A logical formalism for the subjective approach in a multi-agent setting

Guillaume Aucher

Université Paul Sabatier, Toulouse (F)
University of Otago, Dunedin (NZ)
aucher@irit.fr

Abstract. Representing an epistemic situation involving several agents depends very much on the modeling point of view one takes. In fact, the interpretation of a formalism relies quite a lot on the nature of this modeling point of view. Classically, in epistemic logic, the models built are supposed to represent the situation from an external and objective point of view. We call this modeling approach the objective approach. In this paper, we study the modeling point of view of a particular agent involved in the situation with other agents. We propose a logical formalism based on epistemic logic that this agent can use to represent 'for herself' the surrounding world. We call this modeling approach the subjective approach. We then set some formal connections between the subjective approach and the objective approach. Finally we axiomatize our logical formalism and show that the resulting logic is decidable.

Note 1. All the proofs of this paper can be found in the appendix.

1 Introduction

In the literature about epistemic logic, when it comes to model epistemic situations, not much is said explicitly about which modeling point of view is considered. However, modeling an epistemic situation depends very much on the modeling point of view. Indeed, the models built will be quite different whether the modeler is an agent involved in the situation or not. Let us consider the following example. Assume that the agents Ann and Bob are in a room and that there is a coin in a box that both cannot see because the box is closed. Now, assume that Bob cheats, opens the box and looks at the coin. Moreover, assume that Ann does not suspect anything about it and that Bob knows it (Ann might be inattentive or out of the room for a while). On the one hand, if the modeler is an external agent (different from Ann and Bob) knowing everything that has happened, then in the model that this modeler builds to represent this resulting situation Bob knows whether the coin is heads or tails up. On the other hand, if the modeler is Ann then in the model that Ann builds to represent this resulting situation Bob does not know whether the coin is heads or tails up. As we see in this example, specifying the modeling point of view is also quite essential to

Dagstuhl Seminar Proceedings 07351
Formal Models of Belief Change in Rational Agents
http://drops.dagstuhl.de/opus/volltexte/2007/1200

interpret the formal models.

But what kinds of modeling points of view are there ? For a start, we can distinguish whether the modeler is involved in the situation or not.

1. First, consider the case where the modeler is involved in the situation. In other words she has the same status as the other agents and is considered by them on a par. So the modeler should be represented in the formalism because she takes an active part in the situation. This formalism then deals not only with the other agents' beliefs but also with the other agents' beliefs about her own beliefs. This is then an internal point of view, and the models built are supposed to represent the way she perceives the surrounding world. Note that because the modeler is part of the situation, her beliefs might be erroneous. Hence the models she builds might also be erroneous. We call this point of view the *subjective* point of view because the models built are the formal models that the modeler-agent has 'in her mind' in order to represent the surrounding world.
2. Second, consider the case where the modeler is not involved in the situation. In other words, she simply does not exist for the other agents, or at least she is not taken into consideration in their representation of the world. So the modeler should not be represented in the formalism and particularly the agents' beliefs about her own beliefs should also not be represented because they simply do not exist. The models that the modeler builds are supposed to represent the situation from an external and objective point of view. There are then two other possibilities depending on whether or not the modeler has a perfect and omniscient knowledge of the situation.
   (a) In case the modeler has a perfect and omniscient knowledge of the situation then everything that is true in the model that she builds is true in reality and vice versa, everything that is true in reality is also true in the model. This thesis was already introduced in [1]. Basically, the models built by the modeler are perfectly correct. The modeler has access to the mind of the agents and knows perfectly not only what they believe but also what the real state of affairs is. This is a kind of 'divine' point of view and we call it the *objective* point of view.
   (b) In case the modeler does not have a perfect and omniscient knowledge of the situation then, unlike the objective point of view, the models built might be erroneous. The models could also be correct but then, typically, the modeler would be uncertain about which is the actual world (in that sense, she would not have an omniscient knowledge of the situation). What the modeler knows can be obtained for example by observing what the agents say and do by asking them questions . . . We call this point of view the *imperfectly objective* point of view.

We claim that the subjective, the objective and the imperfectly objective point of view are the only three possible points of view when we want to model epistemic situations. From now on we will call them the subjective, the objective and the imperfectly objective approach.

Moreover, the fields of application of these approaches are different. The subjective approach has rather applications in artificial intelligence where autonomous agents like machines or robots acting in the world need to have a formal representation of the surrounding world. The objective and imperfectly objective approach have rather applications in game theory [2], cognitive psychology or distributed systems [3] for example. Indeed, in these fields we need to model situations from an external point of view in order to explain and predict what happens in these situations.

In this paper we will focus only on the subjective approach (and its connections with the objective approach). For a work that deals with reasoning about another agent using an imperfectly objective approach, see [4, 5]. Standard epistemic logic [6] rather follows the objective approach. On the other hand, AGM belief revision theory [7] rather follows the subjective approach. But AGM is designed for a single agent. In fact there is no logical formalism for the subjective approach in a multi-agent setting. However, such a formalism is crucial if we want to design autonomous agents. That is what we are going to propose in this paper.

The paper is organized as follows. In Section 2 we recall epistemic logic. In Section 3 we propose a semantics for the subjective approach. Then in Section 4 we set some connections between the subjective and the objective approach. Finally, in Section 5 we propose an axiomatization of the subjective semantics.

2 Epistemic logic

Epistemic logic is a modal logic [8] that is concerned with the logical study of the notions of knowledge and belief. So what we call an epistemic model is just a particular kind of Kripke model as used in modal logic. The only difference is that instead of having a single accessibility relation we have a set of accessibility relations, one for each agent. This set of agents is noted G and its cardinality N. Besides, Φ is a set of propositional letters.

Definition 1. *An epistemic model M is a triple $M = (W, R, V)$ such that*

- *W is a non-empty set of possible worlds;*
- *$R : G \rightarrow 2^{W \times W}$ assigns an accessibility relation to each agent;*
- *$V : \Phi \rightarrow 2^W$ assigns a set of possible world to each propositional letter.*

If $M = (W, R, V)$ is an epistemic model, a pair (M, w_a) with $w_a \in W$ is called a pointed epistemic model. We also note $R_j := R(j)$ and $R_j(w) := \{w' \in W; wR_jw'\}$, and $w \in M$ for $w \in W$.

Intuitively, a pointed epistemic model (M, w_a) represents from an external point of view how the actual world w_a is perceived by the agents G. This entails that epistemic logic clearly follows the objective approach. The possible worlds W are the relevant worlds needed to define such a representation and the valuation V specifies which propositional facts (such as 'it is raining') are true in these worlds. Finally the accessibility relations R_j models the notion of belief. We set $w' \in R_j(w)$ in case in world w, agent j considers the world w' possible.

Finally, the submodel of M generated by a set of worlds $S \subseteq M$ is the restriction[1] of M to the worlds $\{(\bigcup\limits_{j \in G} R_j)^*(w_S); w_S \in S\}$ (where $(\bigcup\limits_{j \in G} R_j)^*$ is the reflexive transitive closure of $(\bigcup\limits_{j \in G} R_j)$, see [8] for details). In case the submodel of M generated by a set of worlds $S \subseteq M$ is M itself then M is said to be generated by S. Intuitively, the submodel of M generated by a set of worlds S contains all the relevant information in M about these worlds S.

Now inspiring ourselves from modal logic, we can define a language for epistemic models which will enable us to express things about them. The modal operator is then a 'belief' operator, one for each agent.

