of I on $(S, \mathcal{A}_S)$ (in terms of knowledge-type mappings) is a tuple $\langle (\Omega, \mathcal{D}), (t_i)_{i\in I}, \Theta \rangle$ where $(\Omega, \mathcal{D})$ is a κ-measurable space (i.e., a $\mathbf{B}_\kappa$-object) and $(\Theta, (t_i)_{i\in I}) : (\Omega, \mathcal{D}) \to (F(\Omega, \mathcal{D}), \mathcal{F}(\Omega, \mathcal{D}))$ is a κ-measurable mapping (i.e., a $\mathbf{B}_\kappa$-morphism) such that each type mapping t_i satisfies the required introspective (i.e., Truth Axiom, Positive Introspection, and Negative Introspection) and Kripke properties.[A.2]

A knowledge morphism $\varphi : \langle (\Omega, \mathcal{D}), (t_i)_{i\in I}, \Theta \rangle \to \langle (\Omega', \mathcal{D}'), (t'_i)_{i\in I}, \Theta' \rangle$ (where both knowledge spaces reside in the same class of knowledge spaces) can be seen as a κ-measurable mapping $\varphi : (\Omega, \mathcal{D}) \to (\Omega, \mathcal{D})$ such that $(\Theta', (t'_i)_{i\in I}) \circ \varphi = F(\varphi) \circ (\Theta, (t_i)_{i\in I})$. In other words,

$$(\Theta'(\varphi(\omega)), (t'_i(\varphi(\omega))(\cdot))_{i\in I}) = (\Theta(\omega), (t_i(\omega)(\varphi^{-1}(\cdot)))_{i\in I}) \text{ for each } \omega \in \Omega.$$

In the language of category theory, a pair $\langle (\Omega, \mathcal{D}), (\Theta, (t_i)_{i\in I}) \rangle$ is a coalgebra over the endofunctor F. Thus, the category of knowledge(-type) spaces is seen as the full subcategory of F-coalgebras such that each t_i satisfies the given introspective and/or Kripke properties.

Now, we show that if no introspective property nor the Kripke property is imposed in a given class of knowledge spaces then the universal knowledge space $\overrightarrow{\Omega^*}$ studied in Section 6.1 is isomorphic to $S \times M(\Omega^*, \mathcal{D}^*)^I$.

[A.2] We simply denote the product mapping by $(\Theta, t)(\omega) := (\Theta, (t_i)_{i\in I})(\omega) := (\Theta(\omega), (t_i(\omega))_{i\in I}) \in F(\Omega) = S \times M(\Omega)^I$ for each $\omega \in \Omega$.

Proof of Remark 5. The assertion simply follows from the Lambek lemma (Lambek [47]) in category theory. For completeness, we provide a proof below.

$$\begin{tikzcd}
(\Omega^*, \mathcal{D}^*) \arrow[r, "(\Theta^*, t^*)"] \arrow[d, "(\Theta^*, t^*)"'] & F(\Omega^*, \mathcal{D}^*) \arrow[r, "h"] \arrow[d, "F(\Theta^*, t^*)"'] & (\Omega^*, \mathcal{D}^*) \arrow[d, "(\Theta^*, t^*)"] \\
F(\Omega^*, \mathcal{D}^*) \arrow[r, "F(\Theta^*, t^*)"'] & F^2(\Omega^*, \mathcal{D}^*) \arrow[r, "F(h)"'] & F(\Omega^*, \mathcal{D}^*)
\end{tikzcd}$$

First, $\langle (F(\Omega^*), \mathcal{F}(\mathcal{D}^*)), F(\Theta^*, (t_i^*)_{i\in I})\rangle$ is a knowledge space. Since $\overrightarrow{\Omega^*}$ is universal, let $h : F(\Omega^*) \to \Omega^*$ be the description map. Since there is a unique knowledge morphism from $\overrightarrow{\Omega^*}$ into itself and since it is the identity map, we obtain $h \circ (\Theta^*, t^*) = \mathrm{id}_{\Omega^*}$. Next, we also have

$$F(h) \circ F(\Theta^*, t^*) = F(h \circ (\Theta^*, t^*)) = F(\mathrm{id}_{\Omega^*}) = \mathrm{id}_{F(\Omega^*, \mathcal{D}^*)}.$$

Then, we get $(\Theta^*, t^*) \circ h = F(h) \circ F(\Theta^*, t^*) = \mathrm{id}_{F(\Omega^*, \mathcal{D}^*)}$. Hence, $(\Theta^*, t^*) := (\Theta^*, (t_i^*)_{i\in I})$ is a knowledge isomorphism and its inverse is h. □

A.5.3 Section 6.2 (Proof of Theorem 4)

Proof of Theorem 4. Step 1. First, consider the universal knowledge space $\overrightarrow{\Omega^*}$ established in Section 6.1. By definition, Ω^* is a subset of H. The κ-complete algebra $\mathcal{D}^*$ and the mapping Θ^* are defined as in Definition 17. Second, it is also clear that Ω^* respects the required introspective properties of knowledge. Third, we have already shown that Ω^* is coherent in establishing that $\overrightarrow{\Omega^*}$ is universal.

Step 2. Let Ω be a coherent subset of H which respects the required introspective properties. First, we show that $\langle (\Omega, \mathcal{D}), (t_i)_{i\in I}, \Theta\rangle$ is a knowledge space in a given category. To do so, it is enough to show that t_i satisfies the required logical properties of knowledge.

1. Consider No-Contradiction Axiom. We have $t_i(s, \mu)(\emptyset) = \mu_i^\alpha(\emptyset) = 0$.

2. Consider Necessitation. We have $t_i(s, \mu)(\Omega) = \mu_i^\alpha(H^\alpha) = 1$.

3. Consider Consistency. If $t_i(s, \mu)((\pi^\alpha|_\Omega)^{-1}(E^\alpha)) = 1$ and $t_i(s, \mu)(\neg(\pi^\alpha|_\Omega)^{-1}) = 1$, then $\mu_i^\alpha(E^\alpha) = 1$ and $\mu_i^\alpha(\neg E^\alpha) = 1$, which contradicts the assumption that μ_i^α is a knowledge-type satisfying Consistency (i.e., $\mu_i^\alpha \in M(H^\alpha, \mathcal{H}^\alpha)$).

4. Consider Monotonicity. Suppose that $(\pi^{\alpha}|_{\Omega})^{-1}(E^{\alpha}) \subseteq (\pi^{\beta}|_{\Omega})^{-1}(F^{\beta})$. Without loss of generality, we can assume as if $\alpha = \beta$. Since $E^{\alpha} \subseteq F^{\beta}$, we have
$$t_i(s,\mu)((\pi^{\alpha}|_{\Omega})^{-1}(E^{\alpha})) = \mu_i^{\alpha}(E^{\alpha}) \leq \mu_i^{\beta}(F^{\beta}) = t_i(s,\mu)((\pi^{\beta}|_{\Omega})^{-1}(F^{\beta})).$$

5. We consider Non-empty λ-Conjunction. Without loss of generality, suppose that $t_i(t)((\pi^{\alpha}|_{\Omega})^{-1}(E^{\alpha})) = 1$ for all $E^{\alpha} \in \mathcal{E}$, where $\mathcal{E}$ is a non-empty subset of $\mathcal{H}^{\alpha}$ with $|\mathcal{E}| < \lambda$. Then, we obtain
$$t_i(s,\mu)(\bigcap_{E^{\alpha}\in\mathcal{E}} (\pi^{\alpha}|_{\Omega})^{-1}(E^{\alpha})) = t_i(s,\mu)((\pi^{\alpha}|_{\Omega})^{-1}(\bigcap \mathcal{E})) = \mu_i^{\alpha}(\bigcap \mathcal{E}) = 1.$$

Second, we show that the description map $h : \overrightarrow{\Omega} \to \overrightarrow{\Omega^*}$ is an inclusion map. For $\alpha = 0$, we have $h^0 = \pi^0|_{\Omega}$. For a successor ordinal $\alpha = \beta + 1$, since $t(s,\mu) \circ (\pi^{\beta}|_{\Omega})^{-1} = \mu^{\beta}$, we have $h^{\alpha} = \pi^{\alpha}|_{\Omega}$. For a limit ordinal α, by construction, if $h^{\beta} = \pi^{\beta}|_{\Omega}$ for all $\beta < \alpha$ then $h^{\alpha} = \pi^{\alpha}|_{\Omega}$.

□

A.6 Section 7

Proof of Theorem 5. Construct Ω^* by collecting all the dynamic knowledge-belief spaces as in the proof of Theorem 1, where note the following. First, Ω^* is not empty because we can consider a dynamic knowledge-belief space $\overrightarrow{\{s\}}$ as in Section 3.1. Second, it follows from Lemma A.1 that the space $\overrightarrow{\Omega^*}$ satisfies the specified properties of knowledge and belief including perfect recall (when it is assumed). □

Proof of Theorem 6. First, our entire arguments in Section 3.1 establish a universal knowledge-unawareness space, where the universal knowledge-unawareness space inherits the properties of knowledge and unawareness by Lemma A.1. It is non-empty because we can consider a knowledge-unawareness space $\overrightarrow{\{s\}}$ as in Section 3.1.

Second, suppose that, in a given category, there is a knowledge-unawareness space $\overrightarrow{\Omega}$ such that $U_i([\![e]\!]_{\overrightarrow{\Omega}}) \neq \emptyset$ for some $i \in I$ and $e \in \mathcal{L}$. Then, there is $\omega \in U_i([\![e]\!]_{\overrightarrow{\Omega}}) = D^{-1}(U_i^*([e]))$ and thus $D(\omega) \in U_i^*([e])$. Hence, $\overrightarrow{\Omega^*}$ is non-trivial. □

A.7 Further Knowledge Representations in Knowledge Spaces (Supplementary Appendix to Section 4)

A.7.1 Generalized Posibility Correspondences

The Kripke property certainly generalizes the idea of the previous possibility correspondence models of knowledge in the sense that each player's knowledge at each state ω is characterized by her information set $b_{K_i}(\omega)$. The information set, however, is not necessarily an event. Moreover, a player whose knowledge satisfies the Kripke property is logically omniscient.[A.3]

We generalize standard possibility correspondence models in the following ways. First, we retain a spirit of the Kripke property in the sense that each player's knowledge is logically entailed from her information. Second, we require information that represents each player's knowledge to be an object of knowledge. Third, while the first requirement pre-supposes Monotonicity, we aim to dispense with Arbitrary Conjunction (including Necessitation) endemic in the Kripke property.

Specifically, a (generalized) possibility correspondence associates, with each sate of the world ω, a collection of events $\mathcal{B}_i(\omega)$ that can be a source/generator of knowledge at the state ω in the following sense: an agent knows an event E at a state ω if there is an event $F \in \mathcal{B}_i(\omega)$ which is contained in E. In order to distinguish between standard and generalized possibility correspondences, hereafter, we call such a generalized possibility correspondence to be an *information correspondence*. While we require Monotonicity by the very definition, it turns out that this information correspondence approach works in any domain.

At the conceptual level, $\mathcal{B}_i(\omega)$ is no longer interpreted as the set of states considered possible at ω. Each collection $\mathcal{B}_i(\omega)$ of events can be understood as the collection of information available to player i at state ω. Thus, we say that E is i's information set at ω if $E \in \mathcal{B}_i(\omega)$.

Before we go on to the formal analysis, we discuss the following two closely related approaches. First, Fagin and Halpern [27] take the "society-of-mind" (or "local reasoning") approach, where an agent, endowed with a collection of possibility sets $\mathcal{B}_i(\omega)$ at each ω, considers one possibility set possible at each time when she makes inferences.[A.4]

Second, Doignon and Falmagne [22, 23] study what they call "surmise systems" in the psychology literature. A surmise system $\mathcal{B}_i$ encodes all possible (thus not necessarily unique) ways to making inferences about the world at each state.[A.5] See Section A.7.4 for more details.

We start with the following preliminary notation. Throughout the subsection, fix a

[A.3] DLR [21] show that logical omniscience inherent in the standard possibility correspondence models precludes sensible unawareness under certain conditions.

[A.4] See also the references therein. Especially, both approaches are related to the neighborhood (Montague-Scott) systems approach in modal logic, provided that Monotonicity is assumed (Fagin and Halpern [27, Footnote 9]). In our framework, this approach amounts to specifying the collection of events that a player i knows at each state ω. Our exposition in this subsection works by slight modifications if each player's knowledge is given by the collection of events that she knows at each state.

[A.5] In light of inferences, our idea is also related to that of Shin [80]. He identifies the notion of knowledge with the logical provability, where a player knows an event when she can prove it from her "basic knowledge" through use of propositional logic. In our framework, we can also interpret the knowledge of an E as the fact that she can "prove" E from one of her possibility set $F \in \mathcal{B}_i(\omega)$ (which could be incorrect in the sense that $\omega \notin F$), where provability is simply taken as the set inclusion in the form of $F \subseteq E$.

κ-complete algebra $(\Omega, \mathcal{D})$. For any subset Γ of $\mathcal{D}$, we define

$$\uparrow \Gamma := \{E \in \mathcal{D} \mid \text{there is } F \in \Gamma \text{ with } F \subseteq E\}.$$

If Γ is player's information (at a particular state) then $E \in \uparrow \Gamma$ means that there is information $F \in \Gamma$ which entails E. Note that $\uparrow \Gamma$ is closed under Monotonicity in the sense that $\uparrow\uparrow \Gamma \subseteq \uparrow \Gamma$.

An information correspondence on $(\Omega, \mathcal{D})$ is a mapping $\mathcal{B}_i$ from Ω into a subset of $\mathcal{P}(\mathcal{D})$ which satisfies the assumptions on i's knowledge and the following regularity condition: for each $E \in \mathcal{D}$,

$$K_{\mathcal{B}_i}(E) := \{\omega \in \Omega \mid E \in \uparrow \mathcal{B}_i(\omega)\} \in \mathcal{D}. \tag{A.3}$$

$K_{\mathcal{B}_i}(E)$ is interpreted as the set of states at which player i has information to support E. Thus, we define the knowledge operator $K_{\mathcal{B}_i} : \mathcal{D} \to \mathcal{D}$ derived from the information correspondence $\mathcal{B}_i$ through Equation (A.3).

We have three remarks regarding the definition of $K_{\mathcal{B}_i}$. First, observe that $\uparrow \mathcal{B}_i(\omega)$ is exactly the collection of events that player i knows at ω. Put differently, $\omega \in K_{\mathcal{B}_i}(E)$ iff $E \in \uparrow \mathcal{B}_i(\omega)$.

Second, the event that i considers E possible can be written as

$$L_{\mathcal{B}_i}(E) = \{\omega \in \Omega \mid F \cap E^c \neq \emptyset \text{ for all } F \in \mathcal{B}_i(\omega)\}.$$

That is, it is the set of states ω such that i's information is not inconsistent with E.

Third, if $\mathcal{B}_i$ is singleton-valued (i.e., if $\mathcal{B}_i(\cdot) = \{b_i(\cdot)\}$) then it reduces to the (standard) possibility correspondence such (i) that each information/possibility set $b_i(\omega)$ is an event and (ii) that $\mathcal{B}_i$ satisfies the regularity condition (i.e., $\{\omega \in \Omega \mid b_i(\omega) \subseteq E\} \in \mathcal{D}$ for each $E \in \mathcal{D}$). We can dispense with Non-empty Conjunction by having multiple possibility sets while we can dispense with Necessitation by allowing the correspondence to be empty-valued. Suppose, for example, that $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and $\mathcal{D} = \mathcal{P}(\Omega)$. Let $\mathcal{B}_i$ be such that $\mathcal{B}_i(\omega_1) = \{\{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}\}$ and $\mathcal{B}_i(\omega_2) = \mathcal{B}_i(\omega_3) = \emptyset$. Then an agent i knows $\{\omega_1, \omega_2\}$ and $\{\omega_1, \omega_3\}$ at state ω_1 but shes does not (cannot) know $\{\omega_1, \omega_2\} \cap \{\omega_1, \omega_3\}$ at that state. At state ω_j (with $j \in \{2, 3\}$), she does not know anything at all, and thus $K_{\mathcal{B}_i}(\Omega) = \{\omega_1\}$.

As we have established the equivalence between knowledge-operator and knowledge-type approaches in Section 5.1, we formalize a class of knowledge spaces (that satisfy Monotonicity) in terms of information correspondences for a given notion of knowledge in the following three steps.

The first step is to define a collection of information sets that information correspondences can take (i.e., the codomain of information correspondences $\mathcal{B}_i$, which is a subset of $\mathcal{P}(\mathcal{D})$). The arguments are parallel to defining the set of knowledge-types $M(\Omega, \mathcal{D})$ in the knowledge-type mapping approach. To that end, we call a subset Γ of $\mathcal{P}(\mathcal{D})$ to be an information collection.

Definition A.1 (Logical Properties of an Information Collection)**.** *Let Γ be a subset of $\mathcal{P}(\mathcal{D})$.*

1. *Γ satisfies No-Contradiction Axiom if $\emptyset \notin\uparrow \Gamma$.*

2. *Γ is serial (satisfies Consistency) if $E^c \notin\uparrow \Gamma$ for any $E \in\uparrow \Gamma$.*

3. *Γ satisfies Necessitation if $\Omega \in\uparrow \Gamma$.*

4. *Γ satisfies Non-empty λ-Conjunction if $\uparrow \Gamma$ is closed under non-empty λ-intersection (i.e., $\bigcap \mathcal{F} \in\uparrow \Gamma$ for any $\mathcal{F} \subseteq\uparrow \Gamma$ with $0 < |\mathcal{F}| < \lambda(\leq \kappa)$).*

We formalize the logical properties in terms of $\uparrow \Gamma$ (instead of the primitive Γ), because $\uparrow \Gamma$ corresponds to the collection of events that an agent knows by making inferences from Γ. Thus, it is clear that the above definition embodies our intended definitions of logical properties of knowledge. For example, Consistency of Γ means that if an agent "Γ-knows" E then she does not "Γ-know" its negation E^c. Another reason is that the above definition can also be used when each player's knowledge is expressed directly in terms of the collection of events that she knows at each state. Yet we provide a straightforward restatement of logical properties in terms of Γ.

Proposition A.6 (Logical Properties of an Information Collection)**.** *Let Γ be a subset of $\mathcal{P}(\mathcal{D})$.*

1. *Γ satisfies No-Contradiction Axiom iff $\emptyset \notin \Gamma$.*

2. *Γ is satisfies Consistency iff $E \cap F \neq \emptyset$ for any $E, F \in \Gamma$.*

3. *Γ satisfies Necessitation iff $\Gamma \neq \emptyset$.*

4. *Γ satisfies Non-empty λ-Conjunction iff, for any $\mathcal{F} \subseteq \Gamma$ with $0 < |\mathcal{F}| < \lambda(\leq \kappa)$, there is $F \in \Gamma$ with $F \subseteq \bigcap \mathcal{F}$.*

Proposition A.6 states the following. If a given information collection Γ does not contain the empty set ("a contradiction") then no contradiction is entailed from Γ. Γ satisfies Consistency when any pair of information $(E, F) \in \Gamma^2$ is not contradictory. Γ satisfies Necessitation when it contains some information in the sense that any tautology is inferred from it. Γ satisfies Non-empty Conjunction when, for any given family of information, the information collection Γ is rich enough to have another information which implies the conjunction of the given family.

In the second step, given a pre-determined concept of knowledge, let $S(\Omega, \mathcal{D})(= S(\Omega))$ be a family of all information collections which satisfy all the implied logical properties. Given such family $S(\Omega)$, we introduce a κ-complete algebra $\mathcal{S}(\Omega, \mathcal{D})(= \mathcal{S}(\mathcal{D}))$ by $\mathcal{S}(\mathcal{D}) := \mathcal{A}_\kappa(\{\mathcal{S}_E \in \mathcal{P}(S(\Omega)) \mid E \in \mathcal{D}\})$, where

$$\mathcal{S}_E := \{\Gamma \in S(\Omega) \mid E \in\uparrow \Gamma\}.$$

As the third and last step, we define the properties of knowledge on a correspondence $\mathcal{B}_i : (\Omega, \mathcal{D}) \to (S(\Omega), \mathcal{S}(\mathcal{D}))$. First, for all the logical properties of knowledge except for

Monotonicity, we say that $\mathcal{B}_i$ satisfies a given logical property if the information set $\mathcal{B}_i(\omega)$ satisfies it for all $\omega \in \Omega$. For example, $\mathcal{B}_i$ satisfies *No-Contradiction Axiom* if each $\mathcal{B}_i(\omega)$ satisfies it (i.e., $\emptyset \not\in \uparrow \mathcal{B}_i(\omega)$). Consistency, Necessitation, and Non-empty λ-Conjunction of $\mathcal{B}_i$ are defined in the similar way. Next, we define the introspective and Kripke properties of $\mathcal{B}_i$ as follows.

Definition A.2 (Introspective Properties of an Information Correspondence)**.** *Fix a κ-measurable mapping $\mathcal{B}_i : (\Omega, \mathcal{D}) \to (S(\Omega), \mathcal{S}(\mathcal{D}))$.*

1. *$\mathcal{B}_i$ is reflexive (satisfies Truth Axiom) if, $E \in \uparrow \mathcal{B}_i(\omega)$ implies $\omega \in E$ (i.e., $\uparrow \mathcal{B}_i(\omega) \subseteq \{E \in \mathcal{D} \mid \omega \in E\}$) for any $\omega \in \Omega$.*

2. *$\mathcal{B}_i$ is transitive (satisfies Positive Introspection) if $\{\omega' \in \Omega \mid E \in \uparrow \mathcal{B}_i(\omega')\} \in \uparrow \mathcal{B}_i(\omega)$ for any $\omega \in \Omega$ and $E \in \uparrow \mathcal{B}_i(\omega)$.*

3. *$\mathcal{B}_i$ is Euclidean (satisfies Negative Introspection) if the following holds: if $E \not\in \uparrow \mathcal{B}_i(\omega)$ for some $(\omega, E) \in \Omega \times \mathcal{D}$ then $\{\omega' \in \Omega \mid E \not\in \uparrow \mathcal{B}_i(\omega')\} \in \uparrow \mathcal{B}_i(\omega)$.*

4. *$\mathcal{B}_i$ satisfies the Kripke property if $\uparrow \mathcal{B}_i(\omega) = \{E \in \mathcal{D} \mid \bigcap \uparrow \mathcal{B}_i(\omega) \subseteq E\}$ for each $\omega \in \Omega$.*

Again, given that the fact that $E \in \uparrow \mathcal{B}_i(\omega)$ is intended to mean that E is known at ω, it is clear that the above definition embodies the intended introspective properties. Note, however, that $\bigcap \uparrow \mathcal{B}_i(\omega)$ may not be an event. In other words, it is not necessarily the case that $\uparrow \mathcal{B}_i(\omega)$ consists of the super-sets of $\bigcap \uparrow \mathcal{B}_i(\omega)$. Now, we restate the introspective properties in terms of $\mathcal{B}_i$.

Proposition A.7 (Introspective Properties of an Information Correspondence)**.** *Fix a κ-measurable mapping $\mathcal{B}_i : (\Omega, \mathcal{D}) \to (S(\Omega), \mathcal{S}(\mathcal{D}))$.*

1. *$\mathcal{B}_i$ is reflexive iff $E \in \mathcal{B}_i(\omega)$ implies $\omega \in E$ for any $\omega \in \Omega$.*

2. *$\mathcal{B}_i$ is transitive iff, for any $(\omega, E) \in \Omega \times \mathcal{D}$ with $E \in \mathcal{B}_i(\omega)$, there is $F \in \mathcal{B}_i(\omega)$ such that if $\omega' \in F$ then there is $E' \in \mathcal{B}_i(\omega')$ with $E' \subseteq E$.*

3. *$\mathcal{B}_i$ is Euclidean iff the following holds. If there is $(\omega, E) \in \Omega \times \mathcal{D}$ such that $E^c \cap F \neq \emptyset$ for all $F \in \mathcal{B}_i(\omega)$, then there is $F' \in \mathcal{B}_i(\omega)$ such that if $\omega' \in F'$ then $E^c \cap F \neq \emptyset$ for any $F \in \mathcal{B}_i(\omega')$.*

We interpret Proposition A.7 as follows. First, $\mathcal{B}_i$ is reflexive when i's information is always correct at each state. Second, $\mathcal{B}_i$ is transitive when, for any information E available to i at a state, there is another information F available to her at the same state (which can possibly be E itself) such that E is always supported as long as F is true. Third, $\mathcal{B}_i$ is Euclidean when, if i considers E possible at a state for some event E, then she has information F at the same state implying that i considers E possible whenever F is true.

As a further remark, assume the standard case where $\mathcal{B}_i(\cdot) = \{b_i(\cdot)\}$. First, $\mathcal{B}_i$ is reflexive iff $\omega \in b_i(\omega)$. Second, $\mathcal{B}_i$ is transitive iff $b_i(\omega') \subseteq b_i(\omega)$ for any $\omega' \in b_i(\omega)$. Third, $\mathcal{B}_i$ is Euclidean iff $\omega' \in b_i(\omega)$ implies $b_i(\omega) \subseteq b_i(\omega')$.[A.6]

In conclusion, given pre-determined assumptions on players' knowledge, a player i's information correspondence is a κ-measurable mapping $\mathcal{B}_i : (\Omega, \mathcal{D}) \to (S(\Omega), \mathcal{S}(\mathcal{D}))$ which satisfies the pre-determined properties of knowledge.

Now, we prove the equivalence between knowledge-operator and information correspondence approaches under Monotonicity. First, if an information correspondence $\mathcal{B}_i : (\Omega, \mathcal{D}) \to (S(\Omega), \mathcal{S}(\mathcal{D}))$ is given, then it is clear that $K_{\mathcal{B}_i} : \mathcal{D} \to \mathcal{D}$ inherits the all the properties of knowledge imposed on $\mathcal{B}_i$.

Second, we define an information correspondence from a given knowledge operator $K_i : \mathcal{D} \to \mathcal{D}$ which (at least) satisfies Monotonicity in such a way that the knowledge operator induced by $\mathcal{B}_i$ coincides with the original knowledge operator. Formally, we say that, for a given knowledge operator $K_i : \mathcal{D} \to \mathcal{D}$, an information correspondence $\mathcal{B}_i : (\Omega, \mathcal{D}) \to (S(\Omega), \mathcal{S}(\mathcal{D}))$ *induces* K_i (or $\mathcal{B}_i$ is a *generator* of K_i) if $K_i = K_{\mathcal{B}_i}$.

The simplest way to find a generator of the given knowledge operator K_i (which satisfies Monotonicity) is to consider the information correspondence $\mathcal{B}_{K_i} : (\Omega, \mathcal{D}) \to (S(\Omega), \mathcal{S}(\mathcal{D}))$ defined by

$$\mathcal{B}_{K_i}(\omega) := \{E \in \mathcal{D} \mid \omega \in K_i(E)\} \text{ for each } \omega \in \Omega. \tag{A.4}$$

Since K_i satisfies Monotonicity, it follows that $K_i = K_{\mathcal{B}_{K_i}}$. We summarize the entire arguments in the following.

Proposition A.8 (Equivalence Between Information Correspondence and Knowledge Operator under Monotonicity). *Let $\overrightarrow{\Omega}$ be a κ-knowledge space such that Monotonicity is assumed for every player.*

1. *Let $\mathcal{B}_i : (\Omega, \mathcal{D}) \to (S(\Omega), \mathcal{S}(\mathcal{D}))$ be an information correspondence. Then, $K_{\mathcal{B}_i}$ inherits the properties of knowledge imposed on $\mathcal{B}_i$. Furthermore, $\uparrow \mathcal{B}_i(\omega) = \mathcal{B}_{K_{\mathcal{B}_i}}(\omega)$ for all $\omega \in \Omega$, where $\mathcal{B}_{K_{\mathcal{B}_i}}$ is defined as in Equation (A.4).*

2. *Given a knowledge operator K_i, let $\mathcal{B}_i : (\Omega, \mathcal{D}) \to (S(\Omega), \mathcal{S}(\mathcal{D}))$ be a generator of K_i. Then, $\mathcal{B}_i$ inherits the properties of knowledge imposed on K_i. Also, $\mathcal{B}_{K_i}$ defined as in Equation (A.4) is one generator of K_i.*

We make three additional remarks. First, if K_i satisfies Positive Introspection as well as Monotonicity, then we can restrict attention to self-evident events so that the following information correspondence $\mathcal{B}_{K_i}$ is also a generator:

$$\mathcal{B}_{K_i}(\omega) = \{E \in \mathcal{D} \mid \omega \in K_i(E) \text{ and } E \subseteq K_i(E)\}.$$

[A.6] Clearly, the Euclidean property of $\mathcal{B}_i$ is given as follows: if there is $E \in \mathcal{D}$ such that $E^c \cap b_i(\omega) \neq \emptyset$ and if $\omega' \in b_i(\omega)$, then $E^c \cap b_i(\omega') \neq \emptyset$. This is obviously equivalent to the statement in question.

Second, a given knowledge operator K_i generally has multiple generators. Any information correspondences $\mathcal{B}_i$ and $\mathcal{B}'_i$ satisfying $\uparrow \mathcal{B}_i(\cdot) = \uparrow \mathcal{B}'_i(\cdot)$ induce the same knowledge operator $K_{\mathcal{B}_i} = K_{\mathcal{B}'_i}$.

Third, information correspondences characterize knowledgeability as follows. Let $\mathcal{B}_i$ and $\mathcal{B}_j$ be generators of K_i and K_j, respectively. Then, since $\{E \in \mathcal{D} \mid \omega \in K_i(E)\} = \{E \in \mathcal{D} \mid E \in \uparrow \mathcal{B}_i(\omega)\}$, we have $K_i(\cdot) \subseteq K_j(\cdot)$ iff $\uparrow \mathcal{B}_i(\cdot) \subseteq \uparrow \mathcal{B}_j(\cdot)$.

We conclude this subsection by characterizing knowledge morphisms in terms of information correspondences.

Proposition A.9 (Characterization of Knowledge Morphisms in terms of Information Correspondences)**.** *Let $\varphi : \Omega \to \Omega'$ be a mapping between knowledge spaces $\overrightarrow{\Omega}$ and $\overrightarrow{\Omega'}$ satisfying Monotonicity. Condition (3) in Definition 3 is written in terms of information correspondences as follows. Fix $\omega \in \Omega$.*

1. *For any $E' \in \mathcal{B}'_i(\varphi(\omega))$, there is $E \in \mathcal{B}_i(\omega)$ such that $\varphi(E) \subseteq E'$; and*

2. *For any $E \in \mathcal{B}_i(\omega)$ and $E' \in \mathcal{D}'$ with $\varphi(E) \subseteq E'$, there is $F' \in \mathcal{B}'_i(\varphi(\omega))$ such that $F' \subseteq E'$.*

As an immediate corollary, if $\mathcal{B}_i(\cdot) = \{b_i(\cdot)\}$, then the above conditions are re-written as: (i) $\varphi(b_i(\omega)) \subseteq b'_i(\varphi(\omega))$; and (ii) $b'_i(\varphi(\omega)) \subseteq E'$ for any $E' \in \mathcal{D}'$ with $\varphi(b_i(\omega)) \subseteq E'$.

A.7.2 Proofs for Section A.7.1

Proof of Proposition A.6.

1. If $\emptyset \not\in \uparrow \Gamma$, then $\emptyset \not\in \Gamma$ because $\Gamma \subseteq \uparrow \Gamma$. Conversely, if $\emptyset \in \uparrow \Gamma$ then there is $E \in \Gamma$ with $E \subseteq \emptyset$, i.e., $\emptyset \in \Gamma$. Colloquially, if a contradiction is logically entailed from a given information collection then it has to contain a contradiction.

2. Suppose that Γ satisfies Consistency. Suppose to the contrary that there are $E, F \in \Gamma$ with $E \cap F = \emptyset$. Then, since $F \subseteq E^c$, it follows that $E, E^c \in \uparrow \Gamma$, a contradiction. Conversely, if $E \in \uparrow$ then there is $E' \in \Gamma$ such that $E' \subseteq E$. If $E^c \in \uparrow \Gamma$ then there is $F' \in \Gamma$ such that $F' \subseteq E^c$ and thus $E' \cap F' \subseteq E \cap E^c$, which is impossible. Hence, $E^c \not\in \uparrow \Gamma$.

3. If $\Gamma \neq \emptyset$ then there is $E \in \Gamma$. Since $E \subseteq \Omega$, we have $\Omega \in \uparrow \Gamma$. Conversely, if $\Omega \in \uparrow \Gamma$ then there is $E \in \mathcal{D}$ with $E \in \Gamma$, and thus $\Gamma \neq \emptyset$.

4. Suppose that Γ satisfies Non-empty λ-Conjunction. Let $\mathcal{F} \subseteq \Gamma$ be with $0 < |\mathcal{F}| < \lambda$. Then, since $\Gamma \subseteq \uparrow \Gamma$, we have $\bigcap \mathcal{F} \in \uparrow \Gamma$, i.e., there is $F \in \Gamma$ with $F \subseteq \bigcap \mathcal{F}$. Conversely, let $\mathcal{F} \subseteq \uparrow \Gamma$ be with $0 < |\mathcal{F}| < \lambda$. For each $F \in \mathcal{F}$, there is $F' \in \Gamma$ such that $F' \subseteq F$. Let $\mathcal{F}'$ be a collection of such F' for each $F \in \mathcal{F}$. Then, there is $F \in \mathcal{F}$ with $F \subseteq \bigcap \mathcal{F}' \subseteq \bigcap \mathcal{F}$, i.e., $\bigcap \mathcal{F} \in \uparrow \Gamma$.

$\square$

Proof of Proposition A.7. 1. If $\mathcal{B}_i$ is reflexive then it is trivial that $\omega \in E$ for any $\omega \in \Omega$ and $E \in \mathcal{B}_i(\omega)$. The converse is also clear because, if $\omega \in \Omega$ and $E \in\uparrow \mathcal{B}_i(\omega)$ then there is $F \in \mathcal{B}_i(\omega)$ with $\omega \in F \subseteq E$.

2. Suppose that $\mathcal{B}_i$ is transitive. Take any $(\omega, E) \in \Omega \times \mathcal{D}$ with $E \in \mathcal{B}_i(\omega)$. It then follows from the supposition that $\{\omega' \in \Omega \mid E \in\uparrow \mathcal{B}_i(\omega')\} \in\uparrow \mathcal{B}_i(\omega)$. Thus there is $F \in \mathcal{B}_i(\omega)$ such that $F \subseteq \{\omega' \in \Omega \mid E \in\uparrow \mathcal{B}_i(\omega')\}$, clearly showing that if $\omega' \in F$ then there is $E' \in \mathcal{B}_i(\omega')$ such that $E' \subseteq E$. Conversely, suppose that $E \in\uparrow \mathcal{B}_i(\omega)$, i.e., suppose that there is $F \in \mathcal{B}_i$ such that $F \subseteq E$. For this $F \in \mathcal{B}_i(\omega)$, there is $F' \in \mathcal{B}_i(\omega)$ such that if $\omega' \in F'$ then $F \in\uparrow \mathcal{B}_i(\omega')$. Thus, we obtain

$$F' \subseteq \{\omega' \in \Omega \mid F \in\uparrow \mathcal{B}_i(\omega')\} \subseteq \{\omega' \in \Omega \mid E \in\uparrow \mathcal{B}_i(\omega')\}.$$

Since $F' \in \mathcal{B}_i(\omega)$, it follows that $\{\omega' \in \Omega \mid E \in\uparrow \mathcal{B}_i(\omega')\} \in\uparrow \mathcal{B}_i(\omega)$.

3. First observe that, for any $E \in \mathcal{D}$, we have $E \notin \mathcal{B}_i(\omega)$ iff $E^c \cap F \neq \emptyset$ for all $F \in \mathcal{B}_i(\omega)$. Thus, $\mathcal{B}_i$ is Euclidean iff the following holds: if there is $(\omega, E) \in \Omega \times \mathcal{D}$ such that $E^c \cap F \neq \emptyset$ for all $F \in \mathcal{B}_i(\omega)$, then there is $F' \in \mathcal{B}_i(\omega)$ such that $F' \subseteq \{\omega' \in \Omega \mid E \notin\uparrow \mathcal{B}_i(\omega')\}$. This last part is equivalent to the following: if $\omega' \in F'$ then $E^c \cap F \neq \emptyset$ for any $F \in \mathcal{B}_i(\omega')$.

□

Proof of Proposition A.8. 1. (a) If $\omega \in K_{\mathcal{B}_i}(\emptyset)$, then $\emptyset \in\uparrow \mathcal{B}_i(\omega)$. Hence, $K_{\mathcal{B}_i}(\emptyset)$ inherits No-Contradiction Axiom.

(b) We show that $K_{\mathcal{B}_i}$ satisfies Consistency. If $\omega \in K_{\mathcal{B}_i}(E)$ then $E \in\uparrow \mathcal{B}_i(\omega)$. Since $E^c \notin\uparrow \mathcal{B}_i(\omega)$, we have $\omega \in (\neg K_{\mathcal{B}_i})(E^c)$.

(c) Consider Non-empty λ-Conjunction. Take any $\mathcal{F} \subseteq \mathcal{D}$ with $0 < |\mathcal{F}| < \kappa$. If $\omega \in \bigcap_{F \in \mathcal{F}} K_{\mathcal{B}_i}(F)$ then $F \in\uparrow \mathcal{B}_i(\omega)$ for all $F \in \mathcal{F}$. Thus, $\bigcap \mathcal{F} \in\uparrow \mathcal{B}_i(\omega)$, i.e., $\omega \in K_{\mathcal{B}_i}(\bigcap \mathcal{F})$.

(d) Consider Necessitation. For any $\omega \in \Omega$, there is $\Omega \in\uparrow \mathcal{B}_i(\omega)$ so that $\omega \in K_i(\Omega)$.

(e) If $\mathcal{B}_i$ is reflexive, then for any $\omega \in K_{\mathcal{B}_i}(E)$, we have $E \in\uparrow \mathcal{B}_i(\omega)$ and thus $\omega \in E$. Thus, Truth Axiom is established.

(f) Consider Positive Introspection. If $\omega \in K_{\mathcal{B}_i}(E)$, then $E \in\uparrow \mathcal{B}_i(\omega)$. Now, we have $\{\omega' \in \Omega \mid E \in\uparrow \mathcal{B}_i(\omega')\} \in\uparrow \mathcal{B}_i(\omega)$, i.e., $\omega \in K_{\mathcal{B}_i} K_{\mathcal{B}_i}(E)$.

(g) Consider Negative Introspection. If $\omega \in (\neg K_{\mathcal{B}_i})(E)$, then $E \notin\uparrow \mathcal{B}_i(\omega)$. Now, we have $\{\omega' \in \Omega \mid E \notin\uparrow \mathcal{B}_i(\omega')\} \in\uparrow \mathcal{B}_i(\omega)$, i.e., $\omega \in (\neg K_{\mathcal{B}_i}) K_{\mathcal{B}_i}(E)$.

(h) The Kripke property of $K_{\mathcal{B}_i}$ amounts to $E \in\uparrow \mathcal{B}_i(\omega)$ (i.e., $\omega \in K_{\mathcal{B}_i}(E)$) whenever $(\bigcap\{F \in \mathcal{D} \mid \omega \in K_{\mathcal{B}_i}(F)\} =) \bigcap \uparrow \mathcal{B}_i(\omega) \subseteq E$. This simply follows from the Kripke property of $\mathcal{B}_i$.

