Noname manuscript No.
(will be inserted by the editor)

Substantive Assumptions in Interaction: A Logical Perspective

Olivier Roy · Eric Pacuit

Abstract In this paper we study substantive assumptions in social interaction. By substantive assumptions we mean contingent assumptions about what the players know and believe about each other's choices and information. We explain why such assumptions are important, we show how to identify and compare them, formally, and we also show that there exist contexts where no substantive assumptions are being made. Towards the end of the paper we briefly explain the relation between such structures and a number of other "large" structures studied in epistemic game theory.

Introduction

R. Aumann [1987] famously wrote that common knowledge of the partition struture is:

> "*not* an assumption, but a 'theorem', a tautology; [...] implicit in the model itself"

What are such "theorems", implicit in the model itself? Can one distinguish them from *substantive* assumptions, for instance common knowledge of rationality? Are there models where no substantive assumptions are being made?

We provide a mathematical answer to these questions. In Section 1 we explain the importance of precisely distinguishing substantive from what we call *structural* assumptions, and define the concepts that we use later on. In Section 2 we provide a formal, syntactic analysis of substantive assumptions,

Olivier Roy
Munich Center for Mathematical Philosophy
Ludwig-Maximilians-Universität München
Germany E-mail: Olivier.Roy@lmu.de

Eric Pacuit
TiLPS
Tilburg University
The Netherlands E-mail: E.J.Pacuit@uvt.nl

and show that there are indeed models where no substantive assumptions are being made. Section 5 situates this result within the extensive literature on “large” structures.

We focus on qualitative, also called logical models of information in games and social interaction [Fagin et al., 1995, Aumann, 1999, van Ditmarsch et al., 2007]. These are models of *all out*, as opposed to *graded* attitudes.[1] They have proved both historically and conceptually important in epistemic game theory. Furthermore, they naturally lend themselves to a syntactic analysis, such as the one we provide in Section 2. It would be an important task to port the present analysis to models of graded attitudes, for instance Harsanyi type spaces, but we leave this for future work.

1 Motivations and Basic Definitions

One of the fundamental insights behind the epistemic approach to games is that strategic interaction takes place in specific *contexts*, c.f. [Aumann and Dreze, 2008]. A context is a description of the players' information about each other, including information about the other's information - beliefs about others' beliefs, knowledge about the knowledge of others, and so on.

Informally[2], a substantive assumption is an assumption made in a specific (class of) context of interaction, but that could be relaxed. In this paper we are interested in *epistemic* substantive assumptions, bearing on what agents *know* or *believe* in a certain context of strategic interaction. Common knowledge of rationality and common priors are paradigmatic examples of such substantive assumptions. They are often used in epistemic modeling, but one can easily construct models of games where either of these condition fails to obtain.

Substantive assumptions play a key role in the epistemic foundations of game theory.[3] Epistemic characterization results can, arguably, be seen as drawing the behavioral or the normative consequences of making certain substantive assumptions. We have already mentioned common knowledge of rationality, which implies playing rationalizable strategies [Bernheim, 1984, Pearce, 1984]. The more recent characterization of self-admissible strategies in terms of common knowledge of admissibility [Brandenburger et al., 2008] provides another clear example.

Not all assumptions are substantive, though. As pointed out in the quote above, some assumptions[4] are “implicit” or built into the [epistemic] model one is working with. In partitional, a.k.a. S5 models of knowledge,[5] for instance,

[1] This terminology comes from philosophical epistemology. See e.g. [Huber and Schmidt-Petri, 2009].

[2] We give a formal definition below.

[3] See for instance the discussion in [Samuelson, 2004] and the references in [Moscati, 2009].

[4] Of course the quote opens by saying that these are “*not* assumptions”. We rather view them as assumptions of a different kind, hence our terminology.

[5] See again: [Fagin et al., 1995, Aumann, 1999, van Ditmarsch et al., 2007].

positive and negative introspection[6] are assumptions that cannot be relaxed—without leaving that class of models, of course. But some assumptions are even more tenacious: so-called "logical omniscience" or what philosophers call "extensionality" seem to be deeply entrenched in the kind of models that are common in epistemic game theory and epistemic logic.[7]

It is thus important to identify and clearly tell apart substantive from structural assumptions, if only to pinpoint their respective roles in epistemic analysis of strategic interaction. This is what we do in the coming sections, starting with a syntactic approach to all-out attitudes in games.

1.1 Finitary epistemic languages and Kripke Structures

Our approach is primarily syntactic, and so we start by introducing (finitary) languages, and then move to so-called Kripke structures. Syntactic approaches have a long history in logic [Hintikka, 1962] and game theory [Aumann, 1999]. Such languages give a more coarse-grained view of knowledge and beliefs than probabilistic representations, of course. They are nevertheless sufficient to analyze certain solution concepts epistemically, for instance iterated elimination of strictly dominated strategies and Nash equilibrium [van Benthem et al., 2011]. Furthermore, syntax gives us greater generality—formal languages like the one we present below can be interpreted on a large variety of models—and it proves convenient to "compare" different informational states, as we do in the next section.

Let N be a finite set of agents and PROP a countable set of propositions.

Definition 1 (Finitary Epistemic Language)
A finitary epistemic language $\mathcal{L}_{EL}$ is recursively defined as follows:

$$\phi := p \mid \neg\phi \mid \phi \wedge \phi \mid \Box_i \phi \mid \Box^*_G \phi$$

where i ranges over N, p over PROP, and $\emptyset \neq G \subseteq N$.

This is the standard multi-agent epistemic logic with a "common knowledge" modality [Fagin et al., 1995, van Ditmarsch et al., 2007]. Formula $\Box_i\phi$ can be read as "agent i knows that ϕ" or as "agent i believes that ϕ", depending on the properties of this operator.[8] The formula $\Box^*_G\phi$ should be read as "it

[6] Positive introspection means that if an agent knows a certain fact ϕ, then she knows that she knows ϕ. Negative introspection means that if an agent doesn't know that ϕ, then she knows that she doesn't know that ϕ.