Definition 2. *The language $\mathcal{L}$ is defined as follows:*

$$\mathcal{L} : \varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \wedge \varphi \mid B_j\varphi$$

where p ranges over Φ and j over G. Moreover, $\varphi \vee \varphi'$ is an abbreviation for $\neg(\neg\varphi \wedge \neg\varphi')$; $\varphi \rightarrow \varphi'$ is an abbreviation for $\neg\varphi \vee \varphi'$; $\hat{B}_j\varphi$ is an abbreviation for $\neg B_j \neg\varphi$; and $\bot$ is an abbreviation for $\neg\top$.

Now we can give meaning to the formulas of this language by defining truth conditions for these formulas on the class of epistemic models.

Definition 3. *Let $M = (W, R, V)$ be an epistemic model and $w \in W$. $M, w \models \varphi$ is defined inductively as follows:*

- *$M, w \models \top$;*
- *$M, w \models p$ iff $w \in V(p)$;*
- *$M, w \models \neg\varphi$ iff it is not the case that $M, w \models \varphi$;*
- *$M, w \models \varphi \wedge \varphi'$ iff $M, w \models \varphi$ and $M, w \models \varphi'$;*
- *$M, w \models B_j\varphi$ iff for all $v \in R_j(w)$, $M, v \models \varphi$.*

We write $M \models \varphi$ for $M, w \models \varphi$ for all $w \in M$.

So agent j believes φ in world w (formally $M, w \models \varphi$) if φ is true in all the worlds that the agent j considers possible (in world w).

But note that the notion of belief might comply to some constraints (or axioms) such as $B_j\varphi \rightarrow B_jB_j\varphi$: if agent j believes something, she knows that she believes it. These constraints might affect the nature of the accessibility relations R_j which may then comply to some extra properties. We list here some properties that will be useful in the sequel: seriality: for all w, $R_j(w) \neq \emptyset$; transitivity: for all w, w', w'', if $w' \in R_j(w)$ and $w'' \in R_j(w')$ then $w'' \in R_j(w)$; euclidicity: for

[1] Let $M = (W, R, V)$ be an epistemic model. The restriction of M to a set of worlds S is the submodel $M' = (W', R', V')$ of M defined as follows. $W' = W \cap S$; $R'_j := R_j \cap (S \times S)$ for all $j \in G$; and $V'(p) = V(p) \cap S$ for all $p \in \Phi$.

all w, w', w'', if $w' \in R_j(w)$ and $w'' \in R_j(w)$ then $w' \in R_j(w'')$. Then we define the class of $\mathsf{KD45_G}$-models as the class of epistemic models whose accessibility relations are serial, transitive and euclidean. If for all $\mathsf{KD45_G}$-models M, $M \models \varphi$ then φ is said to be $\mathsf{KD45_G}$-valid and it is noted $\models_{KD45_G} \varphi$.

Now we are going to axiomatize this semantics with the help of a particular modal logic. Generally speaking, a modal logic is built from a set of axiom schemes and inference rules, called a *proof system*. Then a formula φ belongs to this logic either if it is an axiom or if it is derived by applying successively some inference rules to some axioms. In that case we say that the formula is *provable*.

Definition 4. *The logic $\mathsf{KD45_G}$ is defined by the following axiom schemes and inference rules:*

Taut	$\vdash_{KD45_G} \varphi$ *for all propositional tautologies* φ
K	$\vdash_{KD45_G} B_j(\varphi \to \psi) \to (B_j\varphi \to B_j\psi)$ *for all* $j \in G$
D	$\vdash_{KD45_G} B_j\varphi \to \hat{B}_j\varphi$
4	$\vdash_{KD45_G} B_j\varphi \to B_jB_j\varphi$
5	$\vdash_{KD45_G} \neg B_j\varphi \to B_j\neg B_j\varphi$
Nec	*If* $\vdash_{KD45_G} \varphi$ *then* $\vdash_{KD45_G} B_j\varphi$ *for all* $j \in G$
MP	*If* $\vdash_{KD45_G} \varphi$ *and* $\vdash_{KD45_G} \varphi \to \psi$ *then* $\vdash_{KD45_G} \psi$.

We write $\vdash_{KD45_G} \varphi$ *in case* φ *belongs to the logic* $\mathsf{KD45_G}$*.*

An interesting feature of epistemic (and modal) logic is that we can somehow match the constraints imposed by the axioms on the belief operator B_j with constraints on the accessibility relations R_j. In other words, the notions of validity and provability coincide. That is what the following theorem expresses.

Theorem 1 (soundness and completeness). *For all* $\varphi \in \mathcal{L}$,

$$\vdash_{KD45_G} \varphi \text{ iff } \models_{KD45_G} \varphi$$

As we said, epistemic logic rather follows the objective approach. So now we are going to propose a formalism for the subjective approach.

3 A semantics for the subjective approach

To define a semantics for the subjective approach in a multi-agent setting, we will start from the standard view of an agent's epistemic state as a set of possible worlds (used in the AGM framework), and then extend it to the multi-agent case. Then we will propose an equivalent formalism which will be used in the rest of this paper.

But first some important notations. As we said, in the subjective approach, the modeler is a given agent involved in the situation. Throughout this paper, this given agent/modeler will be called the agent Y (like You) and we thus have that $Y \in G$. Besides, because a computer cannot easily deal with infinite structures, the set Φ of propositional letters is assumed to be finite.

3.1 Multi-agent possible world and subjective model

In the AGM framework, one considers a single agent Y. The possible worlds are supposed to represent how the agent Y perceives the surrounding world. As she is the only agent, these possible worlds deal only with propositional facts about the surrounding world. Now, if we suppose that there are other agents than agent Y, a possible world for Y in that case should also deal with how the other agents perceive the surrounding world. These “multi-agent” possible worlds should then not only deal with propositional facts but also with epistemic facts. So to represent a multi-agent possible world we need to add a modal structure to our (single agent) possible worlds. We do so as follows.

Definition 5. *A multi-agent possible world (M, w) is a finite pointed epistemic model $M = (W, R, V, w)$ generated by $w \in W$ such that R_j is serial, transitive and euclidean for all $j \in G$, and*

1. $R_Y(w) = \{w\}$;
2. *there is no v and $j \neq Y$ such that $w \in R_j(v)$.*

Let us have a closer look at the definition. Condition 2 will be motivated later, but note that any pointed epistemic model satisfying the conditions of a multi-agent possible world except condition 2 is bisimilar to a multi-agent possible world. Condition 1 ensures that in case Y is the only agent then a multi-agent possible world is a unique possible world with a reflexive arrow indexed by Y. This is very similar to the possible worlds of the AGM theory to which we could perfectly add reflexive arrows indexed by Y. Condition 1 also ensures that in case Y assumes that the situation is correctly represented by the multi-agent possible world (M, w) then for her w is the (only) actual world. In fact the other possible worlds of a multi-agent possible world are just present for technical reasons: they express the other agents' beliefs (in world w). One could get rid of the condition that a multi-agent possible world (M, w) is generated by w but the worlds which do not belong to the submodel generated by w would not have neither philosophical nor technical motivation. Besides, for the same reason that Φ is finite, a multi-agent possible world is also assumed to be finite. Finally, notice that we assume that accessibility relations are serial, transitive and euclidean. This means that the agents' beliefs are consistent and that agents know what they believe and disbelieve. These seem to be very natural constraints to impose on the notion of belief. Intuitively, this notion of belief corresponds for example to the kind of belief in a theorem that you have after having proved this theorem and checked the proof several times. In the literature, this notion of belief corresponds to Lenzen's notion of conviction [9] or to Gardenfors' notion of acceptance [10] or to Voorbraak's notion of rational introspective belief [11]. In fact, in all the agent theories the notion of belief satisfies these constraints: in Cohen and Levesque's theory of intention [12] or in Rao and Georgeff BDI architecture [13] or in Meyer et. al. KARO architecture [14] or in the AOP paradigm [15]. However, one should note that all these agent theories follow the objective approach. This is of course at odds with their intention to implement their theories in machines.

