

Common Knowledge for Rational and Behavioral Agents*

José A. Rodrigues-Neto[†] James Taylor[‡]

February 20, 2019

Abstract

An implicit idea in the classic concept of common knowledge is that all agents possess perfect information processing ability; this is consistent with agents who have partitional knowledge. This paper discusses some of the shortcomings of the classic definition of common knowledge. When agents have non-partitional knowledge, the classic common knowledge operator may not satisfy positive introspection. It also does not properly consider many reasonable knowledge questions that one would expect to see in a common knowledge operator. We formalize the idea of a ‘question’, and construct an alternative definition of common knowledge which solves these problems, and reduces to the classic definition when all agents have partitional knowledge. Features and shortcomings of further alternative definitions of common knowledge are discussed.

Keywords: common knowledge, behavioural economics, positive introspection, knowledge operators

JEL Classification: D70, D82, D83

*We would like to thank Larry Samuelson, Matthew Ryan, Andrew McLennan, John Hillas, Simon Grant, Simona Fabrizi, Steffen Lippert and seminar participants of the 35th Australasian Economic Theory Workshop, the 16th Society for the Advancement of Economic Theory Conference, and the 30th PhD Conference in Economics and Business for comments. All remaining errors are our own responsibility.

[†]Research School of Economics, The Australian National University, Canberra, ACT - 2601, Australia. URL: <http://rodriguesneto.weebly.com/about>. E-mail: jarnwi@yahoo.com.

[‡]Research School of Economics, The Australian National University, Canberra, ACT - 2601, Australia. URL: <http://cbe.anu.edu.au/people/rse/James-Taylor>. E-mail: james.taylor@anu.edu.au.

Contents

1	Introduction	2
2	Kripke Rationality and Operator Knowledge Models	5
3	Classical Common Knowledge Operators	7
3.1	Properties of the Classical Operators	8
3.2	Breakdown of the Classical Operators	10
4	Behavioral Common Knowledge	10
4.1	Defining Behavioral Common Knowledge	11
4.2	Properties of Behavioral Common Knowledge	15
4.3	Behavioral Common Knowledge and the S5 Axioms	16
5	Alternative Ideas of Behavioral Common Knowledge	18
6	Interpreting Behavioral Knowledge	22
7	Conclusion	24
A	Proofs	25

1 Introduction

When considering common knowledge, most of the economics literature follows Aumann [1], defining common knowledge as “Two people, 1 and 2, are said to have common knowledge of an event E if both know it, 1 knows that 2 knows it, 2 knows that 1 knows it, 1 knows that 2 knows that 1 knows it, and so on.” When agents are rational, in the Kripke sense, this notion of common knowledge has all desired properties. However, if agents cease to be Kripke-rational and exhibit behavioral-type knowledge structures, this notion of common knowledge may lose many important properties. In particular, common knowledge may lose the property of positive introspection. This means it is possible that an event is common knowledge, but no-one knows the event is common knowledge.

Another notion introduced by Lewis [21], defines common knowledge as everybody knows, and everybody knows that everybody knows, and so on. In the case of arbitrary knowledge operators this definition can vary substantially from the Aumann definition. They coincide in special cases such as when agents are sufficiently rational in the sense of having Distributive

knowledge operators.¹

Unfortunately, in some situations agents may not be sufficiently rational, in the Kripke sense, for these classical ideas to coincide. Kripke-rationality requires agents to be aware of their own knowledge, and of the limitations of their knowledge. From teaching students, we have observed that, occasionally, students are not entirely aware of the limitations of their knowledge. Sherlock Holmes’ ‘curious incident of the dog in the dog in the night-time’, in Conan Doyle’s “The Adventure of Silver Blaze”, gives a famous fictitious example of a lack of Negative Introspection. Watson notes that the dog did not bark. Holmes replies that the lack of barking, the lack of evidence, can be used to produce conclusions. Holmes uses Negative Introspection, while Watson does not.² The seminal work by Tversky and Kahneman [27] gives examples where participants appear to fail to have Distributive knowledge operators. As agents may not be Kripke-rational, the definition of common knowledge should allow for such agents, while still preserving the nature of ‘common knowledge’.

The primary contribution of this work is to propose an alternative notion of Behavioral Common Knowledge with some key improvements. If all agents are Kripke-rational, then the behavioral alternative reduces to both the Aumann and Lewis notions of common knowledge. However, Behavioral Common Knowledge retains Positive Introspection even when agents may not be Kripke-rational.

The information structures used in this work are based around knowledge operators, which encode the natural idea of when an agent knows, or thinks that they know, an event to have occurred. These ideas rest on a robust foundation using Modal Logic, but here can be explored using only the fairly simple mathematical tools of basic set theory and probability. This mathematical simplicity means that many tools already exist for working with this framework, and hence the framework has the potential to be extended to novel situations.³

A framework of knowledge questions is developed which allows a formal definition of this alternative notion of common knowledge. Knowledge questions are sentences of the form ‘when does agent 3 know that agents 1 and 2 know that the event has occurred’. The set

¹An agent’s knowledge is Distributive if she knows the conjunction of two events exactly when she knows each of the events individually.

²Dunning and Kruger [8] give empirical results on breaches of Negative Introspection.

³Details on Model Logic can be found in Chellas [7].

of knowledge questions is formalized by two equivalent definitions, a recursive and a fixed point one. According to the fixed point definition, the set of knowledge questions is the minimal set which is closed under an operation which creates new knowledge questions. The equivalent recursive definition starts from questions of the form ‘when does agent j know the event has occurred’ and constructions new knowledge questions from existing ones.

Behavioral Common Knowledge is defined in terms of these knowledge questions. An event is common knowledge at some state, under Behavioral Common Knowledge, if the event is known at that state according to every knowledge question. We construct an equivalent concrete set-theoretic representation to guarantee existence, and to aid in proofs.

The interaction of the new Behavioral Common Knowledge with the Kripke rationality axioms is explored. In particular, whether, if all agents satisfy some Kripke axiom, then Behavioral Common Knowledge does so as well. Some Kripke axioms, including Awareness, Truth, and Distributivity are preserved by Behavioral Common Knowledge, while Negative Introspection and Conjunction are not preserved. This largely aligns with the results on Lewis [21] common knowledge which was analyzed in Chapter One.

Three additional possibilities for a definition of common knowledge for behavioral agents are discussed, and a variety of shortcomings are discovered. The first is a combination of the Aumann and Lewis definitions, and is based on arbitrary groups of agents; this definition fails Positive Introspection. The second extends Lewis common knowledge by not just repeating ‘everybody knows’, but also infinitely repeating this infinitely repeating process; this definition has Positive Introspection, but misses some knowledge questions. The third and final alternative is a simple attempt at an axiomatic approach, requiring common knowledge to be less informative than any individual’s knowledge, to have Positive Introspection, and to be the maximal operator with these properties; there may fail to exist an operator with these properties. Additionally, in each case Behavioral Common Knowledge is more demanding, that is, less informative, than any of these alternatives.

As this work is interested in agents with non-rational knowledge structures, some issues of interpretation arise. The modeling assumes that each agent has full understanding of all agents’ knowledge structures, but is incapable of properly utilizing this information. This can be resolved with no adjustment to the model through the Kahneman principle of ‘What

You See Is All There Is’. The issue can also be resolved by modifying the model to extend the set of primitives from the knowledge structure of each agent, to a universal type space-style knowledge structure. While this also resolves the issue, the new notion of common knowledge is then no longer a direct extension of the Aumann common knowledge, and loses Positive Introspection.

Section 2 lays the foundations of Kripke-rationality and the operator knowledge model. Section 3 considers the traditional notions of common knowledge and identifies shortcomings in the case where agents are not Kripke-rational. Section 4 proposes a Behavioral Common Knowledge operator with the desired property of positive introspection, relates this new common knowledge to the traditional formulations, and discusses further properties of this new definition. Section 5 evaluates potential alternative definitions of common knowledge for behavioral agents. Section 6 discusses the concerns in interpretation and proposes some solutions to these concerns. The appendix brings all proofs.

2 Kripke Rationality and Operator Knowledge Models

A state space is a non-empty finite set Ω of outcome-relevant states. An agent’s knowledge operator $K : 2^\Omega \rightarrow 2^\Omega$ represents what the agent knows. For each event $E \subset \Omega$, that KE is the subset of states where the agent knows the event E has occurred.⁴ An agent is said to be Kripke-rational if their knowledge operator K satisfies the following axioms:

- 1) Awareness: $K\Omega = \Omega$
- 2) Distributivity: $K(E \cap F) = KE \cap KF$ for all $E, F \in 2^\Omega$
- 3) Truth: $KE \subset E$ for all $E \in 2^\Omega$
- 4) Positive Introspection: $KE \subset KKE$ for all $E \in 2^\Omega$
- 5) Negative Introspection: $\neg KE \subset K(\neg KE)$ for all $E \in 2^\Omega$

⁴As in Chapter One, the term ‘knows’ is used to mean the agent would claim that they ‘know’ the event. In particular, agents are allowed to be mistaken in their claims of knowledge.

Together these properties compose Kripke's S5 logic system.⁵ Bacharach [3] proves that a knowledge operator K is Kripke-rational if and only if it can be represented by a partition of the state space. For a partition of the state space π , write $\pi(\omega)$ for the element of the partition containing ω . By definition, knowledge operator K is represented by partition π when $KE = \{\omega \in \Omega \mid \pi(\omega) \subset E\}$.

A weaker concept of rationality is correspondence knowledge. A knowledge operator is a correspondence knowledge if it can be represented by a function $\gamma : \Omega \rightarrow 2^\Omega$ such that $KE = \{\omega \in \Omega \mid \gamma(\omega) \subset E\}$. A knowledge operator is a correspondence knowledge if and only if it satisfies the axioms of Awareness and Distributivity, as shown in Chapter One.

Occasionally we will separate the Distributivity axiom into Monotonicity and Conjunction.

2a) Monotonicity: $E \subset F \implies KE \subset KF$ for all $E, F \in 2^\Omega$

2b) Conjunction: $KE \cap KF \subset K(E \cap F)$ for all $E, F \in 2^\Omega$

A knowledge operator is Monotonic if the agent is able to reason that whenever the event E logically implies the event F , and they know the event E , then they must also know the event F . A knowledge operator satisfies Conjunction if the agent is able to reason that if they know E and know F , then they must also know $E \cap F$. Clearly an operator K satisfies Distributivity if and only if it satisfies Monotonicity and Conjunction.

A knowledge model is a collection consisting of a state space, a set of players, and a knowledge operator for each player. In principle, any agent's knowledge operator may or may not satisfy any of the axioms of Kripke-rationality. When agents have sufficiently well-behaved knowledge operators, the classical formulations of common knowledge are effective. However, as seen in Tversky and Kahneman [27] and Dunning and Kruger [8] it is not always reasonable to assume that agents process information rationally.

Definition 1. A tuple $(\Omega, J, \{K_j\}_{j \in J})$ is a knowledge model, where Ω is a finite set of outcome-relevant states, J is a finite set of agents, and each K_j is a function $K_j : 2^\Omega \rightarrow 2^\Omega$ which is the knowledge operator for agent $j \in J$.

⁵Further details on the S5 system and Modal Logic more generally can be found in Chellas [7].

3 Classical Common Knowledge Operators

The two traditional, or classical, formulations of the common knowledge operator are (1) infinitely repeated everybody knows, and (2) infinitely repeated sequences of the form agent j knows that j' knows, and so on. To formally define infinitely repeated everybody knows, first 'everybody knows' is defined.

Definition 2. *Let $(\Omega, J, \{K_j\}_{j \in J})$ be a knowledge model. Everybody knows E , denoted $\lambda(E)$, is defined pointwise by*

$$\lambda(E) = \bigcap_{j \in J} K_j E$$

With this notation, $\lambda^2 E$ represents the subset of states where everybody knows that everybody knows event E . Proceeding recursively, the operator $\lambda^n : 2^\Omega \rightarrow 2^\Omega$ is defined as the composition of n copies of $\lambda : 2^\Omega \rightarrow 2^\Omega$.

The “infinitely repeated everybody knows” operator is the intersection of all repeated compositions of the everybody knows operator.