[7] Work on awareness [Fagin and Halpern, 1987, Modica and Rustichini, 1994, Halpern, 2001, Heifetz et al., 2006], however, have made major steps in lifting these assumptions.

[8] Knowledge is usually assumed to satisfy the so-called K axiom: $\Box_i(\phi \to \psi) \to (\Box_i\phi \to \Box_i\psi)$ and the "necessitation rule": from ϕ a theorem, infer $\Box_i\phi$—although this need not be the case, depending on the underlying class of structures one is working with—as well as the S5 axioms: (T) $\Box_i\phi \to \phi$; (4) $\Box_i\phi \to \Box_i\Box_i\phi$ and (5) $\neg\Box_i\phi \to \Box_i\neg\Box_i\phi$. For beliefs one usually drops (T), beliefs can mistaken, after all, and replace it with (D), $\Box_i\phi \to \neg\Box\neg_i\phi$, ensuring consistent attitudes. Unless stated otherwise, in what follows we use $\to$ for the material implication.

is common knowledge/beliefs among group G that ϕ". Both $\Box_i$ and $\Box^*_G$ have their duals, $\Diamond_i$ and $\Diamond^*_G$, defined respectively as $\neg\Box_i\neg$ and $\neg\Box^*_G\neg$. One could also work with languages containing both (common) knowledge and beliefs operators, as well as with more expressive languages such as Propositional Dynamic Logic [Harel et al., 2000].

This language is finitary in the sense that it allows only finite conjunctions ($\wedge$) and disjunctions, as well as finite, but unbounded stacking of epistemic operators. The $\Box^*_G\phi$ modality has an infinitary character, it being equivalent to the infinite conjunction of $E^{n+1}\phi =_{df} \bigwedge_i (K_i E^n \phi)$ for all $n < \omega$, but it can be finitely axiomatized—c.f. the references above. Some, but not all, of the observations below carry over to infinitary versions of this language, for instance those studied in [Segerberg, 1994, Heifetz, 1999]. We leave this generalization for future work.

Epistemic languages can be interpreted in a wide variety of structures, from "Kripke" or relational structures [Blackburn et al., 2001] to partition structures [Heifetz and Samet, 1998a] and topological spaces and "neighborhoods" [Blackburn et al., 2006, chap.1]. We use here Kripke structures as an illustrative example.

Let, again, PROP be a countable set of atomic propositions.

Definition 2 (Kripke Structure) A *Kripke structure* $\mathcal{M}$ is a tuple $\langle W, N, \mathcal{R}, V\rangle$ where W is a nonempty set of *states*, N is a finite set of *agents*, $\mathcal{R}$ is a collection of *binary relations* on W and $V : W \to 2^{\text{PROP}}$ is a *valuation* function from W to subsets of PROP. Given a relation $R \in \mathcal{R}$ and state $w \in W$, we write $R[w]$ for $\{w' : wRw'\}$. A *pointed Kripke structure* is a pair $(\mathcal{M}, w)$.

It is usually assumed that $\mathcal{R}$ contains at least one relation R_i for each agent $i \in N$. We write R^+_G for the transitive closure of the union of the relations R_i for $i \in G$. This relation is used to interpret the common belief modality. For common knowledge one uses the reflexive-transitive closure R^*_G. The epistemic language is then interpreted in Kripke structure as follows, with $\Box^*_G$ read as "common knowledge":

Definition 3 Interpretation of $\mathcal{L}_{EL}$ in Kripke structures.

$$
\begin{array}{lll}
\mathcal{M}, w \Vdash p & \text{iff} & p \in V(w) \\
\mathcal{M}, w \Vdash \neg\phi & \text{iff} & \mathcal{M}, w \nVdash \phi \\
\mathcal{M}, w \Vdash \phi \wedge \psi & \text{iff} & \mathcal{M}, w \Vdash \phi \text{ and } \mathcal{M}, w \Vdash \psi \\
\mathcal{M}, w \Vdash \Box_i\phi & \text{iff} & \forall v (\text{if } wR_iv \text{ then } \mathcal{M}, v \Vdash \phi) \\
\mathcal{M}, w \Vdash \Box^*_G\phi & \text{iff} & \forall v (\text{if } wR^*_Gv \text{ then } \mathcal{M}, v \Vdash \phi)
\end{array}
$$

Subsets of W are usually called an *event*, and an event E is *definable* in a given language whenever there is a formula ϕ of that language such that $E = \{w : \mathcal{M}, w \Vdash \phi\}$. It is well-known that, in general, not all events are definable by formulas of finitary epistemic languages, but that there are elegant model-theoretic characterizations of classes of definable Kripke structures.[9]

[9] For example, a class of elementary pointed Kripke structures is definable by a (set of) modal formulas iff the class is closed under bisimulations and ultraproducts and its

Kripke structures can be seen as models for games, i.e. contexts, in the obvious way. The propositional variables then range over strategies profiles of the underlying game, making the valuation V the usual "strategy function" in interactive epistemology [Aumann, 1999, Board, 2002, Stalnaker, 1999]. One can also equip oneself with a range of primitive propositions describing the agents' preferences over profiles in the game[10], and whether a given strategy choice is "rational" at a given state.[11]

2 Substantive and Structural Assumptions

We want to pinpoint what are substantive assumptions, and what distinguish them from structural assumptions, that is "theorems" that are built into a specific class of models that one works with. We do so using the standard notions of axiom systems, logical consequence, consistency, and completeness with respect to given classes of (Kripke) structures.[12]

An *axiom system* Λ in $\mathcal{L}_{EL}$ is a set of designated formulas, called *axioms*, together with a set of *inference rules*. For instance, the formulas and the rule mentioned in footnote 8, together with all propositional tautologies and the rule *modus ponens*, form the axiom system known as S5. A *derivation* in an axiom system is a finite sequence of formulas that are either axioms or obtainable from some formulas earlier in the sequence by one of the inference rules. When a formula ϕ can be derived from a set of formulas Σ in a given axiom system Λ, we write $\Sigma \vdash_\Lambda \phi$. Call a non-empty set T of formula in $\mathcal{L}_{EL}$ a *theory*. A theory is *consistent* given an axiom system Λ, or Λ-consistent, for short, if it is not the case that $T \vdash_\Lambda \bot$.[13] T is *maximally consistent* if it is consistent and there is no other theory $T' \supsetneq T$ that is also consistent.

