Epistemic Equivalence and Bisimulation

Jelle Gerbrandy*

January 21, 2005

Abstract

Bisimulation is an equivalence relation over Kripke models. It is often claimed, but little argued for, that bisimulation captures a 'natural' notion of similarity in epistemic semantics, in the sense that the differences between bisimilar models are, from an epistemic point of view, irrelevant. In this paper we examine three types of arguments pertaining to the view that bisimilar models are 'epistemically equivalent.' We start with examining an 'ontological arguments' having to do with what a Kripke model is, or is meant to do, and use a theorem from co-algebra to argue that bisimilar models should, under certain assumptions, represent the same epistemic situations. We then examine the relationship between bisimulation and the expressive power of epistemic logic, and conclude that the two notions do not fit in any precise way. Finally, we identify a number of desirable properties of a notion of epistemic equivalence, and conclude that although bisimulation satisfies all of these properties, the properties themselves do not characterize bisimulation.

Contents

*Dipartimento di Informatica, Universit'a di Torino. Research supported by Lagrange Project of the Fondazione CRT. Comments to `jelle@gerbrandy.com`

1 Introduction

The standard way of providing epistemic logic – the logic of knowledge, belief and information – with a semantics is by using Kripke models. Many Kripke models are similar in structure, and one natural question to ask is which differences between two different models can be, for the purposes of epistemic logic, safely ignored – which models are epistemically equivalent. The assumption that the relation of bisimulation captures this notion of epistemic equivalence is often made, as is exemplified by the following quotes:

- "When are two epistemic models equivalent? When can we consider that two different pointed models (S, s) and (S', s') represent essentially the same epistemic situation? The answer lies in the notion of bisimulation. Bisimilar models capture precisely the same situation." (Baltag (2003))
- van Benthem (2002) remarks about two bisimilar models that "in a natural sense, these are two representations of the same information state."
- van der Meyden (1994) writes about models that "intuitively represent the same information about the state of the world and the beliefs of the agents" and states that bisimulation is "a general notion that captures this sort of invariance."
- "... collapsing distinctions between bisimilar models is harmless." (Gerbrandy (1998))

The question whether two models can be considered 'epistemically equivalent' may seem to be of mostly meta-theoretical or philosophical interest. But the question can be important for how one defines the semantics of certain epistemic operators as well. In dynamic epistemic logic, epistemic actions are modeled as operations on Kripke models. In this context, the question whether an epistemic action has changed the model in any essential way, or whether two actions can be considered equivalent, can only be answered after a decision about what epistemic equivalence is. The definition of a semantics for epistemic actions forces one to be precise about epistemic equivalence between models, and all work in the area, with the exception of Plaza (1989), use bisimulation as its mathematical implementation (e.g. Baltag (2002), Baltag and Moss (2004), Gerbrandy and Groeneveld (1997), van Ditmarsch (2000), and others). Another context in which the question of 'epistemic equivalence' turns up is when defining what distributed knowledge is; we will consider this question later.[1]

To make clear what is at stake, consider a Kripke model Kw_0 that represents the information of two agents, a and b. The model K is an $S5$-model

[1] Also the semantics of 'only knowing,' it is natural to compare different models. Here, it is not so much equivalence that is at stake, but 'having more information,' but some of the questions are similar (cf. van der Hoek and Thijsse (2000))

with four worlds, $\{w_0, w_1, w_2, w_3\}$. The accessibility relation $\rightarrow_a$ for agent a is the smallest equivalence relation for which it holds that $w_0 \rightarrow_a w_1$ and $w_2 \rightarrow_a w_3$, and the accessibility relation for agent b is the smallest equivalence relation with $w_1 \rightarrow_b w_3$ and $w_0 \rightarrow_a w_2$. We assume that there is one propositional variable, p, that is true in w_0 and w_3 only. The model can be pictured as follows:

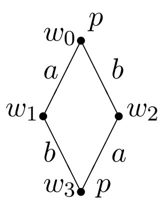

Figure 1:

The models Kw_1 and Kw_2 have exactly the same structure: they are isomorphic. There is a natural sense in which the distinctions between Kw_1 and Kw_2 are irrelevant; if any two models are epistemically equivalent, then surely the models Kw_1 and Kw_2 are.