Definition 3. *Let $(\Omega, J, \{K_j\}_{j \in J})$ be a knowledge model. The first traditional common knowledge operator ‘repeated everybody knows’, denoted $C_e : 2^\Omega \rightarrow 2^\Omega$, is defined pointwise by*

$$C_e E = \bigcap_{n=1}^{\infty} \lambda^n E \tag{1}$$

The second classical common knowledge is ‘infinitely repeated agent j knows that j' knows’ or ‘sequential’ common knowledge. Let $\text{Seq}(J)$ be the set of all finite sequences of elements of J . For each $s \in \text{Seq}(J)$, let $s = s_1, s_2, \dots, s_n$ with each $s_i \in J$. Define the knowledge operator $K_{(s)} : 2^\Omega \rightarrow 2^\Omega$ as $K_{(s)} = K_{s_n} K_{s_{n-1}} \cdots K_{s_1}$.

Definition 4. *Let $(\Omega, J, \{K_j\}_{j \in J})$ be a knowledge model. The second traditional common knowledge operator ‘sequential agents know’, denoted $C_s : 2^\Omega \rightarrow 2^\Omega$, is defined pointwise by*

$$C_s E = \bigcap_{s \in \text{Seq}(J)} K_{(s)} E \tag{2}$$

Remark 1. *Both Definition 3 and 4 allow for arbitrary repetitions for λ or arbitrarily long sequences. However, as we assume Ω is finite, all operations will necessarily end in finitely*

many steps. Many of the results throughout this paper depend crucially on the finiteness of Ω .

3.1 Properties of the Classical Operators

One of the key properties of both traditional concepts of common knowledge is Positive Introspection. Both definitions lead to a knowledge operator which necessarily satisfies Positive Introspection if attention is restricted to those knowledge operators K_j which can be represented as correspondences. That is, if each agent has a sufficiently nice knowledge structure in the sense that it can be represented by a correspondence, then both C_e and C_s will satisfy positive introspection. Moreover, in this case $C_e = C_s$ and this operator is the most informative operator which satisfies positive introspection, and has $C_e E \subset K_j E$ for all agents $j \in J$. Propositions 1 and 2 formalize these results.

Proposition 1. *Let $(\Omega, J, \{K_j\}_{j \in J})$ be an operator knowledge model. Suppose $K_j(E \cap F) = K_j E \cap K_j F$ for all $E, F \in 2^\Omega$ and $j \in J$. Then, C_e satisfies positive introspection. Moreover, the traditional formulations of common knowledge coincide; that is, $C_e = C_s$. Hence, C_s also satisfies positive introspection.*

Proposition 1 only assumes that each K_j , for $j \in J$, satisfies Distributivity. For each operator to be a correspondence would also require that $K_j \Omega = \Omega$, which is not needed here.

Not only do C_e and C_s satisfy positive introspection when each agent's knowledge operator is Distributive; but $C_e = C_s$ is the most informative operator with this property which also has that if an event is common knowledge, then it is known by all agents. That is, when each agents' knowledge operator is Distributive, then both C_e and C_s are characterized by being the maximal operators which satisfy positive introspection, and such that $CE \subset K_j E$ for all agents $j \in J$. This is the content of Proposition 2.

Lemma 1. *Let $(\Omega, J, \{K_j\}_{j \in J})$ be a knowledge model such that each K_j , for each $j \in J$, satisfies Monotonicity. Then repeated everybody knows $\lambda^n : 2^\Omega \rightarrow 2^\Omega$ satisfies Monotonicity, for all $n \in \mathbb{N}$.⁶*

⁶Throughout this work $\mathbb{N} = \{1, 2, 3, \dots\}$.

Proposition 2. *Let $(\Omega, J, \{K_j\}_{j \in J})$ be a knowledge model such that $K_j(E \cap F) = K_j E \cap K_j F$ for all $j \in J$ and $E, F \in 2^\Omega$. Let $C : 2^\Omega \rightarrow 2^\Omega$ be an operator. Under these assumptions, $C = C_e = C_s$ if and only if C satisfies the following properties:*

- i) $CE \subset K_j E$ for all $j \in J$ and $E \in 2^\Omega$,*
- ii) $CE \subset CCE$ for all events $E \in 2^\Omega$, and*
- iii) $C : 2^\Omega \rightarrow 2^\Omega$ is maximal in the sense that for any knowledge operator $D : 2^\Omega \rightarrow 2^\Omega$ satisfying properties (i) and (ii), then $DE \subset CE$ for all events $E \in 2^\Omega$.*

Alternatively, in order to obtain positive introspection for C_e it suffices to assume that some operator K_j satisfies Truth. This assumption is not sufficient to guarantee that C_s satisfies positive introspection.

Proposition 3. *Let $(\Omega, J, \{K_j\}_{j \in J})$ be a knowledge model. Suppose for each event $E \in 2^\Omega$ there exists an agent $j_E \in J$ such that $K_{j_E} E \subset E$. Then, $C_e E \subset C_e C_e E$ for all $E \in 2^\Omega$; in fact $C_e = C_e C_e$. In particular, if K_j satisfies Truth for some $j \in J$, then $C_e = C_e C_e$.*

By contrast, there exists a knowledge model $(\Omega, J, \{K_j\}_{j \in J})$ such that K_j satisfies Truth for all $j \in J$, yet C_s does not satisfy Positive Introspection.

It is possible that C_s does not satisfy Positive Introspection.

The requirement that Ω is finite is essential to all results so far, and indeed without this assumption the results do not hold.⁷ Unfortunately this restriction to operators K satisfying Truth is not sufficient to guarantee the positive introspection of our second classical notion of common knowledge.

Remark 2. *If all knowledge operators satisfy Monotonicity and at least one operator satisfies Truth, then C_e is the unique maximally informative operator such that $CE \subset K_j E$ and $CE \subset CCE$ for all $j \in J$ and $E \in 2^\Omega$. This follows from Proposition 3 and the proof of Proposition 2. Under these assumptions it is possible that $C_e \neq C_s$, so C_s will not have these properties.*

⁷A counterexample would be $\Omega = \mathbb{R}$, $J = \{j\}$ and for $a < -1, b > 1$, let $K_j(a, b) = ((a-1)/2, (b+1)/2)$, $K_j((a, b) \cup E) = ((a-1)/2, (b+1)/2)$, and $K_j E = \emptyset$ otherwise. Then $K_j E \subset E$, and $K_j(E \cap F) = K_j E \cap K_j F$ for all $E, F \in 2^\Omega$, and $C_e((a, b)) = C_s((a, b)) = [-1, 1]$, but $C_e([-1, 1]) = C_s([-1, 1]) = \emptyset$.

3.2 Breakdown of the Classical Operators

In the absence of assumptions on the knowledge operators of agents, the traditional concepts of common knowledge may not have the positive introspection property or other desirable properties. In particular, it is entirely possible for an event to be common knowledge, and for no-one to know the event is common knowledge according to C_e or C_s . Proposition 4 demonstrates this.

Proposition 4. *There exists a knowledge model $(\Omega, J, \{K_j\}_{j \in J})$ such that neither C_e nor C_s satisfy Positive Introspection.*

Equally troubling, it is possible that at some state an event is common knowledge in the traditional C_e or C_s sense, but, for example, agent 1 does not know that agent 2 and 3 know the event has occurred. When we say events are common knowledge, we want all agents to know that all other agents know the event. This is potentially not the case for the traditional notions of common knowledge.

Proposition 5. *There exists a knowledge model $(\Omega, J, \{K_j\}_{j \in J})$, where $J = \{1, \dots, |J|\}$ with $|J| \geq 3$, and event $E \in 2^\Omega$, such that $C_e E \not\subseteq K_1(K_2 E \cap K_3 E)$ and $C_s E \not\subseteq K_1(K_2 E \cap K_3 E)$.*

The classical notions of common knowledge are distinct. It is not the case that one of the classical notions is more or less restrictive than the other. Depending on the knowledge model, it may be the case that $C_e E \subseteq C_s E$, while sometimes $C_s E \subseteq C_e E$. We consider this a downside of the classical notions of common knowledge, as it means there are two competing definitions. We do not have a strong reason to prefer one to the other.

Proposition 6. *There exists a knowledge model $(\Omega, J, \{K_j\}_{j \in J})$ and events $E, \tilde{E} \in 2^\Omega$ such that $C_e E \subsetneq C_s E$ and $C_s \tilde{E} \subsetneq C_e \tilde{E}$.*

4 Behavioral Common Knowledge

While the previous classical notions of common knowledge include standard questions such as ‘when does everybody know the event E has occurred’ (which from now on will be written as ‘everybody knows E ’ and ‘everybody knows that everybody knows E ’, they exclude

considerations of questions like ‘everybody knows that agent 1 knows’. A similar issue holds for the second classical notion of common knowledge, which misses events of the form $K_3K_1(K_2E \cap K_3E)$.

With this in mind, we define a new concept of common knowledge which explicitly takes into account objects like ‘everybody knows that agent 1 knows’, as well as $K_3K_1(K_2E \cap K_3E)$. This new definition works well with any collection of knowledge operators, even those which violate the S5 axioms.

To construct this Behavioral Common Knowledge we will consider what types of knowledge questions are reasonable requirements for common knowledge. In this setting a knowledge question has the form ‘when does agent j know’; or ‘when does agent 3 know that agent 1 knows that agents 2 and 3 know’. The concern then becomes exactly which knowledge questions are reasonable for common knowledge. We will express a knowledge question as an ordered pair. The question ‘when does agent j know’ will be expressed as (j, \emptyset) . The question ‘when does agent 3 know that agent 1 knows that agents 2 and 3 know’ will be expressed as $(3, \{(1, \{(2, \emptyset), (3, \emptyset)\})\})$.

4.1 Defining Behavioral Common Knowledge

Which objects are reasonable to include as knowledge questions? We provide two equivalent definitions for the set of knowledge questions. Definition 5 seeks a fixed point of a set equation to define Behavioral Common Knowledge by making three assumptions on the class of knowledge questions. First, ‘when does agent j know’ is a reasonable knowledge question for each $j \in J$. Second, given any collection of reasonable knowledge questions, then it is reasonable to ask if agent j knows them all. Finally, the set of knowledge questions is the minimal set with these properties. Definition 6 gives a recursive definition for the set of knowledge questions, where new knowledge questions are formed recursively by asking when does agent j know some (finite) collection of previously formed knowledge questions. Proposition 7 shows that Definitions 5 and 6 are equivalent.

Definition 5. *Let J be a set of agents. The set of knowledge questions Q is the minimal set of ordered pairs with the first element an agent $j \in J$, and the second element drawn from*

the collection of sets with atoms J , such that:⁸

Q.1) For all $j \in J$, then $(j, \emptyset) \in Q$, and

Q.2) For all $j \in J$ and $q_1, \dots, q_n \in Q$, then $(j, \{q_1, \dots, q_n\}) \in Q$

Let $\text{Fin}(2^Q)$ be the set of finite subsets of Q . Then Q is the minimal solution of the set equation $Q = J \times \text{Fin}(2^Q)$.

Definition 6. Let J be a set of agents. Define the set Q_n recursively on each $n \in \mathbb{N}$ by

$$Q_1 = J \times \{\emptyset\}$$

$$Q_{n+1} = Q_n \cup \{(j, \{q_1, q_2, \dots, q_m\}) \mid j \in J, m \in \mathbb{N}, q_1, \dots, q_m \in Q_n\}$$

Let

$$\tilde{Q} = \bigcup_{n=1}^{\infty} Q_n$$

Proposition 7. Let J be a set of agents. The set of knowledge questions Q as in Definition 5 is the same as the set \tilde{Q} in Definition 6. That is

$$Q = \tilde{Q}$$

Throughout the remainder of this work, the symbol Q will be used to refer to the set of knowledge questions. The set of knowledge questions depends only on the finite set of agents J , and does not refer at all to the state space Ω , or the knowledge operators K_j .

For ease of notation, the infimum symbol \inf is occasionally used. Let $\inf(E_1, \dots, E_n) = E_1 \cap \dots \cap E_n$. This is the usual definition of the infimum over 2^Ω when the set of events is endowed with the subset partial order.⁹ Similarly, for collections of knowledge operators, let $K = \inf(K_1, \dots, K_n)$ be defined pointwise by $KE = K_1E \cap \dots \cap K_nE$. This is the infimum over the set of functions $2^\Omega \rightarrow 2^\Omega$ with the partial order induced from the order on the codomain.