Finally, we have $\mathcal{B}_{K_{\mathcal{B}_i}}(\omega) = \{E \in \mathcal{D} \mid \omega \in K_{\mathcal{B}_i}(E)\} = \{E \in \mathcal{D} \mid E \in\uparrow \mathcal{B}_i(\omega)\} =\uparrow \mathcal{B}_i(\omega)$ for all $\omega \in \Omega$.

2. Let K_i be a given knowledge operator and let $\mathcal{B}_i$ be a generator of K_i.

   (a) Consider No-Contradiction Axiom. Fix $\omega \in \Omega$. Since $\omega \notin K_i(\emptyset)$, we have $\emptyset \not\in\uparrow \mathcal{B}_i(\omega)$.

   (b) Consider Consistency. Take any $\omega \in \Omega$ and $E \in\uparrow \mathcal{B}_i(\omega)$. Since $\omega \in K_i(F) \subseteq (\neg K_i)(E^c)$, we have $E^c \not\in\uparrow \mathcal{B}_i(\omega)$.

   (c) Consider Non-empty λ-Conjunction. Let $\mathcal{F}$ be a subset of $\mathcal{D}$ with $0 < |\mathcal{F}| < \lambda$. If $F \in\uparrow \mathcal{B}_i(\omega)$ for all $F \in \mathcal{F}$ then $\omega \in \bigcap_{F\in\mathcal{F}} K_i(F) \subseteq K_i(\bigcap \mathcal{F})$. Thus, $\bigcap \mathcal{F} \in\uparrow \mathcal{B}_i(\omega)$.

   (d) Consider Necessitation. For any $\omega \in \Omega$, we have $\omega \in K_i(\Omega)$, i.e., $\Omega \in\uparrow \mathcal{B}_i(\omega)$.

   (e) Suppose that K_i satisfies Truth Axiom. Fix $\omega \in \Omega$ and $E \in\uparrow \mathcal{B}_i(\omega)$. Then, we have $\omega \in K_i(E) \subseteq E$.

   (f) Consider Positive Introspection. Fix $\omega \in \Omega$ and $E \in\uparrow \mathcal{B}_i(\omega)$. Then, $\omega \in K_i(E) \subseteq K_iK_i(E)$, and thus $\{\omega' \in \Omega \mid E \in\uparrow \mathcal{B}_i(\omega')\} \in\uparrow \mathcal{B}_i(\omega)$.

   (g) Suppose that K_i satisfies Negative Introspection. Fix $\omega \in \Omega$ and $E \not\in\uparrow \mathcal{B}_i(\omega)$. Then, we have $\omega \in (\neg K_i)(E) \subseteq K_i(\neg K_i)(E)$, and thus $\{\omega' \in \Omega \mid E \not\in\uparrow \mathcal{B}_i(\omega')\} \in\uparrow \mathcal{B}_i(\omega)$.

   (h) Consider the Kripke property. Observe that $\bigcap \uparrow \mathcal{B}_i(\omega) = \bigcap\{F \in \mathcal{D} \mid \omega \in K_i(F)\}$. The Kripke property of K_i implies that $\bigcap \uparrow \mathcal{B}_i(\omega) \subseteq E$ iff $E \in\uparrow \mathcal{B}_i(\omega)$, establishing the Kripke property of $\mathcal{B}_i$.

   Finally, for each $E \in \mathcal{D}$, we have

$$K_{\mathcal{B}_{K_i}}(E) = \{\omega \in \Omega \mid E \in\uparrow \mathcal{B}_{K_i}(\omega)\} = \{\omega \in \Omega \mid \omega \in K_i(E)\} = K_i(E).$$

$\square$

Proof of Proposition A.9. Take $E' \in \mathcal{B}'_i(\varphi(\omega))$. Since $E' \subseteq E'$, we have $\varphi(\omega) \in K'_i(E')$. Then, $\omega \in \varphi^{-1}(K'_i(E')) = K_i(\varphi^{-1}(E'))$. Now, there is $E \in \mathcal{B}_i(\omega)$ such that $E \subseteq \varphi^{-1}(E')$ and hence $\varphi(E) \subseteq \varphi(\varphi^{-1}(E')) \subseteq E$.

Next, take $E \in \mathcal{B}_i(\omega)$ and $E' \in \mathcal{D}'$ with $\varphi(E) \subseteq E'$. Then, since $E \subseteq \varphi^{-1}(\varphi(E)) \subseteq \varphi^{-1}(E')$, we have $\omega \in K_i(\varphi^{-1}(E')) = \varphi^{-1}(K'_i(E'))$. Thus, we get $\varphi(\omega) \in K'_i(E')$. Then, there is $F' \in \mathcal{B}_i(\varphi(\omega))$ such that $F' \subseteq E'$.

Suppose that (2) holds. If $\omega \in K_i(\varphi^{-1}(E'))$ then there is $E \in \mathcal{B}_i(\omega)$ such that $E \subseteq \varphi^{-1}(E')$, i.e., $\varphi(E) \subseteq \varphi(\varphi^{-1}(E')) \subseteq E'$. By (2), there is $F' \in \mathcal{B}'_i(\varphi(\omega))$ such that $F' \subseteq E'$. Hence, $\varphi(\omega) \in K'_i(E')$, i.e., $\omega \in \varphi^{-1}(K'_i(E'))$.

Suppose that (1) holds. If $\omega \in \varphi^{-1}(K'_i(E'))$ then $\varphi(\omega) \in K'_i(E')$. There is $F' \in \mathcal{B}'_i(\varphi(\omega))$ such that $F' \subseteq E'$. It follows from (1) that there is $F \in \mathcal{B}_i(\omega)$ such that $\varphi(F) \subseteq F'$. Thus, $\varphi(F) \subseteq E'$ and hence $F \subseteq \varphi^{-1}(\varphi(F)) \subseteq \varphi^{-1}(E')$. Then, we get $\omega \in K_i(\varphi^{-1}(E'))$. $\square$

A.7.3 Agreement Theorem

Here, we extend Samet's [77] qualitative agreement theorem to our generalized framework.[A.7] The non-probabilistic agreement theorem says that players cannot have common knowledge of their actions unless their actions are indeed identical. We extend Samet's [77] qualitative agreement theorem independently of assumptions on players' knowledge.

Throughout, fix a κ-complete algebra $(\Omega, \mathcal{D})$, where κ is an infinite (regular) cardinal or $\kappa = \infty$. Let $(K_i)_{i \in I}$ be players' knowledge operators on $\mathcal{D}$. Let $\mathcal{A}$ be a non-empty set of *actions* endowed with a κ-complete algebra $\mathcal{D}_{\mathcal{A}}$. We assume that $\{a\} \in \mathcal{D}_{\mathcal{A}}$ for all $a \in \mathcal{A}$. An action function of player i is a κ-measurable mapping $\alpha_i : (\Omega, \mathcal{D}) \to (\mathcal{A}, \mathcal{D}_{\mathcal{A}})$. We denote by $[\alpha_i = a] := \{\omega \in \Omega \mid \alpha_i(\omega) = a\}$ the event that player i takes an action a.

We denote by $[j \succcurlyeq i]$ the set of states at which player j is *at least as knowledgeable as* player i: $[j \succcurlyeq i] := \bigcap_{E \in \mathcal{D}} ((\neg K_i)(E) \cup K_j(E))$. The set of states at which players i and j are *equally knowledgeable* is defined by $[j \sim i] := [j \succcurlyeq i] \cap [i \succcurlyeq j]$. A player $i \in I$ is said to be an *epistemic dummy* if $[j \succcurlyeq i] = \Omega$ for all $j \in I$.

We make the following two assumptions, by slightly modifying Samet's [77] corresponding assumptions. The first is the *Interpersonal Sure-Thing Principle* (ISTP). Suppose that there is an event E indicating (i) that a player j is at least as knowledgeable as a player i and (ii) that j's action is a. If player i knows E irrespective of a particular form of E in the sense formalized below, we assume that she also takes the action a. Formally, fix any $i, j \in I$ and $a \in \mathcal{A}$. Suppose that there is $E \in \mathcal{D}$ such that $E \subseteq [j \succcurlyeq i] \cap [\alpha_j = a]$ and that $\omega \in K_i(F)$ for all $F \in \mathcal{D}$ with $E \subseteq F \subseteq [j \succcurlyeq i] \cap [\alpha_j = a]$. Then $\omega \in [\alpha_i = a]$.

As a remark, if two agents i and j, who know their own action function, are equally knowledgeable at any state, then the ISTP implies that they take the same action at any state. Indeed, take any $\omega \in \Omega$ and let $a = \alpha_j(\omega)$. The ISTP, together with the assumption that each player knows her own action function, imply that $\omega \in K_j([\alpha_j = a]) = K_i([\alpha_j = a]) = K_i([\alpha_j = a] \cap [j \succcurlyeq i]) \subseteq [\alpha_i = a]$.

The second assumption is the *ISTP Expandability*. Let $\langle (\Omega, \mathcal{D}), (K_i)_{i \in I}, (\alpha_i)_{i \in I} \rangle$ be given, and let i^* be the epistemic dummy whose knowledge corresponds to the common knowledge among I.[A.8] The ISTP Expandability says that there exists an action function α_{i^*} of the epistemic dummy i^* such that the extended action function profile $(\alpha_i)_{i \in I \cup \{i^*\}}$ satisfies the ISTP. Now, we formalize the qualitative agreement theorem.

Proposition A.10 (Qualitative Agreement Theorem)**.** *Let $\langle (\Omega, \mathcal{D}), (K_i)_{i \in I}, (\alpha_i)_{i \in I} \rangle$ be an ISTP expandable. If, for every $i \in I$, it is commonly known among I at a state ω that $\alpha_i(\omega) = a_i$, then players' actions $(a_i)_{i \in I}$ are identical. That is, if $\omega \in \bigcap_{i \in I} C([\alpha_i = a_i])$ then $a_i = a_j$ for all $i, j \in I$.*

Proof of Proposition A.10. Define an epistemic dummy i^* whose knowledge coincides with the common knowledge. Now, by the ISTP Expandability, there is an action function α_{i^*}

[A.7] See also the references in Samet [77] for other qualitative generalizations of the agreement theorem.

[A.8] The knowledge of the epistemic dummy may fail to satisfy some assumptions (e.g., Negative Introspection) imposed on individual players.

such that $(\alpha_i)_{i\in I\cup\{i^*\}}$ satisfies the ISTP. Suppose that $\omega \in \bigcap_{i\in I} C([\alpha_i = a_i]) = \bigcap_{i\in I} K_{i^*}([\alpha_i = a_i])$. It follows from the ISTP that $\omega \in K_{i^*}([\alpha_i = a_i]) = K_{i^*}([i \succcurlyeq i^*] \cap [\alpha_i = a_i]) \subseteq [\alpha_{i^*} = a_i]$ for each $i \in I$. Thus, we obtain $\alpha_{i^*}(\omega) = a_i = a_j$ for all $i, j \in I$. $\square$

It is wothwhile noting that Proposition A.10 holds irrespective of the cardinality of Ω. Samet [76] shows that Aumann's [3] probabilistic agreement theorem may fail to hold in some reflexive and transitive possibility correspondence model when Ω is uncountable. Thus, Aumann's [3] agreement theorem is not necessarily a special case of the qualitative agreement theorem in non-partitional spaces.

We also mention another difference between partitional and non-partitional cases. Henceforth, suppose that every agent knows her own action function. In partitional cases, equally knowledgeable agents at a particular state take the same action at that state under the ISTP (Samet [77, Proposition 1]). However, it is not necessarily the case in non-partitional models.[A.9]

Let $I = \{1, 2\}$, $\Omega = \{\omega_1, \omega_2, \omega_3\}$, and $\mathcal{D} = \mathcal{P}(\Omega)$. We define player 1's knowledge by her possibility correspondence: $b_1(\omega_1) = \{\omega_1\}$ and $b_1(\omega_j) = \Omega$ for $j \in \{2, 3\}$. Likewise, we define player 2's knowledge by $b_2(\omega_j) = \Omega$ and $b_2(\omega_3) = \{\omega_3\}$ for $j \in \{1, 2\}$.

For actions, we let $\mathcal{A} = \{a_1, a_2\}$, $\mathcal{D}_{\mathcal{A}}$, and $\alpha_i(\cdot) = a_i$ for each $i \in I$. Then, each player i knows her action function $\alpha_i : (\Omega, \mathcal{D}) \to (\mathcal{A}, \mathcal{D}_{\mathcal{A}})$. Also, it can be seen that the ISTP is trivially satisfied (i.e., $K_i([j \succcurlyeq i] \cap [\alpha_j = a_j]) = \emptyset$).

It can be seen that, while players 1 and 2 are equally knowledgeable at ω_2 ($[1 \sim 2] = \{\omega_2\}$), there is no state at which they take the same action (i.e., $\bigcup_{a\in\{a_1,a_2\}}([\alpha_1 = a] \cap [\alpha_1 = a]) = \emptyset$). In this particular example, both players are equally knowledgeable at ω_2 because they "mistakenly" ignore their knowledge. If they were to obey partitional knowledge, then player 1's partition cell at ω_2 would be $\{\omega_2, \omega_3\}$ and that of player 2 would be $\{\omega_1, \omega_2\}$.

A.7.4 Relation to a Model of Knowledge in Psychology

Here, we state connections between a model of knowledge in psychology studied by Doignon and Falmagne [22, 23] and our framework.[A.10] Doignon and Falmagne [22, 23] study the knowledge of an agent regarding subsets of Ω, which, in their work, consists of questions (note that we use our corresponding notations in order to make it easier to see the connections). An agent's knowledge is modeled as a pair $(\Omega, \mathcal{J})$, where $\emptyset, \Omega \in \mathcal{J} \subseteq \mathcal{P}(\Omega)$ and $\mathcal{J}$ is closed under arbitrary union. Thus, in terms of our framework, knowledge is defined on $\mathcal{D} = \mathcal{P}(\Omega)$ and $\mathcal{J}$ satisfies the maximality property as well as Necessitation.[A.11] In their work, each set $E \in \mathcal{J}$ is interpreted as a set of questions that an agent is capable of solving.

[A.9] Note that this statement is different from the previously stated result: if two agents are equally knowledgeable at *all* states then they take the same action at all states.

[A.10] It is worth pointing out that Doignon and Falmagne [23] document empirical procedures for asessing the knowledge of an agent in the psychology literature.

[A.11] Recall that the maximality property is equivalent to the closure under arbitrary union on a complete algebra (Proposition 5).

We make the following two connections. First, Doignon and Falmagne [22, 23] show one way to represent an agent's knowledge is to construct $\mathcal{J}_{\preccurlyeq}$ from a pre-ordered set $(\Omega, \preccurlyeq)$. They call the pre-order a surmise relation, and it is interpreted as follows: for any questions $\omega, \omega' \in \Omega$, it is said that ω' is surmised from ω (denoted $\omega \preccurlyeq \omega'$) iff it can be surmised that, from observing a correct response to question ω, a correct response would be given to question ω'. Thus, a question ω' is "at least as informative as" ω for assessing an agent's knowledge when $\omega \preccurlyeq \omega'$. The collection $\mathcal{J}_{\preccurlyeq}$ represents an agent's knowledge in the sense that each $E \in \mathcal{J}_{\preccurlyeq}$ is an upper set with respect to $\preccurlyeq$ (i.e., $\omega' \in E$ for all $\omega \in E$ and $\omega \preccurlyeq \omega'$). It can be seen that $\mathcal{J}_{\preccurlyeq}$ is closed under arbitrary union and intersection (with $\emptyset, \Omega \in \mathcal{J}_{\preccurlyeq}$). Thus, in our framework, knowledge derived from $\mathcal{J}_{\preccurlyeq}$ satisfies: Truth Axiom, Monotonicity, Positive Introspection, Necessitation, and (Non-empty) Conjunction.

Observe that the above properties characterize the reflexive and transitive possibility correspondence models in ∞-knowledge spaces.[A.12] Suppose that a pre-ordered set $(\Omega, \preccurlyeq)$ is given. Viewing Ω as a set of states of the world, each upper contour set of ω with respect to $\preccurlyeq$ defines the reflexive and transitive possibility correspondence $b_{\preccurlyeq}(\omega) = \{\omega' \in \Omega \mid \omega \preccurlyeq \omega'\}$. In relation to modal logic, the surmise relation $\preccurlyeq$ is a (reflexive and transitive) accessibility relation. An upper set $E \in \mathcal{J}_{\preccurlyeq}$ with respect to $\preccurlyeq$ is a self-evident event E in terms of $b_{\preccurlyeq}$: $b_{\preccurlyeq}(\omega) \subseteq E$ (i.e., $\omega \in K_{b_{\preccurlyeq}}(E)$) for all $\omega \in E$.

Second, Doignon and Falmagne [22, 23] introduce a surmise system $(\Omega, \mathcal{B})$ as another way to induce an agent's knowledge. A surmise mapping is a mapping $\mathcal{B} : \Omega \to \mathcal{P}(\mathcal{P}(\Omega))$ with the following interpretation. Each $\mathcal{B}(\omega)$ is interpreted as encoding all possible (not necessarily unique) ways of inferring a correct response to the question ω. Put differently, if an agent is capable of solving a question ω, then there exists $E \in \mathcal{B}(\omega)$ such that she is capable of solving all the questions in E. With this interpretation, Doignon and Falmagne [22, 23] define $\mathcal{J}_{\mathcal{B}}$ as a collection of $E \in \mathcal{P}(\Omega)$ such that for all $\omega \in E$, there is $F \in \mathcal{B}(\omega)$ such that $F \subseteq E$. Thus, a surmise system is closely related to an information correspondence.[A.13]

[A.12] Note that the possibility relation between states is equivalent to the informativeness relation in this situation. See Proposition 12.

[A.13] Doignon and Falmagne [22, 23], however, impose the following minimality condition: if $E, F \in \mathcal{B}(\omega)$ satisfy $E \subseteq F$ then $E = F$. Doignon and Falmagne [22, Theorem 3.7] establish the equivalence between a surmise system $\mathcal{B}$ and a collection $\mathcal{J}_{\mathcal{B}}$ of subsets when an underlying set Ω is finite. More generally, Doignon and Falmagne [23, Theorem 3.10] establish the equivalence between a surmise system and a collection which are "granular" for an arbitrary underlying set. In our framework, we demonstrate the equivalence between a reflexive and transitive information correspondence and a self-evident collection when a player's knowledge satisfies Truth Axiom, Positive Introspection, and Monotonicity without imposing the minimality condition.

Representing Unawareness on State Spaces

Satoshi Fukuda

Abstract

We approach notions of unawareness in terms of the lack of knowledge within the framework of a standard state space model. Our questions are as follows. When and how does a state space model have a sensible form of unawareness? How does unawareness relate to ignorance and possibility? First, notions of unawareness reduce to the following two forms. A strong form of unawareness is stated as the ignorance of the possibility that an agent knows an event. A weak form of unawareness is stated as the ignorance of own knowledge. Second, we show that if an agent is unaware of an event, then she is ignorant of being unaware of it. Third, if an agent faces an infinite number of objects of knowledge, then it is possible that she knows that there is an event of which she is unaware, while she cannot know that she is unaware of any particular event. Fourth, getting more information can cause an agent to become unaware of some event.

JEL Classification: C70, D83
Keywords: Unawareness, State Spaces, Plausibility, KU introspection, AU introspection

Contents

1 Introduction

Since the seminal work by Aumann [1, 2], a state space model of knowledge has been developed to model rational agents who reason interactively with each other. One of the subsequent research agendas in state space models has been to accommodate interactive knowledge among "boundedly rational" agents who lack some form of logical or introspective abilities. Especially, a notion of unawareness has been an active research area in economics since the pioneering work of Modica and Rustichini [24, 25] (hereafter, MR).[1] There, an agent is said to be unaware of a statement if she does not know it and she does not know that she does not know it.

Dekel, Lipman, and Rustichini [8] (hereafter, DLR), however, establish the following negative result. No state space model of knowledge can capture a sensible form of unawareness if an agent is logical and if a given notion of unawareness satisfies the following three axioms: Plausibility, KU Introspection, and AU Introspection. Plausibility states that if an agent is unaware of an event, then she does not know it and she does not know that she does not know it. KU Introspection says that an agent does not know that she is unaware of any particular event. AU Introspection demands that if an agent is unaware of an event, then she is unaware of being unaware of that event.[2] Since then, the research focus on unawareness has shifted to scrutinizing concepts of unawareness and to developing enhanced models capable of representing non-trivial forms of unawareness satisfying desirable features including these three axioms.[3]

This paper studies how state space models can (and cannot) capture a sensible form of unawareness. To that end, we confine ourselves to imposing the following conditions on agents' knowledge and unawareness.

First, we assume that agents are logical and introspective about their own knowledge. Namely, agents' knowledge satisfies at least the following three properties: (i) Truth Axiom (an agent can only know what is true), (ii) Positive Introspection (if an agent knows an event then she knows that she knows it), and (iii) Monotonicity (if an agent knows something then she knows its logical consequence). The first property is an essential property of knowledge, which distinguishes itself from belief. While we drop Negative Introspection (if an agent does not know an event then she knows that she does not know it), we assume that each agent is rational to the extent that she is introspective about her own knowledge.[4] These two properties are often assumed in state space models of knowledge involving boundedly rational agents (i.e., non-partitional (reflexive and transitive) possibility correspondence models).[5]

[1]Other pioneering work include Fagin and Halpern [11] and Pires [27].

[2]MR [25] demonstrate another negative result when unawareness is symmetric (i.e., if an agent is unaware of an event then she is also unaware of its negation).

[3]For example, Heifetz, Meier, and Schipper (hereafter, HMS) [16], [17], and Li [21] are earlier attempts along this line of research in economics. See Schipper [31] for a recent overview.

[4]If an agent exibits Negative Introspection, she is never unaware of any event.

[5]First, the previous studies on such models are: Bacharach [3], Binmore and Brandenburger [4], Bran-

Second, we define unawareness solely in terms of (the lack of) knowledge. We say that an agent is k^n-unaware of an event if she does not know it, she does not know that she does not know it, and so forth n times, including the case of $n = \infty$. Thus, notions of unawareness are derived from given properties of knowledge and a level of the lack of knowledge.

Now, we ask the following two strands of questions. First, we ask conditions on knowledge under which the derived notions of unawareness have a non-trivial (or trivial) form in state space models. Especially, we aim to answer the question raised by DLR [8] and subsequently studied by Chen, Ely, and Luo (hereafter, CEL) [6]: which of the previous three axioms is to be retained to represent an interesting form of unawareness in a state space model?

The second strand of questions is to examine how the derived notions of unawareness are related to other notions derived from knowledge. Specifically, we investigate notions of possibility, knowing-whether, and ignorance. An agent considers an event E possible if she does not know its negation E^c (e.g., Hintikka [18]).[6] An agent knows whether an event E is true when either she knows E or she knows its negation E^c (e.g., Hintikka [18] and Hart, Heifetz, and Samet (hereafter, HHS) [15]). An agent is said to be ignorant of an event E if she does not know E and she does not know its negation E^c (e.g., Lehrer and Samet [19]). That is, she is ignorant of E if shes does not know whether E is true.

Our results are as follows. First, three levels of lack of knowledge imply any higher level of lack of knowledge. Thus, the derived notions of unawareness reduce to the two forms: either two levels of lack of knowledge or infinitely many levels of lack of knowledge. For each form of unawareness, we characterize it in terms of possibility and ignorance (Proposition 1). When unawareness is defined as two levels of lack of knowledge, we show that it is equivalent to the ignorance of own knowledge: an agent is k^2-unaware of an event E if and only if (hereafter, iff) she is ignorant of (not) knowing E. Next, we characterize infinitely many levels of lack of knowledge. We show that an agent is k^∞-unaware of an event E iff she is ignorant of (not) knowing that she does not know E iff she is ignorant of the possibility that she knows E. We also show that an agent is k^∞-unaware of an event E iff k^2-unaware of not knowing E.

Next, we give (in Proposition 2 and Corollary 1) a necessary and sufficient condition for a state space model to be non-trivial in terms of a qualitative feature of knowledge. A simple implication of this characterization is that any properly non-partitional state space model can capture a non-trivial form of unawareness if unawareness is defined as two levels of lack of knowledge.

---

denburger, Dekel, and Geanakoplos [5], Dekel and Gul [7], Geanakoplos [14], Morris [26], Rubinstein and Wolinsky [28], Samet [29, 30] and Shin [32]. Second, unlike possibility correspondence models, we do not necessarily impose Necessitation: an agent knows any form of tautology.

[6] While we introduce possibility as the dual of knowledge, we could also introduce other notions of possibility. For example, MR [25] define possibility in the sense that an agent considers an event E possible if she does not know E and if she is aware (not unaware) of E. We could also say that an agent considers an event E possible if she does not know E and if she knows E or E^c (i.e., she knows a tautology). This last notion of possibility reduces to the original one when Necessitation is imposed.

What properties of unawareness do state space models satisfy? We show (in Proposition 3) that the notions of unawareness involve such properties as Plausibility, KU Introspection, the converse of AU Introspection, and the property which we call JU Introspection.[7] The converse of AU introspection refers to the property that if an agent is unaware of being unaware of an event E then she is unaware of E. JU Introspection is that if an agent is unaware of an event then she is ignorant of being unaware of it.

To restate, in any state space model where agents are logical and introspective and where notions of unawareness are defined in terms of the lack of knowledge, the state space model satisfies Plausibility, KU Introspection, and JU Introspection (instead of AU Introspection). We also examine (in Proposition 4) properties of unawareness (e.g., AU Introspection and Symmetry) which lead to a degenerate form.

Finally, we study the following two properties of unawareness. First, recall that KU Introspection states that there is no state at which an agent knows that she is unaware of a particular event. Can an agent know her own unawareness? That is, can she know that there is an event of which she is unaware?[8] We show (in Proposition 5) that if there is an infinite number of objects of knowledge in a given state space model then an agent may know her self-unawareness, i.e., it can be the case that she knows there is an event of which she is unaware (while she does not know that she is unaware of any particular event). If objects of knowledge are finite in a given state space model, then an agent does not know that there is an event of which she is unaware.

Second, we provide examples in which unawareness is not monotonic in knowledgeability. Specifically, getting more information can cause an agent to become unaware of some event. We also study (in Propositions 6 and 7) the sense in which unawareness and knowledgeability are correlated.

The paper is organized as follows. Section 2 provides the state-space-based framework for studying interactive knowledge and unawareness. Section 3 studies properties of unawareness. In Section 3.1, we restate unawareness in terms of ignorance, knowing-whether, and possibility. In Section 3.2, we characterize a necessary and sufficient condition for non-trivial unawareness. Section 3.3 studies which of the existing axioms of unawareness are to be satisfied and violated in state space models.

Section 4 studies the following properties of unawareness. In Section 4.1, we ask knowledge of self-unawareness. In Section 4.2, we demonstrate non-monotonicity of unawareness in knowledgeability. In Section 4.3, we study possible forms of monotonicity of unawareness in knowledgeability. Proofs are relegated to Appendix A. Appendix B briefly discusses extensions of our framework.

[7] First, DLR [8, Footnote 10] note that a given notion of unawareness satisfies KU Introspection if it satisfies Plausibility and if a given notion of knowledge satisfies Truth Axiom and Monotonicity. Second, the "J" in JU Introspection refers to the knowing-whether operator in HHS [15].

[8] See also Schipper [31, Section 3.5] and the references therein.

2 Information Structures

This section presents our framework, which we call an information structure. Let Ω be a set of *states* (i.e., a *state space*). Let I denote a non-empty set of *agents*.

Next, we define events, which are objects of interactive knowledge and unawareness. For ease of exposition, we assume that a collection of *events*, $\mathcal{D}$, forms a complete algebra (of sets) on Ω. That is, $\mathcal{D}$ is a subset of $\mathcal{P}(\Omega)$, where $\mathcal{P}(\cdot)$ is the power set operation, which is closed under complementation, arbitrary union, and arbitrary intersection.[9] We call $\mathcal{D}$ the *domain*.

Note that the specification of events (i.e., a domain) is an important consideration in establishing a "universal" structure and that we can indeed accommodate a specification of the domain $\mathcal{D}$ that guarantees existence of a universal structure (see the first chapter of this dissertation (Fukuda [13])). Namely, we can carry out most of the analyses when $\mathcal{D}$ is assumed to form a set algebra called a κ-complete algebra, where κ is an infinite cardinal.[10] With these definitions in mind, we now define an information structure.

Definition 1. *An information structure (of I) is a tuple $\mathcal{S} := \langle(\Omega, \mathcal{D}), (K_i, U_i)_{i\in I}\rangle$ with the following properties.*

1. *Ω is a state space, and $\mathcal{D}$ is a complete algebra of events.*
2. *$K_i : \mathcal{D} \to \mathcal{D}$ is an agent i's knowledge operator satisfying (at least) the following.*
   (a) *Truth Axiom: $K_i(E) \subseteq E$ (for all $E \in \mathcal{D}$).*
   (b) *Positive Introspection: $K_i(E) \subseteq K_iK_i(E)$.*
   (c) *Monotonicity: if $E \subseteq F$ then $K_i(E) \subseteq K_i(F)$.*
3. *$U_i : \mathcal{D} \to \mathcal{D}$ is i's unawareness operator.*

Fix any event $E \in \mathcal{D}$. The set $K_i(E)$ is the event that (the set of states at which) i knows E. Likewise, $U_i(E)$ is the event that agent i is unaware of E.

We maintain the common domain assumption that agents' interactive knowledge and unawareness can be described on $(\Omega, \mathcal{D})$. We briefly discuss in Appendix B agent-specific domains where each agent's knowledge and unawareness operators K_i and U_i are defined on agent-specific domain $(\Omega_i, \mathcal{D}_i)$ where Ω_i is a subset of Ω (so that $\mathcal{D}_i$ is identified as a subset of $\mathcal{D}$). One possible justification of the common domain assumption is that we restrict attention to those events in $\mathcal{D} = \bigcap_{i\in I} \mathcal{D}_i$ under heterogeneous domains.[11]

[9] If $\mathcal{E}$ is a subset of $\mathcal{D}$, then $\bigcup \mathcal{E} (:= \bigcup_{E\in\mathcal{E}} E) \in \mathcal{D}$ and $\bigcap \mathcal{E} (:= \bigcap_{E\in\mathcal{E}} E) \in \mathcal{D}$. Also, if $E \in \mathcal{D}$ then $E^c \in \mathcal{D}$, where E^c is the complement of E (we also use $\neg E$ to denote the complement of E). Note that we follow the conventions that $\emptyset = \bigcup \emptyset$ and that, with an underlying set Ω fixed, $\Omega = \bigcap \emptyset$.

[10] First, $\mathcal{D}$ is a κ-complete algebra if the following hold: if $E \in \mathcal{D}$ then $E^c \in \mathcal{D}$; and if $\mathcal{E}$ satisfies $\mathcal{E} \subseteq \mathcal{D}$ and $|\mathcal{E}| < \kappa$ then $\bigcap \mathcal{E} \in \mathcal{D}$ and $\bigcup \mathcal{E} \in \mathcal{D}$. Second, if κ is the least infinite cardinal, we would have to take care of infinite operations in the analyses to follow.

[11] It would still be interesting to explore how each agent processes an operation of negation.

Truth Axiom distinguishes knowledge from "belief" in the sense that an agent can only know what is true while she can believe what is false. Positive Introspection allows an agent to know what she knows. Monotonicity renders an agent a logical inference ability.

We introduce further properties of knowledge operators. First, K_i satisfies Necessitation if $K_i(\Omega) = \Omega$. Second, K_i satisfies Non-empty Conjunction if $\bigcap_{E \in \mathcal{E}} K_i(E) \subseteq K_i(\bigcap \mathcal{E})$ for any non-empty $\mathcal{E} \subseteq \mathcal{D}$. In a similar vein, we define Non-empty Finite (Countable) Conjunction as follows: $\bigcap_{E \in \mathcal{E}} K_i(E) \subseteq K_i(\bigcap \mathcal{E})$ for any non-empty finite (countable) $\mathcal{E} \subseteq \mathcal{D}$. Note that Non-empty Conjunction and Necessitation can be jointly regarded as (Arbitrary) Conjunction: $\bigcap_{E \in \mathcal{E}} K_i(E) \subseteq K_i(\bigcap \mathcal{E})$ for any $\mathcal{E} \subseteq \mathcal{D}$. Third, we define Negative Introspection of K_i to be $(\neg K_i)(\cdot) \subseteq K_i(\neg K_i)(\cdot)$.

Next, we state some of joint postulates on agents' knowledge and unawareness operators. First, we say that (K_i, U_i) is *plausible* if $U_i(\cdot) \subseteq (\neg K_i)(\cdot) \cap (\neg K_i)^2(\cdot)$. Plausibility says that if an agent is unaware of an event E then she does not know E and she does not know that she does not know E. If every (K_i, U_i) is plausible, we say that $\mathcal{S}$ is plausible.

Second, we say that (K_i, U_i) satisfies *KU Introspection* if $K_i U_i(\cdot) = \emptyset$. KU Introspection means that, for any event E, there is no state at which an agent knows that she is unaware of E. Third, we say that (K_i, U_i) satisfies *AU Introspection* if $U_i(\cdot) \subseteq U_i U_i(\cdot)$. AU Introspection says that if an agent is unaware of E then she is unaware of being unaware of E. These three properties are proposed in DLR [8]. Other properties are examined in Section 3.3.

To conclude the exposition of the framework, we relate this framework to possibility correspondence models. A knowledge operator K_i satisfying Monotonicity and Arbitrary Conjunction can be induced from a possibility correspondence. If K_i also satisfies Truth Axiom and Positive Introspection, then it is characterized by a reflexive and transitive possibility correspondence. See Footnote 5 for the literature. If K_i additionally satisfies Negative Introspection, then it is induced by a partition (Aumann [1, 2]). In any plausible information structure $\mathcal{S}$ such that K_i is induced from a partition, however, we have $U_i(\cdot) = \emptyset$.

2.1 Associated Concepts

We define other associated concepts which are derived from knowledge (and unawareness). In this subsection, fix an information structure $\mathcal{S}$ of I.

Derived Operators. We start with defining the following four operators defined on $\mathcal{D}$. For any event $E \in \mathcal{D}$, we let: (i) $L_i(E) := (\neg K_i)(E^c)$, (ii) $\partial_i(E) := (\neg K_i)(E) \cap (\neg K_i)(E^c)$, (iii) $J_i(E) := (\neg \partial_i)(E) (= K_i(E) \cup K_i(E^c))$, and (iv) $A_i(E) := (\neg U_i)(E)$. We call L_i, ∂_i, J_i, and A_i to be *possibility* (e.g., Hintikka [18]), *ignorant* (e.g., Lehrer and Samet [19]), *knowing-whether* (e.g., Hintikka [18] and HHS [15]), and *awareness* (MR [24, 25]) operators, respectively.

First, $L_i(E)$ is regarded as the event that i considers E possible in the sense that i does not know its negation E^c. Second, $\partial_i(E)$ is interpreted as the event that i is ignorant of E in the sense that i does not know E nor E^c.[12] Third, $J_i(E)$ is considered to be the event

[12] We use the symbol "∂" of the boundary operator in a topological space in the sense that K_i satisfies a

that i knows whether E obtains (or not) in the sense that either i knows E or she knows its negation E^c. Fourth, $A_i(E)$ is regarded as the event that i is aware of E in the sense that i is not unaware of E.

Defining Unawareness from Knowledge. We define unawareness operators derived from a given knowledge operator as follows. For $n \in \mathbb{N}_2^\infty := \{n \in \mathbb{N} \mid n \geq 2\} \cup \{\infty\}$, we define $U_i^{(n)}(\cdot) := \bigcap_{r=1}^n (\neg K_i)^r(\cdot)$.[13] We say that an agent i is (k^n-)*unaware* of an event E at a state ω if $\omega \in U_i^{(n)}(E)$. The k^n-awareness operator $A_i^{(n)}$ is defined by $A_i^{(n)}(\cdot) := (\neg U_i^{(n)})(\cdot)$. MR [24] define their unawareness operator by $U_i^{(2)}$ while $U_i^{(\infty)}$ is considered in DLR [8].

Self-evident Collection. We define an event $E \in \mathcal{D}$ to be *self-evident* to an agent i if $E \subseteq K_i(E)$. We denote by $\mathcal{J}_i := \{E \in \mathcal{D} \mid E \subseteq K_i(E)\}$ the collection of events which are self-evident to i. We call $\mathcal{J}_i$ to be i's *self-evident collection.* We have $\mathcal{J}_i \neq \emptyset$ because $\emptyset \in \mathcal{J}_i$. Also, it can be seen that Monotonicity of K_i implies that $\mathcal{J}_i$ is closed under arbitrary union.[14] Denoting $\mathcal{A}_i := \{E \in \mathcal{D} \mid L_i(E) \subseteq E\}$, we have $\mathcal{A}_i = \{E \in \mathcal{D} \mid E^c \in \mathcal{J}_i\}$.

Conversely, given the self-evident collection $\mathcal{J}_i$ which is induced from K_i, we can define the associated knowledge operator $K_{\mathcal{J}_i}(E) := \{\omega \in \Omega \mid \omega \in F \subseteq E \text{ for some } F \in \mathcal{J}_i\} = \bigcup\{F \in \mathcal{J}_i \mid F \subseteq E\}$. It turns out that $K_i = K_{\mathcal{J}_i}$. Moreover, the following can be verified (see the first chapter of this dissertation (Fukuda [13]) for proofs). First, K_i satisfies Non-empty (Finite/Countable) Conjunction iff $\mathcal{J}_i$ is closed under non-empty (finite/countable) intersection. Second, K_i satisfies Necessitation iff $\Omega \in \mathcal{J}_i$. Third, K_i satisfies Negative Introspection iff $\mathcal{J}_i$ is closed under complementation.

Knowledgeability. We denote by $\mathrm{IK}_i(\omega) := \{E \in \mathcal{D} \mid \omega \in K_i(E)\}$ the collection of events that an agent i knows at a state ω. We say that an agent i is *at least as knowledgeable as* an agent j *at* a state ω if $\mathrm{IK}_j(\omega) \subseteq \mathrm{IK}_i(\omega)$. We also say that agents i and j are *equally knowledgeable at* a state ω if $\mathrm{IK}_j(\omega) = \mathrm{IK}_i(\omega)$.

Likewise, an agent i is *at least as knowledgeable as* an agent j if i is at least as knowledgeable as j at any state. Agents i and j are *equally knowledgeable* if i and j are equally knowledgeable at any state. It can be easily seen that this knowledgeable relation is formulated in terms of knowledge operators and self-evident collections: i is at least as knowledgeable as j iff $K_j(\cdot) \subseteq K_i(\cdot)$ iff $\mathcal{J}_j \subseteq \mathcal{J}_i$.

Common Knowledge. We define the notion of common knowledge (e.g., Aumann [1], Friedell [12], Lewis [20], and McCarthy, Sato, Hayashi, and Igarashi [22]) among I as the knowledge that would be possessed by the most knowledgeable agent who is at least as

---

part of the properties satisfied by the interior operator in a topological space.

[13]Note that $\mathbb{N}$ is the set of positive integers and that $U_i^{(\infty)}(\cdot) = \bigcap_{r\in\mathbb{N}} (\neg K_i)^r(\cdot)$.