There is of course a close correspondence between axiom systems and sets of formulas that are "valid" in all Kripke structures satisfying certain properties.[14] S5, for instance, is *sound* and *complete* with respect to the class of Kripke structures where the relations R_i for each agent i are equivalence relations: all axioms are valid and the inference rules preserve validity (soundness) and all valid formulas in that class of structures are provable in S5 (completeness).[15]

---

complement is closed under ultrapowers [Blackburn et al., 2001, Theorem 2.75, pg.107]. Bisimulations are defined later in the paper. See also [Blackburn et al., 2001] for the relevant definitions.

[10] This can also be done in a "modal" way. See [Van Benthem et al., 2009].

[11] See e.g. [de Bruin, 2010].

[12] The treatment in what follows is entirely standard, but for reasons of space is bound to be elliptic. Details can be found in any textbook on modal logic, for instance [Blackburn et al., 2001, van Benthem, 2010].

[13] We write $\bot$ for contradiction, for instance $p \wedge \neg p$.

[14] Call a frame a Kripke model with the valuation V omitted. A formula is valid on a frame if it is true in all states of all models based on that frame. By "valid in all Kripke structures..." we mean valid on a given class of frames satisfying certain properties.

[15] See again [Blackburn et al., 2001] for more on this.

This connection allows to define syntactically what are substantive assumptions in a given class of Kripke structures, provided the logical system we work with is sound and complete with respect to that class. From now on we will assume that we are working with classes of Kripke structures that are completely axiomatizable.

Definition 4 Let Λ be a given axiom system in $\mathcal{L}_{EL}$, T be a maximally Λ-consistent theory and ϕ a formula in T. Then ϕ is a *substantive assumption* if there is another maximally Λ-consistent theory T' that does not contain ϕ. An *epistemic* substantive assumption is a substantive assumption of the form $\Box_i\phi$ for some agent i, or $\Box^*_G\phi$ for some $G \subseteq I$. If $\phi \in T$ is not a substantive assumption we call it a *structural assumption.*

This definition[16] draws a sharp line between substantive and structural assumptions, and does it in a way which, arguably captures the intuitions mentioned in the introduction. We introduced substantive assumptions, informally, as assumptions that hold in a given (class of) models, but that could be relaxed. Now observe, first, that negations of structural assumptions are neither structural nor substantive assumptions. Structural assumptions are logical truths, so to speak, valid in all structures of a given class, and their negations are "logical impossibilities".[17] One is free, on the other hand, to make or drop substantive assumptions. Both are, by definition, consistent or "logically possible" in a given class of structures. This the intuitive idea that we intended to capture.

Examples might help to make these notions concrete. In S5, both positive and negative introspection turn out to be structural assumptions according to this definition. But if one moves to the class of *all* Kripke structures, then these are, of course, substantive assumptions. On the other hand, so-called logical omniscience, embodied by the K axiom and the Necessitation rule, are structural assumptions in a fundamental sense: to drop them one has to move to a different semantics for informational attitudes, for instance neighborhood semantics.

3 Comparing Substantive Assumptions

For epistemic analysis of games the crucial notion is that of an epistemic substantive assumption. Our recurrent example, common knowledge of rationality, is a clear case of such an epistemic, substantive assumption. This example

[16] We thank Kit Fine for pointing out a problem in an earlier version of this definition.

[17] Here the syntactic approach proves again very useful. In the words of [refs EL5Q, pp]: "In a semantic model it's difficult to express what it means for something to be logically impossible. In a semantic model, something that is logically impossible is represented by an empty set; but if an event is represented by an empty set, that doesn't mean it's logically impossible. You have to have a *universal* semantic model to say that, and these models are large and clumsy."

also shows that substantive assumptions can have important behavioral consequences and that, from a methodological point of view, it is important to pinpoint them precisely.

Intuitively, such assumptions correspond to facts that are contingently known to certain agents in a given context. The crucial properties of these assumptions is that they *could* be relaxed, by moving to another context (in the same class). In such a structure, the agents, intuitively, *know/believe less*. How to make this intuition formal?

Extending the state space is the usual technique to find models where agents know less, but this will not work in full generality. Many extensions of a given space do not translate in changes in informational attitudes.[18] The correct notion here requires a quantification over all possible extensions of a given model, a notion which is more naturally captured by our syntactic definition of substantive assumptions.

The task of comparing informativeness, and by the same token substantive assumptions, is, however, not completely trivial at the syntactic level either. The naive procedure would consist in checking whether all the formulas of the form $\Box_i\phi$ contained in one maximally Λ-consistent theory are also present in some other one. This would boil down to check whether everything known in one theory is known in the other. This doesn't work in the general case. If agents are negatively introspective, as in S5 structures for instance, ignorance of a given fact induces knowledge of that very ignorance. One obviously wants to discard this kind of self-knowledge in comparing substantive assumptions. They would make most pairs of different maximally Λ-consistent theories incomparable according to the naive ordering. On the other hand, comparing only first-order knowledge or beliefs, i.e. attitudes bearing to non-epistemic facts in a given context, will not do either, as assumptions about the information of *others* is of crucial importance—think again of common knowledge of rationality here.[19]

To circumvent these difficulties we use a notion of comparison of informativeness that has been put forward by R. Parikh [1991] in order to analyze non-monotonic phenomena tied to epistemic reasoning. Let $\mathcal{L}_{EL}$ be a finitary epistemic language and Λ a logical system in it. We write $sub(\phi)$ for the set of sub-formulas of ϕ, not necessarily proper ones. In this definition it is convenient to assume that all $\Box\phi$ are re-written as $\neg\Diamond\neg\phi$. This can be done without loss of generality.