Each knowledge question is associated with a knowledge operator where the statement ‘When does Agent 3 knows that agent 1 knows that agents 2 and 3 know the event has

⁸The class of sets is normally defined so that all elements are themselves sets. Sets with atoms also allow elements which are not sets. In this case, all elements of J are atoms.

⁹The subset partial order is the ordering where $X \preceq Y$ if and only if $X \subset Y$.

occurred', usually denoted $K_3(K_1(K_2 \cap K_3))$, is associated with the knowledge question $q = (3, \{(1, \{(2, \emptyset), (3, \emptyset)\})\})$. Definition 7 describes the recursive algorithm which makes this association.

Definition 7. Let $(\Omega, J, \{K_j\}_{j \in J})$ be a knowledge model, Q the set of knowledge questions, and $q \in Q$ a knowledge question where $q = (j, \{q_1, q_2, \dots, q_n\})$. Define the operator $K_q : 2^\Omega \rightarrow 2^\Omega$ by

$$K_q = K_j(\inf(K_{q_1}, \dots, K_{q_n})) \quad (3)$$

For a fixed knowledge question $q \in Q$, Equation 3 suggests q can be interpreted as a function which takes as input $|J|$ knowledge operators, and outputs a new knowledge operator K_q . Viewed in this way, each knowledge question q is the restriction of a knowledge aggregator, as in Chapter One, to a fixed input length.

An event is defined to be common knowledge, under Behavioral Common Knowledge, if all knowledge questions are known.

Definition 8. Let $(\Omega, J, \{K_j\}_{j \in J})$ be a knowledge model, and Q the associated set of knowledge questions. Behavioral Common Knowledge $C_B : 2^\Omega \rightarrow 2^\Omega$ is

$$C_B = \inf_{q \in Q} K_q \quad (4)$$

The construction of Behavioral Common Knowledge admits a recursive set-theoretic characterization. This representation will be useful for proving many of the properties of Behavioral Common Knowledge operator. Definition 9 describes the representation, and Proposition 8 shows it to be equivalent to Behavioral Common Knowledge.

Definition 9. Let $(\Omega, J, \{K_j\}_{j \in J})$ be a knowledge model. For each event $E \in 2^\Omega$, define a sequence of collections of events, $(S_i[E])_{i \in \mathbb{N}}$ according to:

$$\begin{aligned} S_1[E] &= \{K_j F \mid j \in J, F = E\}, \text{ and} \\ S_n[E] &= \{K_j F \mid j \in J, F \in S_{n-1}[E]\} \\ &\cup \{F_1 \cap F_2 \mid F_1, F_2 \in S_{n-1}[E]\}, \text{ for all } n > 1 \end{aligned} \quad (5)$$

That is, $S_1[E]$ is the collection of events of the form $K_j E$ for some j ; and given $S_{n-1}[E]$, a new event in $S_n[E]$ is constructed either by (i) prepending K_j to some set in $S_{n-1}[E]$, or (ii) taking the intersection of two events in $S_{n-1}[E]$.¹⁰

Define \tilde{C} by:

$$\tilde{C}E = \bigcap_{n=1}^{\infty} \bigcap_{F \in S_n[E]} F \quad (6)$$

This sequence of collections of events $(S_i[E])_{i \in \mathbb{N}}$ is a non-decreasing, or telescoping, series, as shown in Lemma 2. This fact, combined with the finiteness of 2^Ω , is especially useful when showing that $\tilde{C} : 2^\Omega \rightarrow 2^\Omega$ is the same as the Behavioral Common Knowledge C_B , as in Proposition 8.

Lemma 2. *Let $(\Omega, J, \{K_j\}_{j \in J})$ be a knowledge model. The sequence $(S_k[E])_{k \in \mathbb{N}}$ given by Equation 5 is a non-decreasing sequence. That is, $S_m[E] \subset S_{m+1}[E]$ for all events $E \in 2^\Omega$, and $m \in \mathbb{N}$.*

Proposition 8. *Let $(\Omega, J, \{K_j\}_{j \in J})$ be a knowledge model. Let $C_B : 2^\Omega \rightarrow 2^\Omega$ be Behavioral Common Knowledge given by Equation 4, and $\tilde{C} : 2^\Omega \rightarrow 2^\Omega$ be given by Equation 6. Then, $C_B = \tilde{C}$.*

As a robustness check, Proposition 9 shows that if the set of knowledge questions were made any smaller, this would lead to a different definition of common knowledge. If we remove questions from the question set Q , then clearly the Behavioral Common Knowledge operator becomes weakly more inclusive. Proposition 9 shows that if even a single knowledge question is removed, there are examples of knowledge models with strict changes in the resulting operator.

Proposition 9. *Let J be a set of agents, Q the associated set of knowledge questions, and $\hat{Q} \subsetneq Q$. Then there exists a knowledge model $(\Omega, J, \{K_j\}_{j \in J})$ such that $C_B \neq \inf_{q \in \hat{Q}} K_q$.*

One impact of Proposition 9 is that, absent additional information about the state space Ω , it is not sufficient to define the set of knowledge questions Q as the union $\bigcup_{n=1}^N Q_n$ for some very large N . It is necessary to allow arbitrarily ‘long’ questions in the set of knowledge

¹⁰Taking intersections of two events is sufficient for the construction. Allowing arbitrary finite intersections would change each $S_n[E]$, but would not effect \tilde{C} .

questions in order to have the definition of common knowledge given here. This mirrors the result of Rubenstein [25] on almost common knowledge.

4.2 Properties of Behavioral Common Knowledge

Two of the major deficiencies of the classical notions of common knowledge, identified in Section 3.2, are that the classical common knowledge (i) does not satisfy Positive Introspection, and (ii) does not require all knowledge questions. Both of these failings are rectified by Behavioral Common Knowledge.

It is immediate that Behavioral Common Knowledge requires all knowledge questions in the sense that $C_B E \subset K_q E$ for all events $E \in 2^\Omega$ and questions $q \in Q$. Behavioral Common Knowledge is, by definition, the most informative operator with this property. It is much less obvious that C_B necessarily satisfies Positive Introspection. Proposition 10 shows that C_B satisfies positive introspection for every knowledge model. Unlike the classical notions of common knowledge, the positive introspection result for Behavioral Common Knowledge does not rely on any assumptions on the underlying knowledge operators.

Proposition 10. *Let $(\Omega, J, \{K_j\}_{j \in J})$ be a knowledge model. Behavioral Common Knowledge C_B satisfies positive introspection.*

Proposition 10 cannot be extended to show $C_B = C_B C_B$, as it cannot be guaranteed that $C_B E \in S_M[C_B E]$ even for very large $M \in \mathbb{N}$. An instance where $C_B \neq C_B C_B$ is given in Example 1.

Example 1. *Let $\Omega = \{a, b, c\}$, and $J = \{1\}$. Let $K_1 \emptyset = \{a, b\}$, $K_1 \{a, b\} = \{a, c\}$, $K_1 \{a\} = \{a, b, c\}$, and $K_1 E = E$ for all other $E \in 2^\Omega$.*

Then, $C_B \emptyset = \{a, b\} \cap \{a, c\} = \{a\}$, while $C_B \{a\} = \{a, b, c\}$. Thus, $C_B \emptyset \subsetneq C_B C_B \emptyset$, and $C_B \neq C_B C_B$.

Behavioral Common Knowledge is an extension of the Aumann notion of common knowledge in that, if all agents are Kripke-rational, then $C_B = C_e = C_s$. Moreover, this equivalence holds even if the knowledge operator of each agent is merely Distributive. This equality does not hold if the operators of each agent are only Monotonic, or only satisfy Conjunction.

Regardless of any properties of the knowledge operators of the agents, it is always the case that $C_B E \subset C_e E$ and $C_B \subset C_s$ for all events $E \in 2^\Omega$.

Proposition 11. *Let $(\Omega, J, \{K_j\}_{j \in J})$ be a knowledge model, and $E \in 2^\Omega$. Then*

- i) Behavioral Common Knowledge is always more demanding than the classical notions of common knowledge; $C_B E \subset C_e E$ and $C_B E \subset C_s E$.*
- ii) If all agents' knowledge operators satisfy the Distributive Property, then the classical notions of common knowledge and Behavioral Common Knowledge coincide, $C_B E = C_e E = C_s E$ for all events $E \in 2^\Omega$.*
- iii) It is possible that all agents' knowledge operators satisfy only Monotonicity, yet the classical and Behavioral Common Knowledge do not coincide.*
- iv) It is possible that all agents' knowledge operators satisfy only Conjunction, yet the classical and Behavioral Common Knowledge do not coincide.*

The Aumann definition of common knowledge, C_s , is elegant as it is able to exploit the Distributivity property so that C_s satisfies positive introspection using a relatively simple formula. Once the assumption of Distributivity of the knowledge operators is dropped, we must move to a more complicated expression in order to capture positive introspection.

4.3 Behavioral Common Knowledge and the S5 Axioms

As was seen in Proposition 10, Behavioral Common Knowledge will satisfy positive introspection regardless of any assumptions on the agents' knowledge operators. While this will not be the case for the other axioms of the S5 system, will Behavioral Common Knowledge preserve the S5 axioms? That is, if we know that each agents' knowledge operator has at least some structure, can it be deduced that the Behavioral Common Knowledge will also have that structure. This section will show that Behavioral Common Knowledge preserves Awareness, Distributivity, Monotonicity and Truth. It does not preserve Conjunction or Negative Introspection.

For the Awareness axiom, it is immediate that if all agents have Awareness then Behavioral Common Knowledge will do so as well. Moreover, Behavioral Common Knowledge satisfies Awareness if and only if all agents satisfy Awareness.

Proposition 12. *Let $(\Omega, J, \{K_j\}_{j \in J})$ be a knowledge model. Suppose K_j satisfies Awareness for all $j \in J$. Then Behavioral Common Knowledge C_B satisfies Awareness. Moreover, if C_B satisfies Awareness, then K_j satisfies Awareness for all $j \in J$.*

For the Distributivity axiom, if all agents satisfy Distributivity, then Behavioral Common Knowledge will do so as well.

Proposition 13. *Let $(\Omega, J, \{K_j\}_{j \in J})$ be a knowledge model. Suppose K_j satisfies Distributivity for all $j \in J$. Then Behavioral Common Knowledge C_B satisfies Distributivity.*

Proposition 14 shows that Behavioral Common Knowledge preserves Monotonicity.

Proposition 14. *Let $(\Omega, J, \{K_j\}_{j \in J})$ be a knowledge model. Suppose K_j satisfies monotonicity for all $j \in J$. Then Behavioral Common Knowledge C_B satisfies monotonicity.*

Proposition 15 shows that Behavioral Common Knowledge does not, however, preserve Conjunction. It is quite possible that each agent satisfies conjunction, but Behavioral Common Knowledge does not.

Proposition 15. *There exists a knowledge model $(\Omega, J, \{K_j\}_{j \in J})$ where K_j satisfies Conjunction for all $j \in J$, yet Behavioral Common Knowledge C_B does not satisfy Conjunction.*

As might be expected, Behavioral Common Knowledge does indeed preserve Truth. If each agent can only know truths, then the more restrictive Common Knowledge will also only know truths. Indeed, it is enough that for each event that there is some player who satisfies truth at that event.

Proposition 16. *Let $(\Omega, J, \{K_j\}_{j \in J})$ be a knowledge model. Suppose for each event $E \in 2^\Omega$, that there is some agent $j \in J$ where $K_j E \subset E$. Then C_B satisfies Truth. In particular, if K_j satisfies Truth for some $j \in J$, then Behavioral Common Knowledge C_B satisfies Truth.*

Since Behavioral Common Knowledge will satisfy positive introspection in the absence of any assumptions, it is certainly the case that it will preserve positive introspection. Negative introspection, however, is not necessarily preserved.