[14]In mathematical psychology, Doignon and Falmagne [9, 10] formalize a notion of knowledge by a set algebra, which is related to a self-evident collection. See the first chapter of this dissertation (Fukuda [13]) for the discussion.

less knowledgeable as every agent within I.[15] We say that an event $E \in \mathcal{D}$ is *commonly known/common knowledge* among I at a state $\omega \in \Omega$ if E is inferred from some event F in the sense of $F \subseteq E$, where F is true at ω and where F is self-evident to every agent $i \in I$ (i.e., *publicly evident* (Milgrom [23])). Formally, we define the *common knowledge operator* $C_I : \mathcal{D} \to \mathcal{D}$ by $C_I(E) := \{\omega \in \Omega \mid \omega \in F \subseteq E \text{ for some } F \in \bigcap_{i \in I} \mathcal{J}_i\} = \bigcup\{F \in \bigcap_{i\in I} \mathcal{J}_i \mid F \subseteq E\}$.[16] It can be seen that C_I inherits the properties that every agent knowledge operator satisfies, because taking the intersection of $(\mathcal{J}_i)_{i \in I}$ preserves set-algebraic properties that every $\mathcal{J}_i$ commonly possess.[17]

If E is commonly known among I at ω, then everyone in the group I knows E at ω, everyone in I knows that everyone in I knows E at ω, *ad infinitum*. Let $K_I(\cdot) := \bigcap_{i \in I} K_i(\cdot)$. That is, $K_I(E)$ is the event that everyone in I knows E. Indeed, it can be seen that $C(\cdot) = \bigcap_{n \in \mathbb{N}} K_I^n(\cdot)$ in any information structure satisfying Countable Conjunction, as $\bigcap_{n\in\mathbb{N}} K_I^n(E)$ is the maximal publicly evident event contained in E (the fact that it is publicly evident follows because, $\bigcap_{n\in\mathbb{N}} K_I^n(E) \subseteq \bigcap_{n\in\mathbb{N}} K_I^{n+1}(E) \subseteq \bigcap_{n\in\mathbb{N}} K_i K_I^n(E) = K_i(\bigcap_{n\in\mathbb{N}} K_I^n(E))$ for all $i \in I$ under Countable Conjunction).

Common (Un)awareness. We define the *common awareness* operator (HMS [17]) $CA_I : \mathcal{D} \to \mathcal{D}$ by $CA_I(\cdot) := \bigcap_{r\in\mathbb{N}} A_I^r(\cdot)$, where A_I is a *mutual awareness* operator defined by $A_I(\cdot) := \bigcap_{i\in I} A_i(\cdot)$. By definition, $K_I(\cdot) \subseteq A_I(\cdot)$. Also, Monotonicity of K_I implies that $K_I^r(\cdot) \subseteq A_I^r(\cdot)$ for each $r \in \mathbb{N}$, and hence we have $C_I(\cdot) \subseteq CA_I(\cdot)$.

Likewise, we define the *common unawareness* operator $CU_I : \mathcal{D} \to \mathcal{D}$ by $CU_I(\cdot) := \bigcap_{r\in\mathbb{N}} U_I^r(\cdot)$, where U_I is the *mutual unawareness* operator defined by $U_I(\cdot) := \bigcap_{i\in I} U_i(\cdot)$. It is clear that $U_I(\cdot) \subseteq (\neg C_I)(\cdot)$.

2.2 An Example

We provide an example to illustrate our framework.

Example 1. We let $\Omega = \{\omega_1, \omega_2, \omega_3\}$, $\mathcal{D} = \mathcal{P}(\Omega)$, and $I = \{i_1, i_2, i_3, i_4\}$. Suppose that each K_i is defined as in Table 1. Each $U_i^{(n)}$ (with $n \in \mathbb{N}_2^\infty$) is depicted in Table 1. Note that agent i_1's knowledge coincides with DLR's [8, p. 161] Example 1.

We make the following remarks. First, we have $\mathcal{J}_{i_1} = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \{\omega_1, \omega_2\}, \Omega\}$, $\mathcal{J}_{i_2} = \{\emptyset, \{\omega_1\}, \{\omega_3\}, \{\omega_1, \omega_3\}\}$, $\mathcal{J}_{i_3} = \{\emptyset, \{\omega_1\}, \Omega\}$, and $\mathcal{J}_{i_4} = \{\emptyset, \{\omega_1\}\}$. Thus, agent i_1 is at least as knowledgeable as $j \in \{i_2, i_3, i_4\}$.

[15] Aumann [1] defines the notion of common knowledge from the finest partition which is coarser than every aggent's partition.

[16] We can similarly define the common knowledge operator C_G for any $G \in \mathcal{P}(I)$ with the convention that $C_\emptyset(E) = E$ for all $E \in \mathcal{D}$.

[17] Hence, we can apply the concept of k^n-unawareness to the common knowledge operator. We can say that an event E is n-negatively common knowledge among I at a state ω if $\omega \in U_{C_I}^{(n)}(E) := \bigcap_{r=1}^n (\neg C_I)^r(E)$ for each $n \in \mathbb{N}_2^\infty$. Hence, an event E is twice negatively common knowledge iff E is not common knowledge and the event that E is not common knowledge is not common knowledge.

E	K_{i_1}	$(\neg K_{i_1})$	$(\neg K_{i_1})^2$	$(\neg K_{i_1})^3$	$(\neg K_{i_1})^4$	∂_{i_1}	$U_{i_1}^{(2)}$	$U_{i_1}^{(n)}$
$\emptyset$	$\emptyset$	Ω	$\emptyset$	Ω	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
$\{\omega_1\}$	$\{\omega_1\}$	$\{\omega_2,\omega_3\}$	$\{\omega_1,\omega_3\}$	$\{\omega_2,\omega_3\}$	$\{\omega_1,\omega_3\}$	$\{\omega_3\}$	$\{\omega_3\}$	$\{\omega_3\}$
$\{\omega_2\}$	$\{\omega_2\}$	$\{\omega_1,\omega_3\}$	$\{\omega_2,\omega_3\}$	$\{\omega_1,\omega_3\}$	$\{\omega_2,\omega_3\}$	$\{\omega_3\}$	$\{\omega_3\}$	$\{\omega_3\}$
$\{\omega_3\}$	$\emptyset$	Ω	$\emptyset$	Ω	$\emptyset$	$\{\omega_3\}$	$\emptyset$	$\emptyset$
$\{\omega_1,\omega_2\}$	$\{\omega_1,\omega_2\}$	$\{\omega_3\}$	Ω	$\emptyset$	Ω	$\{\omega_3\}$	$\{\omega_3\}$	$\emptyset$
$\{\omega_1,\omega_3\}$	$\{\omega_1\}$	$\{\omega_2,\omega_3\}$	$\{\omega_1,\omega_3\}$	$\{\omega_2,\omega_3\}$	$\{\omega_1,\omega_3\}$	$\{\omega_3\}$	$\{\omega_3\}$	$\{\omega_3\}$
$\{\omega_2,\omega_3\}$	$\{\omega_2\}$	$\{\omega_1,\omega_3\}$	$\{\omega_2,\omega_3\}$	$\{\omega_1,\omega_3\}$	$\{\omega_2,\omega_3\}$	$\{\omega_3\}$	$\{\omega_3\}$	$\{\omega_3\}$
Ω	Ω	$\emptyset$	Ω	$\emptyset$	Ω	$\emptyset$	$\emptyset$	$\emptyset$

E	K_{i_2}	$(\neg K_{i_2})$	$(\neg K_{i_2})^2$	$(\neg K_{i_2})^3$	$(\neg K_{i_2})^4$	∂_{i_2}	$U_{i_2}^{(2)}$	$U_{i_2}^{(n)}$
$\emptyset$	$\emptyset$	Ω	$\{\omega_2\}$	Ω	$\{\omega_2\}$	$\{\omega_2\}$	$\{\omega_2\}$	$\{\omega_2\}$
$\{\omega_1\}$	$\{\omega_1\}$	$\{\omega_2,\omega_3\}$	$\{\omega_1,\omega_2\}$	$\{\omega_2,\omega_3\}$	$\{\omega_1,\omega_2\}$	$\{\omega_2\}$	$\{\omega_2\}$	$\{\omega_2\}$
$\{\omega_2\}$	$\emptyset$	Ω	$\{\omega_2\}$	Ω	$\{\omega_2\}$	$\{\omega_2\}$	$\{\omega_2\}$	$\{\omega_2\}$
$\{\omega_3\}$	$\{\omega_3\}$	$\{\omega_1,\omega_2\}$	$\{\omega_2,\omega_3\}$	$\{\omega_1,\omega_2\}$	$\{\omega_2,\omega_3\}$	$\{\omega_2\}$	$\{\omega_2\}$	$\{\omega_2\}$
$\{\omega_1,\omega_2\}$	$\{\omega_1\}$	$\{\omega_2,\omega_3\}$	$\{\omega_1,\omega_2\}$	$\{\omega_2,\omega_3\}$	$\{\omega_1,\omega_2\}$	$\{\omega_2\}$	$\{\omega_2\}$	$\{\omega_2\}$
$\{\omega_1,\omega_3\}$	$\{\omega_1,\omega_3\}$	$\{\omega_2\}$	Ω	$\{\omega_2\}$	Ω	$\{\omega_2\}$	$\{\omega_2\}$	$\{\omega_2\}$
$\{\omega_2,\omega_3\}$	$\{\omega_3\}$	$\{\omega_1,\omega_2\}$	$\{\omega_2,\omega_3\}$	$\{\omega_1,\omega_2\}$	$\{\omega_2,\omega_3\}$	$\{\omega_2\}$	$\{\omega_2\}$	$\{\omega_2\}$
Ω	$\{\omega_1,\omega_3\}$	$\{\omega_2\}$	Ω	$\{\omega_2\}$	Ω	$\{\omega_2\}$	$\{\omega_2\}$	$\{\omega_2\}$

E	K_{i_3}	$(\neg K_{i_3})$	$(\neg K_{i_3})^2$	$(\neg K_{i_3})^3$	$(\neg K_{i_3})^4$	∂_{i_3}	$U_{i_3}^{(2)}$	$U_{i_3}^{(n)}$
$\emptyset$	$\emptyset$	Ω	$\emptyset$	Ω	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
$\{\omega_1\}$	$\{\omega_1\}$	$\{\omega_2,\omega_3\}$	Ω	$\emptyset$	Ω	$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$	$\emptyset$
$\{\omega_2\}$	$\emptyset$	Ω	$\emptyset$	Ω	$\emptyset$	$\{\omega_2,\omega_3\}$	$\emptyset$	$\emptyset$
$\{\omega_3\}$	$\emptyset$	Ω	$\emptyset$	Ω	$\emptyset$	$\{\omega_2,\omega_3\}$	$\emptyset$	$\emptyset$
$\{\omega_1,\omega_2\}$	$\{\omega_1\}$	$\{\omega_2,\omega_3\}$	Ω	$\emptyset$	Ω	$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$	$\emptyset$
$\{\omega_1,\omega_3\}$	$\{\omega_1\}$	$\{\omega_2,\omega_3\}$	Ω	$\emptyset$	Ω	$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$	$\emptyset$
$\{\omega_2,\omega_3\}$	$\emptyset$	Ω	$\emptyset$	Ω	$\emptyset$	$\{\omega_2,\omega_3\}$	$\emptyset$	$\emptyset$
Ω	Ω	$\emptyset$	Ω	$\emptyset$	Ω	$\emptyset$	$\emptyset$	$\emptyset$

E	$C_I = K_{i_4}$	$(\neg K_{i_4})$	$(\neg K_{i_4})^2$	$(\neg K_{i_4})^3$	$(\neg K_{i_4})^4$	∂_{i_4}	$U_{i_4}^{(2)}$	$U_{i_4}^{(n)}$
$\emptyset$	$\emptyset$	Ω	$\{\omega_2,\omega_3\}$	Ω	$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$
$\{\omega_1\}$	$\{\omega_1\}$	$\{\omega_2,\omega_3\}$	Ω	$\{\omega_2,\omega_3\}$	Ω	$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$
$\{\omega_2\}$	$\emptyset$	Ω	$\{\omega_2,\omega_3\}$	Ω	$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$
$\{\omega_3\}$	$\emptyset$	Ω	$\{\omega_2,\omega_3\}$	Ω	$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$
$\{\omega_1,\omega_2\}$	$\{\omega_1\}$	$\{\omega_2,\omega_3\}$	Ω	$\{\omega_2,\omega_3\}$	Ω	$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$
$\{\omega_1,\omega_3\}$	$\{\omega_1\}$	$\{\omega_2,\omega_3\}$	Ω	$\{\omega_2,\omega_3\}$	Ω	$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$
$\{\omega_2,\omega_3\}$	$\emptyset$	Ω	$\{\omega_2,\omega_3\}$	Ω	$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$
Ω	$\{\omega_1\}$	$\{\omega_2,\omega_3\}$	Ω	$\{\omega_2,\omega_3\}$	Ω	$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$

Table 1: Illustrations of Agents' Knowledge and Unawareness in Example 1 ($n \in \{n \in \mathbb{N} \mid n \geq 3\} \cup \{\infty\}$)

Second, each K_i satisfies Non-empty Conjunction. Knowledge operators of i_1 and i_3 also satisfy Necessitation.

Third, the collection of publicly evident events is $\bigcap_{i \in I} \mathcal{J}_i = \{\emptyset, \{\omega_1\}\} = \mathcal{J}_{i_4}$, so that we can also identify agent i_4 as a dummy agent whose knowledge corresponds with common knowledge (i.e., $C_I = K_{i_4}$).

Fourth, each pair $(K_i, U_i^{(n)})$ satisfies Plausibility and KU Introspection. While $(K_i, U_i^{(n)})$ satisfies AU Introspection for each $i \in \{i_2, i_4\}$, the other pairs $(K_j, U_j^{(n)})$ $(j \in \{i_1, i_3\})$ do not satisfy AU introspection because $U_j^{(n)} U_j^{(n)}(\cdot) = \emptyset$. Such j's information structure $\langle(\Omega, \mathcal{D}), (K_j, U_j^{(n)})\rangle$ $(n \in \mathbb{N}_2^\infty)$ is considered to be a reflexive and transitive possibility correspondence model. Indeed, j's knowledge can be induced from the possibility correspondence $b_j : \Omega \to \mathcal{D}$, where $b_{i_1}(\omega_1) = \{\omega_1\}$, $b_{i_1}(\omega_2) = \{\omega_2\}$, and $b_{i_1}(\omega_3) = \Omega$; and $b_{i_3}(\omega_1) = \{\omega_1\}$ and $b_{i_3}(\omega_2) = b_{i_3}(\omega_3) = \Omega$. The fact that $\langle(\Omega, \mathcal{D}), (K_j, U_j^{(n)})\rangle$ does not satisfy AU Introspection is consistent with the finding by DLR [8] that there is no possibility correspondence model which satisfies all of Plausibility, KU Introspection, and AU Introspection.

Fifth, $U_I^{(n)}(\cdot) = \emptyset$ so that $CU_I(\cdot) = \emptyset$. Also, $A_I^{(n)}(\cdot) = \{\omega_1\}$ so that $CA_I(\cdot) = \{\omega_1\}$.

Sixth, agent i_2's (resp. i_4's) knowledge operator can be identified as being defined on $\{\omega_1, \omega_3\}$ (resp. $\{\omega_1\}$). In other words, any state $\omega \in \{\omega_2\}$ (resp. $\omega \in \{\omega_2, \omega_3\}$) is deemed "impossible" by agent i_2 (resp. i_4). On the other hand, any event $E \in \mathcal{P}(\{\omega_1, \omega_3\})$ (resp. $E \in \mathcal{P}(\{\omega_1\})$) is self-evident to agent i_2 (resp. i_4). In this example, their unawareness is always determined by "impossible" states.

3 Unawareness on State Spaces

Having defined the basic framework, we now proceed with the main analyses.

3.1 Equivalent Representations

Our first aim is to relate the concepts of ignorance, knowing-whether, and possibility to that of unawareness when they are derived from knowledge. We also study the relations among the notions of k^n-unawareness. To that end, throughout the subsection, we fix an information structure $\mathcal{S} = \langle(\Omega, \mathcal{D}), (K, U)\rangle$ of a single agent.

The first benchmark result is that, under Truth Axiom, Positive Introspection, and Monotonicity, $(\neg K)^2 = (\neg K)^{2n}$ for all $n \in \mathbb{N}$. We prove this fact in the Appendix (Lemma A.1).[18] This preliminary result implies that $U^{(\infty)} = U^{(n)}$ for all $n \geq 3$. That is, if an agent is k^3-unaware of an event E (i.e., the chain of the lack of knowledge holds repeatedly three times), then she is indeed k^∞-unaware of E (i.e., this chain continues *ad infinitum*). Hence, as long as the notions of unawareness are derived from the lack of knowledge, we can restrict

[18]Mathematically, this property is related to the notion of regularly open/closed sets in general topology (see, for example, Willard [33]) in the sense that K satisfies a part of the properties of the interior operator in a topological space.

attention to $U^{(n)}$ with $n \in \{2, \infty\}$ and we can replace $U^{(\infty)}$ with $U^{(3)}$.[19] Note that $U^{(2)}$ and $U^{(\infty)}(= U^{(3)})$ are generally different (e.g., agents i_1 and i_3 in Example 1).[20]

This observation leads to the following restatement of unawareness when knowledge satisfies Truth Axiom, Positive Introspection, and Monotonicity.

Proposition 1. *Fix any* $E \in \mathcal{D}$.

1. $U^{(\infty)}(E) = U^{(2)}(\neg K)(E)$. *Equivalently,* $A^{(\infty)}(E) = A^{(2)}(\neg K)(E)$.
2. $U^{(\infty)}(E) = \partial K(\neg K)(E) = \partial LK(E)$. *Also,* $U^{(\infty)}(E^c) = \partial KL(E)$.
3. $U^{(\infty)}(E) = LK(E) \setminus KLK(E)$. *Also,* $U^{(\infty)}(E^c) = LKL(E) \setminus KL(E)$.
4. $U^{(2)}(E) = \partial K(E) (\subseteq \partial(E))$. *Also,* $U^{(2)}(E^c) = \partial L(E) (\subseteq \partial(E))$.
5. $U^{(2)}(E) = LK(E) \setminus K(E)$. *Also,* $U^{(2)}(E^c) = L(E) \setminus KL(E)$.

The first three statements in Proposition 1 characterize k^∞-unawareness. The first result relates $U^{(\infty)}$ and $U^{(2)}$ in that an agent is k^∞-unaware of an event E iff she is k^2-unaware of not knowing E.

The second and third statements characterize k^∞-unawareness by ignorance and possibility: an agent is k^∞-unaware of an event E iff she does not know whether she knows that she does not know E (i.e., she is ignorant of knowing that she does not know E) iff she does not know whether it is possible that she knows E (i.e., she is ignorant of the possibility that she knows E). To restate further, an agent is k^∞-unaware of an event E iff she considers it possible that she knows E but she does not know that it is possible that she knows E. At an interpretational level, the equivalence between the second and third statements lies in the fact that the notion of possibility is defined by the lack of knowledge (as well as in the given properties of knowledge (Truth Axiom, Positive Introspection, and Monotonicity)).

In the fourth and fifth statements, we characterize k^2-unawareness by ignorance and possibility: an agent is k^2-unaware of an event E iff she does not know whether she knows E (i.e., she is ignorant of (not) knowing E). To restate further, she is k^2-unaware of E iff she considers it possible that she knows E but she does not know E. Also, (k^2-)unawareness implies ignorance: if an agent is (k^2-)unaware of an event E then she is ignorant of E.

3.2 Characterization of Non-triviality

We say that $\mathcal{S} = \langle (\Omega, \mathcal{D}), (K, U) \rangle$ represents a *non-trivial form of unawareness* (or $\mathcal{S}$ is *non-trivial*) if there exists $E \in \mathcal{D}$ such that $U(E) \neq \emptyset$. We say that $\mathcal{S}$ is trivial otherwise. DLR [8, Theorem 1] show that any possibility correspondence model cannot capture unawareness

[19] If $\mathcal{D}$ is an algebra of sets, then we can restrict attention to $U^{(n)}$ with $n \in \{2, 3\}$ among any finite $n \geq 2$.

[20] Indeed, we will show in Proposition 4 that if $U^{(2)} = U^{(\infty)}$ then $U^{(2)}$ is degenerate in the sense that $U^{(2)}(\cdot) = (\neg K)(\Omega)$.

if an unawareness operator satisfies Plausibility, KU Introspection, and AU Introspection. MR [24] show that if unawareness satisfies Symmetry ($U(E) = U(E^c)$) then $\mathcal{S}$ is trivial.

Given DLR's [8] negative result, the following two questions naturally arise. First, when does a state space model represent a non-trivial form of unawareness? Second, what properties have to be retained in order to represent a non-trivial form of unawareness in a state space model (DLR [8, p. 166])?

This subsection examines the above first question by providing a necessary and sufficient condition for an information structure to be non-trivial. Our characterization implies that $\mathcal{S}^{(2)} = \langle(\Omega, \mathcal{D}), (K, U^{(2)})\rangle$ is generically non-trivial even when it is induced from a reflexive and transitive possibility correspondence model, in the sense that $\mathcal{S}$ is non-trivial as long as it is not a partitional model.

Throughout this subsection, we fix an information structure $\mathcal{S}^{(n)} = \langle(\Omega, \mathcal{D}), (K, U^{(n)})\rangle$, where $n \in \{2, \infty\}$. Now, we characterize the non-triviality.

Proposition 2. *1. $U^{(2)}(E) \neq \emptyset$ iff $K(E) \in \mathcal{J} \setminus \mathcal{A}$ (i.e., $K(E) \subsetneq LK(E)$).*

2. $U^{(\infty)}(E) \neq \emptyset$ iff $LK(E) \in \mathcal{A} \setminus \mathcal{J}$ (i.e., $KLK(E) \subsetneq LK(E)$).

Corollary 1. *1. $\mathcal{S}^{(2)}$ is non-trivial iff $\mathcal{J} \setminus \mathcal{A} \neq \emptyset$ iff $\mathcal{A} \setminus \mathcal{J} \neq \emptyset$ iff $\mathcal{A} \triangle \mathcal{J} (:= (\mathcal{J} \setminus \mathcal{A}) \cup (\mathcal{A} \setminus \mathcal{J})) \neq \emptyset$.*

2. $\mathcal{S}^{(\infty)}$ is non-trivial iff $\{F \in \mathcal{J} \setminus \mathcal{A} \mid L(F) \in \mathcal{A} \setminus \mathcal{J}\} \neq \emptyset$.

We make four remarks. First, the triviality of a partitional information structure follows from $\mathcal{J} = \mathcal{A}$. Recall that Negative Introspection implies that $\mathcal{J}$ is closed under complementation.

Second, as an immediate implication of Corollary 1, the failure of Necessitation implies the non-triviality, although the non-triviality is rather degenerate.[21] This follows because $K(\Omega) \subsetneq \Omega = LK(\Omega)$.[22] In general, the unawareness operators satisfy $(\neg K)(\Omega) \subseteq U^{(\infty)}(E) \subseteq U^{(2)}(E)$ for all $E \in \mathcal{D}$. At any state $\omega \in (\neg K)(\Omega)$, an agent does not know anything and she is unaware of everything.

Third, consider $\mathcal{S}^{(2)}$. Since it is non-trivial iff $\mathcal{J}$ is not closed under complementation, it follows that any reflexive and transitive possibility correspondence model is non-trivial as long as it is not partitional. An underlying intuition is very simple: $\mathcal{S}^{(2)}$ is non-trivial iff Negative Introspection is violated. Thus, any (properly) non-partitional model of knowledge represents a non-trivial form of unawareness.

Fourth, consider an information structure $\mathcal{S}^{(n)}$ which satisfies Necessitation. Noting that $K(E) = \emptyset$ implies $LK(E) = \emptyset$, the fact that an agent is (k^n-)unaware of an event $E \in \mathcal{D}$ implies that she knows E at some state ω and she does not know E at another state ω'. If

[21] The converse is not true. Agents i_1 and i_3 in Example 1 are such examples. We will shortly study the (non-)triviality of information structures satisfying Necessitation.

[22] Since $K(\neg K)(\Omega)$ is the smallest self-evident event, we have $K(\neg K)(\Omega) = \emptyset$, i.e., $\Omega = (\neg K)^2(\Omega) = LK(\Omega)$.

she does not know E at any state, then she knows that she does not know E at any state (because the event that she does not know E becomes a tautology), and thus she is not unaware of E at any state.

3.3 Properties of Unawareness

We keep considering a single agent's information structure $\mathcal{S}^{(n)} = \langle(\Omega, \mathcal{D}), (K, U^{(n)})\rangle$ with $n \in \{2, \infty\}$. Here, we first examine "positive" properties of $U^{(n)}$. Next, we study "negative" properties under which unawareness becomes rather degenerate.

First, by definition, any $\mathcal{S}^{(n)}$ satisfies Plausibility. Also, $\mathcal{S}^{(\infty)}$ satisfies Strong Plausibility (HMS [16, 17] and Schipper [31]): $U^{(\infty)}(E) \subseteq \bigcap_{r\in\mathbb{N}}(\neg K)^r(E)$ with equality.[23] Note that the statement that $U^{(n)}(\cdot) \subseteq \partial(\cdot)$ (i.e., unawareness implies ignorance) could also be regarded as a plausibility condition.

Second, any $\mathcal{S}^{(n)}$ satisfies KU Introspection. DLR [8, Footnote 10] note that any pair (K, U) which satisfies Truth Axiom, Monotonicity, and Plausibility also satisfies KU Introspection. Now, we provide other properties that any $\mathcal{S}^{(n)}$ satisfies.

Proposition 3. *Any $\mathcal{S}^{(n)}$ satisfies the following.*

1. $A^{(n)}(E) \subseteq K(\Omega)$. *Also,* $A^{(n)}U^{(n)}(E) = K(\Omega)$.
2. *Reverse AU Introspection:* $U^{(n)}U^{(n)}(E) \subseteq U^{(n)}(E)$. *Also,* $U^{(n)}U^{(n)}U^{(n)}(E) = U^{(n)}U^{(n)}(E)$.
3. *JU Introspection:* $U^{(n)}(E) = \partial U^{(n)}(E)$. *Equivalently,* $A^{(n)}(E) = JA^{(n)}(E)$.
4. *Weak A-Negative Introspection:* $(\neg K)(E) \cap A^{(2)}(E) = K(\neg K)(E)$.
5. *AK Self-Reflection:* $A^{(n)}(E) = A^{(n)}K(E)$. *Equivalently,* $U^{(n)}(E) = U^{(n)}K(E)$.
6. *A-Introspection:* $A^{(n)}(E) = KA^{(n)}(E)$. *Equivalently,* $U^{(n)}(E) = LU^{(n)}(E)$.
7. *Weak AA Self-Reflection:* $A^{(n)}(E) \subseteq A^{(n)}A^{(n)}(E)$ *with equality if* $n = 2$. *Also,* $A^{(n)}A^{(n)}(E) = A^{(n)}A^{(n)}A^{(n)}(E)$.

If K satisfies Finite Conjunction, then the following is also true.

8. *A-Weak Conjunction:* $\bigcap_{E\in\mathcal{E}} A^{(n)}(E) \subseteq A^{(n)}(\bigcap \mathcal{E})$, *where $\mathcal{E}$ is a non-empty finite subset of $\mathcal{D}$.*

Property 1 is a part of Weak Necessitation coined by DLR [8]. Property 2 is the "converse" of AU Introspection. We note that Reverse AU Introspection comes from Plausibility, KU Introspection, and Property 1 (a part of Weak Necessitation). Similarly, the ignorance operator also satisfies $\partial\partial(\cdot) \subseteq \partial(\cdot)$.

[23] We also use other terminologies coined by HMS [16, 17] and Schipper [31] (specifically, Properties 5, 6, and 8) postulated in Proposition 3.

JU Introspection (Property 3) states that if an agent is unaware of an event then she is *ignorant* of being unaware of it. Reverse AU Introspection is also seen as a consequence of JU Introspection since $U^{(n)}(\cdot) \subseteq \partial(\cdot)$ (see Proposition 1).

Weak A-Negative Introspection (Property 4) is proposed by Li [21]. If $n = \infty$, then this is equivalent to Weak Negative Introspection (Fagin and Halpern [11]) for $n = 2$: $(\neg K)(E) \cap A^{(2)}(\neg K)(E) = K(\neg K)(E)$, since $A^{(\infty)}(E) = A^{(2)}(\neg K)(E)$.

Weak A-Negative Introspection for $n = \infty$ (equivalently, Weak Negative Introspection for $n = 2$), however, may not hold (specifically, the "$\subseteq$" part may fail). That is, there may exist a state at which the following hold: an agent does not know an event E and she is (k^2-)aware that she does not know E but she does not know that she does not know E. Indeed, observe that Weak A-Negative Introspection is equivalent to $(\neg K)(\cdot) \cap (\neg K)^2(\cdot) \subseteq U^{(n)}(\cdot)$, and hence this may not hold for $n = \infty$.[24] As a particular example, consider Example 1. We have $K_{i_1}(\neg K_{i_1})(\{\omega_1, \omega_2\}) = \emptyset$ while $(\neg K_{i_1})(\{\omega_1, \omega_2\}) \cap A_{i_1}^{(\infty)}(\{\omega_1, \omega_2\}) = \{\omega_3\}$. Moreover, $(\neg K_{i_1})(\{\omega_1, \omega_2\}) \cap A_{i_1}^{(\infty)}(\neg K_{i_1})(\{\omega_1, \omega_2\}) = \{\omega_3\}$.

Properties 5,6, and 8 are proposed by MR [24, 25]. AK Self-Reflection (Property 5) is equivalently stated as $U^{(n)}(\cdot) = U^{(n)}K(\cdot)$: an agent is unaware of E iff she is unaware of knowing E. On the other hand, A-Introspection (Property 6) is equivalently stated as $U^{(n)} = LU^{(n)}$: an agent is unaware iff she considers it possible that she is unaware. It is equivalent to $U^{(n)} = (\neg K)A^{(n)}$, and hence an agent is unaware iff she does not know that she is aware.

Note that it is not necessarily true that $U^{(n)}(\cdot) = U^{(n)}L(\cdot)$: an agent is unaware of an event E iff she is unaware of the possibility of E.[25] Consider Example 1: $U_{i_1}^{(2)}(\{\omega_1, \omega_2\}) = \{\omega_3\}$ while $U_{i_1}^{(2)} L_{i_1}(\{\omega_1, \omega_2\}) = \emptyset$.

Property 7 (Weak AA Self-Reflection) is based on AA Self-Reflection (MR [24, 25]): $A^{(n)}(E) = A^{(n)}A^{(n)}(E)$. While AA Self-Reflection holds when $n = 2$, only the weak form (Property 6) is true when $n = \infty$. For instance, in Example 1, we have $A_{i_1}^{(\infty)}(\{\omega_1\}) = \{\omega_1, \omega_2\}$ while $A_{i_1}^{(\infty)} A_{i_1}^{(\infty)}(\{\omega_1\}) = \Omega$.

Consider Property 8 (A-Weak Conjunction). We have two remarks. First, it may not necessarily hold with equality. In Example 1, we have $A_{i_1}^{(n)}(\{\omega_1, \omega_3\}) \cap A_{i_1}^{(n)}(\{\omega_2, \omega_3\}) = \{\omega_1, \omega_2\} \subsetneq \Omega = A_{i_1}^{(n)}(\{\omega_3\})$ for $n \in \{2, \infty\}$. Second, A-Weak Conjunction holds for any non-empty collection of events $\mathcal{E}$ with respect to $A^{(2)}$ in any information structure satisfying Non-empty Conjunction. We, however, leave it an open question whether this extends to the case of $n = \infty$.

Let us go back to the question raised by DLR [8, p. 166], which of their three axioms is to be retained in a possibility correspondence model so as to capture a non-trivial

[24]Given an information structure $\langle(\Omega, \mathcal{D}), (K, U)\rangle$, one set of axioms that induces $U = U^{(2)}$ trivially turns out to be Plausibility and Weak A-Negative Introspection. It would be an interesting question to ask whether there are other "non-trivial" combinations of axioms that yield $U = U^{(2)}$ given a pair (K, U).

[25]We have $U^{(2)}L(E) = U^{(2)}(\neg K)(E^c) = U^{(\infty)}(E^c)$ and $U^{(\infty)}L(E) = (\neg K)^3(E^c) \cap (\neg K)^4(E^c) = (\neg K)^3(E^c) \cap (\neg K)^2(E^c) = U^{(\infty)}(E^c)$.

form of unawareness. One implication of Proposition 3 is that any information structure satisfies Plausibility (by definition), KU Introspection, Reverse AU Introspection, and JU Introspection (instead of AU Introspection).

Now, we turn to examining other properties which lead to a degenerate form of unawareness. Especially, these properties lead to trivial unawareness under Necessitation.

Proposition 4. *Let $\mathcal{S}^{(n)}$ be an information structure.*

1. *Let $n = 2$. (a)-(g) are all equivalent to $U^{(2)}(\cdot) = (\neg K)(\Omega)(= L(\emptyset))$.*

2. *Let $n = \infty$. (f)-(i) are all equivalent to $U^{(\infty)}(\cdot) = (\neg K)(\Omega)(= L(\emptyset))$.*

(a) Subjective Negative Introspection: $K(\Omega) \setminus K(E) \subseteq K(K(\Omega) \setminus K(E))$.

(b) If $E \in \mathcal{J}$ then $K(\Omega) \setminus E \in \mathcal{J}$.

(c) Negative Non-Introspection: $(\neg K)(E) \cap (\neg K)^2(E) \subseteq (\neg K)^3(E)$.

(d) $U^{(2)} = U^{(\infty)}$.

(e) Symmetry of $U^{(2)}$: $U^{(2)}(E) = U^{(2)}(E^c)$.

(f) AU introspection of $U^{(n)}$.

(g) Monotonicity of $U^{(n)}$: if $E \subseteq F$ then $U^{(n)}(E) \subseteq U^{(n)}(F)$.

(h) Monotonicity of $A^{(n)}$: if $E \subseteq F$ then $A^{(n)}(E) \subseteq A^{(n)}(F)$ (i.e., $U^{(n)}(F) \subseteq U^{(n)}(E)$).

(i) AA Self Reflection of $A^{(\infty)}$: $A^{(\infty)}(E) = A^{(\infty)}A^{(\infty)}(E)$.

We make the following four remarks. First, while (a) states that $K(\Omega) \setminus K(E)$ is self-evident, (b) states that if E is self-evident then so is $K(\Omega) \setminus E$. The equivalence between (a) and (f) (i.e., AU Introspection) is closely related to CEL [6] in the sense that Subjective Negative Introspection reduces to Negative Introspection under Necessitation.

Second, the terminology, Negative Non-Introspection, is from Schipper [31]. It is clearly equivalent to (d). This part says that unawareness is degenerate when $U^{(2)} = U^{(\infty)}$.

Third, MR [24, Theorem] show the triviality of k^2-unawareness under Symmetry. While Symmetry of $U^{(2)}$ yields a rather degenerate form of unawareness, Symmetry of $U^{(\infty)}$ does not necessarily imply this property (e.g., agent i_1 in Example 1).

Fourth, if $U^{(2)}(\cdot) = (\neg K)(\Omega)$ then we have $U^{(2)} = U^{(\infty)}$. However, $U^{(\infty)}(\cdot) = (\neg K)(\Omega)$ does not necessarily imply $U^{(2)} = U^{(\infty)}$ (e.g., agent i_3 in Example 1).

4 Further Properties of Unawareness

4.1 Knowledge of Self-awareness

We ask the *knowledge of self-unawareness*, i.e., we ask whether there exists a state in which an agent knows that she is unaware of "something" even though KU Introspection requires that she do not know that she is unaware of any *particular* event.[26]

Throughout this subsection, fix $\mathcal{S}^{(n)} = \langle (\Omega, \mathcal{D}), (K, U^{(n)}) \rangle$ with $n \in \{2, \infty\}$. We denote the event that an agent is unaware of something by $\overline{U}^{(n)} := \{\omega \in \Omega \mid \omega \in U^{(n)}(E) \text{ for some } E \in \mathcal{D}\} = \bigcup_{E \in \mathcal{D}} U^{(n)}(E)$. Note that $\overline{U}^{(n)}$ is a well-defined event (i.e., $\overline{U}^{(n)} \in \mathcal{D}$) and that $\mathcal{S}^{(n)}$ is non-trivial iff $\overline{U}^{(n)} \neq \emptyset$.

Proposition 5. *1. Assume Finite Conjunction on K. If $\mathcal{D}$ is finite, then $K(\overline{U}^{(n)}) = \emptyset$ and $A^{(n)}(\overline{U}^{(n)}) = K(\Omega)$ (i.e., $U^{(n)}(\overline{U}^{(n)}) = (\neg K)(\Omega)$). It is possible that $K(\overline{U}^{(n)}) \neq \emptyset$ when $\mathcal{D}$ is infinite.*

2. $A(\overline{U}^{(n)}) \neq \emptyset$ provided that $K(\Omega) \neq \emptyset$. Also, $U^{(n)}(\overline{U}^{(n)}) \subseteq \overline{U}^{(n)}$.

3. If K satisfies Arbitrary Conjunction, then $\overline{U}^{(n)} = L(\overline{U}^{(n)})$.

The first part of Proposition 5 states that, for any information structure satisfying Finite Conjunction, if the domain is finite then an agent never knows that she is unaware of something (i.e., $K(\overline{U}^{(n)}) = \emptyset$). Thus, whenever she knows something, she infers that she never knows that she is unaware of something: she is aware that she is unaware of something (i.e., $K(\Omega) = K(\neg K)(\overline{U}^{(n)}) = A^{(n)}(\overline{U}^{(n)})$). On the other hand, it is possible that an agent knows that she is unaware of something, if a given domain is infinite, even though she never knows that she is unaware of any particular event.

The second part states the following. Suppose that an agent's knowledge is not degenerate in the sense that $K(E) \neq \emptyset$ for some $E \in \mathcal{D}$. Then, there is always a state at which the agent is aware of being unaware of something.

The third part implies that an agent is unaware of something iff she considers it possible that she is unaware of something in any information structure satisfying Arbitrary Conjunction. Note that the statement is true for any information structure satisfying Non-empty Conjunction as long as $\overline{U}^{(n)} \neq \emptyset$. If a given domain is finite then the statement is clearly true for any information structure satisfying Finite Conjunction.

4.2 Non-monotonicity of Unawareness in Knowledge

Proposition 2 and Corollary 1 show that the non-triviality of unawareness hinges on the qualitative feature of knowledge (e.g., whether the lack of knowledge is self-evident when

[26] See also Schipper [31, Section 3.5] and the references therein for syntactic approaches to self-awareness.

unawareness is defined by the two levels of the lack of knowledge). Thus, the non-triviality is not related to knowledgeability.

Here, we take a further look at the non-monotonicity of unawareness in knowledgeability through examples. An underlying intuition is that, while an increase in knowledge enhances awareness through knowledge itself, a decrease in knowledge also enhances awareness through the knowledge of the lack of knowledge. In an extreme case, an agent with her self-evident collection $\mathcal{J} = \{\emptyset, \Omega\}$ is not unaware of any event. This is because she always knows that she does not any non-tautological event.

Now, we examine three cases in which knowledge and unawareness exhibit non-monotonicity. In order to keep the discussion simple, we use Example 1. First, we consider the "monotonicity of awareness in knowledge." One agent is at least as knowledgeable as another at a certain state, and the better informed agent is aware of any event, at that state, of which the less informed agent is aware at that state. Consider agents i_1 and i_4 in Example 1. Agent i_1 is at least as knowledgeable as i_4 at any state, and i_1 is aware of any event of which i_4 is aware at any state.

The second case, on the other hand, exhibits "monotonicity of unawareness in knowledge." One agent is at least as knowledgeable as another at a given state, and the better informed agent is unaware, at that state, of any event of which the less informed agent is unaware at that state. Again, consider agent i_1 in Example 1 and agent whose knowledge is defined by $\{\emptyset, \Omega\}$. As Corollary 1 implies, the latter agent is aware of every event at each state. In this example the less informed agent knows her own ignorance, which leads to more awareness.