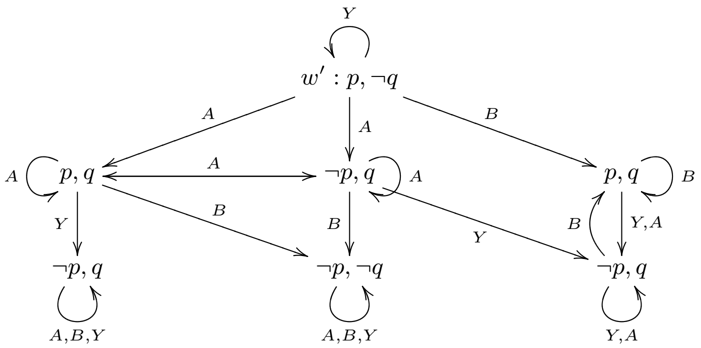
Remark 1. In this paper we deal only with the notion of belief but one could also add the notion of knowledge. Indeed, it might be interesting to express things such as 'the agent Y believes that agent j does not *know* p' (even if this could be rephrased in terms of beliefs). We refrain to do so in order to highlight the main new ideas and because in most applications of the subjective approach the notion of knowledge is not essential.

Example 1. We see in the figure above that a multi-agent possible world is really a generalization of a possible world.

a (single-agent) possible world:

$$w : p, \neg q$$

a multi-agent possible world:



In the single agent case (in AGM belief revision theory), the epistemic state of the agent Y is represented by a finite set of possible worlds. In a multi agent setting, this is very similar: the epistemic state of the agent Y is represented by a (disjoint and) finite set of *multi-agent* possible worlds. We call this a subjective model of type 1.

Definition 6. *A* subjective model of type 1 *is a disjoint and finite union of multi-agent possible worlds.*

A subjective model of type 1 will sometimes be noted $(\mathcal{M}, W_a)$ where W_a are the roots of its multi-agent possible worlds.

Example 2. In Figure 1 is depicted an example of subjective model. In this subjective model, the agent Y does not know wether p is true or not (formally

$\neg B_Y p \wedge \neg B_Y \neg p$). Indeed, in one multi-agent possible world (on the left) p is true at the root and in the other (on the right) p is false at the root. The agent Y also believes that the agent A does not know whether p is true or false (formally $B_Y(\neg B_A p \wedge \neg B_A \neg p)$. Indeed, in both multi-agent possible worlds, $\neg B_A p \wedge \neg B_A \neg p$ is true (at the roots). Finally, the agent Y believes that A believes that she does not know whether p is true or false (formally $B_Y B_A(\neg B_Y p \wedge \neg B_Y \neg p)$) since $B_A(\neg B_Y p \wedge \neg B_Y \neg p)$ is true at the roots of both multi-agent possible worlds.

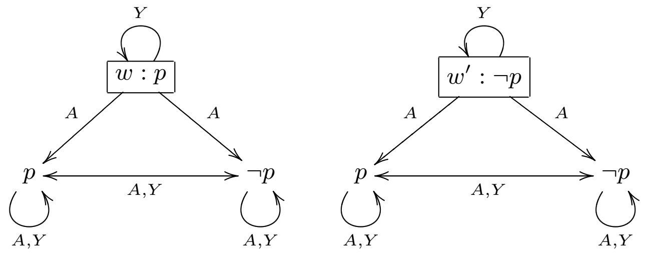


Fig. 1. Example of subjective model of type 1

Thanks to condition 2 in the definition of a multi-agent possible world, we could define the notion of subjective model differently. Indeed, we could perfectly set an accessibility relation between the roots of the multi-agent possible worlds. Figure 2 gives an example of such a process, starting from the example of Figure 1. Condition 2 ensures us that by doing so we do not modify the information present in the original subjective model. Indeed, if condition 2 was not fulfilled then it might be possible that j's beliefs about Y's beliefs (for some $j \neq Y$) might be different between the original subjective model and the new one, due to the creation of these new accessibility relations between the multi-agent possible worlds. This phenomenon will become explicit when we define the language for the subjective models of type 1.

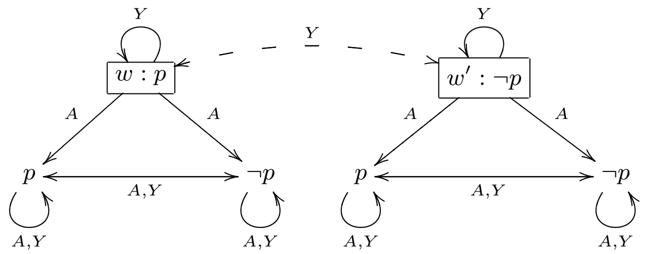


Fig. 2. A new definition of subjective model

Then in this new formalism, one can notice that the former roots of the multi-agent possible worlds form an equivalence class for the accessibility relation indexed by Y. Note also that the accessibility relations are still serial, transitive and euclidean. This leads us to the following new definition of a subjective model.

Definition 7. *A* subjective model of type 2 *is a couple* $(\mathcal{M}, W_a)$ *where* $\mathcal{M}$ *is a finite epistemic model* $\mathcal{M} = (W, R, V)$ *generated by* $W_a \subseteq W$ *such that* R_j *is serial, transitive and euclidean for all* $j \in G$*, and* $R_Y(w_a) = W_a$ *for all* $w_a \in W_a$*.* W_a *is called the* actual equivalence class*.*

So from a subjective model of type 1, one can easily define an equivalent subjective model of type 2. But of course, the other way around, we will see that from a subjective model of type 2 one can also easily define an equivalent subjective model of type 1 (see Proposition 1).

Example 3. In Figure 3 is depicted a subjective model of type 2. The worlds of the actual equivalence class are within boxes. It turns out that this subjective model is bisimilar to the one depicted in Figure 2, which is itself an equivalent representation of the subjective model of type 1 depicted in Figure 1. So the subjective model of type 2 depicted in Figure 3 is an equivalent representation of the subjective model of type 1 depicted in Figure 1. One can indeed check for example that the formulas $\neg B_Y p \wedge \neg B_Y \neg p$, $B_Y(\neg B_A p \wedge \neg B_A \neg p)$ and $B_Y B_A(\neg B_Y p \wedge \neg B_Y \neg p)$ are indeed true. Note that this second representation is much more compact.

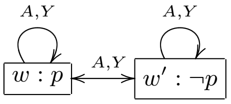


Fig. 3. Example of subjective model of type 2

As we said in the introduction, the subjective approach can be applied in artificial intelligence. In this case, the agent Y is an artificial agent (such as a robot) that has a subjective model 'in her mind'. But to stick with a more standard approach (used in the single agent case), we could perfectly consider that the agent Y has sentences from a particular language 'in her mind' and draws inferences from them. In that respect, this language could also be used by the agent Y in the former approach to perform some model checking in her subjective model in order to reason about the situation or to answer queries. So in any case we do need to define a language.

3.2 Language for the subjective approach

Definition 8.

$$\mathcal{L} : \varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \wedge \varphi \mid B_j\varphi$$

where p ranges over Φ and j over G.

For sake of simplicity and in order to highlight the new results, we do not introduce in this paper a common knowledge operator, but this could be done easily. In fact all the results of this paper still hold if we add the common knowledge operator to the language. Note that the language is identical to the usual language of epistemic logic. If we consider the class of subjective models of type 2 then its truth conditions are also the same and are spelled out in Definition 3. But if we consider the class of subjective models of type 1 then its truth conditions are a bit different and are set out below.