Proposition 17. *There exists a knowledge model $(\Omega, J, \{K_j\}_{j \in J})$ where K_j satisfies Negative Introspection for all $j \in J$, yet Behavioral Common Knowledge C_B does not satisfy Negative Introspection.*

5 Alternative Ideas of Behavioral Common Knowledge

This section discusses alternative ideas of common knowledge for behavioral agents. These ideas seem reasonable, and in some cases have some nice properties, but in each case lack at least some property that is desirable for a definition of common knowledge for behavioral agents.

When looking at repeated everybody knows C_e and sequential common knowledge C_s , we could consider generalizing these ideas to ‘infinitely repeated agents $j \in \tilde{J}$ know that agents $j' \in \hat{J}$ know’.¹¹ To this end, let $\text{Seq}(2^J \setminus \emptyset)$ be the set of finite sequences of non-empty subsets of J . For each subset $\tilde{J} \subset J$ define $K_{\tilde{J}} = \inf_{j \in \tilde{J}} K_j$. For each sequence $S = J_1, J_2, \dots, J_n \in \text{Seq}(2^J \setminus \emptyset)$ define a knowledge operator $K_{(S)}$ as $K_{(S)} = K_{J_n} K_{J_{n-1}} \cdots K_{J_1}$.

Definition 10. *Let $(\Omega, J, \{K_j\}_{j \in J})$ be a knowledge model. The knowledge operator ‘repeated groups of agents know’, denoted $C_g : 2^\Omega \rightarrow 2^\Omega$, is defined pointwise by*

$$C_g E = \bigcap_{S \in \text{Seq}(2^J \setminus \emptyset)} K_{(S)} E \quad (7)$$

Comparing the new operator C_g to previous ideas of common knowledge, clearly $C_g E \subset C_e E$, $C_g E \subset C_s E$, and $C_B E \subset C_g E$ for all events E regardless of the knowledge model. That is, C_g is weakly more restrictive than both traditional notions of common knowledge, and weakly less restrictive than Behavioral Common Knowledge. It is not the case that $C_g = C_B$ in general. Moreover, C_g does not necessarily satisfy positive introspection.

Proposition 18. *There exists a knowledge model $(\Omega, J, \{K_j\}_{j \in J})$ such that $C_g \neq C_B$ and C_g does not satisfy Positive Introspection.*

¹¹Special thanks to Matthew Ryan for suggesting the inclusion of this idea of common knowledge.

An alternative extension of the classical operators, is to apply the operator repetition process more than once. That is, where C_e is infinitely repeated everybody knows; perhaps positive introspection could be obtained by taking infinitely repeated C_e .

Definition 11. Let $(\Omega, J, \{K_j\}_{j \in J})$ be a knowledge model. The knowledge operator ‘repeated C_e ’, denoted $C_2 : 2^\Omega \rightarrow 2^\Omega$, is defined pointwise by

$$C_2 E = \bigcap_{n=1}^{\infty} C_e^n E$$

Unfortunately, this is still not enough to guarantee positive introspection, as shown in Example 2.

Example 2. Let $\Omega = \{a, b, c, d, e\}$, $J = \{1\}$, and $K_1 : 2^\Omega \rightarrow 2^\Omega$ given by

$$\begin{aligned} K_1\{a, b\} &= \{b, c, d\}, & K_1\{b, c, d\} &= \{b, c, e\} \\ K_1\{b, c\} &= \{a, b, d\}, & K_1\{a, b, d\} &= \{a, b, e\} \\ K_1\{b\} &= \emptyset & K_1 E &= E, \quad \text{otherwise} \end{aligned}$$

Then, $C_e\{a, b\} = \{b, c, d\} \cap \{b, c, e\} = \{b, c\}$, while $C_e\{b, c\} = \{a, b, c\} \cap \{a, b, e\} = \{a, b\}$. Therefore, $C_2\{a, b\} = \{a, b\} \cap \{b, c\} = \{b\}$. However, $C_2\{b\} = \emptyset$. Hence

$$C_2\{a, b\} = \{b\} \not\subseteq \emptyset = C_2 C_2\{a, b\}$$

Consequently, C_2 does not satisfy positive introspection in general.

As there is only a single agent in Example 2, this example also serves to show that infinitely repeated C_s , or even C_g will not necessarily satisfy positive introspection.

However, if this idea of repetition is continued ad infinitum, the positive introspection property does indeed emerge.

Definition 12. Let $C_1 : 2^\Omega \rightarrow 2^\Omega$ be defined pointwise by

$$C_1 E = \bigcap_{n \in \mathbb{N}} \lambda^n E$$

Then for each $k \in \mathbb{N}$, $k \geq 2$, define $C_k : 2^\Omega \rightarrow 2^\Omega$ pointwise by

$$C_k E = \bigcap_{n \in \mathbb{N}} C_{k-1}^n E$$

Finally, let $C_* : 2^\Omega \rightarrow 2^\Omega$ be given by

$$C_*E = \bigcap_{k \in \mathbb{N}} C_k E \quad (8)$$

Lemma 3 shows that the sequence of operators $(C_k)_{k \in \mathbb{N}}$ becomes weakly more restrictive as k increases. As Ω is finite, this means the sequence $(C_k)_{k \in \mathbb{N}}$ eventually reaches a fixed point, and the fixed point is C_* . Due to the nature of this process, the resulting operator C_* must satisfy Positive Introspection. Proposition 19 provides this result.

Lemma 3. *Let $(\Omega, J, \{K_j\}_{j \in J})$ be a knowledge model. Then, $C_k E \subset C_{k-1} E$ for all events $E \in 2^\Omega$ and $k \geq 2$.*

Proposition 19. *Let $(\Omega, J, \{K_j\}_{j \in J})$ be a knowledge model. Then, C_* satisfies positive introspection.*

While C_* satisfies positive introspection, it is not necessarily the case that for every knowledge question q , then $K_q E \subset C_* E$. That is, it is possible that $C_* \neq C_B$. Example 3 gives an instance where $\omega \in C_* E$, but agent 1 does not know that agents 2 and 3 know E has occurred.

Example 3. *Let $\Omega = \{a, b, c, d\}$, and $J = \{1, 2, 3\}$. Let $K_1 \Omega = \{a, b, c\}$, $K_2 \Omega = \{a, b, d\}$, $K_3 \Omega = \{a, c, d\}$, $K_1 \{a, d\} = \emptyset$, and $K_j E = E$ otherwise.*

Then $\wedge \Omega = \{a\}$, and $\wedge \wedge \Omega = \{a\}$. Thus $C_1 \Omega = \{a\}$. Moreover, $C_ \Omega = \{a\}$. However,*

$$C_B \Omega \subset K_1(K_2 \Omega \cap K_3 \Omega) = K_1 \{a, d\} = \emptyset$$

Hence, $C_B \Omega = \emptyset \neq C_ E$.*

In Example 3, $C_B \Omega \subset C_* \Omega$. This holds in general in that for every knowledge model and every event $C_B E \subset C_* E$. That is, Behavioral Common Knowledge is more demanding than C_* .

Proposition 20. *Let $(\Omega, J, \{K_j\}_{j \in J})$ be a knowledge model, and $E \in 2^\Omega$. Then, $C_B E \subset C_* E$.*

While C_* does satisfy positive introspection, since it misses many knowledge questions, it is not considered an optimal definition of common knowledge for behavioral agents.

The final possibility considered, drawing from the characterization of common knowledge given in Proposition 2, is to define common knowledge based on a set of desired properties. These are: i) if an event is common knowledge, then each agent knows the event; ii) if an event is common knowledge, then this fact is also common knowledge; and iii) common knowledge is the most informative operator with these properties. Definition 13 formalizes this idea.

Definition 13. *Let $(\Omega, J, \{K_j\}_{j \in J})$ be a knowledge model. A knowledge operator $C_a : 2^\Omega \rightarrow 2^\Omega$ is an axiomatic common knowledge operator if*

- i) $C_a E \subset K_j E$ for all $j \in J$ and $E \in 2^\Omega$,*
- ii) $C_a E \subset C_a C_a E$ for all events $E \in 2^\Omega$, and*
- iii) it is maximal in the sense that for any knowledge operator $D : 2^\Omega \rightarrow 2^\Omega$ satisfying (i) and (ii), then $DE \subset C_a E$ for all events $E \in 2^\Omega$.*

Unfortunately, unlike the case where each agents knowledge operator is distributive, the existence of the axiomatic knowledge operator cannot be guaranteed. While an operator satisfying requirements i) and ii) of Definition 13 can always be found,¹² it is possible that the set of operators with these properties does not admit a maximal element in the sense of part iii) of that definition. Example 4 gives a knowledge model where the axiomatic common knowledge C_a does not exist.

Example 4. *Let $\Omega = \{a, b, c\}$, $J = \{1\}$, and $K_1 : 2^\Omega \rightarrow 2^\Omega$ given by $K_1\{a, b, c\} = \{a, b\}$, $K_1\{a, b\} = \emptyset$, and $K_1 E = E$ otherwise.*

Suppose C_a exists, and we will try, and fail, to construct C_a .

For all events E other than Ω or $\{a, b\}$, we have $K_1 E = E$. Therefore, $C_a E \subset K_1 E = E$.

As $K_1 \Omega = \{a, b\}$, then $C_a \Omega \subset \{a, b\}$. Similarly, $K_1\{a, b\} = \emptyset$ implies $C_a\{a, b\} = \emptyset$. Suppose, by way of contradiction, that $C_a \Omega = \{a, b\}$. As $C_a \Omega \subset C_a C_a \Omega$ then $\{a, b\} \subset C_a\{a, b\}$, which is $\{a, b\} \subset \emptyset$. This contradiction means $C_a\{a, b\} \neq \{a, b\}$.

¹² $CE = \emptyset$ for all events $E \in 2^\Omega$ satisfies these requirements.

So far $C_a\{a, b\} = \emptyset$, $C_a\Omega \subsetneq \{a, b\}$, and $C_aE \subset E$ for all other events E . This leads to two potential ‘maximal’ operators:

$$C_a^1E = \begin{cases} \{a\} & ; \text{for } E = \Omega \\ \emptyset & ; \text{for } E = \{a, b\} \\ E & ; \text{otherwise} \end{cases} \quad C_a^2E = \begin{cases} \{b\} & ; \text{for } E = \Omega \\ \emptyset & ; \text{for } E = \{a, b\} \\ E & ; \text{otherwise} \end{cases}$$

Both these operators satisfy parts *i)* and *ii)* of Definition 13. Moreover, both of these operators are undominated among the operators satisfying parts *i)* and *ii)* of Definition 13. Therefore, there is no operator satisfying requirement *iii)* of Definition 13. In this knowledge model the axiomatic common knowledge operator C_a does not exist.

Similarly to ‘repeated groups of agents know’ C_g , and ‘further repeated everybody knows’ C_* , the new axiomatic common knowledge C_a is not as demanding as Behavioral Common Knowledge C_B . That is, $C_BE \subset C_aE$ for all events $E \in 2^\Omega$ regardless of the underlying knowledge model. This is immediate as C_B satisfies both parts *(i)* and *(ii)* of Definition 13, and C_a , when it exists, is the maximal operator satisfying those properties.

6 Interpreting Behavioral Knowledge

A major concern regarding the interpretation of these results is that, while in the traditional partitional setting it is reasonable to suppose that each agent knows each other agent’s knowledge partition and is able to react accordingly, this is not necessarily the case in the more general framework. That is, in this general framework, is it reasonable to even ask ‘when do I know that you know E ’, and if it is reasonable, why should ‘when I know you know E ’ actually be K_1K_2E ? Essentially, in order to build K_1K_2E it is implicitly assumed that agent 1 knows agent 2’s knowledge operator. Three potential solutions to this concern are discussed.

One solution to this problem relies on the behavioral background to this work. The primary reason for interest in this generalized knowledge framework is that, experimentally, people appear to behave as though their knowledge structures are knowledge operators, rather than correspondences or partitions. Observationally, people frequently violate negative introspection, they often violate positive introspection, they are often incorrect, and

they occasionally violate the distributive property.¹³ Drawing from Kahneman’s principle of ‘What you see is all there is’ (WYSIATI), this paper assumes that each agent is entirely capable of complex thought, personal reflection, and interpersonal modeling; it just does not occur to them to do so without prompting. That is, if $KE \not\subset KKE$, this does not cause internal concerns as it will not have occurred to the agent to ask the question ‘when do I know that I know E ’. However, once the agent has been asked the question ‘when do I know that I know E ’, she has no trouble imagining the set of states when she knows E , and considering those states where she knows the event KE .