In the third case, we consider two agents who are equally knowledgeable at a given state, where one is unaware of an event at that state while the other is aware of it at the state. For example, consider agents i_2 and i_3 in Example 1. At state $\omega \in \{\omega_2, \omega_3\}$, they are equally knowledgeable. Consider state ω_2. Agent i_3 is not unaware of any event $E \in \{\emptyset, \{\omega_2\}, \{\omega_3\}, \{\omega_2, \omega_3\}, \Omega\}$ while agent i_2 is unaware of every event. Consider ω_3. Agent i_3 is not unaware of any event $E \in \{\emptyset, \{\omega_2\}, \{\omega_3\}, \{\omega_2, \omega_3\}, \Omega\}$ while agent i_2 is aware of every event.

Finally, we have two remarks. First, we can view such comparison of agents' knowledge and unawareness as one agent's knowledge and unawareness over time. To that end, let a state space be given by $\Omega = \{\omega_1, \omega_2, \omega_3\}$ as in Example 1. Denote an agent i's knowledge at time t by $\mathcal{J}_{i(t)}$. Specifically, we let $\mathcal{J}_{i(0)} = \mathcal{J}_{i_4}$, $\mathcal{J}_{i(1)} = \mathcal{J}_{i_1}$, and $\mathcal{J}_{i(2)} = \{\emptyset, \Omega\}$. At time 1, getting more information causes agent i to get aware of some event at each realized state. At time 2, on the other hand, she "forgets" some events, and this may make her aware of some events at some states.

Second, the entire discussion also applies to common knowledge. It is possible that if some event is not commonly known then it is commonly known that this is not common knowledge. When each agent receives some events, on the contrary, it may become possible that it is not common knowledge that this is not common knowledge.

4.3 Possible Forms of Monotonicity of Unawareness in Knowledge

We examine possible forms of monotonicity of unawareness in knowledgeability. The key observation is monotonicity of the knowledge and ignorance operators in knowledgeability. Thus, if j is at least as knowledgeable as i, then KU Introspection of $(K_j, U_j^{(n)})$ implies that $K_i U_j^{(n)}(E) = \emptyset$.

Ignorance is "decreasing" in knowledge because ignorance of an event E is expressed in terms of the lack of knowledge of E and its negation E^c. That is, for any event E, if j is ignorant of E then so is i, provided that j is at least as knowledgeable as i. Monotonicity of these operators on knowledgeability implies the following.

Proposition 6. *Let j be at least as knowledgeable as i. Fix $n \in \{2, \infty\}$ and $E \in \mathcal{D}$.*

1. *(a) $\partial_j(K_i E) \subseteq U_i^{(2)}(E)$. Equivalently, $A_i^{(2)}(E) \subseteq J_j(K_i E)$.*
   (b) $\partial_j(L_i K_i E) \subseteq U_i^{(\infty)}(E)$. Equivalently, $A_i^{(\infty)}(E) \subseteq J_j(L_i K_i E)$.
   (c) $\partial_j U_i^{(n)}(E) \subseteq U_i^{(n)}(E)$ and $U_j^{(n)}(E) \subseteq \partial_i U_j^{(n)}(E)$.
2. *(a) $U_j^{(2)}(E) \subseteq \partial_i(K_j E)$. Equivalently, $J_i(K_j E) \subseteq A_j^{(2)}(E)$.*
   (b) $U_j^{(\infty)}(E) \subseteq \partial_i(L_j K_j E)$. Equivalently, $J_i(L_j K_j E) \subseteq A_j^{(\infty)}(E)$.
3. *$A_i^{(n)}(E) = K_j A_i^{(n)}(E) = A_i^{(n)} K_j K_i(E)$. Also, $U_i^{(n)}(E) = L_j U_i^{(n)}(E) = U_i^{(n)} K_j K_i(E)$.*
4. *$A_i^{(n)}(E) \subseteq A_j^{(n)} A_i^{(n)}(E)$ and $U_j^{(n)} U_i^{(n)}(E) \subseteq U_i^{(n)}(E)$.*

Suppose that an agent j is at least as knowledgeable as i. The first two statements are comparative statics of unawareness with respect to ignorance. First, if j is ignorant of i knowing E, then i is (k^2-)unaware of E. Second, on the contrary, if j is (k^2-)unaware of E, then i is ignorant of j knowing E.

The third statement says that i's awareness of an event is self-evident to j. Likewise, we show in the fourth statement that i's awareness of an event E implies j's awareness of i's awareness of E. We also show that if j is unaware of i's unawareness of an event E then i is indeed unaware of E.

While we demonstrated possible forms of monotonicity of unawareness, it is, however, to be noted that the fact that $K_i(\cdot) \subseteq K_j(\cdot)$ does not necessarily imply that $U_j(\cdot) \subseteq U_i(\cdot)$ (Example 1 with $(i, j) = (i_1, i_3)$).

Finally, the first part of the next proposition compares an agent's unawareness at different states, while the second part compares different agents' unawareness at a given state.

Proposition 7. *Fix $n \in \{2, \infty\}$.*

1. *Suppose that state ω' is as informative as ω in the sense that $\mathrm{IK}_i(\omega) \subseteq \mathrm{IK}_i(\omega')$. If $\omega' \in U_i^{(n)}(E)$ then $\omega \in U_i^{(n)}(E)$ (i.e., if $\omega \in A_i^{(n)}(E)$ then $\omega' \in A_i^{(n)}(E)$).*

2. Suppose that j is at least as knowledgeable as i at ω. If $\omega \in U_j^{(n)}(E)$, then $\omega \notin K_i U_j^{(n)}(E)$. If $\omega \in A_i^{(n)}(E)$, then $\omega \in K_j A_i^{(n)}(E)$.

The first part of Proposition 7 says that if i is unaware of an event E at ω' and if ω' ranks no lower than ω in her informativeness relation (i.e. she knows at ω' any event that she knows at ω), then she is also unaware of E at ω. On the other hand, with regard to the relation, at least as knowledgeable as at a state, if an agent j is at least as knowledgeable as an agent i at a state ω and if j is unaware of an event E there, then i does not know that j is unaware of E there. Interestingly, it is not always true that if an agent j is at least as knowledgeable as an agent i at a state ω and if j is unaware of an event E there, then i is unaware of E there.

References

[1] Aumann, R. J. "Agreeing to Disagree." Ann. Statist. 4 (1976), 1236-1239.

[2] Aumann, R. J. "Interactive Epistemology I, II." Int. J. Game Theory. 28 (1999), 263-300, 301-314.

[3] Bacharach, M. "Some extensions of a claim of Aumann in an axiomatic model of knowledge," J. Econ. Theory 37 (1985), 167-190.

[4] Binmore, K. and A. Brandenburger. "Common Knowledge and Game Theory." in K. Binmore. *Essays on the Foundations of Game Theory.* Basil Blackwell, 1990, 105-150.

[5] Brandenburger, A., E. Dekel, and J. Geanakoplos. "Correlated Equilibrium with Generalized Information Structures." Games Econ. Behav. 4 (1992), 182-201.

[6] Chen, Y-C., J. C. Ely, and X. Luo. "Note on Unawareness: Negative Introspection versus AU Introspection (and KU Introspection)." Int. J. Game Theory 41 (2012), 325-329.

[7] Dekel, E., and F. Gul, "Rationality and Knowledge in Game Theory," in Kreps, D. M., and K. F. Wallis (eds.), *Advances in Economics and Econometrics: Theory and Applications*, Volume 1, Seventh World Congress, Cambridge University Press, 1997.

[8] Dekel, E., B. L. Lipman, and A. Rustichini. "Standard State-Space Models Preclude Unawareness." Econometrica 66 (1998), 159-173.

[9] Doignon, J.-P. and J.-C. Falmagne. "Spaces for the Assessment of Knowledge." Int. J. Man Mach. Stud. 23 (1985), 175-196.

[10] Doignon, J.-P. and J.-C. Falmagne. Knowledge Spaces. Springer-Verlag, Berlin Heidelberg, 1999.

[11] Fagin, R. and Halpern, J. Y. “Belief, Awareness, and Limited Reasoning.” Art. Intell. 34 (1988), 39-76.

[12] Friedell, M. F. “On the Structure of Shared Awareness.” Behav. Sci. 14 (1969), 28-39.

[13] Fukuda, S. “The Existence of Universal Knowledge Spaces.” 2017.

[14] Geanakoplos, J. D. “Games theory without Partitions, and Applications to Speculation and Consensus,” Cowles Foundation Discussion Paper No. 914, Yale University, 1989.

[15] Hart, S., A. Heifetz, and D. Samet. ““Knowing Whether,” “Knowing That,” and The Cardinality of State Spaces.” J. Econ. Theory 70 (1996), 249-256.

[16] Heifetz, A., M. Meier, and B. C. Schipper. “Interactive Unawareness.” J. Econ. Theory 130 (2006), 78-94.

[17] Heifetz, A., M. Meier, and B. C. Schipper. “A Canonical Model of Interactive Unawareness.” Games Econ. Behav. 62 (2008), 304-324.

[18] Hintikka, J. *Knowledge and Belief: An Introduction to the Logic of the Two Notions.* Cornell University Press, 1962.

[19] Lehrer, E., and D. Samet. “Agreeing to Agree.” Theoretical Econ. 6 (2011), 269-287.

[20] Lewis, D. K. *Convention: A Philosophical Study.* Cambridge: Harvard University Press, 1969.

[21] Li, J. “Information Structures with Unawareness .” J. Econ. Theory 144 (2009), 977-993.

[22] McCarthy, J., M. Sato, T. Hayashi, and S. Igarashi. (1978) “On the Model Theory of Knowledge.” Stanford Artificial Intelligence Laboratory AIM-312, Computer Science Department STAN-CS-78-657.

[23] Milgrom, P. “An Axiomatic Characterization of Common Knowledge.” Econometrica 49 (1981), 219-222.

[24] Modica S. and A. Rustichini. “Awareness and Partitional Information Structures,” Theory Dec. 37 (1994), 107-124.

[25] Modica S. and A. Rustichini. “Unawareness and Partitional Information Structures,” Games Econ. Behav. 27 (1999), 265-298.

[26] Morris, S. “The Logic of Belief and Belief Change: A Decision Theoretic Approach.” J. Econ. Theory. 69 (1996), 1-23.

[27] Pires, C. P. “Do I know Ω? An Axiomatic Model of Awareness and Knowledge.” A Manuscript, 1994.

[28] Rubinstein, A., and A. Wolinsky. “On the Logic of “Agreeing to Disagree” Type Results.” J. Econ. Theory 51 (1990), 184-193.

[29] Samet, D. “Ignoring Ignorance and Agreeing to Disagree.” J. Econ. Theory 52 (1990), 190-207.

[30] Samet, D. “Agreeing to disagree in infinite information structures.” Int. J. Game Theory 21 (1992), 213-218.

[31] Schipper, B. C. “Awareness,” in: Handbook of Epistemic Logic, Chapter 3, H. van Ditmarsch, J. Y. Halpern, W. van der Hoek and B. Kooi (Eds.), College Publications, London, 2015, 77-146.

[32] Shin, H. S. “Logical Structure of Common Knowledge.” J. Econ. Theory 60 (1993), 1-13.

[33] Willard, S. *General Topology*. New York: Dover Publications, 2004.

A Appendix

A.1 Section 3.1

In order to prove Proposition 1, we consider the following preliminary lemma.

Lemma A.1. *Fix* $E \in \mathcal{D}$.

1. $K(E) \subseteq LK(E) = (\neg K)^2(E) = (\neg K)^{2n}(E) \subseteq (\neg K)(E^c)(= L(E))$.
2. $((\neg L)(E) =)K(E^c) \subseteq (\neg K)^{2n+1}(E) = (\neg K)^3(E) \subseteq (\neg K)(E)$.

Proof of Lemma A.1. It is enough to show (1). First, Truth Axiom and Monotonicity imply that $K(E) \subseteq LK(E) \subseteq L(E)$. Second, by definition, we have $L(E) = (\neg K)(E^c)$ and $LK(E) = (\neg K)^2(E)$. Now, it suffices to show $(\neg K)^2(E) = (\neg K)^4(E)$, i.e., $LK(E) = LKLK(E)$. Since Truth Axiom implies $K(E) \subseteq LK(E)$, Positive Introspection and Monotonicity imply $K(E) \subseteq KK(E) \subseteq KLK(E)$. By Monotonicity, we get $LK(E) \subseteq LKLK(E)$. Conversely, Truth Axiom implies $KLK(E) \subseteq LK(E)$. Monotonicity and Positive Introspection imply $LKLK(E) \subseteq LLK(E) \subseteq LK(E)$. $\square$

Proof of Proposition 1. Part 1. By Lemma A.1, we have $U^{(\infty)}(E) = U^{(3)}(E) = (\neg K)^2(E) \cap (\neg K)^3(E) = U^{(2)}(\neg K)(E)$. Equivalently, $A^{(\infty)}(E) = A^{(3)}(E) = K(\neg K)(E) \cup K(\neg K)^2(E) = A^{(2)}(\neg K)(E)$.

Part 4. First, we have $U^{(2)}(E) = (\neg K)(E) \cap (\neg K)^2(E) = (\neg K)(KE) \cap (\neg K)(\neg KE) = \partial(\neg K)(E) = \partial K(E)$. Second, $\partial KE = (\neg K)KE \cap (\neg K)(\neg K)(E) = (\neg K)E \cap LKE \subseteq (\neg K)E \cap LE = (\neg K)E \cap (\neg K)(E^c) = \partial E$. Third, we have $U^{(2)}(E^c) = \partial K(E^c) = \partial(\neg K)(E^c) = \partial L(E)$. Also, $\partial L(E) = \partial K(E^c) \subseteq \partial(E^c) = \partial(E)$.

Part 5. We have $U^{(2)}(E) = LK(E) \cap (\neg K)(E) = LK(E) \setminus K(E)$. Also, $U^{(2)}(E^c) = (\neg K)(\neg K)(E^c) \cap (\neg K)(E^c) = (\neg K)L(E) \cap L(E) = L(E) \setminus KL(E)$.

Part 2. It follows that $U^{(\infty)}(E) = U^{(2)}(\neg K)(E) = \partial K(\neg K)(E) = \partial(\neg K)^2(E) = \partial LK(E)$. Also, $U^{(\infty)}(E^c) = \partial K(\neg K)(E^c) = \partial KL(E)$.

Part 3. We obtain $U^{(\infty)}(E) = \partial LK(E) = (\neg K)LK(E) \cap (\neg K)(\neg L)K(E) = LLK(E) \setminus KLK(E) = LK(E) \setminus KLK(E)$. Likewise, $U^{(\infty)}(E^c) = \partial KL(E) = (\neg K)KL(E) \cap (\neg K)(\neg K)L(E) = (\neg K)L(E) \cap LKL(E) = LKL(E) \setminus KL(E)$. $\square$

As a remark, we show below further properties of unawareness.

Proposition A.1. *Fix* $E \in \mathcal{D}$.

1. $U^{(2)}(E) \cap U^{(2)}(E^c) \subseteq U^{(\infty)}(E)$.

2. $U^{(2)}(E) \cup U^{(2)}(E^c) \subseteq U^{(2)}J(E)$.

Proof of Proposition A.1. 1. We have the following:

$$\begin{aligned}U^{(2)}(E) \cap U^{(2)}(E^c) &= (\neg K)(E) \cap (\neg K)^2(E) \cap (\neg K)(E^c) \cap (\neg K)^2(E^c)\\ &= (\neg K)(E) \cap (\neg K)^2(E) \cap L(E) \cap (\neg K)L(E)\\ &\subseteq (\neg K)(E) \cap (\neg K)^2(E) \cap (\neg K)LK(E)\\ &= (\neg K)(E) \cap (\neg K)^2(E) \cap (\neg K)^3(E) = U^{(\infty)}(E).\end{aligned}$$

2. Noting that $KJ(E) = J(E)$, we have the following:

$$\begin{aligned}U^{(2)}J(E) &= (\neg K)J(E) \cap (\neg K)^2 J(E) = \partial(E) \cap (\neg K)\partial(E)\\ &= \partial(E) \cap (\neg K)((\neg K)(E) \cap (\neg K)(E^c))\\ &\supseteq \partial(E) \cap \neg(K(\neg K)(E) \cap K(\neg K)(E^c))\\ &= \partial(E) \cap ((\neg K)^2(E) \cup (\neg K)^2(E^c))\\ &= (\partial(E) \cap (\neg K)^2(E)) \cup (\partial(E) \cap (\neg K)^2(E^c))\\ &= (\partial(E) \cap (\neg K)^2(E)) \cup (\partial(E^c) \cap (\neg K)^2(E^c)).\end{aligned}$$

Now, since Lemma A.1 implies that

$$\partial(E) \cap (\neg K)^2(E) = (\neg K)(E) \cap (\neg K)^2(E) \cap (\neg K)(E^c) = U^{(2)}(E),$$

we obtain $U^{(2)}J(E) \supseteq U^{(2)}(E) \cup U^{(2)}(E^c)$. The equality holds when K satisfies Finite Conjunction.

□

A.2 Section 3.2

Proof of Proposition 2. 1. Suppose that there is $E \in \mathcal{D}$ such that $U^{(2)}(E) \neq \emptyset$. Since $U^{(2)}(E) = (\neg K)(E) \cap LK(E)$, we must have $K(E) \neq LK(E)$. Then we have $K(E) \in \mathcal{J} \setminus \mathcal{A}$.

Conversely, suppose that there is $E \in \mathcal{D}$ with $K(E) \in \mathcal{J} \setminus \mathcal{A}$. Then, there is $\omega \in LK(E) \setminus K(E)$. That is, $U^{(2)}(E) = LK(E) \setminus K(E) \neq \emptyset$.

2. Suppose that $\emptyset \neq U^{(\infty)}(E)$ for some $E \in \mathcal{D}$. Since $U^{(\infty)}(E) = LK(E) \setminus KLK(E)$, we have $KLK(E) \neq LK(E)$. We obtain $LK(E) \in \mathcal{A} \setminus \mathcal{J}$.

Conversely, suppose that there is $E \in \mathcal{D}$ such that $LK(E) \in \mathcal{A} \setminus \mathcal{J}$. Then, we have $KLK(E) \subsetneq LK(E)$. That is, $U^{(\infty)}(E) = LK(E) \setminus KLK(E) \neq \emptyset$.

□

Proof of Corollary 1.

1. If $\mathcal{S}^{(2)}$ is non-trivial, then $U^{(2)}(E) \neq \emptyset$ for some $E \in \mathcal{D}$. It follows from Proposition 2 that $F := K(E) \in \mathcal{J} \setminus \mathcal{A}$, i.e., $\mathcal{J} \setminus \mathcal{A} \neq \emptyset$. Conversely, if $\mathcal{J} \setminus \mathcal{A} \neq \emptyset$, then there is $K(E) = E \in \mathcal{J} \setminus \mathcal{A}$, and hence $U^{(2)}(E) \neq \emptyset$, i.e., $\mathcal{S}^{(2)}$ is non-trivial. The rest follows because $E \in \mathcal{J} \setminus \mathcal{A}$ iff $E^c \in \mathcal{A} \setminus \mathcal{J}$.

2. If $\mathcal{S}^{(\infty)}$ is non-trivial, then $U^{(\infty)}(E) \neq \emptyset$ for some $E \in \mathcal{D}$. It follows from Proposition 2 that $F := K(E) \in \mathcal{J} \setminus \mathcal{A}$ such that $L(F) = LK(E) \in \mathcal{A} \setminus \mathcal{J}$. Hence, $\{E \in \mathcal{J} \setminus \mathcal{A} \mid L(E) \in \mathcal{A} \setminus \mathcal{J}\} \neq \emptyset$. Conversely, if $\{E \in \mathcal{J} \setminus \mathcal{A} \mid L(E) \in \mathcal{A} \setminus \mathcal{J}\} \neq \emptyset$, then there is $K(E) = E \in \mathcal{J} \setminus \mathcal{A}$ such that $L(E) = LK(E) \in \mathcal{A} \setminus \mathcal{J}$, and hence $U^{(\infty)}(E) \neq \emptyset$, i.e., $\mathcal{S}^{(\infty)}$ is non-trivial.

□

A.3 Section 3.3

Before we show Proposition 3, we prove an intermediate result, which will be used for proving (the last statement of) Proposition 3 (and Proposition 5).

Lemma A.2. *Let $\mathcal{S} = \langle(\Omega, \mathcal{D}), (K, U)\rangle$ be an information structure satisfying Finite Conjunction. For any $E, F \in \mathcal{D}$, we have $K(E \cup F) \cup L(F) = KE \cup LF$.*

Proof of Lemma A.2. Monotonicity implies $K(E \cup F) \cup L(F) \supseteq KE \cup LF$. Conversely, Finite Conjunction and Monotonicity imply that $K(E \cup F) \cap K(F^c) = K(E \cap F^c) \subseteq KE$. Then, we get $K(E \cup F) \cup LF = (K(E \cup F) \cap K(F^c)) \cup LF \subseteq KE \cup LF$. □

Proof of Proposition 3. Fix $E \in \mathcal{D}$. Since $A^{(n)}(E) = \bigcup_{r=1}^{n} K(\neg K)^{r-1}(E)$ is self-evident, A-Introspection ($A^n(E) = KA^{(n)}(E)$) clearly holds.

1. First, since Monotonicity implies that $K(F) \subseteq K(\Omega)$ for any $F \in \mathcal{D}$, it is clear that $A^{(n)}(E) \subseteq K(\Omega)$. Second, it follows from KU Introspection that $A^{(n)}U^{(n)}(E) \supseteq K(\neg K)U^{(n)}(E) = K(\Omega)$. Hence, $A^{(n)}U^{(n)}(E) = K(\Omega)$.

2. First, we have

$$(\neg K)(\Omega) \subseteq U^{(n)}U^{(n)}(E) \subseteq (\neg K)U^{(n)}(E) \cap (\neg K)^2 U^{(n)}(E) = (\neg K)(\Omega) \subseteq U^{(n)}(E),$$

   where the first set-inclusion is from Property 1 (a part of Weak Necessitation), the second set-inclusion is from Plausibility, the next equality is from KU Introspection, and the last set-inclusion is from Property 1 (a part of Weak Necessitation).[A.1] Second, since $U^{(n)}U^{(n)}(F) = (\neg K)(\Omega)$ for all $F \in \mathcal{D}$, we have $U^{(n)}U^{(n)}U^{(n)}(E) = U^{(n)}U^{(n)}(E)$.

3. JU Introspection follows from KU Introspection and A-Introspection: $\partial U^{(n)}(E) = (\neg K)U^{(n)}(E) \cap (\neg K)A^{(n)}(E) = U^{(n)}(E)$.

[A.1] Alternatively, it follows from $U^{(n)}(\cdot) \subseteq \partial(\cdot)$ (Proposition 1) and JU Introspection that $U^{(n)}U^{(n)}(E) \subseteq \partial U^{(n)}(E) = U^{(n)}(E)$.

4. Weak A-Negative Introspection simply follows from the definition of $A^{(2)}$.

5. It follows from Truth Axiom and Positive Introspection that $K(E) = KK(E)$, and thus $A^{(n)}(KE) = A^{(n)}(E)$.

6. We have already shown A-Introspection.

7. First, let $n = 2$. It follows from A-Introspection and KU Introspection that $U^{(2)}A^{(2)}(E) = (\neg K)A^{(2)}(E) \cap (\neg K)(\neg K)A^{(2)}(E) = U^{(2)}(E) \cap (\neg K)U^{(2)}(E) = U^{(2)}(E)$. Then, we get $A^{(2)}(E) = A^{(2)}A^{(2)}(E)$.

   Next, let $n = \infty$. It also follows from A-Introspection that $U^{(\infty)}A^{(\infty)}(E) \subseteq (\neg K)A^{(\infty)}(E) = U^{(\infty)}(E)$. We get $A^{(\infty)}(E) \subseteq A^{(\infty)}A^{(\infty)}(E)$. Since $A^{(\infty)}A^{(\infty)}(E) = K(\Omega)$, we have $A^{(\infty)}A^{(\infty)}A^{(\infty)}(E) = A^{(\infty)}A^{(\infty)}(E)$.

8. We show that $U^{(n)}(\bigcap \mathcal{E}) \subseteq \bigcup_{E\in\mathcal{E}} U^{(n)}(E)$, where $\mathcal{E}$ is a non-empty finite subset of $\mathcal{D}$. For $n = 2$, it follows from Finite Conjunction (and Monotonicity) that

$$
\begin{aligned}
U^{(2)}(\bigcap \mathcal{E}) = (\neg K)(\bigcap \mathcal{E}) \cap (\neg K)^2(\bigcap \mathcal{E}) &= \bigcup_{E\in\mathcal{E}} (\neg K)(E) \cap (\neg K)^2(\bigcap \mathcal{E}) \\
&\subseteq \bigcup_{E\in\mathcal{E}} (\neg K)(E) \cap \bigcap_{E\in\mathcal{E}} (\neg K)^2(E) \\
&\subseteq \bigcup_{E\in\mathcal{E}} ((\neg K)(E) \cap (\neg K)^2(E)) = \bigcup_{E\in\mathcal{E}} U^{(2)}(E).
\end{aligned}
$$

   Note that the same proof clearly works for any non-empty (countable) set $\mathcal{E}$ if $\mathcal{S}$ satisfies Non-empty (Countable) Conjunction.

   Next, consider $n = \infty$. We show:

$$
\begin{aligned}
U^{(\infty)}(\bigcap \mathcal{E}) &= (\neg K)^2(\bigcap \mathcal{E}) \cap (\neg K)^3(\bigcap \mathcal{E}) \\
&\subseteq \bigcap_{E\in\mathcal{E}} (\neg K)^2(E) \cap (\neg K)^2(\bigcup_{E\in\mathcal{E}} (\neg K)(E)) \\
&\subseteq \bigcap_{E\in\mathcal{E}} (\neg K)^2(E) \cap \bigcup_{E\in\mathcal{E}} (\neg K)^3(E) \\
&\subseteq \bigcup_{E\in\mathcal{E}} ((\neg K)^2(E) \cap (\neg K)^3(E)) = \bigcup_{E\in\mathcal{E}} U^{(\infty)}(E),
\end{aligned}
$$

   where the second line follows from Finite Conjunction (and Monotonicity). Thus, it suffices to show the following for a finite Λ, which implies the third line:

$$
(\neg K)^2(\bigcup_{\lambda\in\Lambda} F_\lambda) \subseteq \bigcup_{\lambda\in\Lambda} (\neg K)^2(F_\lambda), \text{ where } F_\lambda = (\neg K)(E_\lambda).
$$

Without loss of generality, assume $\Lambda = \{1, 2\}$. It follows from Lemma A.2 that $K(F_1 \cup F_2) \subseteq K(F_1) \cup L(F_2) \subseteq LK(F_1) \cup F_2$, where note that $LF_\lambda = F_\lambda$ for each λ. Then, Monotonicity and Finite Conjunction (of K) imply that $LK(F_1 \cup F_2) \subseteq L(LK(F_1) \cup F_2) = LK(F_1) \cup F_2$. On the other hand, it follows from Lemma A.2 that $K(LK(F_1) \cup F_2) \subseteq K(F_2) \cup LLK(F_1) \subseteq LK(F_1) \cup LK(F_2)$, and hence Monotonicity implies $LK(LK(F_1) \cup F_2) \subseteq L(LK(F_1) \cup LK(F_2)) = LK(F_1) \cup LK(F_2)$. Now, recalling that $LKLK(\cdot) = LK(\cdot)$ (see Lemma A.1), we have

$$LK(F_1 \cup F_2) = LKLK(F_1 \cup F_2) \subseteq LK(LK(F_1) \cup F_2) \subseteq LK(F_1) \cup LK(F_2).$$

□

Remark A.1. We have established A-Weak Conjunction in Proposition 3. Here, we provide counterexamples for other forms of Conjunction and Disjunction with regards to $A^{(n)}$ and $U^{(n)}$. Let i_1 be as in Example 1.

1. Consider $A^{(n)}(\bigcup_{\lambda \in \Lambda} E_\lambda) = \bigcup_{\lambda \in \Lambda} A^{(n)}(E_\lambda)$. We have $A^{(n)}_{i_1}(\{\omega_1\}) \cup A^{(n)}_{i_1}(\{\omega_2, \omega_3\}) = \{\omega_1, \omega_2\} \subsetneq \Omega = A^{(n)}_{i_1}(\Omega)$. We also have $A^{(n)}_{i_1}(\{\omega_2\}) \cup A^{(n)}_{i_1}(\{\omega_3\}) = \Omega \supsetneq \{\omega_1, \omega_2\} = A^{(n)}_{i_1}(\{\omega_2, \omega_3\})$.

2. Consider $U^{(n)}(\bigcup_{\lambda \in \Lambda} E_\lambda) = \bigcup_{\lambda \in \Lambda} U^{(n)}(E_\lambda)$. First, consider the "$\subseteq$" part. We have $U^{(n)}_{i_1}(\{\omega_1\}) \cup U^{(n)}_{i_1}(\{\omega_2\}) \cup U^{(n)}_{i_1}(\{\omega_3\}) = \{\omega_3\} \supsetneq \emptyset = U^{(n)}_{i_1}(\Omega)$.

   Second, consider the "$\supseteq$" part. We first assume $n = 2$. Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and $\mathcal{J} = \{\emptyset, \{\omega_1, \omega_2\}, \Omega\}$. We have: $U^{(2)}(\{\omega_1, \omega_2\}) = (\neg K)(\{\omega_1, \omega_2\}) \cap (\neg K)^2(\{\omega_1, \omega_2\}) = \{\omega_3\}$; $U^{(2)}(\{\omega_1\}) = (\neg K)(\{\omega_1\}) \cap (\neg K)^2(\{\omega_1\}) = \emptyset$; and $U^{(2)}(\{\omega_2\}) = (\neg K)(\{\omega_2\}) \cap (\neg K)^2(\{\omega_2\}) = \emptyset$. Hence, $U^{(2)}(\{\omega_1, \omega_2\}) \supsetneq U^{(2)}(\{\omega_1\}) \cup U^{(2)}(\{\omega_2\})$. Next, we assume $n = \infty$. Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ and $\mathcal{J} = \{\emptyset, \{\omega_1\}, \{\omega_2, \omega_3\}, \{\omega_1, \omega_2, \omega_3\}, \Omega\}$. See Table 2 below. Then, it can be seen that $U^{(\infty)}(\{\omega_2, \omega_3, \omega_4\}) = \{\omega_4\} \not\subseteq \emptyset = U^{(\infty)}(\{\omega_2\}) \cup U^{(\infty)}(\{\omega_3\}) \cup U^{(\infty)}(\{\omega_4\})$.

3. Consider $U^{(n)}(\bigcap_{\lambda \in \Lambda} E_\lambda) = \bigcap_{\lambda \in \Lambda} U^{(n)}(E_\lambda)$. We have $U^{(n)}_{i_1}(\{\omega_1, \omega_3\} \cap \{\omega_2, \omega_3\}) = \emptyset \neq \{\omega_3\} = U^{(n)}_{i_1}(\{\omega_1, \omega_3\}) \cap U^{(n)}_{i_1}(\{\omega_2, \omega_3\})$. Also, $U^{(n)}_{i_1}(\Omega \cap \{\omega_1\}) = \{\omega_3\} \supsetneq \emptyset = U^{(n)}_{i_1}(\Omega) \cap U^{(n)}_{i_1}(\{\omega_1\})$.

Proof of Proposition 4. First, we show the equivalence between (a) and (b). If (a) holds, then $K(\Omega) \setminus E = K(\Omega) \setminus K(E) \in \mathcal{J}$ for any $E \in \mathcal{J}$, i.e., (b) holds. Conversely, if (b) holds then $K(E) \in \mathcal{J}$ implies (a).

Second, we show the equivalence between (a) and $U^{(2)}(\cdot) = (\neg K)(\Omega)$. Suppose (a). Since $K(\neg K)(E) = K(\Omega \cap (KE)^c) \supseteq K(K(\Omega) \setminus K(E))$, we have $(\neg K)^2(E) \subseteq (\neg K)(K(\Omega) \setminus K(E)) = (K(\Omega) \setminus K(E))^c$. Then, we get

$$\begin{aligned} U^{(2)}(E) = (\neg K)(E) \cap (\neg K)^2(E) &\subseteq ((\neg K)(\Omega) \cup (K(\Omega) \setminus K(E))) \cap (K(\Omega) \setminus K(E))^c \\ &= (\neg K)(\Omega). \end{aligned}$$

E	K	$(\neg K)$	$(\neg K)^2$	$(\neg K)^3$	∂	$U^{(2)}$	$U^{(n)}$
$\emptyset$	$\emptyset$	Ω	$\emptyset$	Ω	$\emptyset$	$\emptyset$	$\emptyset$
$\{\omega_1\}$	$\{\omega_1\}$	$\{\omega_2,\omega_3,\omega_4\}$	$\{\omega_1,\omega_4\}$	$\{\omega_2,\omega_3,\omega_4\}$	$\{\omega_4\}$	$\{\omega_4\}$	$\{\omega_4\}$
$\{\omega_2\}$	$\emptyset$	Ω	$\emptyset$	Ω	$\{\omega_2,\omega_3,\omega_4\}$	$\emptyset$	$\emptyset$
$\{\omega_3\}$	$\emptyset$	Ω	$\emptyset$	Ω	$\{\omega_2,\omega_3,\omega_4\}$	$\emptyset$	$\emptyset$
$\{\omega_4\}$	$\emptyset$	Ω	$\emptyset$	Ω	$\{\omega_4\}$	$\emptyset$	$\emptyset$
$\{\omega_1,\omega_2\}$	$\{\omega_1\}$	$\{\omega_2,\omega_3,\omega_4\}$	$\{\omega_1,\omega_4\}$	$\{\omega_2,\omega_3,\omega_4\}$	$\{\omega_2,\omega_3,\omega_4\}$	$\{\omega_4\}$	$\{\omega_4\}$
$\{\omega_1,\omega_3\}$	$\{\omega_1\}$	$\{\omega_2,\omega_3,\omega_4\}$	$\{\omega_1,\omega_4\}$	$\{\omega_2,\omega_3,\omega_4\}$	$\{\omega_2,\omega_3,\omega_4\}$	$\{\omega_4\}$	$\{\omega_4\}$
$\{\omega_1,\omega_4\}$	$\{\omega_1\}$	$\{\omega_2,\omega_3,\omega_4\}$	$\{\omega_1,\omega_4\}$	$\{\omega_2,\omega_3,\omega_4\}$	$\{\omega_4\}$	$\{\omega_4\}$	$\{\omega_4\}$
$\{\omega_2,\omega_3\}$	$\{\omega_2,\omega_3\}$	$\{\omega_1,\omega_4\}$	$\{\omega_2,\omega_3,\omega_4\}$	$\{\omega_1,\omega_4\}$	$\{\omega_4\}$	$\{\omega_4\}$	$\{\omega_4\}$
$\{\omega_2,\omega_4\}$	$\emptyset$	Ω	$\emptyset$	Ω	$\{\omega_2,\omega_3,\omega_4\}$	$\emptyset$	$\emptyset$
$\{\omega_3,\omega_4\}$	$\emptyset$	Ω	$\emptyset$	Ω	$\{\omega_2,\omega_3,\omega_4\}$	$\emptyset$	$\emptyset$
$\{\omega_1,\omega_2,\omega_3\}$	$\{\omega_1,\omega_2,\omega_3\}$	$\{\omega_4\}$	Ω	$\emptyset$	$\{\omega_4\}$	$\{\omega_4\}$	$\emptyset$
$\{\omega_1,\omega_2,\omega_4\}$	$\{\omega_1\}$	$\{\omega_2,\omega_3,\omega_4\}$	$\{\omega_1,\omega_4\}$	$\{\omega_2,\omega_3,\omega_4\}$	$\{\omega_2,\omega_3,\omega_4\}$	$\{\omega_4\}$	$\{\omega_4\}$
$\{\omega_1,\omega_3,\omega_4\}$	$\{\omega_1\}$	$\{\omega_2,\omega_3,\omega_4\}$	$\{\omega_1,\omega_4\}$	$\{\omega_2,\omega_3,\omega_4\}$	$\{\omega_2,\omega_3,\omega_4\}$	$\{\omega_4\}$	$\{\omega_4\}$
$\{\omega_2,\omega_3,\omega_4\}$	$\{\omega_2,\omega_3\}$	$\{\omega_1,\omega_4\}$	$\{\omega_2,\omega_3,\omega_4\}$	$\{\omega_1,\omega_4\}$	$\{\omega_4\}$	$\{\omega_4\}$	$\{\omega_4\}$
Ω	Ω	$\emptyset$	Ω	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$

Table 2: Violation of $U^{(n)}(\bigcup_{\lambda\in\Lambda} E_\lambda) \subseteq \bigcup_{\lambda\in\Lambda} U^{(n)}(E_\lambda)$

Since $(\neg K)(\Omega) \subseteq U^{(2)}(E)$, we obtain $U^{(2)}(E) = (\neg K)(\Omega)$. Conversely, if $U^{(2)}(E) = (\neg K)(\Omega)$, then we have $K(\Omega) = K(E) \cup K(\neg K)(E)$. Then, (a) follows because

$$K(\Omega) \cap (KE)^c = (K(E) \cup K(\neg K)(E)) \cap (\neg K)(E) = K(\neg K)(E) \in \mathcal{J}.$$

Third, if $U^{(2)}(E) = (\neg K)(\Omega)$, then $(\neg K)(\Omega) = (\neg K)(E) \cap (\neg K)^2(E)$. Since $(\neg K)(\Omega) \subseteq (\neg K)^3(E)$, we obtain (c). Clearly, (c) implies (d). Now, (d) implies that $U^{(2)}A^{(2)} = U^{(\infty)}A^{(2)}$. It follows from Proposition 3 that $U^{(2)}A^{(2)} = (\neg A^{(2)})A^{(2)} = (\neg A^{(2)}) = U^{(2)}$. On the other hand, it also follows from Proposition 3 that $U^{(\infty)}A^{(2)}(E) = U^{(2)}(\neg K)A^{(2)}(E) = U^{(2)}U^{(2)}(E) = (\neg K)(\Omega)$. Thus, we obtain $U^{(2)}(E) = (\neg K)(\Omega)$.

Fourth, if $U^{(2)}(\cdot) = (\neg K)(\Omega)$, then Symmetry (i.e., (e)) is trivially satisfied. Conversely, if Symmetry holds, then $U^{(2)}(E) = U^{(2)}(E) \cap U^{(2)}(E^c) \subseteq U^{(\infty)}(E) \subseteq U^{(2)}(E)$, where the first set-inclusion follows from Proposition A.1. We get (d).

Fifth, if $U^{(n)}(\cdot) = (\neg K)(\Omega)$, then AU Introspection (i.e., (f)) is clearly satisfied. If AU Introspection (i.e., (f)) holds, then together with Reverse AU Introspection and Property 1 (a part of Weak Necessitation) in Proposition 3, we obtain $U^{(n)}(\cdot) = U^{(n)}U^{(n)}(\cdot) = (\neg K)(\Omega)$.

Alternatively, we show that Symmetry (i.e., (e)) and AU Introspection (i.e., (f)) are equivalent. If (e) holds, then we have $U^{(2)}(E) = (\neg A^{(2)})(E) = (\neg A^{(2)})A^{(2)}(E) = U^{(2)}A^{(2)}(E) = U^{(2)}U^{(2)}(E)$, where (e) is used for the last equality. This implies (f), which, in turn, implies $U^{(2)}(\cdot) = (\neg K)(\Omega)$. Conversely, (f) and Proposition 3 imply that $U^{(2)}(E) = U^{(2)}U^{(2)}(E) = (\neg K)(\Omega)$. Then, (e) trivially holds.