Definition 5 Given T_1, T_2 two maximally Λ-consistent theories, we say that the agents in T_1 *know at least as much* than in T_2, written, $T_1 \geq T_2$, iff for

[18] Put in logical terms, epistemic languages like the one defined above are "invariant" under a good number of model transformations [Blackburn et al., 2001]. The same point can also be made model-theoretically, c.f. the notion of knowledge morphism in [Heifetz and Samet, 1998a].

[19] It should be observed that in non-introspective contexts, cases where agents do have information about their own information are of obvious importance, and will indeed count as substantive assumptions according to our definition.

all formula $\phi \in (T_1 \cup T_2) \setminus (T_1 \cap T_2)$[20], up to logical equivalence, there is a ψ in $sub(\phi)$ such that $\psi = \Diamond_i \chi$ for some i, and $\psi \in T_2 \setminus T_1$. If $T_1 \neq T_2$ and $T_1 \geq T_2$, then we say that the agents know strictly more in T_1 than in T_2, written $T_1 > T_2$.

This ordering discards the self-knowledge about one's own ignorance mentioned above. Suppose that the only difference between T_1 and T_2 is that i knows that ϕ in the first but not in the second. Formally, this means that $\neg\Box_i\phi$ is in T_2, and so is $\Diamond_i\neg\phi$, by maximal Λ-consistency of T_2.[21] Neither of these formulas is in T_1. But if i is negatively introspective then $\Box_i\Diamond_i\neg\phi$ is also in T_2 but not in T_1 and *vice versa* if he is positively introspective: $\Box_i\neg\Diamond_i\neg\phi$ is in T_1 but not in T_2. In this case one can still say that agents know more in T_1 than in T_2 because, even though i knows about his knowledge and ignorance, respectively, since there is a sub-formula of $\Box_i\Diamond_i\neg\phi$, namely $\Diamond_i\neg\phi$ itself, that is in T_2 but not in T_1.

This ordering is consistent with the naive one, mentioned above, for agents that are not introspective. Take two maximally consistent theories T_1 and T_2 that agree on literals. If all formulas $\Box_i\phi$ in T_2 are also in T_1, something that cannot happen in the introspective case, an easy argument shows that it must be the case that $T_1 \geq T_2$.

Given that there is a natural correspondence between maximally Λ-consistent theories and states in structures for which Λ is sound and complete, this ordering also compares epistemic substantive assumptions at the structural level. Suppose that $\mathcal{M}_1, w$ and $\mathcal{M}_2, v$ are two pointed Kripke structures in a class $\mathcal{K}$ for which a given Λ is sound and complete. Let $T_1 = \{\phi : \mathcal{M}_1, w \Vdash \phi\}$ and $T_2 = \{\phi : \mathcal{M}_2, v \Vdash \phi\}$. Then if $T_1 > T_2$ we know that some epistemic substantive assumptions are relaxed by moving from $\mathcal{M}_1, w$ to $\mathcal{M}_2, v$. The relation $>$ thus provides a means of comparing structures and, as we show presently, it can be used to find models where *no* substantive assumptions are being made.

4 Making No Substantive Assumptions

It is important to know whether there exist structures where no epistemic, substantive assumptions are being made. Substantive assumptions about the information of agents can have drastic consequences on what rational players will/should do in a given context. If no such structures exist, then one is never sure whether conclusions drawn from epistemic modeling rest on implicit, perhaps unwarranted substantive assumptions. If, on the other hand, such structures (of a given class) do exist, then they can be seen as weakest possible informational contexts, i.e. contexts where agents have as little information as possible about each other. Such contexts can be used as benchmark cases to

[20] Given two sets X, Y, we write $X \setminus Y$ for the set-theoretic difference between X and Y.

[21] We assume here that Λ contains at least axiom D (c.f. footnote 8). This can be done without loss of generality. If Λ doesn't contain D, then all theories where one agent has inconsistent beliefs become $>$-minimal. All remarks below would then bear on the $>$-minimal *consistent* theories.'

"test" behavioral conclusions - or simply to pinpoint precisely which structural assumptions are being made in that class.

We now show that such structures do exist. We do this in three steps. We start by showing that $\geq$ is a partial order. Building on this we show that any descending $\geq$-chain starting with a given theory T has a minimal element T^*, and that for any theory T' that agrees with T on literals, the agents know at least as T' than as in T^*. By minimality of T^*, and up to logical equivalence, this is enough to show that the class of pointed structures that satisfy all and only the formulas in T^* is the one where no epistemic, substantive assumptions are being made.[22]

Recall that if $T_1 \geq T_2$ then these two theories agree on literals. Formally, let $md(\phi)$ be the modal depth of ϕ, and $|T_i| = \{\phi \in T_i : md(\phi) = 0\}$. If $T_1 \geq T_2$ then $|T_1| = |T_2|$. This simple observation grounds the fact that $\geq$ is indeed a partial order on theories that share the same literals.

Proposition 1 *Let T_1 be a MCS and $\mathcal{T}_1 = \{T_i : T_i$ is a MCS and $|T_i| = |T_1|\}$. Then $\geq$ is a partial order on $\mathcal{T}_1$.*

Proof Reflexivity is obvious. For anti-symmetry, suppose $T_k \neq T_l$ and $T_k \leq T_l$. We show that $T_l \not\leq T_k$. We know by assumption that $((T_k \cup T_l) \setminus (T_k \cap T_l)) \neq \emptyset$. Take a $\phi = \Diamond_i \psi \in T_k \setminus T_l$ such that all proper sub-formulas of ϕ are in $T_k \cap T_l$. We know that such a formula exists because $T_k \leq T_l$. By our choice of ϕ we know that it has no sub-formula $\Diamond_j \chi'$ in T_l but not in T_k, which means that $T_l \not\leq T_k$.

For transitivity, suppose $T_k \geq T_l \geq T_m$. Because $\geq$ is anti-symmetric we can assume that all these three theories are pairwise different. Take $\phi \in ((T_k \cup T_l) \setminus (T_k \cap T_l))$. We can assume WLOG that ϕ is of the form $L_i\psi$. We show by induction on the modal depth of ϕ it has a sub-formula $\chi = \Diamond_j \chi'$ such that $\chi \in T_m$ but not in T_k.