The case with multiple individuals works in a similar manner. It is assumed that agent 1 is fully cognizant of K_2 , she simply have not previously been given any impetus to use this information. If asked, ‘when do I know that you know E ’, agent 1 can easily answer K_1K_2E . Essentially, each agent is fully aware of all agents’ knowledge operators, including their own, it just has not occurred to them to use this information.

Another solution to this concern is to simply change the interpretation of the knowledge operators to be a description of what some identified agent thinks are the other agents’ knowledge operators. If agent 1 is the identified agent then, in this interpretation, K_1 is indeed what agent 1 knows; while for agents $j \neq 1$, operator K_j encodes what agent 1 thinks is the knowledge of agent j . Behavioral Common Knowledge C_B is then what agent 1 thinks is common knowledge.

Finally, if we instead take seriously the criticism that ‘when does agent 1 know that agent 2 knows E ’ should be unrelated to K_1K_2E , then the underlying knowledge model would need to be substantially expanded to allow for this. Definition 14 describes one method of extending the operator knowledge model to allow this. In this extended framework the knowledge operator associated with each knowledge question is defined separately. This allows ‘when does agent 1 know that agent 2 knows E ’ to be unrelated to K_1K_2E . The approach to knowledge operators is similar to that Fagin et al. [9].

Definition 14. *A tuple $(\Omega, J, Q, \{K_q\}_{q \in Q})$ is an extended knowledge operator model, where Ω is a finite set of outcome-relevant states, J is a set of agents, Q is the induced set of knowledge questions, and each K_q is a function from 2^Ω to 2^Ω , which is the knowledge operator for the*

¹³Famously in the Linda and Bill experiments from Kahneman and Tversky (1983).

knowledge question q .

As with the model of Section 4, a notion of common knowledge for behavioral agents can be described in this setting, as in Definition 15.

Definition 15. *Let $(\Omega, J, Q, \{K_q\}_{q \in Q})$ be an extended knowledge operator model. Common knowledge in this extended framework, denoted $C_{eB} : 2^\Omega \rightarrow 2^\Omega$ is defined pointwise by*

$$C_{eB}E = \bigcap_{q \in Q} K_q E$$

This extended common knowledge retains the property that $\omega \in C_{eB}E$ if and only if for each knowledge question q , then $\omega \in K_q E$. However, it will no longer satisfy positive introspection. This is to be expected; if K_q is unrelated to the individual knowledge of the agents we could not expect the notion of common knowledge to have positive introspection.

7 Conclusion

This work proposes a new definition of Common Knowledge which extends the existing definitions for situations where agents may have unusual knowledge structures. This new definition maintains the positive introspection property even for arbitrarily strange knowledge structures, and reduces to the usual Aumann and Lewis definitions when knowledge structures are sufficiently well-behaved. Specifically, when agents' knowledge operators are Distributive, then Behavioral common knowledge is the same as Aumann or Lewis common knowledge.

Key to defining Behavioral Common Knowledge is the construction of Knowledge Questions. These are questions of the form 'Does agent 1 know that agents 2 and 3 know that the event has occurred'. This structure allows a formal definition for Behavioral Common Knowledge which guarantees that all knowledge questions will be answered by the common knowledge. A set-theoretic construction is developed which allows for easy investigation of the properties of Behavioral Common Knowledge. Alternative ways of defining common knowledge for behavioral agents are discussed, but shortcomings are shown for each.

One area of future research would be to investigate the impact of using Behavioral Common Knowledge in games of incomplete information. This may be of particular relevance for

developing the appropriate equilibrium concept when agents have poorly behaved information structures.

Another area of future research is to extend the existing definitions and results for arbitrary state spaces. If the definitions are used as given, except that the state space is allowed to be infinite, then the results of this Chapter do not hold. In particular, Behavioral Common Knowledge ceases to have positive introspection. The inductive definition of the set of knowledge questions would need to be extended to allow for a degree of trans-finite induction.

A Proofs

The appendix contains the remaining proofs.

Proof of Proposition 1:

Let $K_j(E \cap F) = K_j E \cap K_j F$ for all $j \in J$, $E, F \subset \Omega$. The everybody knows operator λ satisfies the distributive property as

$$\lambda(E \cap F) = \bigcap_{j \in J} K_j(E \cap F) = \bigcap_{j \in J} (K_j E \cap K_j F) = \bigcap_{j \in J} K_j E \cap \bigcap_{j \in J} K_j F = \lambda E \cap \lambda F$$

As $C_e E = \bigcap_{n=1}^{\infty} \lambda^n E$, fix a positive integer $m \in \mathbb{N}$ and consider $\lambda^m C_e E$. We have

$$\lambda^m \left(\bigcap_{n=1}^{\infty} \lambda^n E \right) = \bigcap_{n=1}^{\infty} \lambda^m \lambda^n E = \bigcap_{n=m+1}^{\infty} \lambda^n E \supset \bigcap_{n=1}^{\infty} \lambda^n E = C_e E$$

That is, for each $m \in \mathbb{N}$, $\lambda^m C_e E \supset C_e E$. Therefore, $\bigcap_{m=1}^{\infty} \lambda^m C_e E \supset C_e E$ and so $C_e E \subset C_e C_e E$.

To show $C_e = C_s$, let $E \in 2^\Omega$. Then

$$C_e E = \bigcap_{n=1}^{\infty} \lambda^n E = \bigcap_{n=1}^{\infty} \left[\bigcap_{j_1 \in J} K_{j_1} \left(\bigcap_{j_2 \in J} K_{j_2} \left(\cdots \left(\bigcap_{j_n \in J} K_{j_n} E \right) \cdots \right) \right) \right]$$

Using the distributive property to collect the intersection signs,

$$C_e E = \bigcap_{n=1}^{\infty} \left[\bigcap_{j_1 \in J} \bigcap_{j_2 \in J} \cdots \bigcap_{j_n \in J} K_{j_1} K_{j_2} \cdots K_{j_n} E \right] = \bigcap_{n=1}^{\infty} \left[\bigcap_{j_1, \dots, j_n \in J^n} K_{j_1} K_{j_2} \cdots K_{j_n} E \right]$$

By the definition of the set of finite sequence $\text{Seq}(J)$,

$$C_e E = \bigcap_{s \in \text{Seq}(J)} K_{(s)} E = C_s E$$

as required. □

Proof of Lemma 1:

Let $E, F \in 2^\Omega$ and $E \subset F$. For all $j \in J$, $K_j E \subset K_j F$, so $\bigcap_{j \in J} K_j E \subset \bigcap_{j \in J} K_j F$. That is, $\lambda E \subset \lambda F$. By induction we have

$$E \subset F \implies \lambda E \subset \lambda F \implies \lambda^2 E \subset \lambda^2 F \implies \dots \implies \lambda^n E \subset \lambda^n F$$

In conclusion, the operators $\lambda^n : 2^\Omega \rightarrow 2^\Omega$ satisfy monotonicity for every $n \in \mathbb{N}$. □

Proof of Proposition 2:

Let $C : 2^\Omega \rightarrow 2^\Omega$ satisfy assumptions (i) and (ii). From assumption (i) $CE \subset \lambda E$. Suppose, for the purposes of induction, that $CE \subset \lambda^n E$ for some $n \in \mathbb{N}$ and all $E \in 2^\Omega$. By the inductive hypothesis on the event CE , $CCE \subset \lambda^n CE$. As each K_j is monotonic, by Lemma 1, λ^n is monotonic. Therefore, $CE \subset \lambda E$ implies $\lambda^n CE \subset \lambda^n \lambda E$. By assumption (ii), $CE \subset CCE$, so overall

$$CE \subset CCE \subset \lambda^n CE \subset \lambda^{n+1} E$$

Therefore by the principle of induction, $CE \subset \lambda^n E$ for all $n \in \mathbb{N}$; and any operator C satisfying assumptions (i) and (ii) will satisfy

$$CE \subset \bigcap_{n \in \mathbb{N}} \lambda^n E \quad \text{for all } E \in 2^\Omega$$

Since C is the maximal operator satisfying (i) and (ii), if $C_e E = \bigcap_{n \in \mathbb{N}} \lambda^n E$ satisfies (i) and (ii) then $C = C_e$. Certainly C_e satisfies (i), and by Proposition 1 C_e satisfies positive introspection and $C_e = C_s$. Therefore

$$C = C_e = C_s$$

as required. □

Proof of Proposition 3:

As there is some agent $j_E \in J$ where $K_{j_E} E \subset E$, then $\lambda E \subset E$ for all events $E \in 2^\Omega$. The infinite sequence $(E, \lambda E, \lambda \lambda E, \lambda \lambda \lambda E, \dots)$ is non-increasing. As Ω is finite, and

thus 2^Ω is finite, this sequence eventually reaches a fixed point. Therefore for sufficiently large M , the following are equal

$$C_e E = \bigcap_{n=1}^{\infty} \lambda^n E = \lambda^M E = \bigcap_{n=1}^{\infty} \lambda^n \lambda^M E = \bigcap_{n=1}^{\infty} \lambda^n \left(\bigcap_{m=1}^{\infty} \lambda^m E \right) = C_e C_e E$$

Hence, $C_e E = C_e C_e E$.

By contrast, let $\Omega = \{a, b, c\}$, $J = \{1, 2\}$ and K_1, K_2 be given by $K_1 \Omega = \{a, b\}$, $K_1 \{a\} = \emptyset$, $K_2 \Omega = \{a, c\}$, and $K_j E = E$ otherwise. Then $C_s \Omega = \{a, b\} \cap \{a, c\} = \{a\}$, while $C_s \{a\} = \emptyset$. Therefore $C_s \Omega \not\subseteq C_s C_s \Omega$. In particular, C_s does not satisfy Positive Introspection. \square

Proof of Proposition 4:

Let $\Omega = \{a, b, c, d, e\}$, $J = \{1, 2\}$. Define knowledge operators K_1 and K_2 as follows:

$$\begin{aligned} K_1 \{a, b\} &= \{a, c, d\} & K_2 \{a, b\} &= \{a, c, e\} \\ K_1 \{a, c\} &= \{a, b, d\} & K_2 \{a, c\} &= \{a, b, e\} \\ K_1 \{a\} &= K_2 \{a\} = \emptyset \\ K_1 E &= K_2 E = E, & \text{otherwise.} \end{aligned}$$

By the definition of everybody knows, λ :

$$\lambda \{a, b\} = \{a, c\}, \quad \lambda \{a, c\} = \{a, b\}, \quad \lambda \{a\} = \emptyset, \quad \lambda \emptyset = \emptyset$$

Then, $C_e \{a, b\} = \{a\}$, and $C_e \{a\} = \emptyset$. Hence, $C_e \{a, b\} \not\subseteq C_e C_e \{a, b\}$. Therefore, C_e does not satisfy positive introspection.

For any sequence of agents $K_{j_1, \dots, j_n, 1} \{a, b\} = \{a, c, d\}$, and $K_{j_1, \dots, j_n, 2} \{a, b\} = \{a, c, e\}$, so $C_s \{a, b\} = \{a, c\}$. However $K_{j_1, \dots, j_n, 1} \{a, c\} = \{a, b, d\}$, and $K_{j_1, \dots, j_n, 2} \{a, c\} = \{a, b, e\}$, so $C_s \{a, c\} = \{a, b\}$. Hence $C_s \{a, b\} \not\subseteq C_s C_s \{a, b\}$. Therefore, C_s does not satisfy positive introspection. \square

Proof of Proposition 5:

It is sufficient to provide an example. Let $\Omega = \{a, b, c, d\}$, and $J = \{1, 2, 3\}$. Let $K_1 \Omega = \{a, b, c\}$, $K_2 \Omega = \{a, b, d\}$, $K_3 \Omega = \{a, c, d\}$, $K_1 \{a, d\} = \emptyset$, and $K_j E = E$ for all other events $E \in 2^\Omega$ and agents $j \in J$.

Then $\wedge \Omega = \{a, b, c\} \cap \{a, b, d\} \cap \{a, c, d\} = \{a\}$, and $\wedge \wedge \Omega = \wedge \{a\} = \{a\}$. Continuing in this fashion gives $C_e \Omega = \{a\}$.