Consider now the model $K'w'_0$, which results from Kw_0 by identifying the worlds w_1 and w_2 in K: Now, if Kw_1 and Kw_2 are taken to be equivalent,

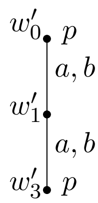

Figure 2:

we have a strong argument that Kw_0 and $K'w'_0$ are epistemically equivalent as well; after all, $K'w'_0$ results from Kw_0 by identifying the worlds w_1 and w_2 that were epistemically equivalent to start with.

We will turn back to these examples later in the paper, and first introduce some formal definitions to make the discussion more precise.

2 Definitions

Mainly to fix notation and terminology, I will quickly review some standard definitions.

Let P be a set of *propositional variables*, and A a finite set of *agents*. A *Kripke model* (for P and A) is a tuple $(W, (\rightarrow_a)_{a \in A}, V, w)$, where W is a set of *possible worlds*, and for each agent $a \in A$, $\rightarrow_a$ is an *accessibility relation* over W. The intended interpretation is that $x \rightarrow_a y$ exactly when an agent a considers y possible in x, i.e. whenever y is consistent with the information that a has in x. The *evaluation function* V assigns to each world in W a function that assigns to each propositional variable a value out of $\{0, 1\}$. Finally, w is an element of W, called the 'the actual world.'

If Kw is a model, we will say that the set of worlds accessible for an agent a, $\{Kv \mid w \rightarrow_a v\}$ is the *information state of* a in w.

When the context allows it, I will confuse the Kripke model Kw with the world w, and write, for example, that $Kw \rightarrow_a Kv$ if $w \rightarrow_a v$. If the Kripke model contains just a single agent, a, I will write $\rightarrow$ instead of $\rightarrow_a$.

Classical modal logic has as its sentences the propositional variables, and is closed under conjunction $\wedge$, negation $\neg$ and modal operators $\Box_a$ for each agent a. We will write $Kw \models \phi$ and say that Kw satisfies ϕ or that ϕ is true in Kw. The interesting clause in the definition of satisfaction is the one for $\Box_a$, which says that a has the information that ϕ just in case ϕ is true in all worlds in the information state of a:

$$w \models \Box_a \phi \quad \text{iff} \quad \text{for all } v \text{ such that } w \rightarrow_a v, v \models \phi$$

Let $\sim$ be a relation between Kripke models. The relation $\sim$ is a *bisimilarity relation* if it has the property that whenever $Kw \sim K'w'$, then (1) $V(w) = V'(w')$, i.e. the worlds satisfy the same propositional variables, and for each agent i, (2) if $w \rightarrow_i v$, then there is a v' such that $w' \rightarrow'_i v'$ and $Kv \sim K'v'$, and, vice versa, (3), if $w' \rightarrow_a v'$ then there is a v such that $w \rightarrow v$ and $Kv \sim K'v'$.

We will say that two models are *bisimilar* just in case there exists a bisimilarity relation connecting the two models; two models are *isomorphic* just in case there is a bisimilarity relation between them that is a one-to-one correspondence. The relation of *bisimulation* is the relation that holds between any two models that are bisimilar.

3 Epistemic logic and bisimulation

There are a number of results that link bisimulation with classical modal logic (cf. Blackburn et al. (2001)), the most central maybe the theorem that bisimilar models cannot be distinguished in classical model logic.

It is sometimes suggested such results provide us with an argument that bisimulation captures a notion of epistemic equivalence. The idea is that elementary equivalence in epistemic logic – satisfying the same sentences of epistemic logic – is as good a notion of epistemic equivalence as any, and that bisimulation corresponds somehow with elementary equivalence.