Definition 9. *Let $(\mathcal{M}, \{w^1, \ldots, w^n\}) = \{(M^1, w^1), \ldots,$ $(M^n, w^n)\}$ be a subjective model of type 1 and let $w \in \mathcal{M}$. Then $w \in M^k$ for some k, with $M^k = (W^k, R^k, V^k)$. $\mathcal{M}, w \models \varphi$ is defined inductively as follows:*

- *$\mathcal{M}, w \models \top$;*
- *$\mathcal{M}, w \models p$ iff $w \in V^k(p)$;*
- *$\mathcal{M}, w \models \neg\varphi$ iff it is not the case that $\mathcal{M}, w \models \varphi$;*
- *$\mathcal{M}, w \models \varphi \wedge \varphi'$ iff $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \varphi'$;*
- *$\mathcal{M}, w \models B_Y\varphi$ iff*
  - *$w \in W_a$ and for all $w^i \in W_a$, $\mathcal{M}, w^i \models \varphi$; or*
  - *$w \in W^k - W_a$ and for all $w' \in R^k_Y(w)$, $\mathcal{M}, w' \models \varphi$*
- *If $j \neq Y$ then $\mathcal{M}, w \models B_j\varphi$ iff for all $w' \in R^k_j(w)$, $\mathcal{M}, w' \models \varphi$.*

Note that the truth condition for the operator B_Y is defined as if there were accessibility relations indexed by Y between the roots of the multi-agent possible worlds. Condition 2 of Definition 5 then ensures that the agents j's beliefs about agent Y's beliefs (with $j \neq Y$) of a given multi-agent possible world stay the same whatever other multi-agent possible world we add to this multi-agent possible world. This would of course be a problem if it was not the case. Condition 2 thus provides a kind of modularity of the multi-agent possible worlds in a subjective model (of type 1).

This truth condition for the operator B_Y is of course completely in line with the truth conditions for the subjective models of type 2. In fact, thanks to the definition of this language, we can show that the two types of subjective models are somehow equivalent.

Definition 10. *Let $(\mathcal{M}, W_a)$ be a subjective model of type 1 and $(\mathcal{M}', W'_a)$ be a subjective model of type 2. $(\mathcal{M}, W_a)$ and $(\mathcal{M}', W'_a)$ are equivalent if and only if*

- *for all $w \in W_a$ there is $w' \in W'_a$ such that for all $\varphi \in \mathcal{L}$, $\mathcal{M}, w \models \varphi$ iff $\mathcal{M}', w' \models \varphi$;*
- *for all $w \in W'_a$ there is $w' \in W_a$ such that for all $\varphi \in \mathcal{L}$, $\mathcal{M}, w \models \varphi$ iff $\mathcal{M}', w' \models \varphi$.*

Proposition 1.

Let $(\mathcal{M}, W_a)$ be a subjective model of type 2. Then there is a subjective model of type 1 $(\mathcal{M}', W'_a)$ which is equivalent to $(\mathcal{M}, W_a)$.

Let $(\mathcal{M}, W_a)$ be a subjective model of type 1. Then there is a subjective model of type 2 $(\mathcal{M}', W'_a)$ which is equivalent to $(\mathcal{M}, W_a)$.

So from now on, by subjective model we mean subjective models of type 2. Even if we do not use subjective models of type 1 anymore, their introduction was useful. Indeed, they helped us to motivate the introduction of subjective models of type 2 and, even if it is beyond the scope of this paper, they turn out to be very useful in a dynamic setting (see conclusion).

Thanks to the truth conditions we can now define the notions of satisfiability and validity of a formula. The truth conditions are defined for any world of the subjective model. However, the satisfiability and the validity should not be defined relatively to any possible world of the subjective model (as it is usually done in epistemic logic) but only to the possible worlds of the actual equivalence class. Indeed, these are the worlds that do count for the agent Y in a subjective model: they are the worlds that agent Y actually considers possible. The other possible worlds are just here for technical reasons in order to express the other agents' beliefs (in these worlds). This leads us to the following definition of satisfiability and validity.

Definition 11. *Let $\varphi \in \mathcal{L}$. The formula φ is subjectively satisfiable if there is a subjective model $(\mathcal{M}, W_a)$ and there is $w \in W_a$ such that $\mathcal{M}, w \models \varphi$. The formula φ is subjectively valid if for all subjective models $(\mathcal{M}, W_a)$, $\mathcal{M}, w \models \varphi$ for all $w \in W_a$. In this last case we write $\models_{\mathsf{Subj}} \varphi$.*

Remark 2. One could define the notions of subjective satisfiability and subjective validity differently. One could say that $\varphi \in \mathcal{L}$ is satisfiable if there is a subjective model $(\mathcal{M}, W_a)$ such that $\mathcal{M}, w \models \varphi$ for all $w \in W_a$. Then, following this new definition, $\varphi \in \mathcal{L}$ is valid if for every subjective model $(\mathcal{M}, W_a)$, there is $w \in W_a$ such that $\mathcal{M}, w \models \varphi$.

This second notion of subjective validity corresponds to Gärdenfors' notion of validity [10]. In fact these two notions of subjective validity correspond to the two notions of validity introduced by Levi and Arlo Costa [16]: they call the first one "positive validity" and the second one "negative validity".

These two notions coincide in the single agent case but not in the multi-agent case. Indeed, the Moore sentence $p \wedge \neg B_Y p$ is positively satisfiable but not negatively satisfiable. Nevertheless there are some connections between them. We can indeed prove that $\varphi \in \mathcal{L}$ is positively valid if and only if $B_Y \varphi$ is negatively valid. Moreover, both have advantages and drawbacks. On the one hand, positive validity is intuitive because it says that a formula φ is valid if in every possible situation, the agent Y believes φ. However positive satisfiability is less intuitive because φ is positively satisfiable if there exists a situation in which the agent Y does not reject φ. On the other hand, negative satisfiability is also intuitive because it says that φ is negatively satisfiable if there exists a situation in which agent Y believes φ. However negative validity is less intuitive because it says that φ is negatively valid if in every situation agent Y does not reject φ.

In modal logic [8] there are two notions of semantic consequence. In the subjective approach we can also define two notions of semantic consequence.

Definition 12. *Let C be a class of subjective models; let Σ be a set of formulas of $\mathcal{L}$ and let $\varphi \in \mathcal{L}$.*

- *We say that φ is a local subjective consequence of Σ over C, noted $\Sigma \models_{\mathsf{C}} \varphi$, if for all subjective models $(\mathcal{M}, W_a) \in \mathsf{C}$ and all $w \in W_a$, if $\mathcal{M}, w \models \Sigma$ then $\mathcal{M}, w \models \varphi$.*
- *We say that φ is a global subjective consequence of Σ over C, noted $\Sigma \models^g_{\mathsf{C}} \varphi$, if and only if for all subjective models $(\mathcal{M}, W_a) \in \mathsf{C}$, if $\mathcal{M}, w \models \Sigma$ for all $w \in W_a$ then $\mathcal{M}, w \models \varphi$ for all $w \in W_a$.*

For example, if we take any class C of subjective models then it is not necessarily the case that $\varphi \models_{\mathsf{C}} B_Y \varphi$, whereas we do have that $\varphi \models^g_{\mathsf{C}} B_Y \varphi$. Moreover, these two notions can be informally associated to the two notions of satisfiability mentioned in Remark 2: local subjective consequence can be associated to positive satisfiability and the global subjective consequence can be associated to negative satisfiability.

4 Some connections between the subjective and the objective approach

Intuitively, there are some connections between the subjective and the objective approach. Indeed, in the objective approach the modeler is supposed to know perfectly how the agents perceive the surrounding world. So from the model she builds we should be able to extract the subjective model of each agent. Likewise, it seems natural to claim that for the agent Y a formula is true if and only if, objectively speaking, the agent Y believes this formula. In this section we are going to formalize these intuitions.