For any sequence $s = (j_1, \dots, j_n) \in \text{Seq}(J)$, as $K_j K_{j_1} \Omega = K_{j_1} \Omega$ for all $j \in J$, then

$$K_s \Omega = K_{j_n} \cdots K_{j_1} \Omega = K_{j_1} \Omega$$

So $C_s \Omega$ is

$$C_s \Omega = \bigcap_{s \in \text{Seq}(J)} K_s \Omega = K_1 \Omega \cap K_2 \Omega \cap K_3 \Omega = \{a\}$$

However, for $K_1(K_2 \Omega \cap K_3 \Omega)$,

$$K_1(K_2 \Omega \cap K_3 \Omega) = K_1(\{a, b, d\} \cap \{a, c, d\}) = K_1\{a, d\} = \emptyset$$

Therefore $K_1(K_2 \Omega \cap K_3 \Omega) \not\subseteq C_e \Omega$ and $K_1(K_2 \Omega \cap K_3 \Omega) \not\subseteq C_s \Omega$. □

Proof of Proposition 6:

Extend the knowledge model from the proof of Proposition 4 so that $\Omega = \{a, b, c, d, e, \tilde{a}, \tilde{b}, \tilde{c}\}$, and $J = \{1, 2\}$. Define the knowledge operators K_1 and K_2 as

$$K_1\{a, b\} = \{a, c, d\} \quad K_2\{a, b\} = \{a, c, e\}$$

$$K_1\{a, c\} = \{a, b, d\} \quad K_2\{a, c\} = \{a, b, e\}$$

$$K_1\{a\} = K_2\{a\} = \emptyset$$

$$K_1\{\tilde{a}, \tilde{b}, \tilde{c}\} = \{\tilde{a}, \tilde{b}\} \quad K_2\{\tilde{a}, \tilde{b}, \tilde{c}\} = \{\tilde{a}, \tilde{c}\}$$

$$K_1\{\tilde{a}\} = K_2\{\tilde{a}\} = \{\tilde{a}\}$$

$$K_1 E = K_2 E = \emptyset, \quad \text{otherwise.}$$

As noted in the proof of Proposition 4, $C_e\{a, b\} = \{a\}$, but $C_s\{a, b\} = \{a, c\}$. Thus, there exists an event E such that $C_e E \subsetneq C_s E$.

Similarly, $C_s\{\tilde{a}, \tilde{b}, \tilde{c}\} = \emptyset$, while $C_e\{\tilde{a}, \tilde{b}, \tilde{c}\} = \{\tilde{a}\}$. Thus, there exists an event \tilde{E} such that $C_s \tilde{E} \subsetneq C_e \tilde{E}$. □

Proof of Proposition 7:

Let Q_n be given as in Definition 6, and $\tilde{Q} = \bigcup_{n=1}^{\infty} Q_n$.

Clearly $Q_1 \subset Q$ as for all $j \in J$, $(j, \emptyset) \in Q$. Suppose, for the purposes of induction, that $Q_n \subset Q$. Let $q \in Q_{n+1}$. Either $q \in Q_n$, or $q = (j, \{q_1, q_2, \dots, q_n\})$ where each $q_i \in Q_n$. If $q \in Q_n$ then by the inductive hypothesis $q \in Q$. If $q = (j, \{q_1, q_2, \dots, q_n\})$ then by the inductive hypothesis each $q_i \in Q$, so by assumption Q.2, $q \in Q$. Therefore $Q_{n+1} \subset Q$. By the induction principle, each $Q_n \subset Q$, so

$$\tilde{Q} = \bigcup_{n=1}^{\infty} Q_n \subset Q$$

Now we show $\tilde{Q} = \bigcup_{n=1}^{\infty} Q_n$ satisfies requirements Q.1 and Q.2. As $Q_1 = J \times \{\emptyset\}$, for all $j \in J$ then $(j, \emptyset) \in \bigcup_{n=1}^{\infty} Q_n$. Let $j \in J$, and $q_1, \dots, q_m \in \bigcup_{n=1}^{\infty} Q_n$. Each q_i is in some Q_i , so write $q_i \in Q_{l_i}$. As the sequence $(Q_n)_{n \in \mathbb{N}}$ is non-decreasing, $q_i \in Q_{\max_k l_k}$ for all i . Therefore $q = (j, \{q_1, q_2, \dots, q_n\}) \in Q_{1+\max_k l_k}$, and in particular $q \in \bigcup_{n=1}^{\infty} Q_i$.

As $\bigcup_{n=1}^{\infty} Q_n \subset Q$, and $\bigcup_{n=1}^{\infty} Q_n$ satisfies the requirements for the set of knowledge questions, it must be the minimal set which satisfies these properties. Therefore $Q = \tilde{Q}$. \square

Proof of Lemma 2:

Let $F \in S_{n-1}[E]$, then $F = F \cap F \in S_n[E]$; so $S_{n-1}[E] \subset S_n[E]$. Therefore the sequence $(S_n[E])_{n \in \mathbb{N}}$ is non-decreasing. \square

Proof of Proposition 8:

First we show $\tilde{C}E \subset C_B E$ for all $E \in 2^\Omega$. Define the length of knowledge question q , denoted $|q|$, by $|(j, \emptyset)| = 1$, and for $q = (j, \{q_1, \dots, q_n\})$, by

$$|q| = 1 + \sum_{i=1}^n |q_i|$$

We prove $\tilde{C}E \subset C_B E$ by induction over the length of knowledge questions. Suppose $|q| = 1$. Then, $q = (j, \emptyset)$ for some $j \in J$, and $K_q E = K_j E \in S_1[E]$.

Now suppose, for the purposes of induction, that for all questions r of length m or less, $K_r E$ is in $S_k[E]$ for some k . Let $|q| = m+1$. As $q = (j, \{q_1, q_2, \dots, q_n\})$ for $j \in J$ and $q_i \in Q$, and $|q| = 1 + \sum_i |q_i|$, then $|q_i| \leq m$ for each i . Therefore, by the inductive hypothesis,

$K_{q_i}E \in S_{k_i}[E]$ for some k_i . From Lemma 2, $(S_k[E])_{k \in \mathbb{N}}$ is a non-decreasing sequence, so $K_{q_i}E \in S_{M'}[E]$ for all $i \in \{1, \dots, n\}$, where $M' = \max_i k_i$. Then $\bigcap_i K_{q_i}E \in S_{M'+n-1}[E]$ as $S_{M'+n-1}[E]$ contains all intersections of n events in $S_{M'}[E]$. Let $M = M' + n - 1$. Therefore:

$$K_qE = K_j(\inf(K_{q_1}E, \dots, K_{q_n}E) \in S_M[E])$$

So by the principle of induction, for every knowledge question q , $q \in S_M[E]$ for some M . Finally, as

$$C_B E = \inf_{q \in Q} K_q E \quad \text{and} \quad \tilde{C} E = \bigcap_{M \in \mathbb{N}} \bigcap_{F \in S_M[E]} F$$

therefore $\tilde{C} E \subset C_B E$.

Conversely, we show $C_B E \subset \tilde{C} E$ for all $E \in 2^\Omega$, again using induction over the sequence $(S_i[E])_{i \in \mathbb{N}}$. Let $F \in S_1[E]$. Then $F = K_j E$ for some $j \in J$, so F can be written as the infimum of a set of knowledge questions, in this case the singleton set $\{(j, \emptyset)\}$.

Suppose, for the purposes of induction, that for all $G \in S_m[E]$, there exists a set of knowledge questions $\{q_1, \dots, q_n\}$ such that $G = \inf(K_{q_1}E, \dots, K_{q_n}E)$. Let $F \in S_{m+1}[E]$. There are two cases: (i) $F = K_j F'$ for some $j \in J$ and $F' \in S_m[E]$, or (ii) $F = F_1 \cap F_2$ for $F_1, F_2 \in S_m[E]$. In case (i), as $F' \in S_m[E]$, by the inductive hypothesis, there exists a set of knowledge questions such that $F' = \inf(K_{q_1}E, \dots, K_{q_n}E)$. Therefore

$$F = K_j F' = K_j(\inf(K_{q_1}E, \dots, K_{q_n}E)) = K_q E$$

where $q = (j, \{q_1, q_2, \dots, q_n\})$. The event F is the infimum over a set of knowledge questions, in this case just the singleton set $\{q\}$. For case (ii), as $F_1, F_2 \in S_m[E]$, by the inductive hypothesis, $F_1 = \inf(K_{q_1}E, \dots, K_{q_n}E)$ and $F_2 = \inf(K_{r_1}E, \dots, K_{r_{n'}}E)$. Therefore:

$$F = F_1 \cap F_2 = \inf(K_{q_1}E, \dots, K_{q_n}E, K_{r_1}E, \dots, K_{r_{n'}}E)$$

Event F can be written as the infimum of a set of knowledge questions. By the principle of induction, for all m and for every element $F \in S_m[E]$, F can be written as the infimum of a set of knowledge questions. Therefore, the intersection over all such F will be larger than the infimum over all knowledge questions, which is $C_B E \subset \tilde{C} E$. \square

Proof of Proposition 9:

Pick $q^* \in Q \setminus \widehat{Q}$. For ease of notation, let $\widehat{C} = \inf_{q \in \widehat{Q}} K_q$. To show $\widehat{C} \neq C_B$, it is sufficient to construct a knowledge model $(\Omega, J, \{K_j\}_{j \in J})$ such that $\widehat{C}\Omega \neq \emptyset$, and $K_{q^*}\Omega = \emptyset$.

Let Q_n be given as in Definition 6, and define the depth of $q \in Q$ as the unique value $n \in \mathbb{N}$ such that $q \in Q_n \setminus Q_{n-1}$. Let $m = \text{depth}(q^*)$, $M = |Q_m|$, and $\Omega = \{1, \dots, M+1\}$. Let $\sigma : Q_m \rightarrow \{1, \dots, M\}$ be bijective such that $\text{depth}(q) < \text{depth}(q')$ implies $\sigma(q) < \sigma(q')$, and $\sigma(q^*) = M$.

We want to define a collection of operators K_j , and thereby operators K_q , so that $M+1 \in K_q\Omega$ for all $q \neq q^*$, and $K_{q^*}\Omega = \emptyset$. Operators K_q are first defined recursively for questions of depth no greater than m , then defined recursively for operators of depth greater than m .

First, for each $q \in Q_m \setminus \{q^*\}$, write $q = (j, \{q_1, \dots, q_n\})$. If $\{q_1, \dots, q_n\} = \emptyset$, then let

$$K_q\Omega = K_j\Omega = \Omega \setminus \{\sigma((j, \emptyset))\} = \Omega \setminus \{\sigma(q)\}$$

Otherwise, let

$$K_j(\Omega \setminus \sigma(\{q_1, \dots, q_n\})) = \{1, \dots, \sigma(q) - 1, \sigma(q) + 1, \dots, M+1\} = \Omega \setminus \{\sigma(q)\}$$

Running through the recursion gives

$$K_q\Omega = \{1, \dots, \sigma(q) - 1, \sigma(q) + 1, \dots, M+1\} = \Omega \setminus \{\sigma(q)\}$$

Also let

$$K_{q^*}\Omega = \emptyset, \quad \text{and} \quad K_j\emptyset = \{1, \dots, M-1, M+1\} \quad \forall j \in J$$

Now, for each $q \notin Q_m$, write $q = (j, \{q_1 \cdots q_n\})$. If $q_i = q^*$ for some i , then $K_q\Omega = K_j\emptyset$ so is already defined. Otherwise, recursively define $K_q\Omega$ such that $|K_q\Omega| = M$, and the element of Ω not in $K_q\Omega$ is the largest number not in one of $K_{q_1}\Omega, \dots, K_{q_n}\Omega$. That is

$$K_q\Omega = K_j(K_{q_1}\Omega \cap \cdots \cap K_{q_n}\Omega) = \Omega \setminus \max_{i=1, \dots, n} (\Omega \setminus K_{q_i}\Omega)$$

For every $q \in Q \setminus \{q^*\}$, we have $M+1 \in K_q\Omega$. So

$$\widehat{C}\Omega = \bigcap_{q \in \widehat{Q}} K_q\Omega \supset \{M+1\}$$

However, $K_{q^*}\Omega = \emptyset$ so that $C_B\Omega = \emptyset$. Therefore $C_B \neq \widehat{C}$ as required. \square

Proof of Proposition 10:

Fix an event $E \in 2^\Omega$. The sequence $(S_n[E])_{n \in \mathbb{N}}$ is a non-decreasing sequence, from Lemma 2. As Ω is finite, this sequence will eventually reach a fixed point. That is, there exists $M \in \mathbb{N}$ such that $S_N[E] = S_M[E]$ for all $N \geq M$. Therefore

$$C_BE = \bigcap_{n=1}^{\infty} \bigcap_{F \in S_n[E]} F = \bigcap_{n=1}^M \bigcap_{F \in S_n[E]} F = \bigcap_{F \in S_M[E]} F$$

As $S_M[E] = S_{M+1}[E]$, the set $S_M[E]$ is closed under intersections. Therefore

$$C_BE = \inf(S_M[E]) \in S_M[E]$$

The sequence $(S_n[C_BE])_{n \in \mathbb{N}}$ is also non-decreasing, so there exists $M' \in \mathbb{N}$ such that $S_{N'}[C_BE] = S_{M'}[C_BE]$ for all $N' \geq M'$. Therefore $C_BC_BE = \inf(S_{M'}[C_BE])$.