- Basic case: $(md(\phi) = 1)$ If $\phi \in T_m \setminus T_k$ we are done. But the other case, $\phi \in T_k \setminus T_m$, is impossible. Suppose indeed that $\phi \in T_k \setminus T_m$. This means that $\Box_i \neg\psi \in T_m$. Now, either $\phi = \Diamond_i \psi$ or $\neg\phi = \Box_i \neg\psi$ is in T_l. In the first case, because $T_l \geq T_m$, there must be a sub-formula χ of ψ of the form $\Diamond_j \chi'$. But this can't be because $md(\phi) = 1$ and thus $md(\psi) = 0$. The reasoning in the other case, viz. when $\Box_i \neg\psi \in T_l$, is the same but uses $T_k \geq T_l$.
- Inductive step. Our inductive hypothesis is that for all formula $\phi \in ((T_k \cup T_m) \setminus (T_k \cap T_m))$, if $md(\phi) \leq n$ then it has a sub-formula $\chi = \Diamond_j \chi'$ such that $\chi \in T_m$ but not in T_k. Take $\phi \in ((T_k \cup T_m) \setminus (T_k \cap T_m))$ with $md(\phi) = n+1$. Observe again that if $\phi \in T_m \setminus T_k$ we are done. Assume then, otherwise, that $\Diamond_i \psi = \phi \in T_k \setminus T_m$. If there is a sub-formula of ψ in either T_k or T_m but not both, we are done by the inductive hypothesis, because $md(\psi) = n$. Suppose then that this is not the case, i.e. $sub(\psi) \subseteq T_k \cap T_m$. Suppose now that $\phi \in T_l$. Because $T_l \geq T_m$ and $\neg\phi \in T_m$, ϕ must have a proper sub-formula $\Diamond_k \chi' \in T_m$ but not in T_l, i.e. $\Box_k \neg\chi' \in T_l$. We know by assumption

[22] Here discarding the agents' information about their own information.

that $\Diamond_k \chi'$ must also be in T_k. But then, because $T_k \geq T_l$, $\Diamond_k \chi'$ must itself have a sub-formula $\Diamond_l \chi'' \in T_l$ such that $\neg\Diamond_l \chi'' \in T_k$, and also in T_m by assumption, and so on... At some point this back-and-forth between T_k and T_m is bound to stop, because there are only finitely many sub-formulas of ψ, which will contradict either $T_k \geq T_l$ or $T_l \geq T_m$. The argument for $\neg\phi \in T_l$ follows the same line.

With this in hand, a slight variation of the constriction of maximally consistent sets in Lindenbaum's Lemma gives us the theory we are looking for.

Proposition 2 *Let T_1 and $\mathcal{T}_1$ as before. $\mathcal{T}_1$ as a $\geq$-minimal element.*

Proof The idea of the proof is to construct T^* step-wise, per modal depth. This is done by considering successive fragments of $\mathcal{L}_{EL}$, adding formulas of each of these fragments, and checking consistency at each step. Within each fragment, an additional sub-division is required, to make sure that the $\Diamond$ formulas are added first.

Let $\mathcal{L}_0$ be the propositional fragment of $\mathcal{L}_{EL}$, that is, is the smallest set of formulas ϕ such that: (with $p \in \mathcal{L}_{EL}$)

$$\phi := p \mid \neg\phi \mid \phi \wedge \phi$$

Then for all $n < \omega$ define $\mathcal{L}^{\Diamond}_{n+1}$ as the smallest set of formulas ϕ such that:

$$\phi := \Diamond_i \psi$$

with $i \in I$ and $\psi \in \mathcal{L}_n$. $\mathcal{L}^{\Box}_{n+1}$ is defined analogously, i.e. as the smallest set of formulas ϕ such that:

$$\phi := \neg\phi \mid \phi \wedge \phi \mid \Box_i \psi$$

with $i \in I$ and $\psi \in \mathcal{L}_n$.

We are now ready to construct inductively T^*:

- $T^*_0 = |T_1|$
- T^*_{n+1} is defined step-wise:
  1. Take a well-ordering $S = \langle \phi_1, ...\rangle$ of $\mathcal{L}^{\Diamond}_{n+1}$. Let $T^0_{n+1} = T^*_n$. For all $\phi_k \in S$ with $k \geq 1$, define :
     - $T^k_{n+1} = T^{k-1}_{n+1} \cup \{\phi_k\}$ if $T^{k-1}_{n+1} \nvdash \neg\phi_k$.
     - $T^k_{n+1} := T^{k-1}_{n+1}$ otherwise.

     Set $T^{\Diamond}_{n+1} = \bigcup_{k<\omega} T^k_{n+1}$

     Claim $T^{\Diamond}_{n+1}$ is consistent and in $\mathcal{T}_1$.

     PROOF OF CLAIM Membership in $\mathcal{T}_1$ is obvious. Suppose $T^{\Diamond}_{n+1}$ is not consistent. Then there is a finite set Ψ and a formula ϕ_i, both in $T^{\Diamond}_{n+1}$, such that $\vdash \bigwedge \Psi \to \neg\phi_i$. We can assume WLOG that for all $\phi_j \in \Psi$, $j < i$. But then by construction $\phi_i \notin T^i_{n+1}$. ◀
  2. Take a well-ordering $S' = \langle \phi_1, ...\rangle$ of $\mathcal{L}^{\Box}_{n+1}$. Let $T^0_{n+1} = T^{\Diamond}_{n+1}$. For all $\phi_k \in S$ with $k \geq 1$, define :
     - $T^k_{n+1} = T^{k-1}_{n+1} \cup \{\phi_k\}$ if $T^{k-1}_{n+1} \nvdash \neg\phi_k$.

  - $T^k_{n+1} := T^{k-1}_{n+1}$ otherwise.