4.1 From objective model to subjective model and vice versa

First we define the notion of objective model. An objective model is a pointed epistemic model $(M, w_a) = (W, R, V, w_a)$ where $w_a \in W$ and the accessibility relations R_j are serial, transitive and euclidean. So what we call an objective model is just a standard pointed epistemic model used in epistemic logic. An objective model is supposed to model truthfully and from an external point of view how all the agents involved in the same situation perceive the actual world (represented formally by w_a). This is thus simply the type of model built by the modeler in the objective approach spelled out in the introduction. The language and truth conditions for these objective models are the same as in epistemic logic and are spelled out in Definitions 8 and 3. The notion of objective validity is also the same as in epistemic logic and we say that $\varphi \in \mathcal{L}$ is *objectively valid*, noted $\models_{\mathsf{Obj}} \varphi$, if for all objective model (M, w), $M, w \models \varphi$ (and similarly for *objective satisfiability*).

Now for a given objective model representing truthfully how a situation is perceived by the agents, we can extract for each agent her subjective model pertaining to this situation.

Definition 13. *Let (M, w_a) be an objective model and let $j \in G$. The model associated to agent j and (M, w_a) is the submodel of M generated by $R_j(w_a)$. Besides $R_j(w_a)$ is its actual equivalence class.*

Because the objective model is supposed to model truthfully the situation, w_a does correspond formally to the actual world. So $R_j(w_a)$ are the worlds that the agent j actually considers possible in reality. In agent j's subjective model pertaining to this situation, these worlds should then be the worlds of the actual equivalence class. Finally, taking the submodel generated by these worlds ensures that the piece of information encoded in the worlds $R_j(w_a)$ in the objective model is kept unchanged in the associated subjective model.

Proposition 2. *Let (M, w_a) be an objective model. The model associated to agent j and (M, w_a) is a subjective model.*

Example 4. In Figure 4 is depicted the 'coin example' after Bob's cheating (see introduction). Here p stands for 'the coin is heads up'; A for Ann and B for Bob. We can check that in the objective model, Ann does not know whether the coin is heads or tails up and moreover believes that Bob does not know neither. This is also true in the subjective model associated to Ann. However, in the objective model, Bob knows that the coin is heads up but this is false in the subjective model associated to Ann and true in Bob's subjective model. Note finally that the subjective model associated to Bob is the same as the objective model. This is because we assumed that Bob perceived correctly the situation and what happened.

So we know from an objective model how to obtain the subjective model of each agent. But the other way round, we could wonder how to get the objective model (of a particular situation) if we suppose given the subjective models of each agent. In that case we must moreover assume that the modeler knows the real state of the world, more precisely she knows what propositional facts are true in the actual world. We can then introduce a single world w_a whose valuation V_a satisfies these propositional facts. The objective model is built by setting accessibility relations indexed by j from w_a to the actual equivalence class of j's subjective model, and so for each agent j.

4.2 A semantic correspondence

As we said in Section 3.2, the language of the subjective approach is the same as that of the objective approach. This enables us to draw easily some connections between the two approaches.

Proposition 3. *For all $\varphi \in \mathcal{L}$, $\models_{\mathsf{Subj}} \varphi$ iff $\models_{\mathsf{Obj}} B_Y \varphi$*

Intuitively, this result is correct: for you φ is true if and only if from an external point of view you believe that φ is the case. (Note that this result does not hold for the notion of negative validity.)

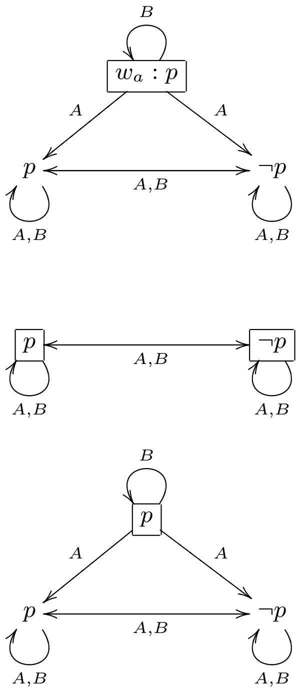


Fig. 4. Objective model (M, w_a) (*up*); Subjective model associated to Ann (*center*), Subjective model associated to Bob (*down*).

As we said earlier, instead of subjective models, agent Y might have formulas 'in her mind' in order to represent the surrounding world. But to draw inferences from them she needs a proof system. In other words, we still need to axiomatize the subjective semantics. That is what we are going to do now.

5 Axiomatization of the subjective semantics

First some notations. Let Obj designate from now on the logic $\mathsf{KD45_G}$. So for all $\varphi \in \mathcal{L}$, $\vdash_{\mathsf{Obj}} \varphi$ iff $\varphi \in \mathsf{KD45_G}$.

Definition 14. *The subjective logic* Subj *is defined by the following axiom schemes and inference rules:*

T $\vdash_{\mathsf{Subj}} B_Y\varphi \to \varphi$;
$\mathsf{S\text{-}O}$ $\vdash_{\mathsf{Subj}} \varphi$ *for all* $\varphi \in \mathcal{L}$ *such that* $\vdash_{\mathsf{Obj}} \varphi$;
MP *if* $\vdash_{\mathsf{Subj}} \varphi$ *and* $\vdash_{\mathsf{Subj}} \varphi \to \psi$ *then* $\vdash_{\mathsf{Subj}} \psi$.

Let us have a closer look at the axiom schemes. The first one tells us that for you, everything you believe is true. This is coherent if we construe the notion of belief as conviction. The second one tells us that you should believe everything which is objectively true, i.e. which is true independently of your own subjectivity. Finally note that the necessitation rule ($\vdash_{\mathsf{Subj}} \varphi$ implies $\vdash_{\mathsf{Subj}} B_j\varphi$ for all j) is not present, which is intuitively correct. Indeed, if for you φ is true (i.e. you believe φ) then in general there is no reason that you should believe that the other agents believe φ as well.

As we announced in Section 3.2, if we add a common knowledge operator to our language then the axiomatization is identical, but in that case formulas φ in the proof system belong to the epistemic language with common knowledge.

Remark 3. In Remark 2, we proposed an alternative definition of validity for the subjective semantics, called negative validity. We do not have a complete axiomatization for the negative validity. However we know that the axiom scheme $\varphi \to B_j\varphi$ is valid but Modus Ponens does not hold anymore.

Theorem 2 (soundness and completeness). *For all $\varphi \in \mathcal{L}$,*

$$\models_{\mathsf{Subj}} \varphi \text{ iff } \vdash_{\mathsf{Subj}} \varphi.$$

From this axiomatization we can prove other nice properties.

Proposition 4. *For all $\varphi \in \mathcal{L}$,*

1. $\vdash_{\mathsf{Subj}} \varphi$ *iff* $\vdash_{\mathsf{Obj}} B_Y\varphi$;
2. $\vdash_{\mathsf{Subj}} \varphi$ *iff* $\vdash_{\mathsf{Subj}} B_Y\varphi$.

Finally the subjective logic Subj has also nice computational properties. Its complexity turns out to be the same as in the objective approach.

Theorem 3. *The validity problem for the subjective logic Subj is decidable and it is $PSPACE$-complete if $N \geq 3$.*

Remark 4. Soon after Hintikka's seminal book was published [6], an issue now known as the logical omniscience problem was raised by Castañeda about Hintikka's epistemic logic: his "senses of 'knowledge' and 'belief' are much too strong [. . .] since most people do not even understand all deductions from premises they know to be true" [17]. It sparked a lot of work aimed at avoiding this problem (such as [18], [19] or [20]).

While we believe that it is indeed a problem when we want to model or describe human-like agents, we nevertheless believe that it is not really a problem for artificial agents. Indeed, these agents are supposed to reason according to our subjective logic and because of its decidability, artificial agents are in fact logically omniscient (even if it will take them some time to compute all the deductions given the complexity of our logic).