Sixth, if $U^{(n)}(\cdot) = (\neg K)(\Omega)$, then $U^{(n)}$ (resp. $A^{(n)}$) is obviously monotonic. Conversely, since $U^{(n)}(\Omega) = (\neg K)(\Omega) = U^{(n)}(\emptyset)$, Monotonicity of $U^{(n)}$ implies that $U^{(n)}(\cdot) = (\neg K)(\Omega)$. Likewise, since $A^{(n)}(\Omega) = A^{(n)}(\emptyset) = K(\Omega)$, Monotonicity of $A^{(n)}$ implies that $A^{(n)}(\cdot) = K(\Omega)$, i.e., $U^{(n)}(\cdot) = (\neg K)(\Omega)$.

Ninth, $U^{(\infty)}(\cdot) = (\neg K)(\Omega)$ clearly implies AA Self-Reflection (i.e., (i)). Conversely, observe that

$$\begin{aligned} A^{(\infty)}A^{(\infty)}(E) &= K(\neg K)A^{(\infty)}(E) \cup K(\neg K)^2 A^{(\infty)}(E) \\ &= KU^{(\infty)}(E) \cup K(\neg K)U^{(\infty)}(E) = K(\Omega). \end{aligned}$$

Then, (i) implies that $A^{(\infty)}(E) = A^{(\infty)}A^{(\infty)}(E) = K(\Omega)$, i.e., $U^{(\infty)}(E) = (\neg K)(\Omega)$. □

A.4 Section 4.1

Proof of Proposition 5. 1. Suppose that $\mathcal{S}^{(n)}$ satisfies Finite Conjunction. First, it follows from Lemma A.2 and Positive Introspection that $K(E \cup F) \subseteq K(KE \cup LF)$ for any $E, F \in \mathcal{D}$. Second, since $\mathcal{D}$ is assumed to be finite, we re-label it by $\mathcal{D} = \{E_1, \dots, E_m\}$, so that $\overline{U}^{(n)} = \bigcup_{r=1}^{m} U^{(n)}(E_r)$. Putting $E = U^{(n)}(E_m)$ and $F = \bigcup_{r=1}^{m-1} U^{(n)}(E_r)$ in the previous statement yields

$$\begin{aligned} K(\bigcup_{r=1}^{m} U^{(n)}(E_r)) &\subseteq K(KU^{(n)}(E_m) \cup L(\bigcup_{r=1}^{m-1} U^{(n)}(E_r))) \\ &= K(KU^{(n)}(E_m) \cup \bigcup_{r=1}^{m-1} U^{(n)}(E_r)) = K(\bigcup_{r=1}^{m-1} U^{(n)}(E_r)), \end{aligned}$$

where observe that $\bigcup_{r=1}^{m-1} U^{(n)}(E_r) = L(\bigcup_{r=1}^{m-1} U^{(n)}(E_r))$ follows from Finite Conjunction. It follows, by induction, that $K(\overline{U}^{(n)}) \subseteq K(U^{(n)}(E_1)) = \emptyset$. Thus, we have $K(\neg K)(\overline{U}^{(n)}) = K(\Omega)$. By Monotonicity of K, we have $A^{(n)}(\overline{U}^{(n)}) = K(\Omega)$.

Next, we provide a counterexample when $\mathcal{D}$ is not finite. Let $\Omega = \mathbb{R}$ and $\mathcal{D} = \mathcal{P}(\Omega)$. Suppose that an agent's self-evident collection is given by the usual Euclidean topology $\mathcal{E}_{\mathbb{R}}$. Since it is closed under arbitrary union and finite intersection, her knowledge satisfies Finite Conjunction. Her knowledge also satisfies Necessitation because $\Omega \in \mathcal{E}_{\mathbb{R}}$. For any $\omega \in \mathbb{R}$, let $E_\omega = (\omega, +\infty)$. Then, we have $U^{(2)}(E_\omega) = \partial KE_\omega = \{\omega\}$ and $U^{(\infty)}(E_\omega) = \partial LKE_\omega = \{\omega\}$. Thus, we have $K(\bigcup_{\omega \in \Omega} U^{(n)}(E_\omega)) = K(\Omega) = \Omega$. This implies that $K(\overline{U}^{(n)}) = \Omega$.

2. If $K(\overline{U}^{(n)}) = \emptyset$, then we have $K(\Omega) \supseteq A^{(n)}(\overline{U}^{(n)}) \supseteq K(\neg K)(\overline{U}^{(n)}) = K(\Omega)$. Thus, $A^{(n)}(\overline{U}^{(n)}) = K(\Omega) \neq \emptyset$. If $K(\overline{U}^{(n)}) \neq \emptyset$, then $A^{(n)}(\overline{U}^{(n)}) \supseteq K(\overline{U}^{(n)}) \neq \emptyset$. Next, it is clear by construction that $U^{(n)}(\overline{U}^{(n)}) \subseteq \bigcup_{E \in \mathcal{D}} U^{(n)}(E) = \overline{U}^{(n)}$.

3. If $\overline{U}^{(n)} = \emptyset$ then Necessitation implies that we have $\overline{U}^{(n)} = \emptyset = L(\emptyset) = L(\overline{U}^{(n)})$. If $\overline{U}^{(n)} \neq \emptyset$, then (Non-empty) Conjunction (of K) and A-Introspection (see Proposition 3) imply that $L(\overline{U}^{(n)}) = L(\bigcup_{E \in \mathcal{D}} U^{(n)}(E)) = \bigcup_{E \in \mathcal{D}} LU^{(n)}(E) = \bigcup_{E \in \mathcal{D}} U^{(n)}(E) = \overline{U}^{(n)}$.

□

A.5 Section 4.3

Proof of Proposition 6.

1. Consider (1a) and (1b). Since $\partial_j(\cdot) \subseteq \partial_i(\cdot)$, substitute $K_i(E)$ and $L_iK_i(E)$. Consider (1c). First, we have $\partial_j U_i^{(n)}(E) = (\neg K_j)U_i^{(n)}(E) \cap (\neg K_j)A_i^{(n)}(E) \subseteq (\neg K_j)A_i^{(n)}(E) = U_i^{(n)}(E)$. Second, we have $\partial_i U_j^{(n)}(E) = (\neg K_i)U_j^{(n)}(E) \cap (\neg K_i)A_j^{(n)}(E) = (\neg K_i)A_j^{(n)}(E) \supseteq (\neg K_j)A_j(E) = U_j^{(n)}(E)$.

2. Substituting $K_j(E)$ and $K_j(\neg K_j)(E)$ into $\partial_j(\cdot) \subseteq \partial_i(\cdot)$ yields the desired results.

3. First, we have $A_i^{(n)}(E) = K_iA_i^{(n)}(E) \subseteq K_jA_i^{(n)}(E) \subseteq A_i^{(n)}(E)$. Second, noting that $K_jK_i(E) = K_i(E)$, we have $A_i^{(n)}(E) = A_i^{(n)}K_i(E) = A_i^{(n)}K_jK_i(E)$.

4. We have $A_i^{(n)}(E) \subseteq K_iA_i^{(n)}(E) \subseteq K_jA_i^{(n)}(E) \subseteq A_j^{(n)}A_i^{(n)}(E)$. Next, by (1c) and Proposition 1, $U_j^{(n)}U_i^{(n)}(E) \subseteq \partial_j U_i^{(n)}(E) \subseteq U_i^{(n)}(E)$.

□

Proof of Proposition 7. Since $U_i^{(\infty)} = U_i^{(3)}$, we can let $n \in \{2, 3\}$.

1. If $\omega' \in U_i^{(n)}(E) = \bigcap_{r=1}^{n}(\neg K_i)^r$, then $K_i(\neg K_i)^{r-1}(E) \notin \mathrm{IK}_i(\omega')$ for all $r \leq n$. Since $\mathrm{IK}_i(\omega) \subseteq \mathrm{IK}_i(\omega')$, we have $K_i(\neg K_i)^{r-1}(E) \notin \mathrm{IK}_i(\omega)$ for all $r \leq n$, that is, $\omega \in U_i^{(n)}(E) = \bigcap_{r=1}^{n}(\neg K_i)^r$.

2. Suppose that $\omega \in U_j^{(n)}(E)$. If $\omega \in K_iU_j^{(n)}(E)$, then we obtain $\omega \in K_iU_j^{(n)}(E) \subseteq K_jU_j^{(n)}(E) = \emptyset$, a contradiction. Hence, $\omega \notin K_iU_j^{(n)}(E)$. Next, suppose that $\omega \in A_i^{(n)}(E)$. Then, we have $\omega \in A_i^{(n)}(E) = K_iA_i^{(n)}(E) \subseteq K_jA_i^{(n)}(E)$.

□

B Extensions to Subjective State Spaces

Let Ω be a state space. Let $\mathcal{D}_i$ and be a complete algebra on *a subset* Ω_i of Ω. Note that $\emptyset = \bigcup \emptyset \in \mathcal{D}$ and $\Omega_i = \bigcap \emptyset \in \mathcal{D}_i$. Now, we define agent i's knowledge and unawareness operators on $\mathcal{D}_i$. An information structure is a tuple $\mathcal{S} = \langle \Omega, (\mathcal{D}_i)_{i \in I}, (K_i, U_i)_{i \in I} \rangle$, where $K_i : \mathcal{D}_i \to \mathcal{D}_i$ and $U_i : \mathcal{D}_i \to \mathcal{D}_i$.

Let $\mathcal{D}$ be a complete algebra on Ω such that $\mathcal{D}_i \subseteq \mathcal{D}$ for all $i \in I$. We re-define agents' knowledge and unawareness operators on this common domain $\mathcal{D}$ from $\mathcal{D}_i$. Specifically, we take the following two approaches. One approach is the logical approach in which we extend agents' knowledge and unawareness operators by rendering agents the logical inference ability to judge whether they know events outside of their domains. The other approach is the naive approach, where each agent is assumed not to know any event which is outside of her original domain $\mathcal{D}_i$.

We make the following three technical remarks. The first is with respect to the difference between Ω and Ω_i. Although we define Ω as the entire state space, objects of knowledge and unawareness, from the standpoint of each agent i, are defined only on Ω_i. Thus, each agent i would consider any state $\omega \in \Omega \setminus \Omega_i$ "impossible" whenever $\Omega \setminus \Omega_i \neq \emptyset$. In other words, each Ω_i is agent i's subjective state space.

Second, from each agent's perspective, the complementation is always taken with respect to Ω_i as a universal set. Hence, if we take the complement of a set with respect to Ω_i, we often append the subscript i by denoting $\neg_i E := \{\omega \in \Omega_i \mid \omega \notin E\}$ for $E \in \mathcal{P}(\Omega_i)$, in contrast to $\neg E := \{\omega \in \Omega \mid \omega \notin E\}$ for $E \in \mathcal{P}(\Omega)$. We also denote $E^{c_i} := \neg_i E$ and $E^c := \neg E$.

Third, the distinction between Ω and Ω_i allows us to examine the implications of Conjunction and Negative Introspection independently of Necessitation. Namely, either Conjunction or Negative Introspection implies $K_i(\Omega_i) = \Omega_i$, which does not always imply Necessitation in the form of $K_i(\Omega) = \Omega$ when $\Omega_i \subsetneq \Omega$. Note that Negative Introspection implies that $(\neg_i K_i)(E) \subseteq K_i(\neg_i K_i)(E)$ (or $\Omega_i \setminus K_i E \subseteq K_i(\Omega_i \setminus K_i E)$) in Ω_i. Conjunction is also taken with respect to Ω_i. Thus, we further introduce the following definition: K_i satisfies Objective Necessitation if $K_i(\Omega) = \Omega$. Note that if Objective Necessitation is assumed, then it is implicitly assumed that $\Omega = \Omega_i \in \mathcal{D}_i$.

B.1 Retaining Logical Ability

Recall that each agent's knowledge is represented by her self-evident collection $\mathcal{J}_i (\subseteq \mathcal{D}_i)$ in the sense that $K_i(E) = \{\omega \in \Omega_i \mid \text{there is } F \in \mathcal{J}_i \text{ such that } \omega \in F \subseteq E\}$. That is, agent i knows an event $E \in \mathcal{D}_i$ at a state $\omega \in \Omega_i$ iff she infers E from a self-evident event $F \in \mathcal{J}_i (\subseteq \mathcal{D}_i)$ which is true at ω. Hence, for any event $E' \in \mathcal{D}$, we define that she knows E' at state $\omega' \in \Omega$ if she infers E' from a self-evident event $F \in \mathcal{J}_i (\subseteq \mathcal{D})$ which is true at ω'.

Formally, we define $\tilde{K}_i : \mathcal{D} \to \mathcal{D}$ by $K_i(E) = \{\omega \in \Omega \mid \text{there is } F \in \mathcal{J}_i (\subseteq \mathcal{D}) \text{ such that } \omega \in F \subseteq E\}$. Then, we have $\tilde{K}_i(E) = K_i(E \cap \Omega_i)$ for any $E \in \mathcal{D}$ such that $\Omega_i \cap E \in \mathcal{D}_i$. For a given knowledge operator K_i, the new operator inherits its properties except for Necessitation (Objective Necessitation has to be explicitly assumed, if necessary).

Our previous arguments can then be applied, and hence this logical extension justifies the common domain assumption within our framework. Note that for single agent's analysis (in Sections 3 and 4.2), the information structure $\mathcal{S} = \langle(\Omega, \mathcal{D}), (K, U^{(n)})\rangle$ should rather be read as $\mathcal{S} = \langle \Omega_i, \mathcal{D}_i, (K_i, U_i^{(n)})\rangle$.

As a specific example, consider agents i_3 and i_4 in Example 1. Suppose that $\mathcal{D} = \mathcal{D}_{i_3} = \mathcal{P}(\Omega)$ and $\mathcal{D}_{i_4} = \{\omega_1\}$. Let $K'_{i_3} : \mathcal{D}_{i_3} \to \mathcal{D}_{i_3}$ be defined by $K'_{i_3} = K_{i_3}$. Let $K'_{i_4} : \mathcal{D}_{i_4} \to \mathcal{D}_{i_4}$ be defined by $K'_{i_4} = (\emptyset) = \emptyset$ and $K'_{i_4}(\{\omega_1\}) = \{\omega_1\}$. Then, it is easily seen that $\tilde{K'}_{i_4} : \mathcal{D} \to \mathcal{D}$ is given by K_{i_4} (the original knowledge operator specified in Example 1). If we assume Objective Necessitation then $\tilde{K'}_{i_4}$ is given by K_{i_3}.

B.2 Introducing Non-monotonicity

The second approach is to assume that each agent simply does not know $E \in \mathcal{D} \setminus \mathcal{D}_i$. This breaks down agents' logical ability in terms of Monotonicity. Formally, we define $\overline{K}_i : \mathcal{D} \to \mathcal{D}$ by

$$\overline{K}_i(E) := \begin{cases} K_i(E) & \text{if } E \in \mathcal{D}_i \\ \emptyset & \text{if } E \in \mathcal{D} \setminus \mathcal{D}_i \end{cases}.$$

Next, we define $\overline{U}_i^{(n)} : \mathcal{D} \to \mathcal{D}$ by $\overline{U}_i^{(n)}(E) := \bigcap_{r=1}^{n} (\neg \overline{K}_i)^r(E)$. Thus, we have:

$$\overline{U}_i^{(n)}(E) = \begin{cases} U_i^{(n)}(E) & \text{if } E \in \mathcal{D}_i \\ \Omega & \text{if } E \in \mathcal{D} \setminus \mathcal{D}_i \end{cases}.$$

From Equals to Despots: The Dynamics of Repeated Decision Making in Partnerships with Private Information*

Vinicius Carrasco† William Fuchs‡ Satoshi Fukuda§

Abstract

This paper considers an optimal renegotiation-proof dynamic Bayesian mechanism in which a group of privately informed agents repeatedly have to take a joint action without resorting to side-payments. We provide a general framework which accommodates as special cases committee decision and collective insurance problems. Thus, we formally connect these separate strands of literature. First, we show that first-best values can be arbitrarily approximated (but not achieved) when the players are sufficiently patient. Second, the optimal mechanism maximizes the weighted sum of agents' utilities while the weights attached on agents vary over time in order to keep intertemporal incentives. We show that the provision of such intertemporal incentives necessarily spreads utility weights over the long run.

JEL Classification: D82, D86, D70
Keywords: Repeated Collective Decision Making, Dynamic Bayesian Mechanism Design, Renegotiation-proofness

*We have benefited from conversations with Suren Basov, Simon Board, Martin Cripps, Sergio Firpo, Drew Fudenberg, Yingni Guo, Johannes Hörner, Andreas Kleiner, Narayana Kocherlakota, Ferdinando Monte, Roger Myerson, Tymofiy Mylovanov, Ezra Oberfield, Aniko Öry, Alessandro Pavan, Phil Reny, Leo Rezende, Yuliy Sannikov, Simone Galperti, Andy Skrzypacz, Nancy Stokey, and Bálazs Szentes. We would like to thank two helpful anonymous referees and an Associate Editor.

†Department of Economics, PUC-Rio.
‡Haas School of Business, UC Berkeley.
§Department of Economics, UC Berkeley.

Contents

1 Introduction

There are many situations in which, repeatedly, a group of agents have to take a common action, cannot resort to side-payments, and each period are privately informed about their preferences. Examples include many supranational organizations such as a monetary union or a common market. In the former, monetary policy must be jointly taken and, in the latter, a common tariff with the outside world must be adopted each period. At the national level, political coalitions must jointly and repeatedly decide on policy issues. Within firms, managers of different divisions often have to make a joint decision repeatedly.

We show agents can improve the efficiency of their repeated collective decisions by using rules that result in time varying decision rights. This captures an informal notion of "political capital," i.e., a player who pushes hard on a given decision spends his political capital and as a result can exert less influence on future decisions.

Formally, we study the long-run properties of an optimal renegotiation-proof dynamic Bayesian mechanism without side-payments. We first show that efficiency can be arbitrarily approximated, but never attained, when agents are sufficiently patient. In a repeated setting, the promise of (equilibrium) continuation values play a similar role to the one side-payments play in static mechanism design problems. Indeed, in order to prove the approximate efficiency result, we construct continuation values that replicate the expected payments proposed by Arrow [3] and d'Aspremont and Gérard-Varet [9], that guarantee efficiency in standard static Bayesian mechanisms *with* transfers.

The difference between side-payments and continuation values is that the latter can only imperfectly transfer utility across agents. In particular, to transfer continuation utility from one agent to another in any period, joint actions for any future periods must be altered. When agents are sufficiently patient (i.e., their (common) discount factor is close to one), their weighted current payoffs become insignificant relative to the continuation values. Hence, in order to guarantee truth-telling in the current period, continuation values have to vary only minimally. Since the (incentive compatible) utility possibilities frontier is locally linear, the associated efficiency losses from the variation in continuation values are arbitrarily small in the limit as the discount factor tends to one. The attainment of full efficiency, however, would call for no variation in continuation values, and this is at odds with the provision of incentives needed for an efficient action to be taken. Hence, full efficiency is not attainable.

Although the limiting efficiency result is of interest, our main focus is on understanding the long-run properties of the optimal mechanism. As a benchmark, it is useful to keep in mind what the first-best mechanism would entail in the environment in which players' private information is their favorite actions. The first-best action would call for a constant weighted average of players' favorite actions. The problem with this decision rule is that agents away from the center have an incentive to exaggerate their positions. If they expect the other types to be to the left (right) of them they would have an incentive to claim to be further right (left) and in that way bring the chosen decision closer to their preferred point. As a result, absent any additional tools to provide incentives, the only way one has to prevent agents from lying is by making the decision rule less sensitive to their reports.[1]

When decisions are made repeatedly, agents care about the future so that they can trade decision power in the current action for decision power in the future. More extreme types are given more weight in the current decision but they pay for it by having less influence on future decisions. In this way, the optimal mechanism can be more sensitive to extreme announcements of preferences.

As known from static mechanism design theory, once incentive compatibility constraints are taken into account, agents' utilities have to be adjusted to incorporate the rents derived from their private information. The adjusted utility is referred to as *virtual* utility, and the optimal mechanism design problem amounts to maximizing the (weighted) sum of agents' virtual utilities for each type (Myerson [21, 22, 24]). In our dynamic setting, virtual utilities also play a key role. We show that the dynamics of efficient collective decision making are fully determined by: (i) a decision rule that, at each period, maximizes the (weighted) sum of agents' virtual utilities and (ii) a process that governs the evolution of the weights given to agents' virtual utilities. While agents' Pareto weights in the first-best problem are constant over time, we show that their weights vary over time in the second-best problem.

The dynamics of the collective decision making process lead to our most interesting result. Continuation values vary from period to period reflecting agents' weights.[2] Indeed, continuation values increase (meaning higher future decision power) for an agent who reports a less extreme announcement, and decrease (meaning lower future decision power) for a player who reports an extreme announcement.

Related Literature

This paper relates to three strands of literature. First, our paper builds on the literature of collective decision making without side-payments. Second, our paper relates to the dynamic insurance literature in that our model can capture it as a special case. Third, on the technical side, our paper is in the literature looking at the use of continuation values as close substitutes for side-payments. We discuss these in order below.

There has been a very extensive and old literature on collective decision making. Within this literature we are most closely related to the models in which agents' values are independent of each other. Thus each agent's preferred action does not directly depend on the private information of the other players. Only recently has this literature started to consider the case where rather than only one decision there are potentially multiple ones to be taken. The question then is how the availability of many decisions is to be taken into account to improve efficiency.

[1] See Carrasco and Fuchs [7] for an analysis of the static problem in which two ex-ante symmetric agents have single-peaked preferences.

[2] Our approach to the analysis of the optimal mechanism in the repeated game relies on the factorization results of Abreu, Pearce, and Stacchetti [1] that an agent's payoff can be split into a current value and a continuation value.

The first paper in this regard is Casella [8] who studies the problem of a committee that each period must take one of two possible decisions and agents have a continuum of preference intensities for these two actions. She proposes the "storable vote" mechanism which endows players with one vote per period. Players can then either cast their vote or store it for the future. This effectively allows them to have more influence on the future decisions since they are taken by majority vote. In a two period setting, the storable vote mechanism can indeed enhance efficiency.[3] Although appealing from a practical standpoint, the storable vote mechanism is not optimal.

Hortala-Vallve [17] allows a continuum of preference intensities over issues in a static environment in which agents know their preferences over all issues when making their decision. He looks at the "qualitative voting" mechanism, which endows players with a large number of votes and allows them to cast their votes over the issues according to their preferences. He shows this mechanism is optimal in the case of two issues and two or three players.

Drexl and Kleiner [10] study committee decision making in which two privately informed agents repeatedly make a binary decision without monetary transfers. They characterize conditions on prior distributions and classes of mechanisms among which a voting rule is optimal.

Jackson and Sonnenschein [18] study a dynamic Bayesian mechanism without side-payments, which they call the "linking mechanism." Agents must budget their representations of preferences so that the frequency of reported preferences across issues mirrors the underlying distribution of preferences. They show that agents' incentives are to satisfy their budget by being as truthful as possible and that when the number of problems being linked goes to infinity this mechanism is arbitrary close to being efficient. Although efficient in the limit, their mechanism, unlike ours, is not optimal. Also, although both their and our mechanisms are approximately efficient, the small inefficiencies are caused by very different aspects of the mechanisms. In theirs, the inefficiency arises from the fact that for any finite number of problems the realized distribution does not exactly match the budget which is based on the underlying preference distribution. The inefficiency in our mechanism arises instead from variations in continuation values to provide incentives. None of these previous papers have a dictatorship result as we do.

While the introduction of dynamics to the collective decision making enhances efficiency, the improved efficiency may be supported by inefficient future decision rules and thus an optimal mechanism may be subject to renegotiation. Hence, we look at an optimal renegotiation-proof mechanism. The problem is that there is no clear-cut definition of renegotiation-proofness in infinitely repeated games, in contrast to the finite horizon games where renegotiation-proofness is well defined by backward induction. Among various concepts of renegotiation-proofness in infinite horizon games, we focus on *(internally) renegotiation-proof* equilibrium, following Ray [28]. The idea that a mechanism is inter-

[3] Skrzypacz and Hopenhayn [29, Section 3.1] use a similar "chips mechanism" to sustain collusion in a repeated auction environment. They numerically demonstrate that the chips mechanism converges to an optimal scheme.

nally renegotiation-proof is captured by the (recursive) condition that the mechanism is (weakly) Pareto efficient such that its continuation mechanism is renegotiation-proof. Hence it extends to the backward-induction definition of renegotiation-proofness in finite horizon games.[4]

A special case of our model can capture the problem of two agents with income shocks trying to self-insure, where the common action is regarded as a transfer from one player to another. The problem between one agent and one principal has been looked at before by Thomas and Worrall [30]. Within the macro literature, this problem has been examined with a continuum of agents by Atkeson and Lucas [5], Green [15], and Phelan [27]. Phelan [26] and Wang [31] study a finite number of players in a dynamic insurance setting.[5] In a related environment, Friedman [11] and Zhao [32] study a repeated moral hazard problem in which a finite number of risk-averse players produce a perishable consumption good each period.

The incentive structure of dynamic insurance problems has special properties. When faced with the first-best contract which would provide full insurance, agents would have an incentive to claim to be of the lowest possible type to obtain the highest insurance payment. Instead, in our general setting, when faced by the first-best mechanism, although agents still have an incentive to exaggerate their types, they would not in general want to claim to be the lowest (or highest) possible type. The reason is that the stage game, at the interim stage, is no longer a zero-sum game. There is a possibility that both agents want a similar action to be taken even after they learn their private information. Suppose for example that two players have their preferred actions drawn from the i.i.d. uniform distribution on $[0, 1]$ and have a quadratic loss utility in the distance between their type and the joint action. An agent with realized type 0.51 expects the other agent to have a lower type and thus wants to exaggerate his type in order to bring the joint decision closer to his preferred action.[6] But he does not want to claim to be of type 1 because if the other player reports a high type this could lead to a joint decision close to 1. The fact that the level of conflict or congruence of preferences is not known allows for some cooperation to take place even in a one-shot game. In contrast, there is no scope for insurance in a one-shot version of the dynamic insurance problem.

Moving on to the technical side, our paper is in the literature, pioneered by the seminal work of Fudenberg, Levine, and Maskin [13], that shows that continuation values are close substitutes for side-payments in repeated settings. Within this literature, the closest paper is Athey and Bagwell [4], who analyze an infinitely repeated Bertrand duopoly game, where each firm privately receives discrete cost shocks. They establish that, for a discount factor strictly less than one, monopoly profits can be *exactly* attained by firms making use of

[4]Zhao [33] formulates internal renegotiation-proofness in terms of a value function in a principal-agent problem. Bergin and MacLeod [6] introduce a similar concept, "weak full recursive efficiency," and compare various notions of renegotiation-proofness.

[5]Phelan [26, Section 3] actually assumes that there is a price-taking representative firm which can separately contract with each agent.

[6]The first-best symmetric decision rule is simply the average of both types in this case.

asymmetric continuation values when cost types are discrete. In our setting, on the other hand, the first best is arbitrary approximated but generally never attained. The difference from our setting is that, for some states that occur with positive probability, firms in their model can transfer profits perfectly.

Finally, we provide a technical overview. We analyze the dynamic renegotiation-proof Bayesian collective decision making problem among a group of two agents. Our setting accommodates any strictly concave utility function in joint action, where preference shocks need not be multiplicative nor separable between type and action. The current action depends on an entire vector of players' reported preference shocks. We first formalize the Bellman equation which characterizes the optimal dynamic mechanism with the promised utility being a state variable, by following Abreu, Pearce, and Stacchetti [1]. We then characterize the optimal dynamic mechanism for each promised value. The optimality conditions show that (i) the optimal current action maximizes the (weighted) sum of agents' instantaneous virtual utilities while (ii) the agents' weights vary over time.

The paper is organized as follows. Section 2 introduces the model. Section 3 characterizes the optimal renegotiation-proof dynamic Bayesian mechanism. Section 4 concludes. All proofs are relegated to Appendix A. Appendix B provides the results of numerical simulations of repeated collective decision and insurance models, respectively.

2 The Model

In each period $t \in \mathbb{N}$, a group of two players (denoted by $N := \{1, 2\}$) must take a joint action a from a compact convex action set $\mathcal{A} := [\underline{a}, \overline{a}]$, where its interior $\mathcal{A}^\circ$ is not empty.[7] Each period, each player i privately receives a preference shock (his private type) from a finite set Θ_i. Preference shocks are drawn from a probability density function $f_i(\cdot) > 0$ on Θ_i. Also, they are assumed to be independent across players and i.i.d. over time. We denote the set of type profiles by $\Theta := \Theta_1 \times \Theta_2$ and the joint density f on Θ by $f(\theta) := f_1(\theta_1) \times f_2(\theta_2)$.

Each player's instantaneous utility $u_i : \mathcal{A} \times \Theta_i \to \mathbb{R}$ depends on a common action and his own private type, i.e., we deal with the case of private values, and side-payments are not allowed. We assume that u_i is continuously differentiable and strictly concave in common action $a \in \mathcal{A}$ for any given type. Each i has his unique best action $a_i(\theta_i)(\in \mathcal{A})$ which maximizes his utility u_i for each type θ_i, where $a_i(\theta_i)$ may lie on the boundary of $\mathcal{A}$.[8]

Our model captures the following three standard environments as special cases: (i) a collective decision making problem with single-peaked preferences, (ii) a dynamic insurance partnership model, and (iii) a multiplicative preference shock model. In the first

[7] One can extend the action space to a (compact convex) multi-dimensional space.

[8] This occurs when $u_i(\cdot, \theta_i)$ is monotone. We further assume that $a_1(\theta_1)$ and $a_2(\theta_2)$ are not identical with positive probability (i.e., $a_1(\theta_1) \neq a_2(\theta_2)$ for some (θ_1, θ_2)). This implies that the players cannot simultaneously achieve their maximum payoffs $\mathbb{E}_\theta [u_i(a_i(\theta_i), \theta_i)]$. In other words, the Pareto frontier of the feasible payoff set is not a singleton set.

environment, agents have to take a joint action repeatedly, where they have single-peaked preferences with their types being their most preferred actions, i.e., $a_i(\theta_i) = \theta_i$. A typical example is that each player has a quadratic utility $u_i(a, \theta_i) = -(a - \theta_i)^2$ with $\Theta_1 = \Theta_2 \subseteq \mathcal{A}$.

Second, two agents insure themselves against private endowment shocks. We can reinterpret the common action as the "transfer" from agent 1 to 2 (this is possible because the mechanism itself does not allow side-payments). Unlike Atkeson and Lucas [5], Green [15], and Thomas and Worrall [30], transfers are restricted in the compact set $\mathcal{A} = [\underline{a}, \overline{a}]$ and dependent on an entire vector of agents' preference shocks.[9] Player 1's (resp. 2's) transfer is restricted to at most $\overline{a}$ (resp. $-\underline{a}$) each time. Note that Wang [31] has a more stringent feasibility condition that each player can only transfer his own endowments, that is, $-\theta_2 \leq a(\theta_1, \theta_2) \leq \theta_1$ for each $(\theta_1, \theta_2) \in \Theta$. In our setting, the set $\mathcal{A}$ of feasible transfers from player 1 to 2 can either contain or be contained in a set of feasible transfers allowed in Wang [31].

Agents' utility functions are written as $u_1(a, \theta_1) = \tilde{u}_1(\theta_1 - a)$ and $u_2(a, \theta_2) = \tilde{u}_2(\theta_2 + a)$, where each $\tilde{u}_i$ is a strictly increasing and strictly concave function defined on an interval.[10] By construction, both agents consume the aggregate endowments each time, i.e., the date-by-date resource constraint is satisfied.

The third environment is a multiplicative preference shock model. It can be interpreted in the contexts of both the collective decision and the insurance settings. Each agent's utility is $u_i(a, \theta_i) := \theta_i \tilde{u}_i(a)$, where $\tilde{u}_i$ is strictly concave and $\Theta_i \subseteq \mathbb{R}_{++}$.[11] The multiplicative preference shocks affect the (marginal) utility of a common action. Each agent has his (fixed) favorite action $a_i(\theta_i) = a_i \in \mathcal{A}$ with $a_1 \neq a_2$. In the repeated collective decision making context, multiplicative preference shocks are understood as preference intensities being private information. In the insurance partnership, if each player has a CARA utility, then endowment shocks can be seen as multiplicative preference shocks.

It is worth pointing out that the collective decision making environment has a distinct feature from the insurance partnership in that agents can possibly make a better decision by sharing more information. Namely, although agents have an incentive to exaggerate their private information, they do not know how aligned their preferences are. Hence it is possible even in the static model to have an efficient allocation dependent on the agents' types in an incentive compatible way. On the other hand, in the static insurance model, it is impossible to have an efficient allocation depend on agents' types in an incentive compatible manner.

Following Fudenberg, Levine, and Maskin [13], we restrict attention to Perfect Public Equilibria (PPE). That is, each player conditions only on the "public" histories of the

[9] Posing a restriction on possible transfers in an insurance model makes the problem technically rather challenging. It limits the degree to which each player can be punished in a single period. Also, an optimal transfer may lie on the boundary of feasible transfers.

[10] Our results could be extended to the case where one agent's utility is affine while the other's is strictly concave.

[11] Whenever the utility function is written as $u_i(a, \theta_i) = r_i(\theta_i)\tilde{u}_i(a)$, where $r_i : \Theta_i \to \mathbb{R}_{++}$ and $\Theta_i \subseteq \mathbb{R}_{++}$, we can redefine it as $\theta_i k_i \tilde{u}_i(a)$, where the density function is $\tilde{f}(\theta_i) := (r_i(\theta_i)/(k_i\theta_i))f_i(\theta_i)$ with $k_i := \sum_{\theta_i \in \Theta_i}(r_i(\theta_i) \cdot f_i(\theta_i)/\theta_i)(> 0)$.

game. Letting the initial history be h^0, a public history at time $t \in \mathbb{N}$, h^t, is a sequence of (i) past announcements of the players (they make reports $(\hat{\theta}_i)_{i\in N} \in \Theta$ after they observe their preference shocks), (ii) past realized actions, and (iii) possibly realizations of a public randomization device. Let H^t be the set of all possible public histories h^t, and let $H^0 = \{h^0\}$.

Given the current reports and the past public history of the game, history-dependent allocations are determined according to an enforceable contract, which is a sequence of functions $a = (a^t)_{t\in\mathbb{N}}$, where $a^t : H^{t-1} \times \Theta \to \mathcal{A}$. The contract is enforceable in the sense that the players commit themselves to it a priori (at time "0") before they learn their preference shocks.

A public strategy for player i is a sequence of functions $\hat{\theta}_i = (\hat{\theta}_i^t)_{t\in\mathbb{N}}$, where $\hat{\theta}_i^t : H^{t-1} \times \Theta_i \to \Theta_i$. Each strategy profile $\hat{\theta} = (\hat{\theta}_i)_{i\in N}$ induces a probability distribution over public histories. Letting $\delta \in (0,1)$ be the common discount factor, player i's ex-ante expected discounted average payoff is given by

$$\mathbb{E}\left[(1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}u_i\left(a^t(h^{t-1},\hat{\theta}^t(h^{t-1},\theta^t)),\theta_i^t\right)\right].$$

3 Properties of the Optimal Mechanisms

We analyze the repeated game using the recursive methods developed by Abreu, Pearce, and Stacchetti [1]. Specifically, letting $\mathcal{W} \subseteq \mathbb{R}^2$ be the set of pure strategy PPE payoff profiles, we can decompose the players' payoffs into (i) a profile of current utilities $(u_i(a(\theta),\theta_i))_{i\in N}$ and (ii) a profile of continuation values $(w_i(\theta))_{i\in N} \in \mathcal{W}$. In other words, any PPE payoff profile is written as $(\mathbb{E}_\theta[(1-\delta)u_i(a(\theta),\theta_i) + \delta w_i(\theta)])_{i\in N}$ and it can be summarized by a current decision rule $a(\hat{\theta})$ and equilibrium continuation values $w(\hat{\theta}) \in \mathcal{W}$ as a function of a profile of announcements $\hat{\theta}$. A *mechanism* (a, w) is a pair of current action function $a : \Theta \to \mathcal{A}$ and continuation value function $w : \Theta \to \mathbb{R}^2$.

The set of PPE payoffs has the following recursive structure. For any candidate equilibrium payoff value set $W \subseteq \mathbb{R}^2$, a mechanism (a, w) is *admissible* with respect to W if it satisfies the following (Bayesian) incentive constraints and self-enforceability condition:

$$\begin{aligned}&\mathbb{E}_{\theta_{-i}}\left[(1-\delta)u_i\left(a(\theta),\theta_i\right)+\delta w_i(\theta)\right]\\&\qquad \geq \mathbb{E}_{\theta_{-i}}[(1-\delta)u_i(a(\hat{\theta}_i,\theta_{-i}),\theta_i)+\delta w_i(\hat{\theta}_i,\theta_{-i})] \text{ for all } i\in N \text{ and } \theta_i,\hat{\theta}_i\in\Theta_i; \text{ and} \quad (1)\\&w(\theta)\in W \text{ for all } \theta\in\Theta. \quad (2)\end{aligned}$$

Expression (1) requires each player to prefer to tell the truth given that the other player is expected to tell the truth. Expression (2) is the self-enforceability condition that continuation values must be in the candidate equilibrium value set. Defining the mapping B from the collection of candidate equilibrium value sets (namely, the power set $\mathcal{P}(\mathbb{R}^2)$ of $\mathbb{R}^2$) into itself by the set of payoffs attainable by admissible mechanisms, the set of PPE payoffs is given by the largest bounded fixed point of the operator B, where we assume that the players can utilize some public randomization device.[12]

Formally, $B : \mathcal{P}(\mathbb{R}^2) \to \mathcal{P}(\mathbb{R}^2)$ is defined as follows: for any subset W of $\mathbb{R}^2$,

$$B(W) := \mathrm{co}\left(\left\{(\mathbb{E}_\theta\left[(1-\delta)u_i(a(\theta),\theta_i) + \delta w_i(\theta)\right])_{i\in N}\,\middle|\,(a,w) \text{ is admissible w.r.t. } W\right\}\right).$$

With slight modifications to Abreu, Pearce, and Stacchetti [1], it can be shown that the monotone mapping B preserves compactness and that the PPE value set $\mathcal{W}$ is a compact convex set. Hence, we define $\overline{v}_i$ (resp. $\underline{v}_i$) as player i's maximum (resp. minimum) equilibrium payoff. Indeed, $\overline{v}_i$ is player i's expected payoff when his preferred action is always taken: $\overline{v}_i = \mathbb{E}_\theta[u_i(a_i(\theta_i),\theta_i)]$.[13]

To study the dynamics of efficient mechanisms, we formalize the Bellman equation that characterizes the frontier of PPE values. Denote each player's possible PPE value set by $\mathcal{W}_i := \{v_i \in \mathbb{R} \mid \text{there exists } v_{-i} \in \mathbb{R} \text{ such that } (v_i, v_{-i}) \in \mathcal{W}\} = [\underline{v}_i, \overline{v}_i]$. For a given expected utility $v \in \mathcal{W}_2$ promised to player 2, define $V(v)$ as the highest value delivered to player 1. The value function $V : \mathcal{W}_2 \to \mathcal{W}_1$ satisfies the following functional equation.

$$V(v) = \max_{a(\cdot)\in\mathcal{A},\, w(\cdot)} \mathbb{E}_\theta\left[(1-\delta)u_1(a(\theta),\theta_1) + \delta V(w(\theta))\right]$$

subject to

$$\mathbb{E}_\theta\left[(1-\delta)u_2(a(\theta),\theta_2) + \delta w(\theta)\right] = v; \tag{PK}$$

$$\begin{aligned}&\mathbb{E}_{\theta_{-1}}[(1-\delta)u_1(a(\theta),\theta_1) + \delta V(w(\theta))]\\ &\quad \geq \mathbb{E}_{\theta_{-1}}[(1-\delta)u_1(a(\hat{\theta}_1,\theta_{-1}),\theta_1) + \delta V(w(\hat{\theta}_1,\theta_{-1}))] \text{ for all } \theta_1,\hat{\theta}_1 \in \Theta_1;\end{aligned} \tag{IC1}$$

$$\begin{aligned}&\mathbb{E}_{\theta_{-2}}[(1-\delta)u_2(a(\theta),\theta_2) + \delta w(\theta)]\\ &\quad \geq \mathbb{E}_{\theta_{-2}}[(1-\delta)u_2(a(\theta_1,\hat{\theta}_2),\theta_2) + \delta w(\theta_1,\hat{\theta}_2)] \text{ for all } \theta_2,\hat{\theta}_2 \in \Theta_2; \text{ and}\end{aligned} \tag{IC2}$$

$$w(\theta) \in \mathcal{W}_2 \text{ for all } \theta \in \Theta. \tag{SE}$$

The promise-keeping constraint (PK) requires that player 2's expected discounted utility delivered by a mechanism must be the promised value v. Expressions (ICi) are the agents' (Bayesian) incentive compatibility constraints: each agent i prefers to tell the truth, taking into account of the future consequences of his announcement, given that the other player is expected to be truthful. The self-enforceability condition (SE) requires the continuation values to be in the equilibrium value set.