Clearly, the truth of this claim depends on what is understood by 'epistemic logic.' In the present context, we mean by 'the language of epistemic logic' a formal language that is interpreted in Kripke models and that contains all operators that can, in some way, be argued to be relevant to notions such as knowledge, belief, or information. This is somewhat loosely formulated, but classical modal logic is clearly contained in it, and so are languages with operators for 'only knowing', 'distributed knowledge', 'common knowledge', and operators that represent 'epistemic actions' such as obtaining new information. The 'backwards-looking' modal operator P from temporal logic does not seem to have any natural epistemic interpretation, and so is not part of epistemic logic.[2]

Bisimulation does not seem to be closely related to the text-book definitions of epistemic logic at all; it can be argued that bisimulation both weaker and stronger than elementary equivalence in epistemic logic.

Bisimulation is too strong, in the sense that some elementary equivalent models are not bisimilar – that is, if we take it that epistemic logic is less expressive than infinitary model logic. (The same holds for classical modal logic.)

More interesting is the fact that bisimulation is too weak: it fails to distinguish models that can be distinguished in epistemic logic. There are two kinds of examples that support this claim; the first concerns certain operators from dynamic epistemic logic; lack of space does not allow us to explain the problem here. The other concerns the text-book definition of distributed knowledge.

Distributed knowledge refers to the information that is present in a interpreted system in a 'distributed' way: one agent may know one thing, another agent something else. Combining this information, we might obtain a third piece of information, which is said to be distributed knowledge between the two agents. This information is usually defined as being represented by the set of worlds that are considered possible by each of the agents (Fagin et al. (1995), Meyer and van der Hoek (1995)).

$$Kw \models D\phi \quad \text{iff} \quad \text{for all } v \text{ such that } w \to_b v \text{ for each } b\text{: } Kv \models \phi$$

What interests us here is that the operator D is not preserved over bisimulation: there are sentences containing this operator that are true in one model, but false in a bisimilar model.

[2]Note that we take a somewhat limited view of epistemic logic here: we are only interested in operators that can be given a semantics in terms of Kripke models, and ignore, for example, probability logic, theories of belief revision or of default reasoning, first order epistemic logic, etcetera. Such languages come with their own semantics, and need their own version of bisimulation.

To see this, consider the example of figure 1. The models Kw_0 and $K'w'_0$ are bisimilar, but in Kw_0 it is distributed knowledge that p (since only w'_0 is in the information state of both of the agents), while $K'w'_0 \not\models Dp$.[3]

To some readers, it might seem absurd to say that the information distributed over a and b in the model Kw_0 does not contain a world where p is false. The information of a is represented by a set $\{Kw_0, Kw_1\}$ and that of b by the set $\{Kw_0, Kw_2\}$; the models Kw_1 and Kw_2 are isomorphic; therefore, the information states representing their respective information are isomorphic as well: the two agents have exactly the same information.

Defining distributed knowledge presupposes a certain ontological commitment: a decision about which worlds are 'really' different. The question whether the difference between two isomorphic models is of importance is exactly the problem discussed in the next section – and whether p is distributed knowledge in the model Kw_0 depends on its answer.

4 What is a Kripke model?

4.1 What is a possible world?

It is common usage in mathematical logic to ignore distinctions between isomorphic models, and to consider isomorphic models as, basically, the same. When a class of models is introduced, we postulate (often implicitly) that certain aspects of these models represent genuine logical elements, and that other aspects are irrelevant, artifacts from the set-theoretical tools we use to create the model. Saying that isomorphism (or bisimulation) implies epistemic equivalence is saying that the identity of possible worlds in a model is irrelevant.

However, identifying bisimilar (or isomorphic) Kripke models contrasts with the way these structures are used in economics or computer science, where possible worlds are often endowed with an internal structure that forms an essential part of the meaning of the model.