As $S_{M'}[C_BE]$ is formed by taking repeated applications of intersections and operators K_j , and as $\{C_BE\} \subset S_M[E]$, then $S_{M'}[C_BE] \subset S_{M+M'}[E] = S_M[E]$. As $S_{M'}[C_BE] \subset S_M[E]$ therefore

$$C_BC_BE = \inf(S_{M'}[C_BE]) \supset \inf(S_M[E]) = C_BE$$

which is $C_BE \subset C_BC_BE$, as required. \square

Proof of Proposition 11:

i) First we show $C_BE \subset C_eE$. By construction, $K_jE \in S_1[E]$, for all $j \in J$. Therefore $K_{j_1}E \cap K_{j_2}E \in S_2[E]$, and so on up to $\wedge E = \bigcap_{j \in J} K_jE \in S_{|J|}[E]$. Similarly, $\wedge^n E \in S_{n|J|}[E]$ for all $n \in \mathbb{N}$. Therefore

$$C_BE = \bigcap_{n \in \mathbb{N}} \bigcap_{F \in S_n[E]} F \subset \bigcap_{n \in \mathbb{N}} \wedge^n E = C_eE$$

Next we show $C_BE \subset C_sE$. Let $s = j_1, j_2, \dots, j_{|s|} \in \text{Seq}(J)$. As $K_{j_{|s|}}E \in S_1[E]$, and $K_{j_{|s|-1}}K_{j_{|s|}}E \in S_2[E]$, and so on, up to $K_{j_1} \cdots K_{j_{|s|}}E \in S_{|s|}[E]$, therefore

$$C_BE = \bigcap_{n \in \mathbb{N}} \bigcap_{F \in S_n[E]} F \subset \bigcap_{s \in \text{Seq}(J)} K_{(s)}E = C_sE$$

ii) We want that if $K_j(F \cap G) = K_jF \cap K_jG$ for all agents $j \in J$ and events $F, G \in 2^\Omega$, then $C_BE = C_eE = C_sE$. By Proposition 1, it is enough to show $C_BE = C_sE$. We already have $C_BE \subset C_sE$, so only need $C_sE \subset C_BE$. Consider a knowledge question $q = (j, \{q_1, q_2, \dots, q_n\})$. As each individual knowledge operator K_j is distributive, $K_qE = K_jK_{q_1}E \cap \dots \cap K_jK_{q_n}E$. Define the length of q by $|q| = 1 + \sum_{i=1}^n |q_i|$ as in the proof of Proposition 8. Suppose, for the purposes of induction, that for each knowledge question r of length at most m , there exists some $A \subset \text{Seq}(J)$ such that $K_rE = \bigcap_{s \in A} K_{(s)}E$. Let $|q| = m + 1$. Then there exists $A_1, \dots, A_n \subset \text{Seq}(J)$ such that

$$\begin{aligned} K_qE &= K_jK_{q_1}E \cap \dots \cap K_jK_{q_n}E \\ &= K_j \left(\bigcap_{s \in A_1} K_{(s)}E \right) \cap \dots \cap K_j \left(\bigcap_{s \in A_n} K_{(s)}E \right) \\ &= \left(\bigcap_{s \in A_1} K_jK_{(s)}E \right) \cap \dots \cap \left(\bigcap_{s \in A_n} K_jK_{(s)}E \right) \\ &= \bigcap_{s \in (A_1 \cup \dots \cup A_n)} K_jK_{(s)}E \end{aligned}$$

So there exists $A = A_1 \cup \dots \cup A_n \subset \text{Seq}(J)$ such that $K_qE = \bigcap_{s \in A} K_{(s)}E$. Therefore

$$K_qE = \bigcap_{s \in A} K_{(s)}E \supset \bigcap_{s \in \text{Seq}(J)} K_{(s)}E = C_sE$$

for each knowledge question q , and so

$$C_BE = \bigcap_{q \in Q} K_qE \supset C_sE$$

So $C_BE = C_sE = C_eE$ for any event $E \in 2^\Omega$, as required.

iii) It is sufficient to provide an example. Let $\Omega = \{a, b, c\}$, $J = \{1, 2\}$. Define knowledge operator K_1 by $K_1\{a, b\} = \{a, c\}$, $K_1\{a, c\} = \{b, c\}$, $K_1\{a\} = K_1\{b\} = K_1\{c\} = \emptyset$, and $K_1E = E$ for all other events $E \in 2^\Omega$. Let $K_2 = K_1$.

Since $K_1 = K_2$, then $\lambda = K_1$, so $C_eE = C_sE$ for all events $E \in 2^\Omega$. Therefore

$$C_eE = C_sE = \bigcap_{s \in \text{Seq}(J)} K_{(s)}E = \bigcap_{n=1}^{\infty} K_1^n E$$

and

$$\begin{aligned} C_e\{a, b\} &= C_s\{a, b\} = \{a, c\} \cap \{b, c\} \cap \{b, c\} \cap \dots = \{c\} \\ C_e\{c\} &= C_s\{c\} = \emptyset \cap \emptyset \cap \emptyset \cap \dots = \emptyset \end{aligned}$$

Therefore, $C_e\{a, b\} \not\subset C_e C_e\{a, b\}$, so C_e does not satisfy positive introspection. Hence C_e not equal to C_B . In this instance, $C_B\{a, b\} = \emptyset$.

Moreover, K_1 and K_2 satisfy Monotonicity. Monotonicity requires that $E \subset F \implies KE \subset KF$. In this case, if $E \subsetneq F$, then either $KE = \emptyset$ or $KF = \Omega$. So in any case K_1 and K_2 are Monotonic.

Therefore, it is possible that even though all agents' operators are Monotonic, nonetheless $C_B \neq C_e$ and $C_B \neq C_s$.

iv) It is sufficient to provide an example. Let $\Omega = \{a, b, c\}$, $J = \{1, 2\}$. Define knowledge operator K_1 by $K_1\{a, b\} = \{a, c\}$, $K_1\{a, c\} = \{a, b\}$, $K_1\Omega = \emptyset$, and $K_1E = \Omega$ for all other events $E \in 2^\Omega$. Let $K_2 = K_1$.

Since $K_1 = K_2$, then $\lambda = K_1$, so $C_e E = C_s E$ for all events $E \in 2^\Omega$. Therefore

$$C_e E = C_s E = \bigcap_{s \in \text{Seq}(J)} K_{(s)} E = \bigcap_{n=1}^{\infty} K_1^n E$$

and

$$\begin{aligned} C_e\{a, b\} &= C_s\{a, b\} = \{a, c\} \cap \{a, b\} \cap \{a, c\} \cap \dots = \{a\}, \\ C_e\{a\} &= C_s\{a\} = \Omega \cap \emptyset \cap \Omega \cap \dots = \emptyset \end{aligned}$$

Therefore, $C_e\{a, b\} \not\subset C_e C_e\{a, b\}$, so C_e does not satisfy positive introspection, and so is not equal to C_B . In this instance $C_B E = \emptyset$ for all events $E \in 2^\Omega$.

Moreover, K_1 satisfies Conjunction. Let $E, F \in 2^\Omega$. For $E \subset F$, or $F \subset E$, it is automatic that $KE \cap KF \subset K(E \cap F)$. For $E \cap F = \emptyset$, $K(E \cap F) = K\emptyset = \Omega$, so $KE \cap KF \subset K(E \cap F)$. Finally for E, F such that $E \not\subset F$, $F \not\subset E$ and $E \cap F \neq \emptyset$, then $|E \cap F| = 1$, so $K(E \cap F) = \Omega$, and $KE \cap KF \subset K(E \cap F)$. So in every case $KE \cap KF \subset K(E \cap F)$.

Therefore, it is possible that even though all agents' operators satisfy Conjunction, nonetheless $C_B \neq C_e$ and $C_B \neq C_s$. \square

Proof of Proposition 12:

Let $K_j \Omega = \Omega$ for all $j \in J$, and $S_i[E]$ be defined as in Definition 9. We immediately have $S_1[\Omega] = \{\Omega\}$, and indeed $S_n[\Omega] = \{\Omega\}$ for all $n \in \mathbb{N}$. Therefore $C_B \Omega = \Omega$ and C_B satisfies Awareness.

Conversely, if $K_j\Omega \neq \Omega$ for some $j \in J$, then $C\Omega \subset K_j\Omega \subsetneq \Omega$, so $C\Omega \neq \Omega$. \square

Proof of Proposition 13:

Let $K_j(E \cap F) = K_jE \cap K_jF$ for all agents $j \in J$ and events $E, F \in 2^\Omega$. By Proposition 11, $C_B = C_s$. For any $E, F \in 2^\Omega$,

$$C_B(E \cap F) = C_s(E \cap F) = \bigcap_{s \in \text{Seq}(J)} K_{(s)}(E \cap F)$$

Fix a sequence $s = s_1, s_2, \dots, s_n \in \text{Seq}(J)$. As $K_{s_n}(E \cap F) = K_{s_n}E \cap K_{s_n}F$, we have $K_{s_{n-1}}K_{s_n}(E \cap F) = K_{s_{n-1}}K_{s_n}E \cap K_{s_{n-1}}K_{s_n}F$, and so on up to

$$K_{(s)}(E \cap F) = K_{s_1} \cdots K_{s_n}(E \cap F) = K_{s_1} \cdots K_{s_n}E \cap K_{s_1} \cdots K_{s_n}F$$

Therefore

$$\bigcap_{s \in \text{Seq}(J)} K_{(s)}(E \cap F) = \bigcap_{s \in \text{Seq}(J)} (K_{(s)}E \cap K_{(s)}F) = \bigcap_{s \in \text{Seq}(J)} K_{(s)}E \cap \bigcap_{s \in \text{Seq}(J)} K_{(s)}F$$

and as $C_BE = \bigcap_{s \in \text{Seq}(J)} K_{(s)}E$, then

$$C_B(E \cap F) = C_BE \cap C_BF$$

as required. \square

Proof of Proposition 14:

Define the depth of a knowledge question recursively, where $\text{depth}(q) = 1$ for $q = (j, \emptyset)$ and for $q = (j, \{q_1, q_2, \dots, q_n\})$,

$$\text{depth}((j, \{q_1, q_2, \dots, q_n\})) = \max\{\text{depth}(q_1), \dots, \text{depth}(q_n)\} + 1$$

In particular, $\text{depth}(q) = m$ if and only if $q \in Q_m \setminus Q_{m-1}$.