6 Conclusion

In the introduction, we have identified what we claim to be the only three possible modeling approaches by proceeding by successive dichotomies. Afterwards, we have focused on the subjective approach for which a logical formalism is missing in a multi-agent setting, although such a formalism is crucial if we want to design autonomous agents. We have proposed one by generalizing the possible world semantics of the AGM belief revision theory. This formalism enabled us to draw some formal links between the objective and the subjective approach which are in line with our intuitions of these two approaches. Finally, we have provided an axiomatization of our formalism whose axioms are also in line with our intuitions of the subjective approach.

However, there are still open problems to be solved. First we do not have an axiomatization for the notion of negative validity. Second, we still need to prove that the subjective logic Subj is $PSPACE$-complete for $N = 2$.

Finally, we have not considered whether and how we could add dynamics to our formalism. In fact one can show that our formalism, and more precisely the notion of multi-agent possible world, allows for a straightforward transfer of the AGM results to the multi-agent case. To do so, we follow the belief base approach and represent a belief base in a multi-agent setting by an epistemic formula ψ. Then we replace possible worlds in the AGM theory by multi-agent possible worlds and we replace the propositional language of AGM theory by the epistemic language. This means that the models of the epistemic formula ψ are the multi-agent possible worlds that satisfy ψ. Then the definition of a faithful ordering $\leq_\psi$ on multi-agent possible worlds for a given epistemic formula ψ is the same as the definition of a faithful ordering $\leq_{\psi'}$ on possible worlds for a given propositional formula ψ'. Intuitively, $(M, w) \leq_\psi (M', w')$ means that the multi-agent possible world (M, w) is closer to ψ than the multi-agent possible world (M', w'). Likewise, the rationality postulates for belief revision in a multi-agent setting are the same as in the AGM theory except that propositional formulas are replaced by epistemic formulas. Then we can show, as in the AGM theory, that a revision operator satisfies these postulates in a multi-agent setting if and only if the models of the revision of the belief base ψ by the epistemic formula φ are the multi-agent possible worlds that satisfy φ and which are minimal with respect to $\leq_\psi$.

References

1. Baltag, A., Moss, L.: Logic for epistemic program. Synthese **139** (2004) 165–224
2. Battigalli, P., Bonanno, G.: Recent results on belief, knowledge and the epistemic foundations of game theory. Research in Economics **53** (1999) 149–225
3. Fagin, R., Halpern, J.Y., Moses, Y., Vardi, M.Y.: Reasoning about knowledge. MIT Press (1995)
4. Nittka, A.: A Method for Reasoning about other Agents' Beliefs from Observations. PhD thesis, University of Leipzig (2008)

5. Booth, R., Nittka, A.: Reconstructing an agent's epistemic state from observations about its beliefs and non-beliefs. Journal of Logic and Computation (2007) accepted for publication.
6. Hintikka, J.: Knowledge and Belief, An Introduction to the Logic of the Two Notions. Cornell University Press, Ithaca and London (1962)
7. Alchourrón, C.E., Gärdenfors, P., Makinson, D.: On the logic of theory change: Partial meet contraction and revision functions. J. Symb. Log. **50** (1985) 510–530
8. Blackburn, P., de Rijke, M., Venema, Y.: Modal Logic. Volume 53 of Cambridge Tracts in Computer Science. Cambridge University Press (2001)
9. Lenzen, W.: Recent Work in Epistemic Logic. Acta Philosophica 30. North Holland Publishing Company (1978)
10. Gärdenfors, P.: Knowledge in Flux (Modeling the Dynamics of Epistemic States). Bradford/MIT Press, Cambridge, Massachusetts (1988)
11. Voorbraak, F.: As Far as I know. Epistemic Logic and Uncertainty. PhD thesis, Utrecht University (1993)
12. Cohen, P.R., Levesque, H.J.: Intention is choice with commitment. Artificial intelligence **42** (1990) 213–261
13. Rao, A.S., Georgeff, M.P.: Modeling rational agents within a BDI-architecture. In Fikes, R., Sandewall, E., eds.: Proceedings of Knowledge Representation and Reasoning (KR & R-91), Morgan Kaufmann Publishers (1991) 473–484
14. van der Hoek, W., van Linder, B., Meyer, J.J.C.: logic of capabilities. In Nerode, A., Matiyasevich, Y., eds.: Proceedings of the Third International Symposium on the Logical Foundations of Computer Science (LFCS'94). Volume 813 of LNAI., Springer Verlag (1994) 366–378 extended abstract.
15. Shoham, Y.: Agent-oriented programming. Artificial Intelligence **60** (1993) 51–92
16. Arlo Costa, H., Levi, I.: Two notions of epistemic validity (epistemic models for Ramsey's conditionals). Synthese **109** (1996) 217–262
17. Castañeda, H.N.: Review of 'knowledge and belief'. Journal of Symbolic Logic **29** (1964) 132–134
18. Levesque, H.: A logic of implicit and explicit knowledge. In: AAAI-84, Austin Texas (1984) 198–202
19. Fagin, R., Halpern, J.Y.: Belief, awareness, and limited reasoning. Artificial Intelligence **34** (1988) 39–76
20. Duc, H.N.: Resource-Bounded Reasoning about Knowledge. PhD thesis (2001)

Proposition 5 (Proposition 1).

Let $(\mathcal{M}, W_a)$ be a subjective model of type 2. Then there is a subjective model of type 1 $(\mathcal{M}', W_a')$ which is equivalent to $(\mathcal{M}, W_a)$.

Let $(\mathcal{M}, W_a)$ be a subjective model of type 1. Then there is a subjective model of type 2 $(\mathcal{M}', W_a')$ which is equivalent to $(\mathcal{M}, W_a)$.

Proof. We only prove the first part. The second part is obvious given what has been said so far.

Let $(\mathcal{M}, W_a) = (W, R, V)$ be a subjective model of type 2. For each $w \in W_a$ we define a corresponding multi-agent possible world (M^w, w) as follows: for all $k \neq Y$, let $M^k := (W^k, R^k, V^k)$ be the submodel of M generated by $R_k(w)$. The multi-agent possible world $(M^w, w) = (W^w, R^w, V^w, w)$ is then defined as follows.

- $W^w = \{w\} \cup \bigcup_{k \neq Y} W^k$;
- $R_j^w = (R_j \cup \bigcup_{k \neq Y} R_j^k) \cap W^w \times W^w$ for all $j \in G$;
- $V^w(p) = (V(p) \cup \bigcup_{k \neq Y} V^k(p)) \cap W^w$.

Then one can easily show that (M^w, w) is a multi-agent possible world and that $\mathcal{M}' := \{(M^w, w); w \in W_a\}$ is a subjective model of type 1 which is equivalent to $(\mathcal{M}, W_a) := (W, R, V)$.

Proposition 6 (Proposition 2). *Let (M, w_a) be an objective model. The model associated to agent j and (M, w_a) is a subjective model.*

Proof. Let $(\mathcal{M}', W_a')$ be the subjective model associated to agent j and (M, w_a) (with $\mathcal{M}' = (W', R', V')$).

Obviously, $\mathcal{M}'$ is generated by W_a. By the generated submodel property, R_j' is serial, transitive and euclidean for all j. Finally, because R_j is euclidean, for all $w \in W_a (= R_j(w_a))$, $R_j(w) = R_j(w_a) = W_a$.

So $(\mathcal{M}', W_a)$ is indeed a subjective model.