[12]This assumption is introduced to convexify the set of PPE payoffs. While we implicitly introduce a public randomization device, we will explicitly formulate, in Appendix A, a stochastic mechanism as a collection of probability distributions on the current actions for each announcement of preference shocks.

[13]When player i's preferred action is always taken, he cannot do any better and the opponent is ignored and hence has no incentive to lie. In other words, a dictatorship is incentive compatible. Also, we assume that, in a one-shot game, the players can achieve a (static) equilibrium payoff profile v that Pareto-dominates the payoff profile associated with some random dictatorial allocation (i.e., a randomization of $a_1(\cdot)$ and $a_2(\cdot)$). This ensures that the set of PPE values has a non-empty interior.

A mechanism (a, w) is *incentive compatible* (IC) if it satisfies (ICi) for all $i \in N$. A mechanism (a, w) is *incentive feasible* at $v \in \mathcal{W}_2$ if it satisfies (PK), (ICi) for all $i \in N$, and (SE) when the promised value is v.

Since player i's best and $-i$'s worst Pareto efficient payoff is attained in PPE, the PPE value function passes through the payoff profile $(\overline{v}_i, \underline{v}_{-i}^P)$, where $\underline{v}_{-i}^P := \mathbb{E}_\theta[u_{-i}(a_i(\theta_i), \theta_{-i})]$, for each $i \in N$. Thus, the Pareto frontier of PPE values is always characterized by $(v, V(v))$ with $v \in \mathcal{W}_2^* := [\underline{v}_2^P, \overline{v}_2]$ (i.e., the frontier of PPE values is well-defined and downward-sloping on $\mathcal{W}_2^*$). Whenever $\underline{v}_2 < \underline{v}_2^P$, on the other hand, any frontier of PPE value on $[\underline{v}_2, \underline{v}_2^P]$ is upward-sloping, put differently, inefficient. This can occur, for example, when the players can take extreme actions in the quadratic preference collective decision making problem (i.e., $\mathcal{A}$ is "large" relative to Θ_i).

3.1 Renegotiation-proof PPE

We have so far characterized the frontier of the full-commitment PPE values and we have seen that some PPE values may be inefficient. In the collective decision making environment, for example, the players may be able to take an efficient joint action today by using an inefficient continuation contract (which may be associated with taking extreme actions from tomorrow on). An efficient full-commitment contract may be susceptible to renegotiation.

Hence, we study renegotiation-proof (RP, for short) PPE. In infinite horizon games like ours, however, there is no clear-cut definition of RP equilibrium, in contrast to the finite horizon games where renegotiation-proofness is well defined by backward induction. Among various concepts of renegotiation-proofness in infinite horizon games, we focus on *(internally) renegotiation-proof* PPE, following the concept of "internally renegotiation-proof" equilibrium sets by Ray [28].[14] The idea that a mechanism is internally renegotiation-proof is captured by the condition that it is (weakly) Pareto efficient such that its continuation contract is renegotiation-proof. It extends to the backward-induction definition of renegotiation-proofness in finite horizon games.

The key observation is that the payoff vector $(\overline{v}_i, \underline{v}_{-i}^P)$, the one associated with player i's best and $-i$'s worst Pareto efficient feasible payoff, is supported by an equilibrium of the one-shot game. Any PPE payoff vector v (where $v_i < \underline{v}_i^P$ for some $i \in N$) is Pareto-dominated. On the other hand, consider the frontier of the largest bounded self-generating set W (i.e., $W \subseteq B(W)$) contained in $[\underline{v}_2^P, \overline{v}_2] \times [\underline{v}_1^P, \overline{v}_1]$, where note that the frontier of W is downward sloping. The set W can be obtained as the largest fixed point of an auxiliary operator $B^* : \mathcal{P}(\mathbb{R}^2) \to \mathcal{P}(\mathbb{R}^2)$ defined as follows: for any $W \in \mathcal{P}(\mathbb{R}^2)$,

$$B^*(W) := B(W \cap ([\underline{v}_2^P, \overline{v}_2] \times [\underline{v}_1^P, \overline{v}_1])) \cap ([\underline{v}_2^P, \overline{v}_2] \times [\underline{v}_1^P, \overline{v}_1]).$$[15]

[14] See also Zhao [33] for the concept of "self-Pareto-generation," which formulates internal renegotiation-proofness in terms of a value function. In our analysis, as is standard, renegotiation can occur at the beginning of any given period.

[15] Alternatively, W can be characterized as the largest set satisfying $W = B(W) \cap ([\underline{v}_2^P, \overline{v}_2] \times [\underline{v}_1^P, \overline{v}_1])$.

We now characterize the frontier of W by the graph of the value function V. The graph of V has the property that the Pareto-frontier of the generated value function coincides with the original value function V. Thus, the RP-PPE value function $V : \mathcal{W}_2^* \to \mathcal{W}_1^*$ (where $\mathcal{W}_1^* := [\underline{v}_1^P, \overline{v}_1])$ satisfies the following Bellman equation. For each $v \in \mathcal{W}_2^*$,

$$V(v) = \max_{(a,w)(\cdot)\in\mathcal{A}\times\mathcal{W}_2^*} \mathbb{E}_\theta \left[(1-\delta)u_1(a(\theta),\theta_1) + \delta V(w(\theta))\right] \tag{OPT}$$
$$\text{subject to (PK) and (IC}i\text{) for each } i \in N.$$

We abuse notation to denote V for (full-commitment) PPE and RP-PPE value functions. Since V is strictly decreasing on $\mathcal{W}_2^*$, the promise-keeping constraint in the Bellman equation can be replaced with the weak inequality. Also, the frontier of W is the maximal fixed point of the Bellman equation.

The difference between the full-commitment and RP value functions is that renegotiation-proofness imposes a lower limit on player 2's continuation values in (OPT) as in the limited-commitment problem. As stressed by Zhao [33], however, the difference between the renegotiation-proof and limited-commitment problems is that while an agent can unilaterally walk away from an ongoing contract in limited-commitment environments, renegotiation-proofness requires that both agents mutually agree on any change in an ongoing contract. Hence, whenever $\underline{v}_2 = \underline{v}_2^P$ as in the insurance model, no restriction is imposed.

The Bellman equations characterizing RP (as well as full-commitment) PPE value functions have a special feature that the value function itself enters into the agent 1's incentive constraint, as in Wang [31] and Zhao [32]. Thus, for any two given candidate value functions, the sets of incentive compatible mechanisms under these value functions would be distinct. This implies that the one-step operator (defined by the right-hand side of the Bellman equation) may not be monotone, and hence the standard technique (i.e., Blackwell condition (see, for example, Aliprantis and Border [2, Theorem 3.53])) to verify that the one-step operator is a contraction mapping cannot readily be invoked. Despite this, we have established that the the RP-PPE value function is well defined. We also have an algorithm to compute the RP-PPE value function of our model by iterating the auxiliary operator B^*.[16]

Moreover, by using this observation, we show in Appendix A that the largest fixed point $W = W(\delta)$ of B^* is monotonically non-decreasing in the discount factor δ. Thus, the RP (as well as the full-commitment) PPE value functions are monotonically non-decreasing in the discount factor.

Note also that internal renegotiation-proofness allows a current action to be ex post inefficient. It requires an optimal mechanism to be (weakly) Pareto efficient subject to the continuation mechanism being "internally renegotiation-proof," and hence ex post inefficiency may occur due to the cost of keeping incentive constraints. Thus, for example, in the quadratic collective decision problem, internal renegotiation-proofness is weaker

[16] In contrast, an internally renegotiation-proof equilibrium does not necessarily exist. See Ray [28] and Zhao [33].

than the requirement that the current action be between players' reported types (i.e., $a(\theta_1, \theta_2) \in [\min(\theta_1, \theta_2), \max(\theta_1, \theta_2)]$). Indeed, its numerical simulation in Appendix B demonstrates that an optimal current action can be ex post inefficient.

3.2 Approximate Efficiency

The first-best allocation maximizes a weighted sum of players' instantaneous utilities. Formally, the first-best value function on $\mathcal{W}_2^*$, $V^{\text{FB}} : \mathcal{W}_2^* \to \mathcal{W}_1^*$, is characterized by the highest feasible payoff delivered to player 1 given that player 2' expected value is at least $v \in \mathcal{W}_2^*$:

$$V^{\text{FB}}(v) = \max_{a(\cdot) \in \mathcal{A}} \mathbb{E}_\theta \left[u_1(a(\theta), \theta_1) \right] \text{ subject to } \mathbb{E}_\theta \left[u_2(a(\theta), \theta_2) \right] \geq v. \tag{FB}$$

Denote by $a^{\text{FB}}(\cdot|v) : \Theta \to \mathcal{A}$ the first-best allocation, i.e., the solution to the first-best Pareto problem (FB).

In the quadratic collective decision model, the first-best action is a weighted average of the agents' types: $a^{\text{FB}}(\theta|v) = \sqrt{\frac{v}{\underline{v}_2^P}}\theta_1 + \left(1 - \sqrt{\frac{v}{\underline{v}_2^P}}\right)\theta_2$, where $v \in [\underline{v}_2^P, 0]$ and $\underline{v}_2^P = -\mathbb{E}_\theta[(\theta_1 - \theta_2)^2]$. The difficulty of implementing the first-best decision rule is that, whenever an agent's preference shock is different from his opponent's average type, he would have an incentive to exaggerate his report towards the extremes.

When $\delta > 0$, continuation values can be used as an additional instrument to get agents to report their types truthfully. The key is to present the agents with a trade-off between the benefit of a larger influence on the current decision and the loss they will incur in future continuation values. This allows the mechanism to take into account the intensity of the agents' preferences, which, in turn, leads to efficiency gains when compared to a static problem.

Continuation values play a similar role to the one side-payments play in standard static incentive problems. The difference between side-payments and continuation values is that the latter can only imperfectly transfer utilities across players. In particular, to transfer continuation utility from one player to another in any period, the decisions for subsequent periods must be altered in an incentive compatible way.[17] This, together with the lack of observability, implies that generally the players cannot attain *exact* efficiency as an equilibrium outcome. Indeed, exact efficiency would call for an equilibrium in which for all histories future decisions would not respond to current announcements. Hence, truth-telling would have to be a static best response for the agents, and this cannot be attained in an incentive compatible manner except for rather trivial cases in which the first-best allocation is achieved even in a one-shot game.[18]

Although full efficiency cannot generally be attained, one can arbitrarily approximate it as the players become patient.

[17] Moreover, continuation values must be drawn from the equilibrium value set.

[18] An example is a quadratic collective decision making problem among ex ante symmetric players with preference shocks less than or equal to three.

Proposition 1 (Approximate Efficiency)**.** *For any $\varepsilon > 0$, there exists $\overline{\delta} \in (0,1)$ such that, if $\delta \in (\overline{\delta}, 1)$ then the RP-PPE value $V_\delta^{\mathrm{SB}}(v) := V(v)$ is within ε of the first-best value $V^{\mathrm{FB}}(v)$ for all $v \in \mathcal{W}_2^*$. Put differently, V_δ^{SB} converges uniformly to V^{FB} as δ approaches 1. Moreover, the convergence of V_δ^{SB} is monotone in the discount factor δ.*

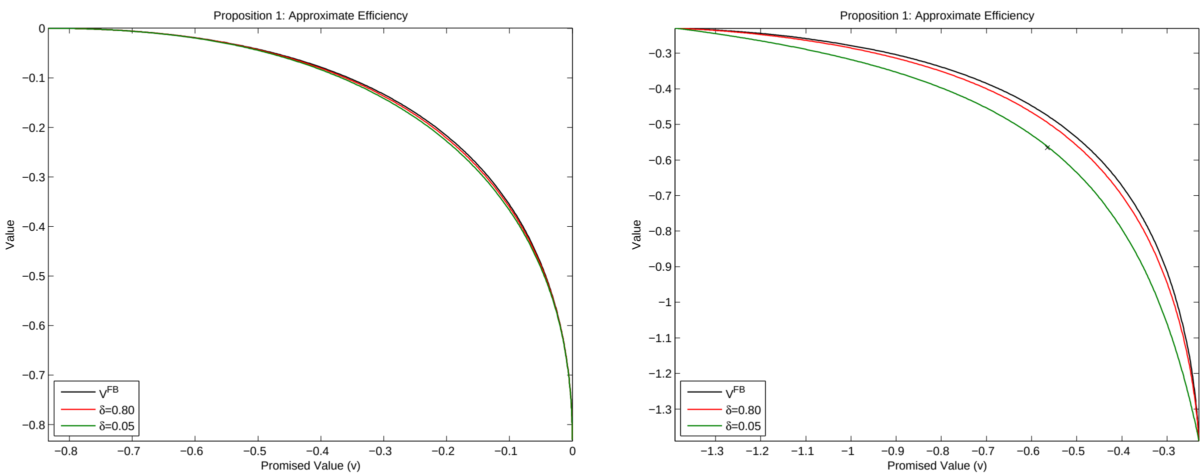


Figure 1: The Value Functions V_δ^{SB} and V^{FB} in the Quadratic Collective Decision Problem (Left) and in the CARA Insurance Problem (Right). The "×" in the right panel denotes the symmetric static second-best values (i.e., the values with no transfer).

We illustrate this result in Figure 1 in which we plot the RP-PPE value functions V_δ^{SB} for different values of δ and compare them with the first-best value function V^{FB}. We do this for the quadratic collective decision problem on the left and for the CARA insurance problem on the right.[19] In both cases it can be observed that, as stated in Proposition 1, V_δ^{SB} uniformly converges to V^{FB} as δ approaches 1. If we were to zoom in we could also see that they actually never quite attain the first-best values for interior v (see also Appendix B, which reports the social welfare loss in %). In addition, it is worth noting that the behavior of both problems is quite different for low values of δ. In particular, the efficiency losses due to the private information are much larger in the insurance problem than in the collective decision problem. This is due to the fact that in the collective decision problem even in the static setting $\delta = 0$ the agents can credibly convey some of their private information since they are not in a zero-sum game as is the case in the static insurance problem. It is also worth noting that the approximate efficiency of the RP-PPE value function immediately implies that of the (full-commitment) PPE value function (on $\mathcal{W}_2^*$).

The intuition behind the approximate efficiency result is as follows. As the discount factor δ approaches unity, the utility value of the current period, which is weighted by $(1-\delta)$,

[19]For the quadratic collective decision making problem, we set $\Theta_1 = \Theta_2 = \{-1+0.25k \mid k \in \{0,1,\ldots,8\}\}$ with the uniform density. The action space is $\mathcal{A} = [-1.1, 1.1]$. For the CARA insurance problem, we let $u_1(a,\theta_1) = -\exp(-(\theta_1 - a))$ and $u_2(a,\theta_2) = -\exp(-(\theta_2 + a))$. We set $\Theta_i = \{0, 0.75, 1.5\}$ with the uniform density. The set of feasible transfers is $\mathcal{A} = [-0.9, 0.9]$. See Appendix B for more details.

becomes insignificant relative to the continuation values. Hence, in order to guarantee truth-telling in the current period, continuation values have to vary only minimally. Since the value function turns out to be locally linear, the associated losses from the variation in continuation values become negligible. It is worth noting that the trade-off between current allocation and continuation values is indeed less steep for high values of δ. When δ is close to unity, it is easier to provide incentives and there will be less variability in future allocations.

Similarly to Athey and Bagwell [4], we construct continuation values that replicate the expected payments of the expected externality mechanism proposed by Arrow [3] and d'Aspremont and Gérard-Varet [9], that guarantee efficiency in standard static Bayesian mechanism design problems with transfers.[20] This approach differs from Fudenberg, Levine, and Maskin [13].[21]

In the context of collective decision problems, our approximate efficiency result can be contrasted with Jackson and Sonnenschein [18]. In their linking mechanism, the agents are allowed to report each possible type a fixed number of times according to the frequency with which that type should statistically be realized. Jackson and Sonnenschein [18, Corollary 2] demonstrate that, for any $\varepsilon > 0$, their linking mechanism is less than ε inefficient relative to the first best if players are patient and face sufficiently many identical problems. The sources of the efficiency losses in their mechanism and in our scheme are thus quite different. In the linking mechanism, when the last periods get closer, agents may not be able to report their types truthfully, as they might have run out of their budgeted reports for a particular type. Then, the linking mechanism forces them to lie. Instead, in our setting, the inefficiency arises from the variations in continuation values over time.

3.3 The Dynamics of Collective Decision Making

For ease of exposition, we take the *first-order approach* to solving the second-best problem (OPT). By the first-order approach, we consider a relaxed problem where each agent's relevant incentive constraints are local incentive constraints. Appendix A establishes our main result even when we consider the full problem (OPT). It also provides a condition under which the first-order approach is valid for the following standard environments: quadratic, CARA, and multiplicative preferences.

To state the first-order approach, we first define $U_i(\theta_i, \hat{\theta}_i|a, w)$ as agent i's expected utility of a mechanism (a, w) when his type is θ_i and he announces $\hat{\theta}_i$:

$$U_i(\theta_i, \hat{\theta}_i|a, w) := \begin{cases} \mathbb{E}_{\theta_{-1}}\left[(1-\delta)u_1(a(\hat{\theta}_1, \theta_2), \theta_1) + \delta V(w(\hat{\theta}_1, \theta_2))\right] & \text{if } i = 1 \\ \mathbb{E}_{\theta_{-2}}\left[(1-\delta)u_2(a(\theta_1, \hat{\theta}_2), \theta_2) + \delta w(\theta_1, \hat{\theta}_2)\right] & \text{if } i = 2 \end{cases}.$$

[20] In the infinitely repeated Bertrand duopoly game studied by Athey and Bagwell [4], in contrast to our approximate efficiency result, monopoly profits can be exactly attained for a discount factor strictly less than one by firms making use of asymmetric continuation values. The key difference from our model is that, for some states that occur with positive probability, firms in their model can allocate profits without any costs.

[21] Furthermore, their conditions for the Folk theorem are not satisfied in our model.

We denote by $U_i(\theta_i|a,w) = U_i(\theta_i,\theta_i|a,w)$ or simply by $U_i(\theta_i)$ whenever it is clear from the context. Second, we assume that each $\Theta_i = \{\theta_i^{(k_i)}\}_{k_i=1}^{m_i}$ is in $\mathbb{R}$ such that $\theta_i^{(1)} < \theta_i^{(2)} < \cdots < \theta_i^{(m_i)}$.[22] Now, each player's local upward incentive constraints (ICi-UP) and local downward incentive constraints (ICi-DW) are written as follows.

$$U_i\left(\theta_i^{(k_i)}\middle| a,w\right) \geq U_i\left(\theta_i^{(k_i)},\theta_i^{(k_i+1)}\middle| a,w\right) \text{ for all } k_i \in \{1,\ldots,m_i-1\}; \text{ and} \tag{ICi-UP}$$

$$U_i\left(\theta_i^{(k_i)}\middle| a,w\right) \geq U_i\left(\theta_i^{(k_i)},\theta_i^{(k_i-1)}\middle| a,w\right) \text{ for all } k_i \in \{2,\ldots,m_i\}. \tag{ICi-DW}$$

We find the optimal mechanism by solving the following relaxed second-best problem (OPT*) for each promised value $v \in \mathcal{W}_2^*$:

$$V(v) = \max_{\substack{a(\cdot)\in\mathcal{A} \\ w(\cdot)\in\mathcal{W}_2^*}} \mathbb{E}_\theta\left[(1-\delta)\,u_1\left(a(\theta),\theta_1\right) + \delta V(w(\theta))\right] \tag{OPT*}$$

$$\text{subject to (PK) and (IC}i\text{-UP) and (IC}i\text{-DW) for each } i \in N.$$

The optimal dynamic mechanism design problem is reduced to maximizing the following Lagrangian $\mathcal{L}(\cdot|\lambda,\gamma,v)$ at each promised value $v \in \mathcal{W}_2^*$:

$$\begin{aligned}\mathcal{L}\left(a,w|\lambda,\gamma,v\right) = &\sum_{i=1}^{2}\gamma_i\mathbb{E}_{\theta_i}\left[U_i(\theta_i|a,w)\right] - \gamma_2 v \\ &+\sum_{i=1}^{2}\sum_{k_i=1}^{m_i-1}\lambda_{i,k_i}^{\text{UP}}\left\{U_i\left(\theta_i^{(k_i)}\middle| a,w\right) - U_i\left(\theta_i^{(k_i)},\theta_i^{(k_i+1)}\middle| a,w\right)\right\} \\ &+\sum_{i=1}^{2}\sum_{k_i=2}^{m_i}\lambda_{i,k_i}^{\text{DW}}\left\{U_i\left(\theta_i^{(k_i)}\middle| a,w\right) - U_i\left(\theta_i^{(k_i)},\theta_i^{(k_i-1)}\middle| a,w\right)\right\},\end{aligned} \tag{LAG}$$

where the vector of Lagrange multipliers $\lambda = ((\lambda_{i,k_i}^{\text{UP}})_{k_i=1}^{m_i-1}, (\lambda_{i,k_i}^{\text{DW}})_{k_i=2}^{m_i})_{i\in N}$ and $\gamma = (\gamma_1,\gamma_2)$ denote the following. The multipliers $(\lambda_{i,k_i}^{\text{UP}})_{k_i}$ and $(\lambda_{i,k_i}^{\text{DW}})_{k_i}$ are those on the player i's local upward and downward incentive compatibility constraints (ICi-UP) and (ICi-DW), respectively. The multipliers γ_1 and γ_2 play the role of Pareto weights. The multiplier γ_2 is the one on the player 2's promise-keeping constraint (PK), and γ_1 is that on the objective function (player 1's utility function). Since $(\gamma_1,\gamma_2) \neq 0$, we normalize it by $\gamma_1+\gamma_2 = 1$ (alternatively, we could normalize γ_1 by $\gamma_1 = 1$ whenever $\gamma_1 > 0$).

Formally, there exists a vector of non-negative Lagrange multipliers (λ,γ) with $\gamma_1+\gamma_2 = 1$ such that the following two conditions characterize an optimal mechanism.

1. The Lagrangian $\mathcal{L}(\cdot,\cdot|\lambda,\gamma,v)$ is maximized at an optimal mechanism (a,w), i.e.,

$$\mathcal{L}(a,w|\lambda,\gamma,v) \geq \mathcal{L}(\tilde{a},\tilde{w}|\lambda,\gamma,v) \text{ for all functions } (\tilde{a},\tilde{w}) : \Theta \to \mathcal{A}\times\mathcal{W}_2^*.$$

[22] This assumption can be dropped when we take care of the entire IC constraints. See Appendix A.

2. The complementary slackness conditions are satisfied:

   - $\gamma_2\left(\mathbb{E}_\theta[(1-\delta)u_2(a(\theta),\theta_2)+\delta w(\theta)]-v\right)=0$;
   - For each $i \in N$ and $k_i \in \{1,\dots,m_i-1\}$,

$$\lambda^{\mathrm{UP}}_{i,k_i}\left[U_i\left(\theta_i^{(k_i)}\middle| a,w\right)-U_i\left(\theta_i^{(k_i)},\theta_i^{(k_i+1)}\middle| a,w\right)\right]=0;$$

   - For each $i \in N$ and $k_i \in \{2,\dots,m_i\}$,

$$\lambda^{\mathrm{DW}}_{i,k_i}\left[U_i\left(\theta_i^{(k_i)}\middle| a,w\right)-U_i\left(\theta_i^{(k_i)},\theta_i^{(k_i-1)}\middle| a,w\right)\right]=0.$$

Following Myerson [21, 22, 24], we view the Lagrangian (LAG) as representing the sum of the agents' *virtual* utilities. To that end, we introduce the dummy multipliers $\lambda^{\mathrm{UP}}_{i,0}=\lambda^{\mathrm{UP}}_{i,m_i}=\lambda^{\mathrm{DW}}_{i,1}=\lambda^{\mathrm{DW}}_{i,m_i+1}=0$ and dummy types $\theta_i^{(0)}$ and $\theta_i^{(m_i+1)}$. We let $\gamma_i\tilde{u}_i(a,\theta_i)$ be player i's instantaneous virtual utility at θ_i with respect to γ_i and λ_i:

$$\begin{aligned}
&\gamma_i\tilde{u}_i\left(a,\theta_i^{(k_i)}\right)\\
&:=\gamma_i\left\{u_i\left(a,\theta_i^{(k_i)}\right)\left(1+\frac{\tilde{\lambda}^{\mathrm{UP}}_{i,k_i}+\tilde{\lambda}^{\mathrm{DW}}_{i,k_i}}{f_i(\theta_i^{(k_i)})}\right)-u_i\left(a,\theta_i^{(k_i-1)}\right)\frac{\tilde{\lambda}^{\mathrm{UP}}_{i,k_i-1}}{f_i(\theta_i^{(k_i)})}-u_i\left(a,\theta_i^{(k_i+1)}\right)\frac{\tilde{\lambda}^{\mathrm{DW}}_{i,k_i+1}}{f_i(\theta_i^{(k_i)})}\right\}\\
&=\gamma_i\left\{u_i\left(a,\theta_i^{(k_i)}\right)\left(1+\frac{\tilde{\lambda}^{\mathrm{UP}}_{i,k_i}-\tilde{\lambda}^{\mathrm{UP}}_{i,k_i-1}+\tilde{\lambda}^{\mathrm{DW}}_{i,k_i}-\tilde{\lambda}^{\mathrm{DW}}_{i,k_i+1}}{f_i(\theta_i^{(k_i)})}\right)\right.\\
&\quad\left.+\left\{u_i\left(a,\theta_i^{(k_i)}\right)-u_i\left(a,\theta_i^{(k_i-1)}\right)\right\}\frac{\tilde{\lambda}^{\mathrm{UP}}_{i,k_i-1}}{f_i(\theta_i^{(k_i)})}+\left\{u_i\left(a,\theta_i^{(k_i)}\right)-u_i\left(a,\theta_i^{(k_i+1)}\right)\right\}\frac{\tilde{\lambda}^{\mathrm{DW}}_{i,k_i+1}}{f_i(\theta_i^{(k_i)})}\right\},
\end{aligned}\tag{VU}$$

where the vector $\tilde{\lambda}$ of multipliers is defined by

$$\tilde{\lambda}^{\mathrm{UP}}_{i,k_i}=\begin{cases}\dfrac{\lambda^{\mathrm{UP}}_{i,k_i}}{\gamma_i} & \text{if } \gamma_i>0\\ 0 & \text{if } \gamma_i=0\end{cases}\text{ and }\tilde{\lambda}^{\mathrm{DW}}_{i,k_i}=\begin{cases}\dfrac{\lambda^{\mathrm{DW}}_{i,k_i}}{\gamma_i} & \text{if } \gamma_i>0\\ 0 & \text{if } \gamma_i=0\end{cases}.$$

The first term of $\tilde{u}_i(a,\theta_i^{(k_i)})$ (in the first expression) is the utility $u_i(a,\theta_i^{(k_i)})$ adjusted by the cost of incentives associated with type $\theta_i^{(k_i)}$ lying upward $\theta_i^{(k_i+1)}$ and downward $\theta_i^{(k_i-1)}$. The second term is the utility $u_i(a,\theta_i^{(k_i-1)})$ adjusted by the type $\theta_i^{(k_i-1)}$ imitating $\theta_i^{(k_i)}$. The third term is the utility $u_i(a,\theta_i^{(k_i+1)})$ adjusted by the type $\theta_i^{(k_i+1)}$ imitating $\theta_i^{(k_i)}$. If we do not take the first-order approach, then we incorporate the costs of incentives associated with type $\theta_i^{(k_i)}$ imitating all the other types and all the other types imitating $\theta_i^{(k_i)}$. See Appendix A.

Now, we re-write the Lagrangian (LAG) as the sum of the agents' virtual utilities:

$$\begin{aligned}\mathcal{L}(a, w|\lambda, \gamma, v) = \mathbb{E}_\theta \Bigg[\Bigg\{ &\sum_{i=1}^{2} (1-\delta)\gamma_i \tilde{u}_i \left(a(\theta_i^{(k_i)}, \theta_{-i}), \theta_i^{(k_i)} \right) \\ &+ \gamma_1 \delta V(w(\theta_1^{(k_1)}, \theta_2^{(k_2)})) \left(1 + \frac{\tilde{\lambda}^{\mathrm{UP}}_{1,k_1} - \tilde{\lambda}^{\mathrm{UP}}_{1,k_1-1} + \tilde{\lambda}^{\mathrm{DW}}_{1,k_1} - \tilde{\lambda}^{\mathrm{DW}}_{1,k_1+1}}{f_1(\theta_1^{(k_1)})} \right) \\ &+ \gamma_2 \delta w(\theta_1^{(k_1)}, \theta_2^{(k_2)}) \left(1 + \frac{\tilde{\lambda}^{\mathrm{UP}}_{2,k_2} - \tilde{\lambda}^{\mathrm{UP}}_{2,k_2-1} + \tilde{\lambda}^{\mathrm{DW}}_{2,k_2} - \tilde{\lambda}^{\mathrm{DW}}_{2,k_2+1}}{f_2(\theta_2^{(k_2)})} \right) \Bigg\} \Bigg] - \gamma_2 v. \end{aligned} \tag{LAG*}$$

The optimal current action maximizes the weighted sum of the agents' *virtual* instantaneous utilities $\tilde{u}_i$ with the weight given to agent i being γ_i.[23] The optimal continuation value function $w(\cdot)$, on the other hand, maximizes the weighted sum of players' continuation virtual utilities with the weight given to agent i being γ_i:

$$\begin{aligned}\mathbb{E}_\theta \Bigg[&\gamma_1 \delta V(w(\theta_1^{(k_1)}, \theta_2^{(k_2)})) \left(1 + \frac{\tilde{\lambda}^{\mathrm{UP}}_{1,k_1} - \tilde{\lambda}^{\mathrm{UP}}_{1,k_1-1} + \tilde{\lambda}^{\mathrm{DW}}_{1,k_1} - \tilde{\lambda}^{\mathrm{DW}}_{1,k_1+1}}{f_1(\theta_1^{(k_1)})} \right) \\ &+ \gamma_2 \delta w(\theta_1^{(k_1)}, \theta_2^{(k_2)}) \left(1 + \frac{\tilde{\lambda}^{\mathrm{UP}}_{2,k_2} - \tilde{\lambda}^{\mathrm{UP}}_{2,k_2-1} + \tilde{\lambda}^{\mathrm{DW}}_{2,k_2} - \tilde{\lambda}^{\mathrm{DW}}_{2,k_2+1}}{f_2(\theta_2^{(k_2)})} \right) \Bigg].\end{aligned}$$

Again, agent i's continuation utility is adjusted by the costs of incentives associated with type $\theta_i^{(k_i)}$ imitating $\theta_i^{(k_i+1)}$ and $\theta_i^{(k_i-1)}$ and types $\theta_i^{(k_i+1)}$ and $\theta_i^{(k_i-1)}$ imitating $\theta_i^{(k_i)}$. If we do not take the first-order approach, each agent's continuation utility is adjusted by the costs of incentives associated with type $\theta_i^{(k_i)}$ imitating all the other types and all the other types imitating $\theta_i^{(k_i)}$. We remark that while we can solve the optimal current allocation and the continuation value function separately, the Pareto weights attached on the players' virtual instantaneous utilities and virtual continuation utilities are the same: γ_1 and γ_2.

Hence, the dynamics of repeated decision making are fully determined by (i) a decision rule that, at each period, maximizes the weighted sum of the agents' instantaneous *virtual* utilities, and (ii) the process that governs the evolution of the weights attached on the agents' virtual utilities. While the Pareto weights (Lagrange multipliers) of the first-best problem are time-invariant, we show that each player's relative Pareto weights always vary with a positive probability in a way that it increases in expectation over time.

While our general model accommodates a dynamic insurance partnership model as a special environment, a repeated collective decision problem has distinct features from a pure dynamic insurance model. First, in the collective decision models, while agents have an incentive to claim that their type is more extreme than it actually is, they do not always have an incentive to exaggerate their preferences all the way to the extremes. This is

[23]Precisely, the player i's virtual utility itself (the left-hand side of (VU)) depends on his weight γ_i. That is, the optimal current allocation maximizes the sum of the agents' instantaneous virtual utilities $\gamma_i \tilde{u}_i$.

in contrast to the dynamic insurance models, where agents have an incentive to claim to be poorer than they actually are and those facing the first-best contract indeed have an incentive to misreport all the way to the lowest possible endowment. Moreover, a relevant direction in which the IC constraints bind depends on whether one player's favorite action is above or below the other player's average type. The relevant constraints for a player whose favorite action is above his opponent's average are those that ensure he does not want to lie upwards. Conversely, for the case in which a player whose favorite action is below his opponent's average type, the relevant constraints are the local downward ones.

Second, unlike the insurance models, variations in the continuation values are not necessary in order to provide *some* insurance. For example, in Thomas and Worrall [30], when $\delta = 0$, there is no way the principal can provide any insurance to the agent that gets a low income realization. This is also true in the dynamic insurance partnership environment. In collective decision making settings, however, since agents do not know how aligned their interests are, it is possible, even in the static case, to have the allocation depend on agents' types in an incentive compatible way.

In the optimal dynamic mechanism it will always be efficient to have continuation values vary over time for any given promised value $v \in (\underline{v}_2^P, \overline{v}_2)$. As our model accommodates the insurance models, the intuition behind this result is similar to the insurance models. Continuation values allow agents with an extreme type in the current period (poor agents in the insurance models) to get more weight in the current allocation choice (higher current consumption) in exchange for forgoing future decision rights (future consumption).

Proposition 2. *Let $v \in (\underline{v}_2^P, \overline{v}_2)$ be such that $w(\theta|v) \in (\underline{v}_2^P, \overline{v}_2)$ for all $\theta \in \Theta$. Let (γ_1, γ_2) be the agents' Pareto weights associated with v, and let $(\gamma_1'(\theta), \gamma_2'(\theta))$ be the agents' Pareto weights associated with $w(\theta|v)$. Then, each agent i's relative Pareto weights satisfies*

$$\frac{\gamma_i}{\gamma_{-i}} \leq \mathbb{E}_\theta \left[\frac{\gamma_i'(\theta)}{\gamma_{-i}'(\theta)} \right].$$

Proposition 3 (Spreading of Values)**.** *Let $v \in (\underline{v}_2^P, \overline{v}_2)$ be such that $w(\theta|v) \in (\underline{v}_2^P, \overline{v}_2)$ for all $\theta \in \Theta$. Then, for each $i \in N$, there is positive probability of both $\frac{\gamma_i'(\cdot)}{\gamma_{-i}'(\cdot)} > \frac{\gamma_i}{\gamma_{-i}}$ and $\frac{\gamma_i'(\cdot)}{\gamma_{-i}'(\cdot)} < \frac{\gamma_i}{\gamma_{-i}}$.*

We remark that similar martingale properties for marginal continuation values also hold in dynamic insurance models with one-sided asymmetric information. In Thomas and Worrall [30], the principal's marginal continuation value (thus the agent's relative Pareto weights) follows a martingale process with respect to the agent's (endowment) shock distribution. In fact, our result would generalize that of Thomas and Worrall [30] in that if player 1 had no preference shock then his marginal value with respect to player 2's preference shocks would follow a martingale process with respect to player 2's type distribution in the sense that the expression in Proposition 2 would hold with equality.

Figure 2 numerically illustrates the dynamics of the optimal mechanism. The left panel of Figure 2 illustrates a sample path of players' continuation values in the quadratic preference collective decision making problem with the parameter values specified in Footnote 19.[24] The right panel of Figure 2 illustrates that the degree of inequality measured by the share of values captured by the better-off agent (precisely, $\frac{\max(\overline{v}_1 - V_\delta^{\mathrm{SB}}(v^t), \overline{v}_2 - v^t)}{(\overline{v}_1 - V_\delta^{\mathrm{SB}}(v^t)) + (\overline{v}_2 - v^t)}$) increases at a faster rate for lower δ. For each discount factor, the share of values captured by the better-off agent is calculated as the average of 500 simulations. Intuitively, as δ decreases, the continuation values matter less so that the agents trade off more continuation values in order to influence the current decision.

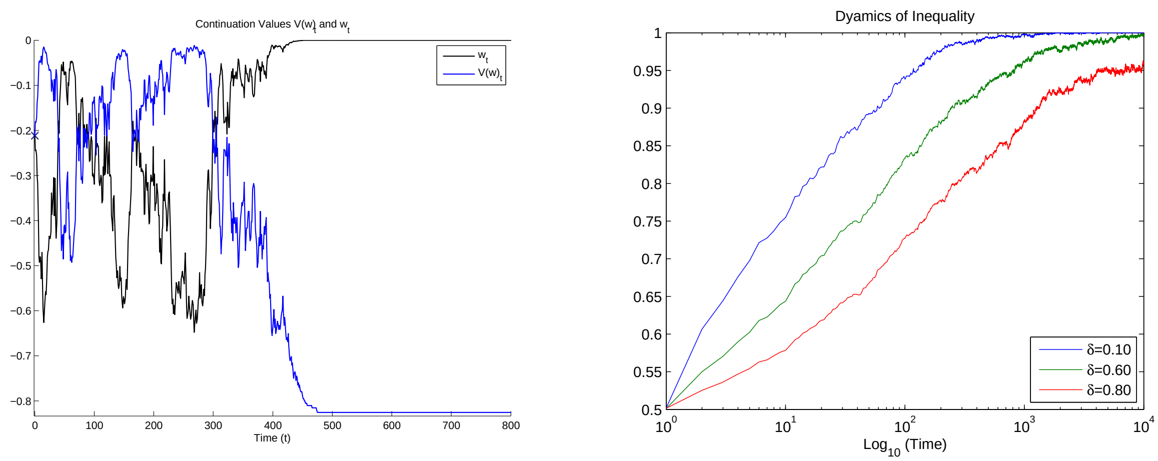


Figure 2: A Sample Path of $(V(v^t), v^t)$ (Left) and Dynamics of Inequality (Right).