  Set $T_{n+1} = \bigcup_{k<\omega} T^k_{n+1}$. T_{n+1} is consistent and in $\mathcal{T}_1$, by a similar argument as before.
- Set $T^* = \bigcup_{n<\omega} T_{n+1}$

Claim T^* is a MCS in $\mathcal{T}_1$.

Proof of Claim Maximality and membership in $\mathcal{T}_1$ are again obvious. If T^* is not consistent, then either there is a finite set Ψ and a formula χ, such that $\vdash \bigwedge \Psi \to \neg\chi$, or inconsistency arises from the breakdown of compactness for the reflexive-transitive closure modality $\Box^*_G$. The first case can't be, since it would contradict the consistency of one of the T^*_n. The second case can be taken care of axiomatically, through introducing an infinitary inference rule for the $\Box^*_G$ modality. ◀

Claim $T_j \geq T^*$, for all $T_j \in \mathcal{T}_1$.

Proof of Claim If $T_j = T^*$ then we're done. Suppose $T_j \neq T^*$ and take ϕ in either T_j or T^* but not both. We can assume WLOG that ϕ is of the form $\Diamond_i\psi$ or $\Box_i\psi$. We show by induction on the modal depth of ϕ that it has a sub-formula ψ of the form $\Diamond_i\chi$ such that $\psi \in T^*$ but not in T_j.

- Basic case ($md(\phi) = 1$). Then ϕ itself must be either $\Diamond_i\chi$ or $\Box_i\chi$ with $md(\chi) = 0$. But by construction[23] it must be that $\Diamond_i\chi$ is in T^*, and so $\Box_i\neg\psi \in T_j$, as required.
- Inductive step. Our inductive hypothesis is that for all ϕ in either T_j or T^* but not both, if ϕ is of modal depth n then we can find the required sub-formula. Take ϕ of modal depth $n+1$ in either T_j or T^* but not both. Now observe that ϕ is of the form $\Diamond_i\psi$ or $\Box_i\psi$ with $md(\psi) = n$, which means that we're done by the inductive hypothesis.

◀

This result makes formal the claim that, for any Kripke structure, one can maximally relax its substantive assumption, *up to equivalence* in the given finitary epistemic languages that we work with. This proviso is of importance, given a number of known results regarding the non-existence of so-called "universal knowledge structures." In the next section we explain the relation between these and the present result, and situate the present analysis with respect to the literature on such "large" structures.

5 Connection with known results on the (non-)existence of large structures

The reader acquainted with the extensive literature on what may be called *large* (or *rich*) type structures[24]— so-called universal or complete structures—

[23] Assuming that the underlying logic is at least KD, so $\Box_i\phi \to \Diamond_i\phi$ is an axiom scheme.

[24] We do not attempt a complete overview of this interesting literature here (see [Brandenburger and Keisler, 2006, Section 11], [Siniscalchi, 2008, Section 3] and Pintér [2005] for discussion and pointers to the relevant results).

can legitimately wonder about the relationship between these structures and those that minimize substantive assumptions as defined in this paper. In this section, we briefly explain this relationship.

5.1 Universal Knowledge Structures

The most direct connection is with so-called *universal knowledge structures* (cf., [Heifetz and Samet, 1998a, Meier, 2005] and the references therein). A *knowledge structure*, as defined in [Heifetz and Samet, 1998a], is essentially a Kripke structure as defined above, except that the players' knowledge and beliefs are defined in terms of set-theoretic operators. So, the essential difference with the setting in this paper is that we restrict attention to a finitary language to express properties of the players' knowledge (and beliefs). A *universal* knowledge structure, as defined in the two papers just quoted, is one that contains the bounded morphic image[25] of any other knowledge structure.[26]

It is not hard to see that a universal knowledge structure (if it exists) must minimize substantive assumptions. If a model contains states that satisfy all and only the formulas in a $>$-minimal element, then no epistemic substantive assumptions will be valid in that model. But since, by definition, that model will have a bounded morphic image in the universal structure, the latter will not validate any epistemic substantive assumptions either. Universal knowledge structures will, in fact, validate only *structural* assumptions.

Observation 1 *Suppose $\mathcal{M}_U$ is a universal knowledge structure. Then $\mathcal{M}_U \Vdash \phi$ iff ϕ is a structural assumption.*

Proof The right to left direction is automatic. For the left to right, suppose that ϕ is not a structural assumption. This means that there is a maximally consistent theory T such that $\phi \notin T$. Consider a pointed model $\mathcal{M}_T, w$ such that $\mathcal{M}_T, w \Vdash \psi$ iff $\psi \in T$. By definition there is a bounded morphism f from $\mathcal{M}_T$ to $\mathcal{M}_U$, and so $M_U, f(w) \Vdash \neg\phi$, which means $M_U \nVdash \phi$.

[25] Heifetz and Samet [1998a] use the notion of "knowledge morphism", which in the case of Kripke structures can easily be shown to be equivalent to the notion of bounded morphism, familiar to modal logicians—see [Blackburn et al., 2001, Definition 2.10, pg. 59]. A *bounded morphism*, also called a *p-morphism*, from a Kripke model $\mathcal{M} = \langle W, \mathcal{R}, V\rangle$ to $\mathcal{M}' = \langle W', \mathcal{R}', V'\rangle$ is a (total) function $f : W \to W'$ such that (1) for all $p \in \textsc{prop}$, $p \in V(w)$ iff $p \in V'(f(w))$, (2) for all players i and for all $w, v \in W$, wR_iv implies $f(w)R'_if(v)$ and (3) for all players i and for all $w \in W$ and $v' \in W'$, if $f(w)R'_iv'$ then there is a $v \in W$ such that wR_iv and $f(v) = v'$. Knowledge morphisms are mappings that preserve not only basic propositional truth (condition 1), but also each agents' information (conditions 2 & 3). More precisely, a well-known observation is that if there is a knowledge morphism f from $\mathcal{M}$ to $\mathcal{M}'$, then for all states w in $\mathcal{M}$, and all formulas $\phi \in \mathcal{L}_{EL}$, $\mathcal{M}, w \models \phi$ iff $\mathcal{M}', f(w) \models \phi$.