Consider, for example, the notion of a *type space* that is used in game theory (Harsanyi (1968), Aumann (1999)), or the notion of a *interpreted system* from Fagin et al. (1995). These models contain descriptions of the way the world – the environment of the agents – could be, together with descriptions of possible states the agents could be in (the 'type' or 'local state' of the agents). A possible world is a tuple consisting of a description of the environment together with a type for each of the agents, and a type space (or an interpreted system) is a set of possible worlds.

Given a type space, one may define a Kripke structure by postulating that $w \rightarrow_i v$ just in case i has the same type in w and in v, and choosing a set

[3] See also van der Hoek et al. (1999), who discuss how this definition relates to other notions of group knowledge, and give a similar example.

of propositional variables that describe certain aspects of the environment. In a construction such as this, possible worlds are first class citizens, and a notion of equivalence that abstracts from this internal structure seems to make little sense.[4] If two worlds, for example, differ with respect to type of one of the agents, they represent situations in which that agent has different information, and it seems wrong to identify the two worlds just because they happen to generate similar Kripke structures.

There are two ways of looking upon Kripke models. For a logician or a mathematician, the set of possible worlds in a Kripke model is just an arbitrary set; asking about the internal structure of these worlds seems to be beside the point. For an economist or a computer scientist working with the constructions mentioned, Kripke models are used as a tool to analyze knowledge and information in a type space or interpreted system. Abstracting from the internal structure of possible worlds, and considering models that are bisimilar or isomorphic as essentially the same, makes sense in the former view, but not much in the second. We have seen in section 3 how the distinction between these two viewpoints is important for the definition of epistemic semantics – particularly, for the definition of distributed knowledge.

4.2 Representing epistemic situations

In this section we argue that the first way of viewing Kripke models, i.e. considering the identity of possible worlds to be irrelevant, implies that bisimilar models represent the same epistemic situations, and can, therefore, be considered epistemically equivalent. The argument basically gives an epistemic twist to a proof from Aczel (1988), who proves that from his axiom of anti-foundation it follows that bisimilar graphs represent the same set.

The basic idea is that we take the idea of a Kripke model being a model of something seriously. We show that certain assumptions about this relation of 'being a model of' imply that two models are models of the same situations just in case they are bisimilar. There are certain aspects of an electronic diagram that are important for the diagram being a represention of, say, a particular implementation of an OR-gate (for example, how the transistors are connected to each other), while there are other elements that we consider irrelevant (for example, how the different transistors are placed relative to each other, or the paper that the diagram is printed on). Similarly, for a

[4]Also in the philosophical literature on epistemic logic it is sometimes assumed that possible worlds have an internal structure that is not arbitrary. Hintikka's model sets (Hintikka (1962), in which possible worlds are identified with sets of formula's, are a case in point. We should also mention the 'modal realism' of Lewis (1986), which is not directly of importance for epistemic logic, who argues that we should think about possible worlds as objects that 'really' exist.

Kripke model, there are some aspects that we consider to be relevant for the Kripke model being a model of a certain epistemic situation, and others that are not.

- Whether a Kripke model Kw is a model of a certain situation is determined by i) the truth value of the propositional variables in w and ii) what is modeled by those Kv that are in the information states of the respective agents.

This statement tells us what the aspects of a Kripke model are that make it a model of one situation rather than another: it is the truth of the propositional variables, and the class of situations modeled by the information states assigned to the agents that are important (and not, for example, the predecessors of a world in the accessibility relation, or the identity of the points in the model).

Consider now this claim:

- A Kripke model is a model of a unique class of situations.

What is meant is that the class of possible situations modeled by a Kripke model is unambiguously determined by the model; we know, as it were, what the model means. Just as our schema for an OR-gate unambiguously carves out the class of actual and possible OR-gates that are built according to this schema, a Kripke model unambiguously determines which states of affairs it is a model of, and of which it is not.