This result contrasts with Wang [31]'s result on existence of non-degenerate limiting wealth distribution. The crucial difference between his and our models is existence of endogenous feasibility constraint. In Wang [31], the agents can transfer only their endowments so that *any* promised value has to vary each time. In contrast, in our model, once a promised value hits some player's maximum payoff, his favorite action must be taken from then on.

4 Final Remarks

We have shown that, in an optimal renegotiation-proof dynamic Bayesian mechanism, the provision of intertemporal incentives spreads utility weights on the agents. Our model leaves several interesting avenues for future research. The first is to incorporate endogenous, type dependent, limited commitment. In environments with endogenous participation constraints, such as Fuchs and Lippi [12], the threat of abandoning the partnership puts a limit on the extent to which one of the agents can dominate the decision process. The second is persistence of private information and correlation of shocks among the agents. Although more realistic, such generalizations will pose significant analytical challenges since when

[24]We set the discount factor $\delta = 0.4$. Also, we discretize the set $\mathcal{W}_2^* = [\underline{v}_2^P, \overline{v}_2]$, where $\underline{v}_i^P = -0.833333$ and $\overline{v}_i = 0$. See Appendix B for more details.

considering a deviation in the current report agents must also consider the effects such a deviation has on the distribution of beliefs for future periods.[25] An additional element to consider in such a generalized environment is the way information is communicated, because it will affect the amount of learning that will take place on equilibrium path. For example, one could have agents confidentially report to a central mechanism that simply reports back the recommended action or a public announcement where all players learn the other players' reports. Despite the difficulties we believe that it is worth trying to incorporate these considerations in future work.

References

[1] D. Abreu, D. Pearce, and E. Stacchetti. "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring". Econometrica 58 (1990), 1041–1063.

[2] C. D. Aliprantis and K. Border. Infinite Dimensional Analysis: A Hitchhiker's Guide. Third Edition. Springer-Verlag Berlin Heidelberg, 2006.

[3] K. J. Arrow. "The Property Rights Doctrine and Demand Revelation under Incomplete Information". Economics and Human Welfare. Ed. by M. J. Boskin. Academic Press, 1979, 23–39.

[4] S. Athey and K. Bagwell. "Optimal Collusion with Private Information". RAND J. Econ. 32 (2001), 428–465.

[5] A. Atkeson and R. E. Lucas Jr. "On Efficient Distribution With Private Information". Rev. Econ. Stud. 59 (1992), 427–453.

[6] J. Bergin and W. MacLeod. "Efficiency and Renegotiation in Repeated Games". J. Econ. Theory 61 (1993), 42–73.

[7] V. Carrasco and W. Fuchs. "Dividing and Discarding: A Procedure for Taking Decisions with Non-transferable Utility". 2009.

[8] A. Casella. "Storable Votes". Games Econ. Behav. 51 (2005), 391–419.

[9] C. d'Aspremont and L.-A. Gérard-Varet. "Incentives and Incomplete Information". J. Public Econ. 11 (1979), 25–45.

[10] M. Drexl and A. Kleiner. "Preference Intensities in Repeated Collective Decision-Making". 2014.

[11] E. Friedman. Risk Sharing and the Dynamics of Inequality. Tech. rep. Discussion Paper 1235, Center for Mathematical Studies of Economics and Management Sciences, Northwestern University, 1998.

[25] The recent work by Pavan, Segal, and Toikka [25] might prove useful in implementing such an extension. Guo and Hörner [16] study a dynamic principal-agent problem without monetary transfers in which the valuations of a good follow a two-state Markov chain.

[12] W. Fuchs and F. Lippi. “Monetary Union with Voluntary Participation”. Rev. Econ. Stud. 73 (2006), 437–457.

[13] D. Fudenberg, D. Levine, and E. Maskin. “The Folk Theorem with Imperfect Public Information”. Econometrica 62 (1994), 997–1039.

[14] M. Goltsman et al. “Mediation, Arbitration and Negotiation”. J. Econ. Theory 144 (2009), 1397–1420.

[15] E. J. Green. “Lending and the Smoothing of Uninsurable Income”. Contractual Arrangements for Intertemporal Trade. Ed. by E. C. Prescott and N. Wallace. Minneapolis: The University of Minnesota Press, 1987, 3–25.

[16] Y. Guo and J. Hörner. “Dynamic Mechanism Design without Money”. 2015.

[17] R. Hortala-Vallve. “Qualitative voting”. J. Theor. Polit. 24 (2012), 526–554.

[18] M. O. Jackson and H. F. Sonnenschein. “Overcoming Incentive Constraints by Linking Decisions”. Econometrica 75 (2007), 241–257.

[19] E. Kováč and T. Mylovanov. “Stochastic mechanisms in settings without monetary transfers: The regular case”. J. Econ. Theory 144 (2009), 1373–1395.

[20] B. Lockwood and J. Thomas. “Asymptotic Efficiency in Principal-agent Models with Hidden Information”. Econ. Letters 30 (1989), 297–301.

[21] R. B. Myerson. “Bayesian Equilibrium and Incentive-Compatibility: an Introduction”. Social Goals and Social Organization: Essays in Memory of Elisha Pazner. Ed. by L. Hurwicz, D. Schmeidler, and H. Sonnenschein. Cambridge University Press, 1985, 229–259.

[22] R. B. Myerson. “Cooperative Games with Incomplete Information”. Int. J. Game Theory 13 (1984), 69–96.

[23] R. B. Myerson. “Incentive Compatibility and the Bargaining Problem”. Econometrica 47 (1979), 61–73.

[24] R. B. Myerson. “Optimal Auction Design”. Math. Oper. Res. 6 (1981), 58–73.

[25] A. Pavan, I. Segal, and J. Toikka. “Dynamic Mechanism Design: A Myersonian Arroach”. Econometrica 82 (2014), 601–653.

[26] C. Phelan. “On the Long Run Implications of Repeated Moral Hazard”. J. Econ. Theory 79 (1998), 174–191.

[27] C. Phelan. “Repeated Moral Hazard and One-Sided Commitment”. J. Econ. Theory 66 (1995), 488–506.

[28] D. Ray. “Internally Renegotiation-Proof Equilibrium Sets: Limit Behavior with Low Discounting”. Games Econ. Behav. 6 (1994), 162–177.

[29] A. Skrzypacz and H. Hopenhayn. “Tacit Collusion in Repeated Auctions”. J. Econ. Theory 114 (2004), 153–169.

[30] J. Thomas and T. Worrall. "Income Fluctuation and Asymmetric Information: An Example of a Repeated Principal-agent Problem". J. Econ. Theory 51 (1990), 367–390.

[31] C. Wang. "Dynamic Insurance with Private Information and Balanced Budgets". Rev. Econ. Stud. 62 (1995), 577–595.

[32] R. R. Zhao. "Dynamic Risk-sharing with Two-sided Moral Hazard". J. Econ. Theory 136 (2007), 601–640.

[33] R. R. Zhao. "Renegotiation-proof Contract in Repeated Agency". J. Econ. Theory 131 (2006), 263–281.

A Appendix A

In this Appendix, we explicitly formulate stochastic mechanisms, which are history-dependent stochastic allocations (i.e., probability distributions on the joint actions $\mathcal{A}$).[26] To obtain the recursive representation of the problem, we first define the following notations. Let $\Delta(\mathcal{A})$ be the set of all probability distributions on the Borel sets on $\mathcal{A}$. For a given subset $\mathcal{W}$ of $\mathbb{R}^2$, a *stochastic (recursive) mechanism* is defined to be a pair of mappings $(P, w) : \Theta \ni \theta \mapsto (P(\cdot|\theta), w(\theta)) \in \Delta(\mathcal{A}) \times \mathcal{W}$. We denote by $\mathcal{P} := (\Delta(\mathcal{A}))^{\Theta}$ the set of stochastic current allocations.

The set $\mathcal{W} \subseteq \mathbb{R}^2$ of PPE payoffs associated with stochastic mechanisms has the following recursive structure. Defining the mapping $B : \mathcal{P}(\mathbb{R}^2) \to \mathcal{P}(\mathbb{R}^2)$ by the set of payoffs generated by an admissible stochastic mechanism (where admissibility is defined as in the main text), the set of PPE payoffs is given by the largest bounded fixed point of the operator B. Formally, for any $W \in \mathcal{P}(\mathbb{R}^2)$, we define $B(W) \in \mathcal{P}(\mathbb{R}^2)$ by

$$B(W) := \left\{ \left(\mathbb{E}_\theta \left[\int (1-\delta) u_i(a, \theta_i) dP(a|\theta) + \delta w_i(\theta) \right] \right)_{i \in N} \middle| (P, w) \text{ is admissible w.r.t. } W \right\}.$$

It can be shown that the monotone mapping B preserves compactness and that the PPE value set $\mathcal{W} = B(\mathcal{W})$ is a compact convex set.

The frontier of the PPE value set (associated with stochastic mechanisms) has the recursive representation with player 2's promised value being a state variable. Define $V(v)$ as the highest value delivered to player 1 given that player 2's promised value is $v \in \mathcal{W}_2$ by stochastic mechanisms. The value function $V : \mathcal{W}_2 \to \mathcal{W}_1$ is characterized as follows.

$$V(v) = \max_{(P,w) \in \mathcal{P} \times \mathcal{W}_2} \mathbb{E}_\theta \left[\int (1-\delta) u_1(a, \theta_1) dP(a|\theta) + \delta V(w(\theta)) \right] \quad \text{(OPT-S)}$$

subject to

$$\mathbb{E}_\theta \left[\int (1-\delta) u_2(a, \theta_2) dP(a|\theta) + \delta w(\theta) \right] = v; \quad \text{(PK-S)}$$

$$\begin{aligned} &\mathbb{E}_{\theta_{-1}} \left[\int (1-\delta) u_1(a, \theta_1) dP(a|\theta) + \delta V(w(\theta)) \right] \\ &\quad \geq \mathbb{E}_{\theta_{-1}} \left[\int (1-\delta) u_1(a, \theta_1) dP(a|\hat{\theta}_1, \theta_2) + \delta V(w(\hat{\theta}_1, \theta_2)) \right] \text{ for all } \theta_1, \hat{\theta}_1 \in \Theta_1; \text{ and} \end{aligned} \quad \text{(IC1-S)}$$

$$\begin{aligned} &\mathbb{E}_{\theta_{-2}} \left[\int (1-\delta) u_2(a, \theta_2) dP(a|\theta) + \delta w(\theta) \right] \\ &\quad \geq \mathbb{E}_{\theta_{-2}} \left[\int (1-\delta) u_2(a, \theta_2) dP(a|\theta_1, \hat{\theta}_2) + \delta w(\theta_1, \hat{\theta}_2) \right] \text{ for all } \theta_2, \hat{\theta}_2 \in \Theta_2. \end{aligned} \quad \text{(IC2-S)}$$

[26] Myerson [23] studies stochastic mechanisms in a static Bayesian collective decision problem. Goltsman et al. [14] and Kováč and Mylovanov [19] consider stochastic mechanisms in static mechanism design problems without side-payments in the contexts of cheap talk communication and delegation problems, respectively.

The renegotiation-proof value function V is characterized by the Bellman equation (OPT-S) where we replace $\mathcal{W}_2$ with $\mathcal{W}_2^* = [\underline{v}_2^P, \overline{v}_2]$. Note that, by abusing the notation, we denote by V both of the (full-commitment) PPE and renegotiation-proof PPE value functions.

A.1 Renegotiation-proof PPE

We use the similar argument to Abreu, Pearce, and Stacchetti [1, Theorem 6] to show that the largest bounded fixed point of B^* is non-decreasing in the discount factor.[27] Fix a discount factor $\delta \in (0,1)$ and the largest bounded fixed point $\mathcal{W}^*(\delta)$ of $B_\delta^* := B^*$. Fix $v \in \mathcal{W}^*(\delta)$ (note that $v_2 \in \mathcal{W}_2^*(\delta) = [\underline{v}_2, \overline{v}_2]$). Choose an admissible pair (a^δ, w^δ) that delivers v. Then, for a new discount factor $\delta' \in [\delta, 1)$, we consider the following continuation value function of player $i \in N$:

$$w_i'(\cdot) = \frac{\delta}{1-\delta}\frac{1-\delta'}{\delta'}w_i^\delta(\cdot) + \frac{\delta'-\delta}{(1-\delta)\delta'}v_i(\in [\underline{v}_i^P, \overline{v}_i]).$$

Now, observe that

$$\mathbb{E}_\theta\left[(1-\delta')u_i(a^\delta(\theta),\theta_i) + \delta' w_i'(\theta)\right] = \frac{1-\delta'}{1-\delta}\mathbb{E}_\theta\left[(1-\delta)u_i(a^\delta(\theta),\theta_i) + \delta w^\delta(\theta)\right] + \frac{\delta'-\delta}{1-\delta}v_i = v_i.$$

Then, it can be seen that the mechanism (a^δ, w') is IC and delivers the expected utility v when the discount factor is δ'. Hence, $\mathcal{W}^*(\delta) \subseteq B_{\delta'}^*(\mathcal{W}^*(\delta))$. This implies that $v \in \mathcal{W}^*(\delta')$, and hence $\mathcal{W}^*(\delta) \subseteq \mathcal{W}^*(\delta')$. Then, we obtain $V_\delta^{\mathrm{SB}}(v_2) \leq V_{\delta'}^{\mathrm{SB}}(v_2)$. Since $v_2 \in [\underline{v}_2^P, \overline{v}_2]$ is arbitrary, we obtain the desired result.

A.2 Approximate Efficiency

Proof of Proposition 1 (Approximate Efficiency). We argue that it is sufficient to prove that the RP-PPE value function V_δ^{SB} approaches point-wise the first-best value function V^{FB} as δ approaches unity. Indeed, since V_δ^{SB} monotonically converges pointwise to V^{FB} on the compact set $\mathcal{W}_2^*$, Dini's theorem (see, for example, Aliprantis and Border [2, Theorem 2.66]) implies that the convergence is uniform. We also make the following three remarks. First, since $V_\delta^{\mathrm{SB}}(\cdot)(\leq V^{\mathrm{FB}}(\cdot))$ is non-decreasing in δ, it follows that $V_\delta^{\mathrm{SB}}(\cdot)$ converges uniformly to a concave function. Thus, we show that it is V^{FB}. Second, since the limit function and the first-best value functions are continuous, it is sufficient to show that the point-wise convergence is almost everywhere. Third, the first-best value $V^{\mathrm{FB}}(v)$ cannot exactly be attained for any particular $\delta \in (0,1)$ when $v \in (\underline{v}_2^P, \overline{v}_2)$. This follows because that would call for the decision $a^{\mathrm{FB}}(\cdot|v)$ to be taken in every period and after every history, so that future decisions would not respond to current announcements. Hence, truth-telling

[27] Since the PPE value set is convex, it immediately follows from Abreu, Pearce, and Stacchetti [1, Theorem 6] that the (full-commitment) PPE value function is monotonically non-decreasing in δ.

would have to be a static best response for the players. This cannot be attained in an incentive compatible manner.

Now, we show that the RP-PPE value function V_δ^{SB} approaches point-wise the first-best Pareto frontier V^{FB} as δ tends to 1. If $v = \underline{v}_2^P$ ($v = \overline{v}_2$), then player 1's (2's) best action $a_1(\theta_1)$ ($a_2(\theta_2)$) is always implemented, and hence $V^{\mathrm{FB}}(v) = V_\delta^{\mathrm{SB}}(v)$ at such v irrespective of the discount factor δ. Thus, fix $v \in (\underline{v}_2^P, \overline{v}_2)$. Suppose to the contrary that $V_\delta^{\mathrm{SB}}(v) \leq \lim_{\delta \to 1} V_\delta^{\mathrm{SB}}(v) < V^{\mathrm{FB}}(v)$ for all $\delta \in (0,1)$.

Recall that the Lagrangian of the first-best problem (FB) (when player 2's promised value is v) is $\mathcal{L} = \mathbb{E}_\theta\left[\sum_{i=1}^2 \gamma_i u_i(a(\theta), \theta_i)\right] - \gamma_2 v$. With slight abuse of notation, the static first-best allocation $a^{\mathrm{FB}}(\cdot) = a^{\mathrm{FB}}(\cdot|v)$ is uniquely given by $a^{\mathrm{FB}}(\theta|v) := \operatorname{argmax}_{a \in \mathcal{A}} \sum_{i=1}^2 \gamma_i u_i(a, \theta_i)$.

Consider the following auxiliary continuation value function $w(\cdot) = w(\cdot|\delta, v)$ which replicates the expected payments of the expected externality mechanism proposed by Arrow [3] and d'Aspremont and Gérard-Varet [9]:

$$w(\theta_1, \hat{\theta}_2|\delta, v) := \left(\frac{1-\delta}{\delta}\right)\left[\xi_2(\hat{\theta}_2|v) - \mathbb{E}_{\theta_2}\left[\xi_2(\theta_2|v)\right]\right] + \mathbb{E}_\theta\left[u_2(a^{\mathrm{FB}}(\theta|v), \theta_2)\right], \tag{A.1}$$

where $\xi_2(\theta_2|v)$ is defined by $\xi_2(\theta_2|v) := \mathbb{E}_{\theta_1}\left[\frac{\gamma_1}{\gamma_2} u_1(a^{\mathrm{FB}}(\theta_1, \theta_2|v), \theta_1)\right]$. Note that w does not depend on player 1's announcement.

Player 2's IC constraint, if a^{FB} is implemented and he faces w as a continuation value function, is

$$\theta_2 \in \operatorname*{argmax}_{\hat{\theta}_2 \in \Theta_2} \mathbb{E}_{\theta_1}\left[(1-\delta)u_2(a^{\mathrm{FB}}(\hat{\theta}_2, \theta_1), \theta_2) + \delta w(\hat{\theta}_2, \theta_1)\right] \text{ for each } \theta_2 \in \Theta_2.$$

Since it is only the first term $\xi_2(\hat{\theta}_2)$ that depends on player 2's announcement $\hat{\theta}_2$ in Expression (A.1), we obtain, for each $\theta_2 \in \Theta_2$,

$$\begin{aligned} &\operatorname*{argmax}_{\hat{\theta}_2 \in \Theta_2} \mathbb{E}_{\theta_1}\left[(1-\delta)u_2(a^{\mathrm{FB}}(\hat{\theta}_2, \theta_1), \theta_2) + \delta w(\hat{\theta}_2, \theta_1)\right] \\ =&\operatorname*{argmax}_{\hat{\theta}_2 \in \Theta_2} (1-\delta)\mathbb{E}_{\theta_1}\left[\sum_{j=1}^2 \gamma_j u_j(a^{\mathrm{FB}}(\theta_1, \hat{\theta}_2), \theta_j)\right]. \end{aligned}$$

It follows that the announcement $\hat{\theta}_2 = \theta_2$ is optimal for player 2.

Next, consider the payoff profile associated with the constructed mechanism $(a^{\mathrm{FB}}, (V_\delta^{\mathrm{SB}}(w(\cdot)), w))$:

$$\left(\mathbb{E}_\theta\left[(1-\delta)u_1(a^{\mathrm{FB}}(\theta), \theta_1) + \delta V_\delta^{\mathrm{SB}}(w(\theta))\right], v\right).$$

We show that player 1's payoff is arbitrary approximated as δ tends to unity in an incentive compatible way. Indeed, recalling that w depends only on θ_2, we re-define player 1's continuation value function as

$$w_1(\theta) := V_\delta^{\mathrm{SB}}(w(\theta_2)) + \frac{1-\delta}{\delta}\frac{\gamma_2}{\gamma_1}\left(u_2(a^{\mathrm{FB}}(\theta), \theta_2) - \max_{\theta \in \Theta} u_2(a^{\mathrm{FB}}(\theta), \theta_2)\right).$$

Then, it can be seen that the mechanism $(a^{\mathrm{FB}}, (w_1, w_2(= w)))$ approximates the above payoff profile as $\delta \to 1$ (note that (w_1, w_2) are self-enforceable when δ is close to unity).

Now, this implies that

$$V_\delta^{\mathrm{SB}}(v) \geq V^{\mathrm{FB}}(v) - \frac{\delta}{1-\delta}\left(V_\delta^{\mathrm{SB}}(v) - \mathbb{E}_\theta\left[V_\delta^{\mathrm{SB}}(w(\theta))\right]\right) - (1-\delta)\frac{\gamma_2}{\gamma_1}\left(\max_{\theta\in\Theta} u_2(a^{\mathrm{FB}}(\theta), \theta_2) - v\right).$$

Hence, we show that the second and third terms of the right-hand side of the above expression go to 0 as δ tends to one. That is, as in Lockwood and Thomas [20, Theorem] and Thomas and Worrall [30, Proposition 4], we show that the efficiency loss $\left(V_\delta^{\mathrm{SB}}(v) - \mathbb{E}_\theta\left[V_\delta^{\mathrm{SB}}(w(\theta))\right]\right)$ goes to 0 faster than δ goes to 1:

$$\lim_{\delta\to 1}\frac{\delta}{1-\delta}\left(V_\delta^{\mathrm{SB}}(v) - \mathbb{E}_\theta\left[V_\delta^{\mathrm{SB}}(w(\theta))\right]\right) = 0.$$

To see this, we let $V_1 = \lim_{\delta\to 1} V_\delta^{\mathrm{SB}}$. Note that V_1 is concave and continuous and is differentiable almost everywhere. Without loss of generality, we can assume that the derivative of V_1 at v, $\frac{d}{dw}V_1(v)$, exists. Now, we have

$$\begin{aligned}
\frac{\delta}{1-\delta}\mathbb{E}_\theta\left[V_\delta^{\mathrm{SB}}(v) - V_\delta^{\mathrm{SB}}(w(\theta))\right] &\leq \frac{\delta}{1-\delta}\mathbb{E}_\theta\left[\frac{d}{dw}V_\delta^{\mathrm{SB}}(w(\theta))(v - w(\theta))\right] \\
&= \mathbb{E}_\theta\left[\frac{d}{dw}V_\delta^{\mathrm{SB}}(w(\theta))\left(\xi_2(\theta_2) - \mathbb{E}_{\theta_2}\left[\xi_2(\theta_2)\right]\right)\right] \\
&= \mathbb{E}_\theta\left[\left(\frac{d}{dw}V_\delta^{\mathrm{SB}}(w(\theta)) - \frac{d}{dw}V_1(v)\right)\left(\xi_2(\theta_2) - \mathbb{E}_{\theta_2}\left[\xi_2(\theta_2)\right]\right)\right] \\
&\leq \kappa\mathbb{E}_\theta\left[\left|\frac{d}{dw}V_\delta^{\mathrm{SB}}(w(\theta)) - \frac{d}{dw}V_1(v)\right|\right],
\end{aligned}$$

where $\frac{d}{dw}V_\delta^{\mathrm{SB}}$ denotes the one-sided derivative and $\kappa := \max_{\hat{\theta}_2\in\Theta_2}\left|\xi_2(\hat{\theta}_2) - \mathbb{E}_{\theta_2}\left[\xi_2(\theta_2)\right]\right| < +\infty$.[28] Now, it can be verified that the right-hand side of the above expression goes to 0 as $\delta \to 1$. □

A.3 The Dynamics of Collective Decision Making

A.3.1 Validity of the First-Order Approach

Next, we provide a condition which validates the first-order approach. For a given stochastic recursive mechanism (P, w), we define

$$\widetilde{U}_i\left(\theta_i, \hat{\theta}_i \middle| P, w\right) := \begin{cases} \mathbb{E}_{\theta_{-1}}\left[\int (1-\delta)u_1(a, \theta_1)dP(a|\hat{\theta}_1, \theta_2) + \delta V(w(\hat{\theta}_1, \theta_2))\right] & \text{if } i = 1 \\ \mathbb{E}_{\theta_{-2}}\left[\int (1-\delta)u_2(a, \theta_2)dP(a|\theta_1, \hat{\theta}_2) + \delta w(\theta_1, \hat{\theta}_2)\right] & \text{if } i = 2 \end{cases}.$$

[28]We note that it can be shown that a concave function V_δ^{SB} is differentiable (and thus continuously differentiable) everywhere on $(\underline{v}_2^P, \overline{v}_2)$.

Thus, $\widetilde{U}_i(\theta_i^{(k_i)}, \theta_i^{(k_i')}|P, w)$ is the expected utility of player i with type $\theta_i^{(k_i)}$ who reports $\theta_i^{(k_i')}$ under the given stochastic mechanism. Then, agent i's expected utility of the mechanism at θ_i is given by $\widetilde{U}_i(\theta_i|P, w) := \widetilde{U}_i(\theta_i, \theta_i|P, w)$. We often omit the stochastic mechanism (P, w) in question when it is clear from the context.

We show that the following expected monotonicity condition on marginal utility (MONi-MU-S) guarantees that the global IC constraints are replaced by the local (upward and downward) IC constraints. To that end, as in the main text, we let each $\Theta_i = \{\theta_i^{(k_i)}\}_{k_i=1}^{m_i}$ be in $\mathbb{R}$ such that $\theta_i^{(1)} < \theta_i^{(2)} < \cdots < \theta_i^{(m_i)}$. We also assume that u_i is continuously differentiable with respect to θ_i on an interval containing Θ_i.

$$\mathbb{E}_{\theta_{-i}}\left[\int \frac{\partial}{\partial \theta_i} u_i\left(a, \tau\right) dP\left(a|\theta_i, \theta_{-i}\right)\right] \text{ is non-decreasing in } \theta_i \text{ at every } \tau, \tag{MONi-MU-S}$$

We also show that, in the following standard environments (quadratic single-peaked preferences, CARA preferences, and multiplicative preferences), the global incentive constraints (ICi-S) can be replaced with the local incentive constraints (ICi-UP-S) and (ICi-DW-S) and the above expected monotonicity condition (MONi-MU-S). Moreover, if each player's utility function is quadratic, then the monotonicity condition (MONi-MU-S) can be replaced by the expected monotonicity condition on the current allocation.

Proposition A.1 (Incentive Compatibility)**.** *A given stochastic mechanism (P, w) satisfies the (i's) IC constraints if it satisfies (MONi-MU-S) and the following local IC constraints:*

$$\widetilde{U}_i\left(\theta_i^{(k_i)}\middle| P, w\right) \geq \widetilde{U}_i\left(\theta_i^{(k_i)}, \theta_i^{(k_i+1)}\middle| P, w\right) \text{ for each } k_i \in \{1, \ldots, m_i - 1\}; \text{ and} \tag{ICi-UP-S}$$

$$\widetilde{U}_i\left(\theta_i^{(k_i)}\middle| P, w\right) \geq \widetilde{U}_i\left(\theta_i^{(k_i)}, \theta_i^{(k_i-1)}\middle| P, w\right) \text{ for each } k_i \in \{2, \ldots, m_i\}. \tag{ICi-DW-S}$$

In either of the following standard environments, the converse is also true: (i) each agent's preference shocks are multiplicative (i.e., $u_i(a, \theta_i) = \theta_i \tilde{u}_i(a)$); (ii) it is of the CARA form in the insurance setting (i.e., $u_i(a, \theta_i) = -\exp(-r_i(\theta_i + (-1)^i a))$ with $r_i > 0$); or (iii) it is quadratic (i.e., $u_i(a, \theta_i) = -(a - \theta_i)^2$).

Moreover, under the quadratic case, (MONi-MU-S) can be replaced with the following monotonicity condition:

$$\mathbb{E}_{\theta_{-i}}\left[\int a dP\left(a|\theta_i, \theta_{-i}\right)\right] \text{ is non-decreasing in } \theta_i \in \Theta_i. \tag{MONi-AL-S}$$

Proof of Proposition A.1. The proof consists of two steps.
Step 1. Fix $(k_i, k_i') \in \{1, 2, \ldots, m_i\}^2$ with $k_i > k_i'$. First, observe that (ICi-UP-S) implies that, for all ℓ_i with $k_i' \leq \ell_i < k_i$,

$$\begin{aligned}\widetilde{U}_i(\theta^{(\ell_i)}) &\geq \widetilde{U}_i(\theta^{(\ell_i+1)}) + \widetilde{U}_i(\theta_i^{(\ell_i)}, \theta_i^{(\ell_i+1)}) - \widetilde{U}_i(\theta_i^{(\ell_i+1)}, \theta_i^{(\ell_i+1)}) \\ &= \widetilde{U}_i(\theta^{(\ell_i+1)}) - (1 - \delta) \int_{\theta_i^{(\ell_i)}}^{\theta_i^{(\ell_i+1)}} \mathbb{E}_{\theta_{-i}}\left[\int \frac{\partial}{\partial \theta_i} u_i(a, \tau) dP(a|\theta_i^{(\ell_i+1)}, \theta_{-i})\right] d\tau.\end{aligned}$$

By summing the above expression up for each ℓ_i with $k_i' \leq \ell_i \leq k_i - 1$, we have

$$(1-\delta)\sum_{\ell_i=k_i'}^{k_i-1}\int_{\theta_i^{(\ell_i)}}^{\theta_i^{(\ell_i+1)}} \mathbb{E}_{\theta_{-i}}\left[\int \frac{\partial}{\partial\theta_i}u_i(a,\tau)dP(a|\theta_i^{(\ell_i+1)},\theta_{-i})\right]d\tau \geq \widetilde{U}_i(\theta^{(k_i)}) - \widetilde{U}_i(\theta^{(k_i')}).$$

Now, it follows from the expected monotonicity condition (MONi-MU-S) that

$$\widetilde{U}_i(\theta^{(k_i')}) \geq \widetilde{U}_i(\theta^{(k_i)}) - (1-\delta)\mathbb{E}_{\theta_{-i}}\left[\int \left\{u_i(a,\theta_i^{(k_i)}) - u_i(a,\theta_i^{(k_i')})\right\}dP(a|\theta_i^{(k_i)},\theta_{-i})\right] = \widetilde{U}_i(\theta_i^{(k_i')},\theta_i^{(k_i)}).$$

Second, observe also that (ICi-DW-S) implies that, for all ℓ_i with $k_i' \leq \ell_i < k_i$,

$$\begin{aligned}\widetilde{U}_i(\theta^{(\ell_i+1)}) &\geq \widetilde{U}_i(\theta^{(\ell_i)}) + \widetilde{U}_i(\theta_i^{(\ell_i+1)},\theta_i^{(\ell_i)}) - \widetilde{U}_i(\theta_i^{(\ell_i)},\theta_i^{(\ell_i)}) \\ &= \widetilde{U}_i(\theta^{(\ell_i)}) + (1-\delta)\int_{\theta_i^{(\ell_i)}}^{\theta_i^{(\ell_i+1)}} \mathbb{E}_{\theta_{-i}}\left[\int \frac{\partial}{\partial\theta_i}u_i(a,\tau)dP(a|\theta_i^{(\ell_i)},\theta_{-i})\right]d\tau.\end{aligned}$$

By summing the above expression up for each ℓ_i with $k_i' \leq \ell_i \leq k_i - 1$, we have

$$\widetilde{U}_i(\theta^{(k_i)}) - \widetilde{U}_i(\theta^{(k_i')}) \geq (1-\delta)\sum_{\ell_i=k_i'}^{k_i-1}\int_{\theta_i^{(\ell_i)}}^{\theta_i^{(\ell_i+1)}} \mathbb{E}_{\theta_{-i}}\left[\int \frac{\partial}{\partial\theta_i}u_i(a,\tau)dP(a|\theta_i^{(\ell_i)},\theta_{-i})\right]d\tau.$$

Thus, it follows from the expected monotonicity condition (MONi-MU-S) that

$$\widetilde{U}_i(\theta^{(k_i)}) \geq \widetilde{U}_i(\theta^{(k_i')}) + (1-\delta)\mathbb{E}_{\theta_{-i}}\left[\int \left\{u_i(a,\theta_i^{(k_i)}) - u_i(a,\theta_i^{(k_i')})\right\}dP(a|\theta_i^{(k_i')},\theta_{-i})\right] = \widetilde{U}_i(\theta_i^{(k_i)},\theta_i^{(k_i')}).$$

Step 2. First, suppose that each agent has a multiplicative preference shock, i.e., $u_i(a,\theta_i) = \theta_i\tilde{u}_i(a)$. Observing that $\frac{\partial}{\partial\theta_i}u_i(a,\theta_i) = \tilde{u}_i(a)$, if the mechanism (P,w) is IC then for any $\theta_i, \hat{\theta}_i \in \Theta_i$,

$$0 \leq \left(\theta_i - \hat{\theta}_i\right)\left\{\mathbb{E}_{\theta_{-i}}\left[\int \frac{\partial}{\partial\theta_i}u_i\left(a,\tau\right)dP(a|\theta_i,\theta_{-i})\right] - \mathbb{E}_{\theta_{-i}}\left[\int \frac{\partial}{\partial\theta_i}u_i\left(a,\tau\right)dP(a|\hat{\theta}_i,\theta_{-i})\right]\right\}.$$

Hence, we obtain the expected monotonicity condition (MONi-MU-S).

Second, suppose that each player's utility function is of the CARA form in the insurance setting: $u_i(a,\theta_i) = -\exp(-r_i(\theta_i + (-1)^i a))$ with $r_i > 0$. Then, we can treat his utility function as if it were $\theta_i\tilde{u}_i(a)$ with $\tilde{u}_i(a) = k_i(-1)\exp(-r_i(-1)^i a)$ and his type distribution is given by $\tilde{f}_i(\theta_i) = \frac{\exp(-r_i\theta_i)}{k_i\theta_i}f_i(\theta_i)(>0)$, where $k_i = \int_{\underline{\theta}_i}^{\overline{\theta}_i}\frac{\exp(-r_i\theta_i)}{\theta_i}f_i(\theta_i)d\theta_i(>0)$.

Third, suppose that each agent's utility is quadratic, i.e., $u_i(a,\theta_i) = -(a-\theta_i)^2$. If the mechanism (P,w) is IC, then for any $\theta_i, \hat{\theta}_i \in \Theta_i$,

$$0 \leq 2\left(\theta_i - \hat{\theta}_i\right)\left\{\mathbb{E}_{\theta_{-i}}\left[\int adP(a|\theta_i,\theta_{-i})\right] - \mathbb{E}_{\theta_{-i}}\left[\int adP(a|\hat{\theta}_i,\theta_{-i})\right]\right\}.$$

Hence, we obtain the expected monotonicity condition on the current allocation (MONi-AL-S). Now, noting that $\frac{\partial}{\partial\theta_i}u_i(a,\theta_i) = 2(a-\theta_i)$, (MON$i$-MU-S) and (MON$i$-AL-S) are equivalent. □

We make two remarks. First, Lemma A.1 suggests that if the optimal mechanism satisfies (MONi-MU) in a particular problem then the first-order approach is valid in that problem.

Second, for deterministic mechanisms, the expected monotonicity condition on marginal utilities is:

$$\mathbb{E}_{\theta_{-i}}\left[\frac{\partial}{\partial\theta_i}u_i(a(\theta_i,\theta_{-i}),\tau)\right] \text{ is non-decreasing in } \theta_i \in \Theta_i \text{ at every } \tau \in \Theta_i. \qquad \text{(MON}i\text{-MU)}$$

Likewise, the expected monotonicity condition on the current allocation is:

$$\mathbb{E}_{\theta_{-i}}\left[a(\theta_i,\theta_{-i})\right] \text{ is non-decreasing in } \theta_i \in \Theta_i. \qquad \text{(MON}i\text{-AL)}$$

A.3.2 The Lagrangian Associated with the Full Problem

We have taken the first-order approach throughout the main text. Here we sketch how our results hold even when we keep track of every incentive constraint. Also, each $\Theta_i = \{\theta_i^{(k_i)}\}_{k_i=1}^{m_i}$ can simply be a finite set. For every $\theta_i^{(k_i)}, \theta_i^{(k_i')} \in \Theta_i$, the incentive constraint $U_i(\theta_i^{(k_i)}|a,w) \geq U_i(\theta_i^{(k_i)}, \theta_i^{(k_i')}|a,w)$ can be written as follows:

$$U_i(\theta_i^{(k_i)}) - U_i(\theta_i^{(k_i')}) - (1-\delta)\mathbb{E}_{\theta_{-i}}\left[u_i(a(\theta_i^{(k_i')},\theta_{-i}),\theta_i^{(k_i)}) - u_i(a(\theta_i^{(k_i')},\theta_{-i}),\theta_i^{(k_i')})\right] \geq 0.$$

We formulate the Lagrangian $\hat{\mathcal{L}}(a,w|\lambda,\gamma,v)$ associated with the full problem (OPT) as:

$$\begin{aligned}\hat{\mathcal{L}}(a,w|\lambda,\gamma,v) &= \sum_{i=1}^{2}\gamma_i\sum_{k_i=1}^{m_i}f_i(\theta_i^{(k_i)})U_i(\theta_i^{(k_i)}) + \sum_{i=1}^{2}\gamma_i\sum_{k_i=1}^{m_i}\sum_{k_i'=1}^{m_i}\tilde{\lambda}^i_{k_i,k_i'}\left\{U_i(\theta_i^{(k_i)}) - U_i(\theta_i^{(k_i')})\right\}\\ &+ (1-\delta)\sum_{i=1}^{2}\gamma_i\sum_{k_i=1}^{m_i}\sum_{k_i'=1}^{m_i}\tilde{\lambda}^i_{k_i,k_i'}\mathbb{E}_{\theta_{-i}}\left[u_i(a(\theta_i^{(k_i')},\theta_{-i}),\theta_i^{(k_i')}) - u_i(a(\theta_i^{(k_i')},\theta_{-i}),\theta_i^{(k_i)})\right].\end{aligned}$$

If the Lagrange multiplier of the non-local incentive constraints are zero, then the above Lagrangian clearly reduces to the one (LAG) associated with the first-order approach.

It can be verified that the Lagrangian is written as follows.

$$\begin{aligned}\hat{\mathcal{L}}(a,w|\lambda,\gamma,v) &= \sum_{i=1}^{2}\gamma_i\sum_{k_i=1}^{m_i}f_i(\theta_i^{(k_i)})U_i(\theta_i^{(k_i)})\left(1+\frac{\sum_{k_i'=1}^{m_i}(\tilde{\lambda}^i_{k_i,k_i'} - \tilde{\lambda}^i_{k_i',k_i})}{f_i(\theta_i^{(k_i)})}\right)\\ &+ (1-\delta)\sum_{i=1}^{2}\gamma_i\sum_{k_i=1}^{m_i}f_i(\theta_i^{(k_i)})\sum_{k_i'=1}^{m_i}\frac{\tilde{\lambda}^i_{k_i,k_i'}}{f_i(\theta_i^{(k_i)})}\mathbb{E}_{\theta_{-i}}\left[u_i(a(\theta_i^{(k_i')},\theta_{-i}),\theta_i^{(k_i')}) - u_i(a(\theta_i^{(k_i')},\theta_{-i}),\theta_i^{(k_i)})\right].\end{aligned}$$

Equivalently, we have:

$$\hat{\mathcal{L}}(a,w|\lambda,\gamma,v) = \sum_{i=1}^{2}\gamma_i \sum_{k_i=1}^{m_i} f_i(\theta_i^{(k_i)})U_i(\theta_i^{(k_i)})\left(1+\frac{\sum_{k_i'=1}^{m_i}(\tilde{\lambda}^i_{k_i,k_i'}-\tilde{\lambda}^i_{k_i',k_i})}{f_i(\theta_i^{(k_i)})}\right)$$

$$+(1-\delta)\sum_{i=1}^{2}\gamma_i\sum_{k_i=1}^{m_i} f_i(\theta_i^{(k_i)})\sum_{k_i'=1}^{m_i}\frac{(\tilde{\lambda}^i_{k_i,k_i'}-\tilde{\lambda}^i_{k_i',k_i})}{f_i(\theta_i^{(k_i)})}\mathbb{E}_{\theta_{-i}}\left[u_i(a(\theta_i^{(k_i')},\theta_{-i}),\theta_i^{(k_i')})-u_i(a(\theta_i^{(k_i')},\theta_{-i}),\theta_i^{(k_i)})\right]$$

$$+(1-\delta)\sum_{i=1}^{2}\gamma_i\sum_{k_i=1}^{m_i} f_i(\theta_i^{(k_i)})\sum_{k_i'=1}^{m_i}\frac{\tilde{\lambda}^i_{k_i',k_i}}{f_i(\theta_i^{(k_i)})}\mathbb{E}_{\theta_{-i}}\left[u_i(a(\theta_i^{(k_i')},\theta_{-i}),\theta_i^{(k_i')})-u_i(a(\theta_i^{(k_i')},\theta_{-i}),\theta_i^{(k_i)})\right].$$

Note that if the Lagrange multiplier of the non-local incentive constraints are zero, then we obtain the Lagrangian (LAG*) associated with the first-order approach.