[26] That is, a knowledge structure $\mathcal{M}_U$ is *universal* iff for any knowledge structure $\mathcal{M}$, there is a knowledge morphism from $\mathcal{M}$ into $\mathcal{M}_U$. In the language of category theory, such structures are also called **terminal** objects.

This observation is only of limited interest since, in general, universal knowledge structures do not exist. The most salient result here is Martin Meier's [2005] proof that universal Kripke structures do not exist.[27]

Of course, it does not necessarily follow from this negative result that there are no structures where only structural assumptions are valid. The *canonical model*[28] for a given logical system is a clear example, and the argument for this is simply that this model is constructed from *all* maximally consistent theories of a given logical system[29]. Universality is thus related to the minimization of substantive assumptions, but the two notions are different.

Our results in this paper show precisely in what sense the *canonical Kripke structure*, familiar in the modal logic literature, minimizes substantive assumptions. This construction of the canonical model depends on the underlying language $\mathcal{L}$ and logical system Λ. The key aspect of this construction is that for any Λ-consistent set Γ of formulas from $\mathcal{L}$, there is a state in the canonical model that satisfies all formulas in Γ. When this language has the expressive resources to describe higher-order knowledge and beliefs, this means that the structure represents all consistent hierarchies of knowledge and beliefs. This is also the key idea behind the classic construction of the so-called *canonical type space* of Mertens and Zamir [1985].

The situation is much better behaved in the probabilistic setting. The central result of Brandenburger and Dekel [1993] shows that (under suitable topological assumptions) the canonical type space is, in fact, a *universal* type structure. The result has been generalized in numerous ways[30].

For all-out attitudes specified in finitary languages, like the one studied in the present paper, the canonical model minimizes substantive assumptions, but is not universal. But, as already observed by [Heifetz, 1999][31], for instance, if one redefines universality in terms of truth preservation in a given language, then universal knowledge structures do exist. Of course the question then becomes one of motivating the choice of a specific *language* to describe the agents' attitudes. This is a difficult question, to which we come back briefly in the Conclusion, but for now it is sufficient to point out that it is not unlike the one of motivating certain topological assumptions on type structures.

[27] This result is based on an earlier non-existence result of Heifetz and Samet [1998a] for knowledge structures whose relations are partitions. For other non-existence results, this time for Harsanyi type spaces, see Pintér [2010]. These results are best viewed in the context of the general theory of *final coalgebras* (see [Venema, 2006, Section 10] for a discussion and pointers to the relevant results).

[28] For the precise definition of the canonical model construction, also known as Henkin model in model theory, see the references in footnote 12.

[29] Some care is needed here for logical systems that are not *compact*, such as epistemic logic with a common knowledge. See [Blackburn et al., 2001, Chapter 4] for a discussion. These technical issues are not crucial for the general point we are making here.

[30] We do not discuss these generalizations here. See Heifetz and Samet [1998b], Meier [2008, 2006], Pintér [2005] for results and discussion of the relevant literature.

[31] Cf. also the extensive discussion in Fagin et al. [1999].

5.2 Complete structures

A structure is said to be *assumption-complete* if, for each subset X in a given set of subsets of that structure and each agent i, there is a state where i "assumes" X. Assumes is taken here to mean strongest belief. In a Kripke structure, for instance, a set X is assumed by i at a state w if $R_i[w] = X$.

A simple counting argument shows that there cannot exist a complete structure where the set of conjectures is *all* subsets of the set of states (types) [Brandenburger, 2003]. A deeper result is the *impossibility theorem* from [Brandenburger and Keisler, 2006, Theorem 5.4] showing that a complete structure does not exist even if the set of events is restricted to first-order definable sets[32]. Some positive results are in sight as well: Mariotti et al. [2005] constructs a complete structure where the set of conjectures are compact subsets of some well-behaved topological space.

The relation between assumption-complete structures and those that minimize substantive assumptions can be encapsulated as a quantifier switch. Let B_{-i} be a consistent set of formulas of the form $B_j\phi$ for $i \neq j$, and $\mathcal{B}_{-i}$ be the set of all such B_{-i}. If an assumption-complete structure exists for the language at hand, then for all B_{-i} there is a state where agent i assumes $\{w : w \Vdash B_j\phi$ for all $B_j\phi \in B_{-i}\}$. Minimization of epistemic substantive assumptions, on the other hand, means that there is a state w where the agents assumes "all" $B_{-i} \in \mathcal{B}_{-i}$ or, to be more precise: for all such B_{-i} there is a state in $R_i[w]$ that satisfies all formulas in B_{-i}. Assumption-complete structures, when they exist, are thus structures that contain states where epistemic substantive assumptions are minimized, but the other way around is not true.

6 Conclusion

In this paper we studied substantive assumptions in interactive situations. We have explained why they are important, shown how to identify and compare them, formally, and shown that there exist contexts where no substantive assumptions are being made. Towards the end of the paper we briefly explained the relation between such structures and a number of other "large" structures studied in the literature.

Our approach was primarily syntactic, and this was of great importance in determining what count as assumptions at all, substantive or structural, in a given class of structures. Properties of structures that are not definable in the language at hand are simply off the radar for our notion of structural and/or substantive assumptions. This raise a broader conceptual question: which granularity of epistemic analysis is needed or desirable? Or, the other way around, why would one choose to ignore details in favor of more coarse-grained languages? Issues of computational complexity speak in favor of the second approach, while notions of behavioral equivalence seem to point towards

[32] See [Abramsky and Zvesper, 2010] for an extensive analysis and generalization of this result.

the first. We do not take a stance on this question here, but rather leave it as an open question, phrased in the words of the founding fathers of analytic philosophy: *Of what one cannot speak, must one pass over in silence?*

References

S. Abramsky and J. A. Zvesper. From Lawvere to Brandenburger-Keisler: interactive forms of diagonalization and self-reference. *CoRR*, abs/1006.0992, 2010.