Proposition 7 (Proposition 3). *For all $\varphi \in \mathcal{L}$, $\models_{\mathsf{Subj}} \varphi$ iff $\models_{\mathsf{Obj}} B_Y \varphi$*

Proof. For all $\varphi \in \mathcal{L}$, $\models_{\mathsf{Subj}} \varphi$ iff $\models_{\mathsf{Obj}} B_Y \varphi$ amounts to prove that for all $\varphi \in \mathcal{L}$, φ is subjectively satisfiable iff $\hat{B}_Y \varphi$ is objectively satisfiable.

Assume that φ is subjectively satisfiable. Then there is a subjective model $(\mathcal{M}, W_a)$ and $w \in W_a$ such that $\mathcal{M}, w \models \varphi$. But $w \in R_Y(w)$, so $\mathcal{M}, w \models \hat{B}_Y \varphi$. Besides, $(\mathcal{M}, w)$ can be viewed as an objective model. So $\hat{B}_Y \varphi$ is objectively satisfiable.

Assume that $\hat{B}_Y \varphi$ is objectively satisfiable. Then there is an objective model (M, w_a) such that $M, w_a \models \hat{B}_Y \varphi$. Then there is $w \in R_Y(w_a)$ such that $M, w \models \varphi$. Let $(\mathcal{M}', W_a)$ be the subjective model associated to (M, w_a) and agent Y. Then $w \in W_a$ and $\mathcal{M}', w \models \varphi$ by the generated submodel property. So φ is subjectively satisfiable.

Theorem 4 (soundness and completeness). *For all $\varphi \in \mathcal{L}$,*

$$\models_{\mathsf{\mathit{Subj}}} \varphi \textit{ iff } \vdash_{\mathsf{\mathit{Subj}}} \varphi.$$

Proof. Proving the soundness of the axiomatic system is straightforward. We only focus on the completeness proof.

Let φ be a Subj-consistent formula. We need to prove that there is a subjective model $(\mathcal{M}_{\mathsf{Subj}}, W_a)$, there is $w \in W_a$ such that $\mathcal{M}_{\mathsf{Subj}}, w \models \varphi$.

Let $Sub^+(\varphi)$ be all the subformulas of φ with their negations. Let W_{Subj} be the set of maximal Subj-consistent subsets of $Sub^+(\varphi)$. Let W_{Obj} be the set of maximal Obj-consistent subsets of $Sub^+(\varphi)$. For all $\Gamma, \Gamma' \in W_{\mathsf{Subj}} \cup W_{\mathsf{Obj}}$, let $\Gamma/B_j := \{\psi; B_j\psi \in \Gamma\}$ and $B_j\Gamma := \{B_j\psi; B_j\psi \in \Gamma\} \cup \{\neg B_j\psi; \neg B_j\psi \in \Gamma\}$.

We define the epistemic model $M := (W, R, V)$ as follows:

- $W := W_{\mathsf{Subj}} \cup W_{\mathsf{Obj}}$;
- for all $j \in G$ and $\Gamma, \Gamma' \in W$, $\Gamma' \in R_j(\Gamma)$ iff $\Gamma/B_j = \Gamma'/B_j$ and $\Gamma/B_j \subseteq \Gamma'$;
- $\Gamma \in V(p)$ iff $p \in \Gamma$.

We are going to prove the *truth lemma*, i.e. for all $\psi \in Sub^+(\varphi)$, all $\Gamma \in W$

$$M, \Gamma \models \psi \text{ iff } \psi \in \Gamma$$

We prove it by induction on ψ. The case $\psi = p$ is fulfilled by the definition of the valuation. The cases $\psi = \neg\chi, \psi = \psi_1 \wedge \psi_2$ are fulfilled by the induction hypothesis. It remains to prove the case $\psi = B_j\chi$.

- Assume $\psi \in \Gamma$. Then $\chi \in \Gamma/B_j$. So for all Γ' such that $\Gamma' \in R_j(\Gamma)$, $\chi \in \Gamma'$. So for all Γ' such that $\Gamma' \in R_j(\Gamma)$ $M, \Gamma' \models \chi$ by induction hypothesis. So $M, \Gamma \models B_j\chi$, i.e. $M, \Gamma \models \chi$.
- Assume $M, \Gamma \models B_j\psi$. Then $B_j\Gamma \cup \Gamma/B_j \cup \{\neg\psi\}$ is not Obj-consistent. Assume on the contrary that $B_j\Gamma \cup \Gamma/B_j \cup \{\neg\psi\}$ is Obj-consistent. Then there is $\Gamma' \in W_{\mathsf{Obj}}$ such that $B_j\Gamma \cup \Gamma/B_j \cup \{\neg\psi\} \subseteq \Gamma'$. So $\Gamma/B_j = \Gamma'/B_j$ and $\Gamma/B_j \subseteq \Gamma'$. Then $\Gamma' \in R_j(\Gamma)$ and $\neg\psi \in \Gamma'$, i.e. $\Gamma' \in R_j(\Gamma)$ and $M, \Gamma' \models \neg\psi$ by induction hypothesis. So $M, \Gamma \models \neg B_j\psi$, which is impossible by assumption.
  So $B_j\Gamma \cup \Gamma/B_j \cup \{\neg\psi\}$ is not Obj-consistent. Now we consider two cases: first $\Gamma \in W_{\mathsf{Subj}}$ and then $\Gamma \in W_{\mathsf{Obj}}$.
  1. $\Gamma \in W_{\mathsf{Subj}}$. Then there are $\varphi_1, \ldots, \varphi_n \in \Gamma/B_j$, $\varphi'_1, \ldots, \varphi'_m \in B_j\Gamma$ such that
     $\vdash_{\mathsf{Obj}} \varphi_1 \rightarrow (\varphi_2 \rightarrow \ldots \rightarrow (\varphi_n \rightarrow (\varphi'_1 \rightarrow (\varphi'_2 \rightarrow \ldots \rightarrow (\varphi'_m \rightarrow \psi))))))$. So
     $\vdash_{\mathsf{Obj}} B_j[\varphi_1 \rightarrow (\varphi_2 \rightarrow \ldots \rightarrow (\varphi_n \rightarrow (\varphi'_1 \rightarrow (\varphi'_2 \rightarrow \ldots \rightarrow (\varphi'_m \rightarrow \psi))))))]$
     by the necessitation rule of Obj. So
     $\vdash_{\mathsf{Obj}} B_j\varphi_1 \rightarrow (B_j\varphi_2 \rightarrow \ldots \rightarrow (B_j\varphi_n \rightarrow (B_j\varphi'_1 \rightarrow (B_j\varphi'_2 \rightarrow \ldots \rightarrow (B_j\varphi'_m \rightarrow B_j\psi))))))$. i.e.
     $\vdash_{\mathsf{Obj}} B_j\varphi_1 \rightarrow (B_j\varphi_2 \rightarrow \ldots \rightarrow (B_j\varphi_n \rightarrow (\varphi'_1 \rightarrow (\varphi'_2 \rightarrow \ldots \rightarrow (\varphi'_m \rightarrow B_j\psi))))))$ because for all $i \vdash_{\mathsf{Obj}} \varphi'_i \leftrightarrow B_j\varphi'_i$. So

$\vdash_{\mathsf{Subj}} B_j\varphi_1 \to (B_j\varphi_2 \to \ldots \to (B_j\varphi_n \to (\varphi'_1 \to \ldots \to (\varphi'_m \to B_j\psi))))$ by axiom scheme (S-O).
But $B_j\varphi_1, \ldots, B_j\varphi_n, \varphi'_1, \ldots, \varphi'_m \in \Gamma$. So $B_j\psi \in \Gamma$.