We will show K_q satisfies Monotonicity by strong induction over the depth of q . By assumption each K_j satisfies Monotonicity, so K_q satisfies Monotonicity for all questions q of depth one. Suppose, for the purposes of induction, that all knowledge questions of depth at most m satisfy monotonicity. Let q be of depth $m + 1$, so that $q = (j, \{q_1, q_2, \dots, q_n\})$ where

each q_i has depth at most m . By the inductive hypothesis, each q_i is Monotonic. Therefore, for any events $E, F \in 2^\Omega$ where $E \subset F$,

$$K_{q_1}E \cap \cdots \cap K_{q_n}E \subset K_{q_1}F \cap \cdots \cap K_{q_n}F$$

As K_j is monotonic,

$$K_j(K_{q_1}E \cap \cdots \cap K_{q_n}E) \subset K_j(K_{q_1}F \cap \cdots \cap K_{q_n}F)$$

so $K_qE \subset K_qF$ as required. Therefore, by strong induction, K_q is monotonic for every knowledge question q .

Now let $E \subset F$. Then

$$\bigcap_{q \in Q} K_qE \subset \bigcap_{q \in Q} K_qF$$

which is to say $C_BE \subset C_BF$, as required. \square

Proof of Proposition 15:

It is sufficient to provide an example. Let $\Omega = \{a, b, c\}$, and $J = \{1\}$. Define the knowledge operator K by $K\{a, b\} = \{a, b\}$, $K\{a, c\} = \{a, c\}$, $K\{a\} = \{a, b, c\}$, and $KE = \emptyset$ for all other events $E \in 2^\Omega$.

By exhaustive search, K satisfies conjunction, $KE \cap KF \subset K(E \cap F)$ for all events E, F . However Behavioral Common Knowledge is

$$\begin{aligned} C_B\{a, b\} &= \{a, b\}, & C_B\{a, c\} &= \{a, c\} \\ C_BE &= \emptyset, & \text{otherwise} \end{aligned}$$

Thus $C_B\{a, b\} \cap C_B\{a, c\} = \{a\}$, but $C_B(\{a, b\} \cap \{a, c\}) = C_B\{a\} = \emptyset$. Therefore C_B does not satisfy Conjunction. \square

Proof of Proposition 16:

Fix an event $E \in 2^\Omega$, and suppose there is some agent $j \in J$ such that $K_jE \subset E$. Without loss of generality, let $K_1E \subset E$. As $C_BE \subset K_jE$ for each agent $j \in J$, in particular $C_BE \subset K_1E$. But $K_1E \subset E$, so $C_BE \subset E$, as required. \square

Proof of Proposition 17:

It is sufficient to provide an example. Let $\Omega = \{a, b, c\}$, and $J = \{1, 2\}$. Define the knowledge operators K_1 and K_2 as

E	K_1E	K_2E	C_BE
\emptyset	\emptyset	\emptyset	\emptyset
$\{a\}$	$\{a, b\}$	$\{a, c\}$	$\{a\}$
$\{b\}$	$\{b\}$	$\{b\}$	$\{b\}$
$\{c\}$	$\{c\}$	$\{c\}$	$\{c\}$
$\{a, b\}$	$\{a, b\}$	$\{a, b\}$	$\{a, b\}$
$\{a, c\}$	$\{a, c\}$	$\{a, c\}$	$\{a, c\}$
$\{b, c\}$	$\{b\}$	$\{c\}$	\emptyset
Ω	Ω	Ω	Ω

By exhaustive search, K_1 and K_2 satisfy negative introspection, $\neg KE \subset K\neg KE$ for all events E . However Behavioral Common Knowledge does not, as for the event $\{a\}$;

$$\neg C_B\{a\} = \{b, c\}, \quad C_B\neg C_B\{a\} = \emptyset$$

Thus there exists some event E such that $\neg C_BE \not\subset C_B\neg C_BE$. Consequently, the Behavioral Common Knowledge operator C_B does not satisfy negative introspection. \square

Proof of Proposition 18:

It is sufficient to provide an example. As in the proof of Proposition 11(iv), let $\Omega = \{a, b, c\}$, $J = \{1, 2\}$. Define knowledge operator K_1 by $K_1\{a, b\} = \{a, c\}$, $K_1\{a, c\} = \{a, b\}$, $K_1\Omega = \emptyset$, and $K_1E = \Omega$ for all other events $E \in 2^\Omega$. Let $K_2 = K_1$.

As $K_j = K_i$ for all $i, j \in J$, then $C_e = C_s = C_g$. As noted in the proof of Proposition 11(iv), $C_e\{a, b\} \not\subset C_eC_e\{a, b\}$. Thus, as $C_g = C_e$, the group common knowledge C_g does not satisfy Positive Introspection. By Proposition 10, Behavioral Common Knowledge C_B does satisfy Positive Introspection. Therefore $C_g \neq C_B$. \square

Proof of Lemma 3:

Let $k \geq 2$ and $E \in 2^\Omega$. Then

$$C_kE = \bigcap_{n \in \mathbb{N}} C_{k-1}^n E \subset \bigcap_{n=1} C_{k-1}^n E = C_{k-1}E$$

□

Proof of Proposition 19:

By Lemma 3, the sequence of events $(C_1E, C_2E, C_3E, \dots)$ is non-increasing. As Ω is finite, and thus 2^Ω is finite, this sequence eventually reaches a fixed point. Therefore for sufficiently large M , the following are equal

$$C_*E = \bigcap_{k \in \mathbb{N}} C_kE = C_ME = C_{M+1}E = \bigcap_{n \in \mathbb{N}} C_M^n E$$

As $C_ME = C_{M+1}E \subset C_M^n E$ for all $n \in \mathbb{N}$, in particular $C_ME \subset C_MC_ME$. As $C_* = C_M$, then $C_*E \subset C_*C_*E$. □

Proof of Proposition 20:

The structure of the proof is to show that $C_*E \in S_n[E]$ for some large n , and thereby conclude that $C_BE \subset C_*E$.

Fix an event E . Let $M = |2^\Omega|$ so that for any function $K : 2^\Omega \rightarrow 2^\Omega$, $\bigcap_{n \in \mathbb{N}} K^n E = \bigcap_{n=1}^M K^n E$. Following the proof of Proposition 11, $\lambda^n E \in S_{n|J|}[E]$ for all $n \in \mathbb{N}$. In particular $\lambda^n E \in S_{M|J|}$ for all $n \leq M$, so $\bigcap_{n=1}^M \lambda^n E \in S_{M|J|+M}[E]$. As

$$C_1E = \bigcap_{n \in \mathbb{N}} \lambda^n E = \bigcap_{n=1}^M \lambda^n E$$

therefore $C_1E \in S_{M|J|+M}[E]$.

Similarly, $\lambda^n C_1E \in S_{M|J|+M+n|J|}[E]$ for all $n \in \mathbb{N}$. In particular $\lambda^n C_1E \in S_{M|J|+M+M|J|}$ for all $n \leq M$, so $\bigcap_{n=1}^M \lambda^n C_1E \in S_{M|J|+M+M|J|+M}[E]$. As

$$C_1C_1E = \bigcap_{n \in \mathbb{N}} \lambda^n C_1E = \bigcap_{n=1}^M \lambda^n C_1E$$

therefore $C_1C_1E \in S_{2M|J|+2M}[E]$. Continuing in this manner gives

$$C_1^k \in S_{kM|J|+kM}[E] \quad , \text{ for } k \in \mathbb{N}$$

Then

$$C_2E = \bigcap_{n \in \mathbb{N}} C_1^n E = \bigcap_{n=1}^M C_1^n E \in S_{M^2|J|+M^2+M}$$

Each movement from C_k to C_{k+1} involves multiplying the subscript of S by M then adding M as seen above. In general

$$C_k E \in S_{M^k | J | + M^k + M^{k-1} + \dots + M} [E]$$

As $C_* = C_K$ for sufficiently large K ,

$$C_* E = C_K E \in S_{M^K | J | + M^K + M^{K-1} + \dots + M} [E]$$

as $C_B E = \bigcap_{n \in \mathbb{N}} \bigcap_{F \in S_n[E]} F$, and $C_* E \in S_n[E]$ for some n , therefore $C_B E \subset C_* E$. □

References

- [1] Aumann, R., (1976) Agreeing to Disagree. *Annals of Statistics*, 4(6), 1236-1239.
- [2] Aumann, R., & Brandenburger, A., (1995) Epistemic Conditions for Nash Equilibrium. *Econometrica*, 63(5), 1161-180.
- [3] Bacharach, M., (1985) Some extensions to a claim of Aumann in an axiomatic model of knowledge. *Journal of Economic Theory*, 37(1), 167-190.
- [4] Baston C., & Capretta V. (2018) The Coinductive Formulation of Common Knowledge. In: Avigad J., Mahboubi A. (eds) *Interactive Theorem Proving. ITP 2018. Lecture Notes in Computer Science*, vol 10895. Springer, Cham.
- [5] Belle, V., & Lakemeyer, G., (2015) Only knowing meets common knowledge. *Proceedings of the 24th International Conference on Artificial Intelligence*. 2755-2761.
- [6] Brandenburger, A., Dekel, E., & Geanakoplos, J. D., (1992) Correlated Equilibrium with Generalized Information Structures. *Games and Economic Behavior*, 4(2), 182-201.
- [7] Chellas, B. F., (1980) *Modal Logic: An Introduction*. Cambridge University Press.
- [8] Kruger, J., & Dunning, D., (1999) Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments. *Journal of Personality and Social Psychology*. 77(6), 1121-1134.

- [9] Fagin, R., Geanakoplos, J., Halpern, J.Y., & Vardi, M.Y., (1999) The hierarchical approach to modeling knowledge and common knowledge. *International Journal of Game Theory* 28(3), 331-365.
- [10] Friedell, M., (1969) On the Structure of Shared Awareness. *Behavioral Science* 14, 28-39.
- [11] Fukuda, S., (2016) *The Existence of Universal Knowledge Spaces*. Retrieved from https://websfukuda.files.wordpress.com/2015/03/universal_knowledge_space_jmp_oct_28_2016.pdf
- [12] Geanakoplos, J. D., (1989) Game Theory without Partitions, and Applications to Speculation and Consensus. Cowles Foundation Discussion Paper No. 914.
- [13] Geanakoplos, J. D., (1992) Common Knowledge. *Handbook of Game Theory* (pp. 1437-1496). North Holland: Elsevier.
- [14] Gintis, H., (2010) Rationality and Common Knowledge. *Rationality and Society*, 22(3), 259-282.
- [15] Halpern, J., (2003) *Reasoning about Uncertainty*. The MIT Press.
- [16] Harsanyi, J. C., (1967-8) Games with Incomplete Information Played by Bayesian Agents I-III. *Management Science*, 14, 159-182, 320-334, 486-502.
- [17] Hart, S., Heifetz, A., & Samet, D., (1996) “Knowing Whether”, “Knowing That”, and The Cardinality of State Spaces. *Journal of Economic Theory*, 70(1), 249-256.
- [18] Hughes, G. E., & Cresswell, M. J., (1968) *An Introduction to Modal Logic*. London: Methuen.
- [19] Kahneman, D., (2011) *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- [20] Kripke, S., (1963) Semantical Analysis of modal logic I: normal modal propositional calculi. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 9, 67-96.
- [21] Lewis, D., (1969) *Convention: A philosophical study*. Harvard University Press.

- [22] Lismont, L., & Mongin, P., (1994) On the logic of common belief and common knowledge. *Theory and Decision*, 37(1), 75-106.
- [23] Mertens, J.F. & Zamir, S. (1985) Formulation of Bayesian analysis for games with incomplete information. *International Journal of Game Theory*, 14(1), 1-29.
- [24] Rubinstein, A., (1988) *Modelling Bounded Rationality*. MIT Press.
- [25] Rubinstein, A., (1989) The Electronic Mail Game: Strategic Behavior Under “Almost Common Knowledge”. *The American Economic Review* 79(3), 385-391.
- [26] Samuelson, L., (2004) Modeling Knowledge in Economic Analysis. *Journal of Economic Literature*, 42(4), 367-403.
- [27] Tversky, A., & Kahneman, D., (1983) Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychological Review*, 90(4), 293-315.
- [28] van Benthem, J., Eijck, J., & Kooi, B., (2005) Common knowledge in update logics. *Proceedings of the 10th conference on Theoretical aspects of rationality and knowledge*. 253-261.
- [29] van Ditmarsch, H., Halpern, J.Y., van der Koek, J., & Kooi, B. (Ed.). (2015) *Handbook of Epistemic Logic*. Location: Milton Keynes, UK.