R. Aumann and J. Dreze. Rational expectations in games. *American Economic Review*, 98:72 – 86, 2008.

R.J. Aumann. Correlated equilibrium as an expression of bayesian rationality. *Econometrica*, 55(1-18), 1987.

R.J. Aumann. Interactive epistemology I: Knowledge. *International Journal of Game Theory*, 28:263–300, 1999.

D. Bernheim. Rationalizable strategic behavior. *Econometrica*, 52:1007–1028, 1984.

P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, Cambirdge, 2001.

P. Blackburn, J. van Benthem, and F. Wolter, editors. *Handbook of Modal Logic*. Elsevier, November 2006.

O. Board. Knowledge, beliefs and game-theoretic solution concepts. *Oxford Review of Economic Policy*, 18:433 – 445, 2002.

A. Brandenburger. On the existence of a 'complete' possibility structure. In M. Basili, N. Dimitri, and I. Gilboa, editors, *in Cognitive Processes and Economic Behavior*, pages 30–34. Routledge, 2003.

A. Brandenburger and E. Dekel. Hierarchies of beliefs and common knowledge. *Journal of Economic Theory*, 59, 1993.

A. Brandenburger and H.J. Keisler. An impossibility theorem on beliefs in games. *Studia Logica*, 84(2):211–240, 2006.

A. Brandenburger, A. Friedenberg, and H. J. Keisler. Admissibility in games. *Econometrica*, 76:307–352, 2008.

B. de Bruin. *Explaining Games: The Epistemic Programme in Game Theory*, volume 346 of *Synthese Library*. Springer, 2010.

R. Fagin and J.Y. Halpern. Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34(1):39–76, 1987.

R. Fagin, J.Y. Halpern, Y. Moses, and M. Vardi. *Reasoning about Knowledge*. MIT Press, 1995.

R. Fagin, J. Halpern, J. Geanakoplos, and M. Vardi. The hierarchical approach to modeling knowledge and common knowledge. *International Journal of Game Theory*, 28(3):331 – 365, 1999.

J.Y. Halpern. Alternative semantics for unawareness. *Games and Economic Behavior*, 37(2):321–339, 2001.

D. Harel, D. Kozen, and J. Tiuryn. *Dynamic Logic*. The MIT Press, 2000.

A. Heifetz. How canonical is the canonical model? A comment on Aumann's interactive epistemology. *International Journal of Game Theory*, 28(3):435–442, 1999.

A. Heifetz and D. Samet. Knowledge spaces with arbitrarily high rank. *Games and Economic Behavior*, 22(2):260–273, 1998a.

A. Heifetz and D. Samet. Topology-free typology of beleifs. *Journal of Economic Theory*, 82:324 – 341, 1998b.

A. Heifetz, M. Meier, and B.C. Schipper. Interactive unawareness. *Journal of Economic Theory*, 130(1):78–94, 2006.

J. Hintikka. *Knowledge and Belief: An Introduction to the Logic of Two Notions*. ornell University Press, Ithaca, N.Y., 1962.

F. Huber and C. Schmidt-Petri. *Degrees of Belief*, volume 342. Springer, Synthese Library, 2009.

T. Mariotti, M. Meier, and M. Piccione. Hierarchies of beliefs for compact possibility models. *Journal of Mathematical Economics*, 41:303 – 324, 2005.

M. Meier. Universal knowledge-beleif structures. *Game and Economic Behavior*, 62:53 – 66, 2008.

M. Meier. On the nonexistence of universal information structures. *Journal of Economic Theory*, 122(1):132–139, 2005.

Martin Meier. Finitely additive beliefs and universal type spaces. *Annals of Probability*, 34(1):386 – 422, 2006.

J.F. Mertens and S. Zamir. Formulation of Bayesian analysis for games with incomplete information. *International Journal of Game Theory*, 14(1):1–29, 1985.

S. Modica and A. Rustichini. Awareness and partitional information structures. *Theory and Decision*, 37(1):107–124, 1994.

I. Moscati. Interactive and common knowledge in the state-space model. Cesmep working papers, University of Turin, 2009. URL `http://econpapers.repec.org/RePEc:uto:cesmep:200903`.

R. Parikh. Monotonic and non-monotonic logics of knowledge. *Fundamenta Informaticae*, XV:255–274, 1991.

D. Pearce. Rationalizable strategic behavior and the problem of perfection. *Econometrica*, 52:1029–1050, 1984.

M. Pintér. Type space on a purely measurable parameter space. *Economic Theory*, 26(1):129–139, 2005.

M. Pintér. The non-existence of a universal topological type space. *Journal of Mathematical Economics*, 46(2):223–229, 2010.

L. Samuelson. Modeling knowledge in economic analysis. *Journal of Economic Literature*, 42(2):367–403, 2004.

K. Segerberg. A model existence theorem in infinitary propositional modal logic. *Journal of Philosophical Logic*, 23(4):337–367, 1994.

M. Siniscalchi. Epistemic game theory: beliefs and types. In S. Durlauf and L. Blume, editors, *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, Basingstoke, 2008.

R. Stalnaker. Extensive and strategic forms: Games and models for games. *Research in Economics*, 53:293 – 319, 1999.

J. van Benthem. *Modal Logic for Open Minds*. CSLI Publications, 2010.

J. Van Benthem, P. Girard, and O. Roy. Everything else being equal: A modal logic for ceteris paribus preferences. *Journal of philosophical logic*, 38(1): 83–125, 2009.

J. van Benthem, E. Pacuit, and O. Roy. Toward a theory of play: A logical perspective on games and interaction. *Games*, 2(1):52–86, 2011. ISSN 2073-4336. doi: 10.3390/g2010052. URL `http://www.mdpi.com/2073-4336/2/1/52/`.

H. van Ditmarsch, W. van de Hoek, and B. Kooi. *Dynamic Epistemic Logic*, volume 337 of *Synthese Library Series*. Springer, 2007.

Y. Venema. Algebra and coalgebra. In *Handbook of Modal Logic [Blackburn et al., 2006]*, pages 331 – 426. Elsevier, Amsterdam, 2006.