Now, somewhat surprisingly, we do not need to be more specific about what exactly the range of the 'being-a-model-of' relation is to prove the following claim.

Proposition 1 Two Kripke models are models of the same situations if and only if they are bisimilar

proof: See the appendix. □

5 What is epistemic equivalence?

The concept of epistemic equivalence is a underspecified one. We have examined two ways of viewing epistemic equivalence: as saying something about models, and as being related to the epistemic logic. In this third part, we take a more classical philosophical approach, and try to identify some properties that a notion of epistemic equivalence should probably satisfy to deserve the name.

5.1 Extensionality

One thing that seems clear is that 'epistemic equivalence' should be an equivalence relation – a relation that is reflexive, transitive and symmetric.

Another property that seems uncontroversial is that when two models are epistemically equivalent, they should satisfy the same propositional variables. It seems natural to extend this property to information states, and assume if two models w and w' are equivalent, then the information state assigned to an agent a in w should be equivalent to the one assigned to her in w'.

- **extensionality** An equivalence relation $\sim$ is *(epistemically) extensional* if it has the property that two worlds Kw and $K'w'$ are equivalent exactly when the same propositions are true in Kw and $K'w'$, and the information state of each agent in Kw is equivalent to her information state in $K'w'$. This definition gives some leeway with regard to how we define equivalence between sets of models on the basis of an equivalence relation between the models, but a natural way is to lift a given equivalence relation to a relation between sets is by saying that the set A is equivalent to B if for each element of A we can find an equivalent element in B, and vice versa.[5]

Saying that $\sim$ is extensional implies, of course, that $\sim$ is a bisimulation, although not all bisimilarity relations are extensional. Bisimulation is in fact the largest extensional relation, i.e. it is the relation that identifies all Kripke models that possibly can be identified, given the principe of extensionality.

Even if extensionality seems to be a natural property for any epistemic equivalence relation, it is easy to find a natural candidate for epistemic equivalence that is not extensional. The relation of satisfying the same sentences of classical modal logic is not extensional, and neither are plain identity or isomorphy.

Another interesting example of an extensional relation is the relation that coincides with identity on all $S5$-models, and with bisimulation on all others. This example shows that with regard to $S5$-models, the assumption of extensionality tells us very little about epistemic equivalence.

5.2 Quotients

In the example in section 1, we obtained a new model by identifying two equivalent worlds. Generalizing this, it should hold that if we take a model, and transform this model into a new one in which all epistemically equivalent worlds are identified, the result should be equivalent to the model we started with.

[5] We use the term 'extensionality' because it is similar to the extensionality principle in set theory, where two sets are identified iff their elements are. Compare also footnote 6 in Baltag (2002), who states exactly this same principle: "... a minimal requirement for the relation R to be acceptable as a good notion of observational equivalence between epistemic states: if we identify two states (via R) then we should identify their content (via =) and their [information state] (via $\breve{R}$)," where $\breve{R}$ is the lifting as defined here.

To define this idea in a precise way, we need a formal definition of what it means to identify two worlds in a model. We will borrow the definition of a quotient of a model under an equivalence relation.

- **preservation under quotients** Let $K = (W, \rightarrow, V)$ be a model, and let $\sim$ be an equivalence relation over W. The *quotient of K under $\sim$*, written as $K^\sim$, is the model $(W^\sim, \rightarrow^\sim, V^\sim)$, where $W^\sim$ consists of all $\sim$-equivalence classes $w^\sim$ of worlds w in W, and we set $w^\sim \rightarrow^\sim v^\sim$ iff there are $w' \in w^\sim$ and $v' \in v^\sim$ with $w' \rightarrow v'$. Finally, we set $V^\sim(w^\sim)(p) = 1$ iff for all $w' \in w^\sim$ it holds that $V(w')(p) = 1$.

  We say that an equivalence relation $\sim$ between Kripke models is *preserved under quotients* just in case $Kw \sim K^\sim w^\sim$ for each Kw in the domain of $\sim$.