Moreover, these expressions yield the agent's (general) virtual utility (which also incorporates all the non-local IC constraints). We denote by $\gamma_i\tilde{u}_i(a,\theta_i^{(k_i)})$ the agent i's instantaneous virtual utility at $\theta_i^{(k_i)}$ with respect to γ_i and λ_i as follows:

$$\gamma_i\tilde{u}_i\left(a,\theta_i^{(k_i)}\right):=\gamma_i\left\{u_i\left(a,\theta_i^{(k_i)}\right)\left(1+\frac{\sum_{k_i'=1}^{m_i}\tilde{\lambda}^i_{k_i,k_i'}}{f_i(\theta_i^{(k_i)})}\right)-\frac{\sum_{k_i'=1}^{m_i}u_i\left(a,\theta_i^{(k_i')}\right)\tilde{\lambda}^i_{k_i',k_i}}{f_i(\theta_i^{(k_i)})}\right\}$$

$$=\gamma_i\left\{u_i\left(a,\theta_i^{(k_i)}\right)\left(1+\frac{\sum_{k_i'=1}^{m_i}(\tilde{\lambda}^i_{k_i,k_i'}-\tilde{\lambda}^i_{k_i',k_i})}{f_i(\theta_i^{(k_i)})}\right)+\sum_{k_i=1}^{m_i}\frac{\tilde{\lambda}^i_{k_i',k_i}}{f_i(\theta_i^{(k_i')})}\mathbb{E}_{\theta_{-i}}\left[u_i(a,\theta_i^{(k_i)})-u_i(a,\theta_i^{(k_i')})\right]\right\}.$$

Hence, the Lagrangian $\hat{\mathcal{L}}=\hat{\mathcal{L}}(a,w|\lambda,\gamma,v)$ can be expressed as

$$\hat{\mathcal{L}}=\mathbb{E}_\theta\left[\sum_{i=1}^{2}(1-\delta)\gamma_i\tilde{u}_i(a(\theta),\theta_i)\right.$$
$$\left.+\delta\gamma_1 V(w(\theta))\left(1+\frac{\sum_{k_1'=1}^{m_1}(\tilde{\lambda}^1_{k_1,k_1'}-\tilde{\lambda}^1_{k_1',k_1})}{f_1(\theta_1^{(k_1)})}\right)+\delta\gamma_2 w(\theta)\left(1+\frac{\sum_{k_2'=1}^{m_2}(\tilde{\lambda}^2_{k_2,k_2'}-\tilde{\lambda}^2_{k_2',k_2})}{f_2(\theta_2^{(k_2)})}\right)\right].$$

A.3.3 The Long-run Properties

Proof of Proposition 2. We prove the result without taking the first-order approach. Under the first-order approach, one needs to replace $\sum_{k_i'=1}^{m_i}(\tilde{\lambda}^i_{k_i,k_i'}-\tilde{\lambda}^i_{k_i',k_i})$ with $\tilde{\lambda}^{\mathrm{UP}}_{i,k_i}-\tilde{\lambda}^{\mathrm{UP}}_{i,k_i-1}+\tilde{\lambda}^{\mathrm{DW}}_{1,k_i}-\tilde{\lambda}^{\mathrm{DW}}_{i,k_i+1}$ (or assign zero to the Lagrange multipliers of all the non-local IC constraints).

The first-order condition with respect to $w(\theta_1^{(k_1)},\theta_2^{(k_2)})=w(\theta_1^{(k_1)},\theta_2^{(k_2)}|v)$ is:

$$V'(w(\theta_1^{(k_1)},\theta_2^{(k_2)}))\left(1+\frac{\sum_{k_1'=1}^{m_1}(\tilde{\lambda}^1_{k_1,k_1'}-\tilde{\lambda}^1_{k_1',k_1})}{f_1(\theta_1^{(k_1)})}\right)=-\frac{\gamma_2}{\gamma_1}\left(1+\frac{\sum_{k_2'=1}^{m_2}(\tilde{\lambda}^2_{k_2,k_2'}-\tilde{\lambda}^2_{k_2',k_2})}{f_2(\theta_2^{(k_2)})}\right).$$

In other words,

$$\frac{\gamma'_{-i}(\theta_1^{(k_1)},\theta_2^{(k_2)})}{\gamma'_i(\theta_1^{(k_1)},\theta_2^{(k_2)})}=\frac{\gamma_{-i}}{\gamma_i}\frac{1+\frac{\sum_{k'_{-i}=1}^{m_{-i}}(\tilde{\lambda}^{-i}_{k_{-i},k'_{-i}}-\tilde{\lambda}^{-i}_{k'_{-i},k_{-i}})}{f_{-i}(\theta_{-i}^{(k_{-i})})}}{1+\frac{\sum_{k'_i=1}^{m_i}(\tilde{\lambda}^{i}_{k_i,k'_i}-\tilde{\lambda}^{i}_{k'_i,k_i})}{f_i(\theta_i^{(k_i)})}}.$$

Thus, taking expectation on both sides, we obtain

$$\frac{\gamma_i}{\gamma_{-i}}=\mathbb{E}_{\theta_i}\left[\frac{1}{\mathbb{E}_{\theta_{-i}}\left[\frac{\gamma'_{-i}(\theta)}{\gamma'_i(\theta)}\right]}\right]\le\mathbb{E}_\theta\left[\frac{\gamma'_i(\theta)}{\gamma'_{-i}(\theta)}\right],$$

where the last inequality follows from Jensen's inequality. □

Proof of Proposition 3. First, assume that $\frac{\gamma'_i(\theta)}{\gamma'_{-i}(\theta)}\le\frac{\gamma_i}{\gamma_{-i}}$ for all $\theta\in\Theta$. Since Proposition 2 implies $\mathbb{E}_\theta\left[\frac{\gamma'_i(\theta)}{\gamma'_{-i}(\theta)}\right]\ge\frac{\gamma_i}{\gamma_{-i}}$, it follows that $\frac{\gamma'_i(\theta)}{\gamma'_{-i}(\theta)}=\frac{\gamma_i}{\gamma_{-i}}$ for all $\theta\in\Theta$. Player i's relative Pareto weights is constant, which is a contradiction to the statement that full efficiency is not attainable at $v\in(\underline{v}_2^P,\overline{v}_2)$.

Next, assume that $\frac{\gamma'_i(\theta)}{\gamma'_{-i}(\theta)}\ge\frac{\gamma_i}{\gamma_{-i}}$ for all $\theta\in\Theta$. Then, $\frac{\gamma'_{-i}(\theta)}{\gamma'_i(\theta)}\le\frac{\gamma_{-i}}{\gamma_i}$ for all $\theta\in\Theta$. As in the previous argument, this implies that player $-i$'s relative Pareto weights is constant, a contradiction. □

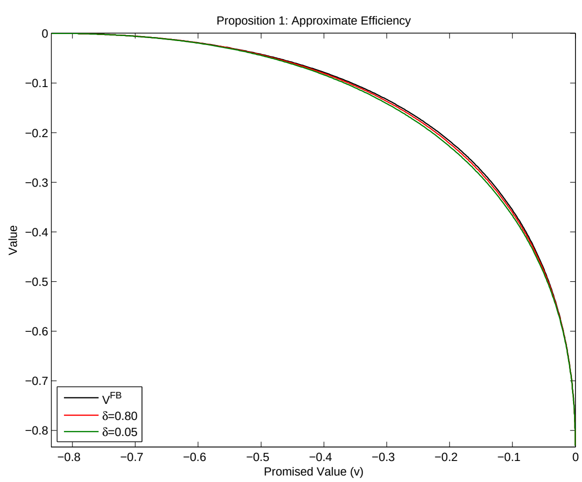
B Appendix B

This appendix documents the results of two sets of numerical simulations: quadratic collective decision making and CARA insurance problems. In both simulations, we solve the relaxed second-best problem (OPT*) without introducing a public randomization device.

B.1 Quadratic Collective Decision Making Problem

Suppose that two (ex ante symmetric) players repeatedly take a joint action. Their utility functions are quadratic $u_i(a,\theta_i) = -(a-\theta_i)^2$ and their types follow the uniform density $f_i(\cdot) = \frac{1}{9}$ on $\Theta_i = \{-1, -0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75, 1\}$. The set of feasible actions is $\mathcal{A} = [-1.1, 1.1]$. Under these parametric settings, we have $\overline{v}_i = \mathbb{E}_\theta\left[u_i(\theta_i,\theta_i)\right] = 0$ and $\underline{v}_i^P = \mathbb{E}_\theta\left[u_i(\theta_{-i},\theta_i)\right] = -0.833333$. Note that the first-best value function is $V^{\text{FB}}(v) = -\left(\sqrt{-v} - \sqrt{-\underline{v}_2^P}\right)^2$. We divide the set of continuation values $[\underline{v}_2^P, \overline{v}_2]$ into 500 grid points by using the Chebyshev nodes.

B.1.1 Approximate Efficiency



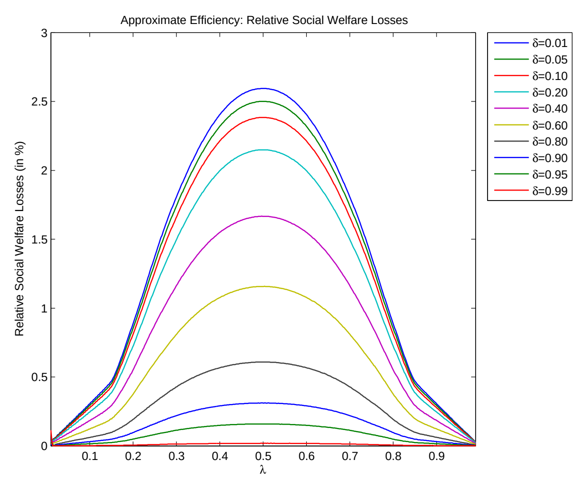


Figure 3: The Value Functions V_δ^{SB} and V^{FB} (Left) and Relative Losses (in %) (Right).

The left panel of Figure 3 depicts the RP-PPE value function V_δ^{SB} as well as the first-best value function V^{FB}. The right panel of Figure 3 depicts the relative social welfare loss (in %) in terms of the weights attached on agents instead of player 2's promised values. To do that, for each given weight $\gamma \in (0,1)$ (we consider 500 uniform grid points), we compare $\max_{v\in\mathcal{W}_2^*} \gamma V^{\text{FB}}(v) + (1-\gamma)v$ (which turns out to be $\gamma(1-\gamma)\underline{v}_2^P$) and $\max_{v\in\mathcal{W}_2^*} \gamma V_\delta^{\text{SB}}(v) + (1-\gamma)v$ (implicit in this computation is the strict concavity of the value function V_δ^{SB}).

For more uneven Pareto weights (for γ closer to the extreme points 0 and 1), the current decision takes into account of an announcement by the agent with higher weight and the continuation values are closer to his best payoff. At the extreme weights, the current decision reflects only one agent's announcement irrespective of the discount factor. On the other hand, for γ close to half, the current decision reflects both players' announcements, and the incentive costs are higher. Moreover, the lower the discount factor is, the higher the incentive costs are.

B.1.2 Optimal Mechanisms

We report the optimal mechanism $(a(\cdot|v), w(\cdot|v))$ for each continuation value $v \in [\underline{v}_2^P, \overline{v}_2]$. Unless otherwise specified, the common discount factor is set to $\delta = 0.4$.

Figure 4 illustrates the continuation value functions $w(\theta_1^{(k_1)}, \theta_2^{(k_2)}|v)$ with player 1's shock $\theta_1^{(k_1)}$ fixed (where $k_1 \in \{1, 5\}$). First, while the continuation value functions are time-invariant for each promised value in the first-best problem, they vary with reported types in the second-best problem. Second, the left panel reports that when player 1's shock is the extreme shock $\theta_1^{(1)}$, player 2's continuation values are above 45° line for a wide range of player 2's promised values. Third, in the right panel, player 2's continuation values $w(\theta_1^{(5)}, \theta_2^{(5)}|v)$ when both players' shocks are the middle one are above (below) the 45° line when his promised value v_2 is above (below) the symmetric point w_0 with $w_0 = V(w_0) \approx 0.21$.

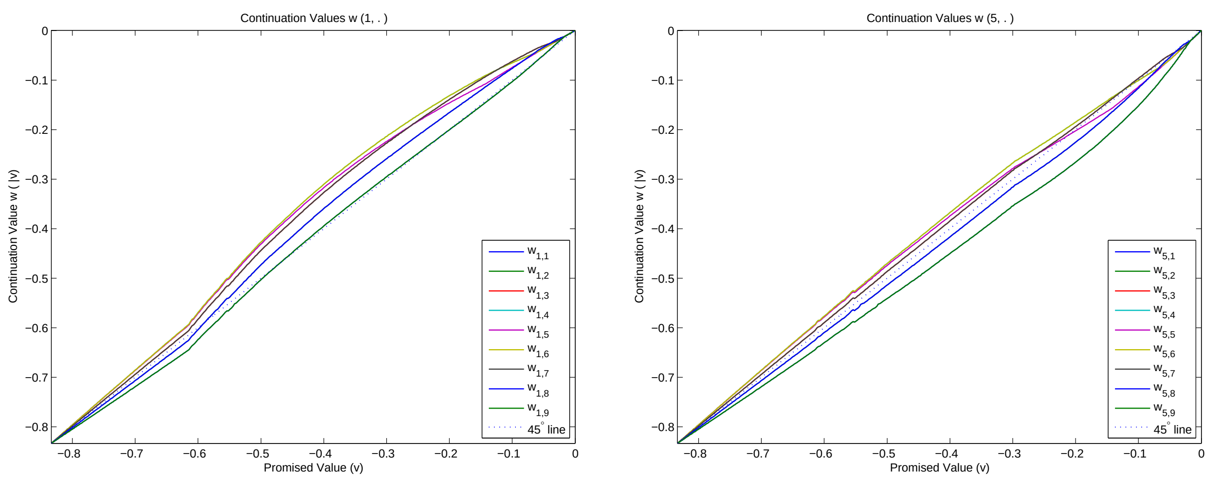


Figure 4: Continuation Values: $w(\theta_1^{(1)}, \cdot|v)$ (Left) and $w(\theta_1^{(5)}, \cdot|v)$ (Right).

Figures 5 and the left panel of Figure 6 illustrate the optimal first-best and second-best current actions $a^{\text{FB}}(\theta|v)$ and $a(\theta|v)$, respectively. We only report $a^{\text{FB}}(\theta_1^{(k_1)}, \cdot|v)$ and $a(\theta_1^{(k_1)}, \cdot|v)$ with $k_1 \in \{1, 3, 5\}$. Given symmetry, the results of numerical simulations exhibit $a^{\text{FB}}(\theta_1^{(10-k_1)}, \cdot|v) = a^{\text{FB}}(\theta_1^{(k_1)}, \cdot|v)$ and $a(\theta_1^{(10-k_1)}, \cdot|v) = a(\theta_1^{(k_1)}, \cdot|v)$ for $k_1 \in \{1, \ldots, 5\}$.

It is to be noted that the second-best current action requires the players to take an action which is lower than -1 (resp. larger than 1) when both players announce the lowest type $\theta_i^{(1)} = -1$ (resp. the highest type $\theta_i^{(9)} = 1$). The right panel of Figure 6 illustrates

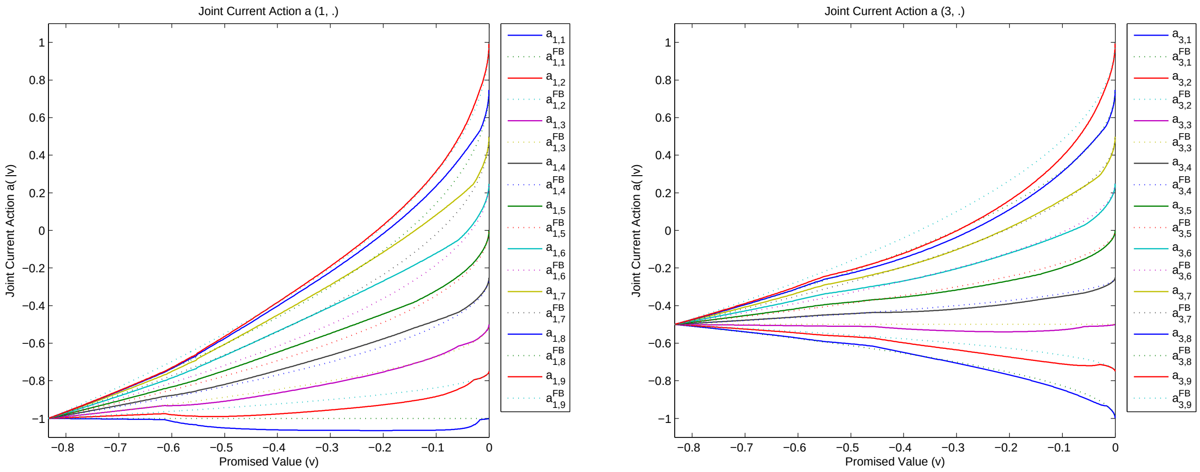


Figure 5: Current Allocations: $a(\theta_1^{(1)}, \cdot|v)$ (Left) and $a(\theta_1^{(3)}, \cdot|v)$ (Right).

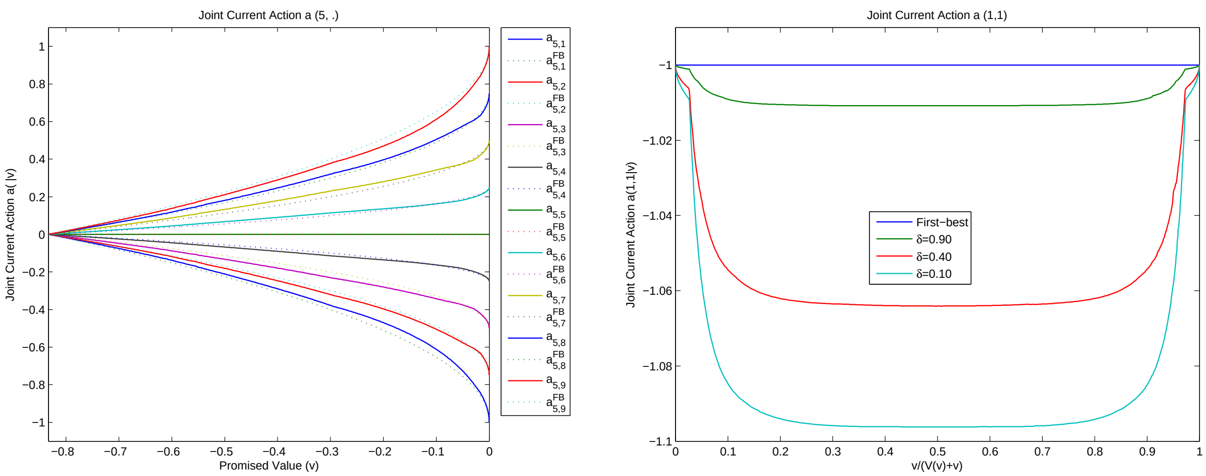


Figure 6: Current Allocations. The left panel depicts $a(\theta_1^{(5)}, \cdot|v)$ while the right panel depicts $a(\theta_1^{(1)}, \theta_2^{(1)}|v)$ ($\delta \in \{0.1, 0.4, 0.9\}$) with x-axis being $v/(V_\delta^{SB}(v) + v)$.

this phenomena, where the x-axis is adjusted to $v/(V_\delta^{SB}(v) + v)$. Also, the degree of such inefficiency increases as the discount factor decreases. Intuitively, the lower δ is, the more the current allocation has to take care of players' incentives. See also the left panel of Figure 5 for the current allocation $a(\theta_1^{(1)}, \theta_2^{(1)}|v)$ for each $v \in \mathcal{W}_2^*$.

Making the current allocation $a(-1, -1)$ be less than the lowest type -1 itself deters each agent from exaggerating his announcement. Consider the second-best joint action when player i's announcement is the lowest one. While making $a(-1, -1)$ less than -1 entails certain costs (harmful to both players), it relaxes the incentive constraints in the following way. If the other player also announces the lowest type -1, then the players will eventually take $a(-1, -1) < -1$. Hence, player i, expecting that the other player might also announce the lowest type -1 with certain probability, is less inclined to exaggerate his announcement by announcing $-1 = \theta_i^{(1)}$ when his true preference shock is $-0.75 = \theta_i^{(2)}$, because his announcement of $-1 = \theta_i^{(1)}$ induces the joint action further away from his preference shock $-0.75 = \theta_i^{(2)}$. While such an incentive scheme is inefficient for both players ex post, the ex ante loss from such "money burning" turns out to be second order compared to the first order gain from relaxing the incentive costs. Accordingly, the second-best current action exhibits $a(-1,-1) < -1$. Likewise, the second-best current action exhibits $a(1,1) > 1$.

B.1.3 Incentive Constraints

The numerical simulation confirms that each player's local upward (downward) constraints are binding when his preference shock is larger (lower) than the other player's average type. Unlike the repeated insurance models, in which the local downward incentive constraints would always be binding and the local upward incentive constraints would always be slack, the relevant incentive constraints can be both downward and upward given the other player's average preference shock.

Another important aspect of the second-best problem is the monotonicity condition. Figure 7 illustrates the expected monotonicity condition on the current allocations of player i: the expected current allocation $\mathbb{E}_{\theta_{-i}}\left[a(\theta_i^{(k_i)}, \theta_{-i})\right]$ is non-decreasing in type $\theta_i^{(k_i)}$. Hence the relevant incentive constraints are the local ones. Indeed, it is directly confirmed that the optimal mechanism found in the relaxed problem satisfies the global incentive constraints.

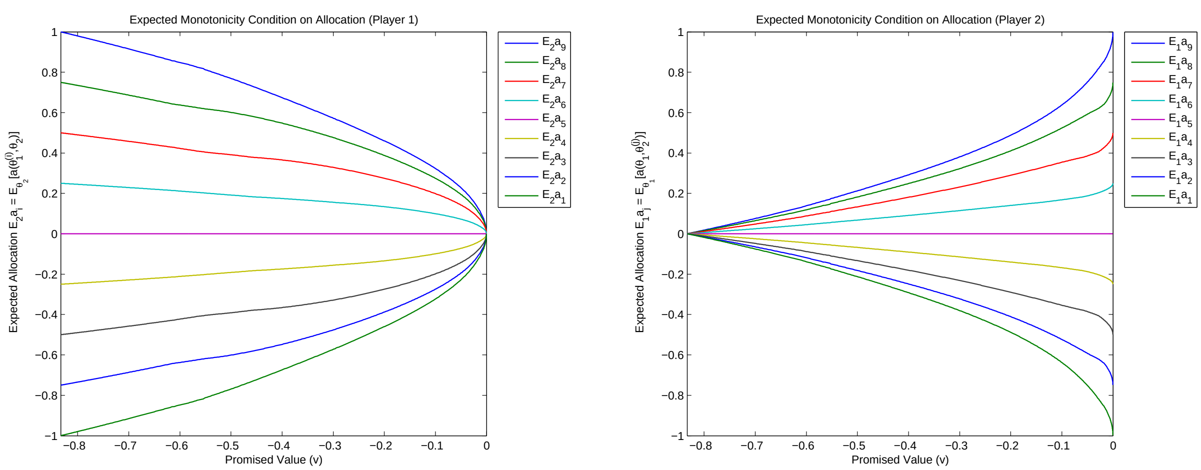


Figure 7: Expected Monotonicity: Player 1 (Left) and Player 2 (Right). $\mathbb{E}_j a_{k_i}$ denotes $\mathbb{E}_{\theta_j}[a(\theta_{-j}^{(k_i)}, \theta_j)]$.

B.2 CARA Insurance Problem

Let us now turn to a repeated CARA insurance problem. We set the coefficient of absolute risk aversion being equal to one, i.e, $u_1(a,\theta_1) = -\exp(-(\theta_1 - a))$ and $u_2(a,\theta_2) = -\exp(-(\theta_2 + a))$. We set $\Theta_1 = \Theta_2 = \{0, 0.75, 1.5\}$ and the uniform distribution $f_i(\cdot) = \frac{1}{3}$. The set of feasible transfers is set to $\mathcal{A} = [-0.9, 0.9]$. These parametric settings yield

$$\underline{v}_1 = \mathbb{E}_\theta\left[u_1(1,\theta_1)\right] = -1.390083 \text{ and } \overline{v}_1 = \mathbb{E}_\theta\left[u_1(-1,\theta_1)\right] = -0.229779.$$

We also have $\underline{v}_2^P = \underline{v}_2 = \underline{v}_1 = \underline{v}_1^P$ and $\overline{v}_2 = \overline{v}_1$. We divide the set of continuation values $[\underline{v}_2, \overline{v}_2]$ into 500 grid points by using the Chebyshev nodes. For each promised value $v \in [\underline{v}_2, \overline{v}_2]$, we consider the relaxed problem in which the relevant incentive constraints are local downward ones. We show that the optimal mechanism found in the relaxed problem satisfies the global incentive constraints.

B.2.1 Approximate Efficiency

We first demonstrate that the second-best value function V_δ^{SB} approaches the first-best value function V^{FB} as δ approaches unity. The left panel of Figure 8 illustrates the first-best and second-best value functions, V^{FB} and V_δ^{SB}. The right panel of Figure 8 depicts the relative social welfare loss (in %) defined as follows: for each weight $\gamma \in (0,1)$ (we consider 500 uniform grid points), we compare the first-best value, $\max_{v\in[\underline{v}_2,\overline{v}_2]} \gamma V^{\mathrm{FB}}(v) + (1-\gamma)v$, and the second-best value $\max_{v\in[\underline{v}_2,\overline{v}_2]} \gamma V_\delta^{SB}(v) + (1-\gamma)v$ (implicit in this computation is the strict concavity of the value function V_δ^{SB}). It is to be noted that the relative social loss is 0 when γ is close to either 0 or 1. This is due to the fact that the marginal value functions (especially, the first-best value function) are bounded in our insurance model. If γ is close to 0, then the relative weight $\frac{1-\gamma}{\gamma}$ is close to infinity, so that it is optimal to choose $v = \overline{v}_2$ in both problems. Likewise, if γ is close to 1, then the relative weight $\frac{1-\gamma}{\gamma}$ is close to zero, so that it is optimal to choose $v = \underline{v}_2$ in both problems. Hence, the right panel does not mean that the first-best allocation is attainable at an interior $v \in (\underline{v}_2, \overline{v}_2)$.

B.2.2 Optimal Mechanisms

Hereafter, the common discount factor is set to $\delta = 0.4$. For each continuation value $v \in [\underline{v}_2, \overline{v}_2]$, the optimal mechanism $(a(\cdot|v), w(\cdot|v))$ is found.

We first report the continuation values. Figure 9 illustrates the continuation value functions $w(\theta_1^{(k_1)}, \theta_2^{(k_2)}|v)$ with player 1's shock $\theta_1^{(k_1)}$ fixed at $k_1 \in \{1,2\}$. First, with player 1's endowment shock fixed, the larger player 2's endowment is, the larger player 2's continuation value becomes. Second, with player 2's endowment shock fixed, the larger player 1's endowment is, the smaller player 2's continuation value becomes. Third, the continuation value functions vary with reported types for each promised value in the second-best problem.

Next, Figure 10 illustrates the optimal first-best and second-best current actions (transfers) $a^{\mathrm{FB}}(\theta|v)$ and $a(\theta|v)$, respectively. Given that the feasible transfer is limited, the optimal

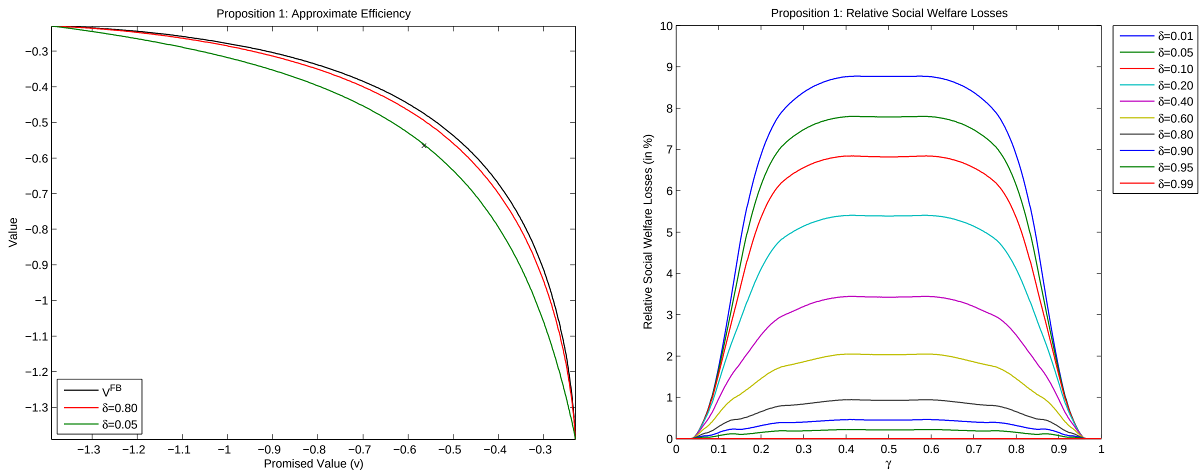


Figure 8: The Value Functions V_δ^{SB} and V^{FB} (Left) and Relative Losses (in %) (Right). The "$\times$" in the left panel denotes the symmetric static second-best values (i.e., the values with no transfer).

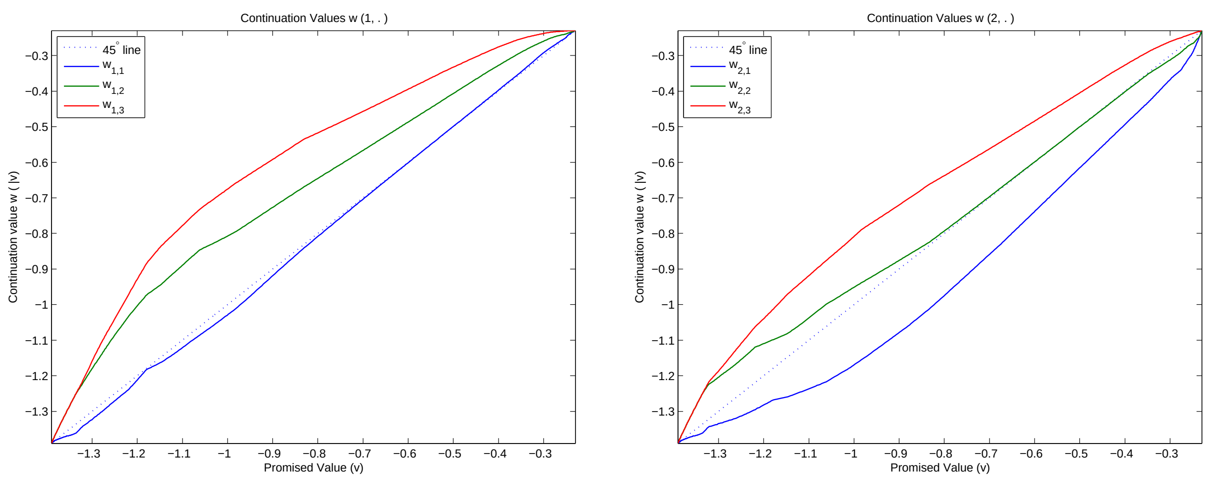


Figure 9: Continuation Values: $w(\theta_1^{(1)}, \cdot|v)$ (Left) and $w(\theta_1^{(2)}, \cdot|v)$ (Right).

current actions exhibit corner solutions for some promised values. Such a feature makes it difficult to find the closed-form solutions.

B.2.3 Incentive Constraints

The numerical simulation shows that, at the optimal mechanism of the relaxed problem, the local downward constraints are binding while the others are slack. Indeed, Figure 11 illustrates the expected monotonicity conditions on marginal utilities. It shows that the expected marginal utility $\mathbb{E}_{\theta_{-i}}\left[\frac{\partial}{\partial\theta_i}u_i(a(\theta_i^{(k_i)}, \theta_{-i}), \tau)\right]$ is non-decreasing in type $\theta = \theta_i^{(k_i)}$ at

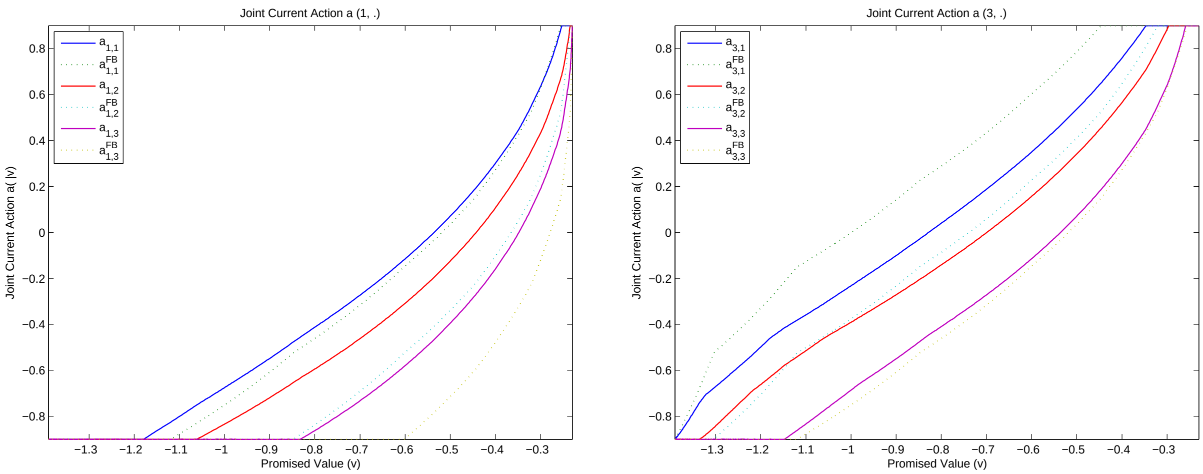


Figure 10: Current Allocations: $a(\theta_1^{(1)}, \cdot)$ (Left) and $a(\theta_1^{(3)}, \cdot)$ (Right).

any τ.

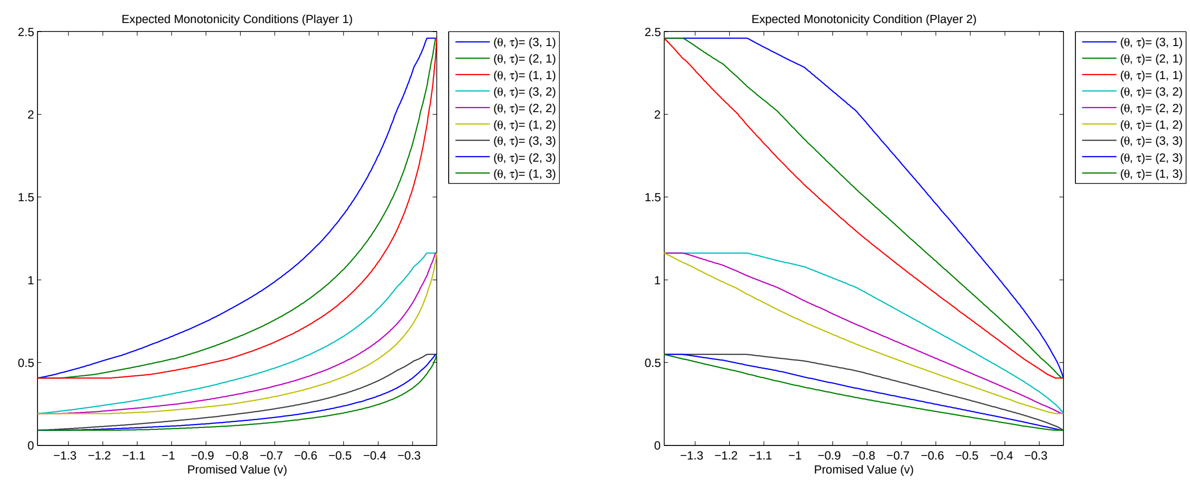


Figure 11: Expected Monotonicity: Player 1 (Left) and Player 2 (Right).

B.2.4 Dynamics

Finally, the left panel of Figure 12 numerically illustrates that one player's continuation value converges to his best value. It illustrates a sample path of stochastic processes $(V(v_t))_t$ and $(v_t)_t$. We choose a point v_0 which is as close as possible to the symmetric point $v = V(v)$. In our numerical grids, $(V(v_0), v_0) = (-0.536721, -0.535245)$. The right panel of Figure 12 illustrates the dynamics of inequality. We measure the share of values captured by the better-off agent at each time t by $\frac{\max(\overline{v}_1 - V_\delta^{\mathrm{SB}}(v^t), \overline{v}_2 - v^t)}{(\overline{v}_1 - V_\delta^{\mathrm{SB}}(v^t)) + (\overline{v}_2 - v^t)}$, and we plot the shares for various δ. We find that the degree of inequality increases at a faster rate for lower δ.

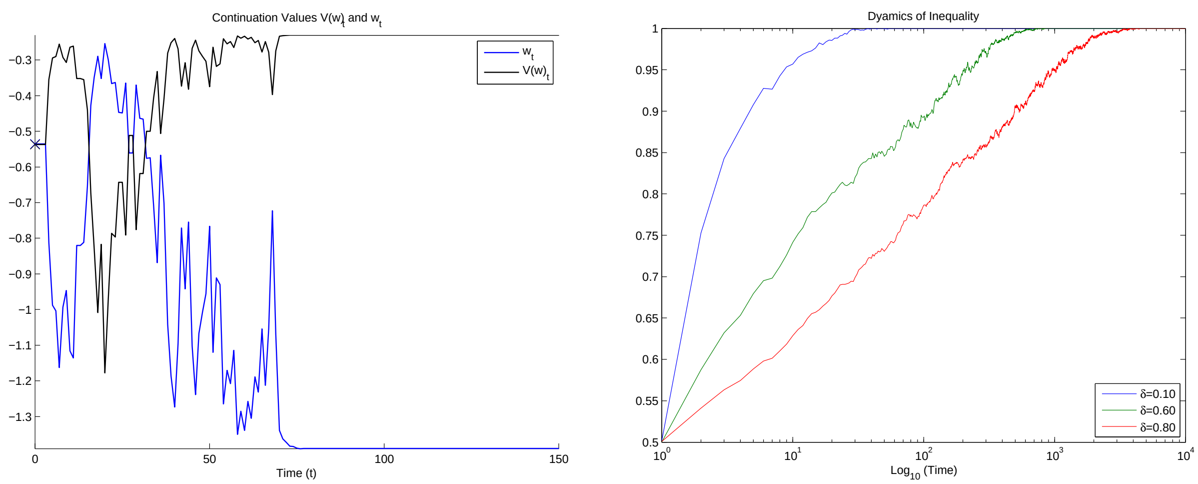


Figure 12: A sample path of $(V(v^t), v^t)$ (Left) and the dynamics of inequality (Right).