2. $\Gamma \in W_{\mathsf{Obj}}$. Then there are $\varphi_1, \ldots, \varphi_n \in \Gamma/B_j$ and $\varphi'_1, \ldots, \varphi'_m \in B_j\Gamma$ such that
$\vdash_{\mathsf{Obj}} \varphi_1 \to (\varphi_2 \to \ldots \to (\varphi_n \to (\varphi'_1 \to (\varphi'_2 \to \ldots \to (\varphi'_m \to \psi))))))$. So
$\vdash_{\mathsf{Obj}} B_j[\varphi_1 \to (\varphi_2 \to \ldots \to (\varphi_n \to (\varphi'_1 \to (\varphi'_2 \to \ldots \to (\varphi'_m \to \psi))))))]$ by the necessitation rule of Obj. So
$\vdash_{\mathsf{Obj}} B_j\varphi_1 \to (B_j\varphi_2 \to \ldots \to (B_j\varphi_n \to (B_j\varphi'_1 \to (B_j\varphi'_2 \to \ldots \to (B_j\varphi'_m \to B_j\psi))))))$. i.e.
$\vdash_{\mathsf{Obj}} B_j\varphi_1 \to (B_j\varphi_2 \to \ldots \to (B_j\varphi_n \to (\varphi'_1 \to (\varphi'_2 \to \ldots \to (\varphi'_m \to B_j\psi))))))$ because for all i $\vdash_{\mathsf{Obj}} \varphi'_i \leftrightarrow B_j\varphi'_i$.
But $B_j\varphi_1, \ldots, B_j\varphi_n, \varphi'_1, \ldots, \varphi'_m \in \Gamma$. So $B_j\psi \in \Gamma$.

Finally we have shown that in all cases $B_j\psi \in \Gamma$.

So we have proved the truth lemma. Now we need to prove that the accessibility relations R_j are serial, transitive and euclidean.

- Transitivity. Assume that $\Gamma' \in R_j(\Gamma)$ and $\Gamma'' \in R_j(\Gamma')$. i.e. $\Gamma'/B_j = \Gamma''/B_j$ and $\Gamma'/B_j \subseteq \Gamma''$; and $\Gamma/B_j = \Gamma'/B_j$ and $\Gamma/B_j \subseteq \Gamma'$. Then clearly $\Gamma/B_j = \Gamma''/B_j$ and $\Gamma/B_j \subseteq \Gamma''$. i.e. $\Gamma'' \in R_j(\Gamma)$.
- Euclidicity. Assume that $\Gamma' \in R_j(\Gamma)$ and $\Gamma'' \in R_j(\Gamma)$. i.e. $\Gamma/B_j = \Gamma'/B_j$ and $\Gamma/B_j \subseteq \Gamma'$; and $\Gamma/B_j = \Gamma''/B_j$ and $\Gamma/B_j \subseteq \Gamma''$. Then clearly $\Gamma'/B_j = \Gamma''/B_j$ and $\Gamma'/B_j \subseteq \Gamma''$. i.e. $\Gamma'' \in R_j(\Gamma')$.
- Seriality. We only prove the case $\Gamma \in W_{\mathsf{Subj}}$. The case $\Gamma \in W_{\mathsf{Obj}}$ is similar. We are going to show that $B_j\Gamma \cup \Gamma/B_j$ is Obj-consistent.
Assume the contrary. Then there are $\varphi_1, \ldots, \varphi_n \in \Gamma/B_j$ and $\varphi'_1, \ldots, \varphi'_m \in B_j\Gamma$ such that
$\vdash_{\mathsf{Obj}} \varphi_1 \to (\varphi_2 \to \ldots \to (\varphi_n \to (\varphi'_1 \to \ldots \to (\varphi'_{m-1} \to \neg\varphi'_m))))$. So
$\vdash_{\mathsf{Obj}} B_j[\varphi_1 \to (\varphi_2 \to \ldots \to (\varphi_n \to (\varphi'_1 \to \ldots \to (\varphi'_{m-1} \to \neg\varphi'_m))))]$. So
$\vdash_{\mathsf{Obj}} B_j\varphi_1 \to (B_j\varphi_2 \to \ldots \to (B_j\varphi_n \to (B_j\varphi'_1 \to \ldots \to (B_j\varphi'_{m-1} \to B_j\neg\varphi'_m))))$. So
$\vdash_{\mathsf{Obj}} B_j\varphi_1 \to (B_j\varphi_2 \to \ldots \to (B_j\varphi_n \to (\varphi'_1 \to \ldots \to (\varphi'_{m-1} \to \neg\varphi'_m))))$.
But $B_j\varphi_1, \ldots, B_j\varphi_n, \varphi'_1, \ldots, \varphi'_m \in \Gamma$. So $\neg\varphi'_m \in \Gamma$ which is impossible because $\varphi'_m \in \Gamma$.
Finally $B_j\Gamma \cup \Gamma/B_j$ is Obj-consistent. So there is $\Gamma' \in W_{\mathsf{Obj}}$ such that $B_j\Gamma \cup \Gamma/B_j \subseteq \Gamma'$. i.e. there is $\Gamma' \in W$ such that $\Gamma' \in R_j(\Gamma)$

Finally we prove that for all $\Gamma \in W_{\mathsf{Subj}}$, $\Gamma \in R_Y(\Gamma)$ (*).

Let $\Gamma \in W_{\mathsf{Subj}}$. For all $B_Y\varphi \in \Gamma$, $\varphi \in \Gamma$ by axiom scheme (T). So $\Gamma/B_j \subseteq \Gamma$. So $\Gamma \in R_Y(\Gamma)$.

φ is a Subj-consistent formula so there is $\Gamma \in W_{\mathsf{Subj}}$ such that $\varphi \in \Gamma$, i.e. $M, \Gamma \models \varphi$. Let $\mathcal{M}_{\mathsf{Subj}}$ be the submodel generated by $R_Y(\Gamma)$. Then clearly $(\mathcal{M}, W_a)$ with $W_a := R_Y(\Gamma)$ is a subjective model. Finally, because $\Gamma \in R_Y(\Gamma)$ by (*), there is $\Gamma \in W_a$ such that $\mathcal{M}_{\mathsf{Subj}}, \Gamma \models \varphi$.

Theorem 5. *The subjective logic* Subj *is decidable and its validity problem is PSPACE-complete if $N \geq 3$.*

Proof.

- The decidability of Subj can be proved in two ways. First, because Obj is decidable, Subj is also decidable by Proposition 3. Second, because Subj has the finite model property (see proof of Theorem 2), Subj is decidable.
- Because the validity problem is PSPACE-complete for Obj if $N \geq 2$ then the validity problem for Subj is in PSPACE by Proposition 3.
  Besides, as a corollary of the lemma below, we get that the validity problem for Subj is PSPACE-complete if $N \geq 3$ because the validity problem for Obj is PSPACE-complete if $N = 2$.

  Lemma 1. *Assume $\{Y, i, j\} \subseteq G$ and let $\varphi \in \mathcal{L}$ dealing only with agents Y and j. Then,*

  $$\models_{Obj} \varphi \text{ iff } \models_{Subj} t(\varphi)$$

  where $t(\varphi)$ is the formula obtained by replacing every occurence of Y by i.

  Proof. Assume $\varphi \in \mathcal{L}$ dealing only with agents Y and j is objectively satisfiable. Then clearly $t(\varphi)$ is also objectively satisfiable. Let $M = (W, R, V)$ be an objective model generated by $w \in M$ such that $M, w \models t(\varphi)$. Let M' be the epistemic model obtained from M by replacing the accessibility relation R_Y by $R'_Y = \{(v, v); v \in W\}$. Then clearly $M', w \models t(\varphi)$ and (M', w) is a multi-agent possible world. So $t(\varphi)$ is subjectively satisfiable.
  Finally, if $\models_{Obj} \varphi$ then clearly $\models_{Obj} t(\varphi)$. So $\models_{Subj} t(\varphi)$ by axiom S-O.