The definition works for any equivalence relation, but it makes not much sense to identify two worlds in a model in which, for example, the propositional variables have different truth values.

Bisimulation is preserved under quotients. This fact is used a lot in the literature, sometimes in an implicit way, but sometimes explicitly as well (for didactic or presentational reasons, as it allows for small or elegent representations, as in de Lavalette and van Ditmarsch (2002) or in van Benthem (2002), or for efficient representations in 'epistemic theorem proving', as in van Eijck (2004))).

Another example of a relation that is preserved under quotients is that of satisfying the same sentences of classical modal logic.

Relations that are not preserved under quotients are isomorphy and the relation of satisfying the same sentences of the language of classical modal logic with the operator for distributed knowledge.

Together, the properties of extensionality and preservation under quotients seem to be close to characterizing bisimulation; we do not know of any 'natural' equivalence relation – a relation that is defined in a simple and intuitively plausible way – that satisfies both principles. However, the two conditions do not single out bisimulation.

Proposition 2 There is an extensional equivalence relation that is preserved under quotients, but that is strictly weaker than bisimulation.

proof: See the appendix. □

6 Conclusions

There are two ways of looking upon a Kripke model as a model for epistemic logic. The first viewpoint, common in logic and mathematics, takes the definition of a Kripke model at face value. From this viewpoint, it is natural to consider the identity of the possible worlds to be irrelevant, and the argument of section 4.2 leads to the conclusion that bisimilar models must represent the same epistemic situations. The second viewpoint sees the Kripke model as a tool to reason about knowledge in an interpreted system or a type space. If two worlds in a type space are different, they are different for a reason (otherwise, there would have been one world instead of two different ones), and identifying worlds that are bisimilar or even isomorphic in the Kripke model used to reason about the type space seems to make little sense.

We used the case of distributed knowledge to point out that bisimulation does not seem to be related to the expressive power of epistemic logic in any obvious way – one cannot argue that bisimulation captures epistemic equivalence on the basis of the expressive power of epistemic logic alone.

Bisimulation does have some properties that seem natural for a relation of epistemic equivalence: it is extensional, and it is preserved under quotients. Plain old identity does not satisfy both properties, neither does isomorphy, and neither does elementary equivalence.

What is the practical import of these observations? For the greater part of the literature on epistemic semantics, the notion of epistemic equivalence plays a marginal role at best: the question after epistemic equivalence can be left unanswered, and the two ways of viewing Kripke models can peacefully coexist. But there are at least two cases for which the questions discussed in this paper are of some importance: for defining distributed knowledge, and for defining dynamic epistemic semantics. As we noted in the introduction, in most definitions of dynamic epistemic semantics, the claim that bisimilar models can be identified is an essential part of the definitions, and there does not seem to be any way to get around this aspect. This leads to two problems: first of all, it makes it hard to see how these definitions can be combined with the type-space model or the multi-agent model, and secondly, it means that the classical definitions of distributed knowledge cannot be straightforwardly added to these systems.

With regard to the definition of distributed knowledge, another remark is in order. We can use our observations on epistemic equivalence as constraints on what kind of epistemic operators make sense (compare Baltag (2003)). The argument from 4.2 that bisimilar models represent the same situations, and the observation that elementary equivalence in the language with distributed knowledge is not preserved under quotients can both be seen as arguments that the standard definition of distributed knowledge in terms of intersection must be mistaken.

References

Aczel, P.: 1988, *Non-well-founded sets.* CSLI Publications, Stanford.

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

Baltag, A.: 2002, 'A logic for suspicious players: epistemic actions and belief-updates in games'. *Bulletin of Economic Research* **54**(1), 0307–3378.

Baltag, A.: 2003, 'Logics for Communication: reasoning about information flow in dialogue games.'. Lecture notes for NASSLLI 2003.

Baltag, A. and L. S. Moss: 2004, 'Logics for Epistemic Programs'. *Synthese* **139**(2), 165–224.

Blackburn, P., M. de Rijke, and Y. Venema: 2001, *Modal Logic*, Theoretical Tracts in Computer Science. Cambridge University Press.

de Lavalette, G. R. R. and H. van Ditmarsch: 2002, 'Epistemic actions and minimal models'. In: P. Balbiani, N.-Y. Suzuki, and F. Wolter (eds.): *Advances in Modal Logic (conference proceedings).* pp. 77 – 90, Institut de Recherche en Informatique de Toulouse IRIT.

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

Gerbrandy, J.: 1998, 'Bisimulations on Planet Kripke'. Ph.D. thesis, Universiteit van Amsterdam. ILLC Dissertation Series DS-1999-01.

Gerbrandy, J. and W. Groeneveld: 1997, 'Reasoning about Information Change'. *Journal of Logic, Language and Information* **6**(2), 147–169.

Harsanyi, J. C.: 1967-1968, 'Games with incomplete information played by Bayesian players'. *Management Science* **14**(3), 159–182, 320–334, 486–502.

Hintikka, J.: 1962, *Knowledge and Belief.* Cornell U. Press.

Lewis, D.: 1986, *On the Plurality of Worlds.* Blackwell.

Meyer, J.-J. C. and W. van der Hoek: 1995, *Epistemic Logic for AI and Computer Science.* Cambridge University Press.

Plaza, J.: 1989, 'Logics of Public Communications'. In: M. Emrich, M. Pfeifer, M. Hadzikadic, and Z. Ras (eds.): *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems.* New York: Academic Press, pp. 201–216.

van Benthem, J.: 2002, ''One is a Lonely Number': on the logic of communication'. Manuscript.

van der Hoek, W. and E. Thijsse: 2000, 'A General Approach To Multi-Agent Minimal Knowledge'. In: *Proceedings of the European Workshop on Logics in Artificial Intelligence*. pp. 254 – 268.

van der Hoek, W., B. van Linder, and J.-J. Meyer: 1999, 'Group knowledge is not always distributed (neither is it always implicit)'. *Mathematical Social Sciences* **38**(2), 215–240.

van der Meyden, R.: 1994, 'Mutual Belief Revision (Preliminary Report)'. In: *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning*. pp. 595–606.

van Ditmarsch, H.: 2000, 'Knowledge games'. Ph.D. thesis, Groningen University. ILLC Dissertation Series 2000-06.

van Eijck, J.: 2004, 'Dynamic Epistemic Modelling'. Technical report, Centrum voor Wiskunde en Informatica. Report SEN-E0424.

Appendix

Proposition 1 Two Kripke models are models of the same situations if and only if they are bisimilar

proof: The proof is a straightforward adaptation from the proof in Aczel (1988) that the 'axiom of anti-foundation' implies that two bisimilar pictures are pictures of the same set. To keep matters simple, we prove the statement for models with just a single agent, but the proof can be generalized straighforwardly.

We are interested in the relation 'being-a-model-of', which is a relation between Kripke models and situations. This relation 'being-a-model-of' defines a function, δ, that assigns to each Kripke model the class of situations that are modeled by it. Condition (1) can be rephrased using this function (we are a bit loose here with our use of set-theoretical notation for what could be classes of classes of situations):

(1) $\delta(w) = \delta(v)$ iff $V(w) = V(v)$ and $\{\delta(w') \mid w \to w'\} = \{\delta(v') \mid v \to v'\}$

Following terminology of Aczel (1988), we will call a function δ that satisfies the above condition a *decoration* of the Kripke model.

From (1) it follows that if $\delta(w) = \delta(v)$, then w and v are bisimilar. This is relatively easy to show, by defining a relation $\sim$ by setting $a \sim b$ iff $\delta(a) = \delta(b)$, and showing that this relation is a bisimilarity relation.

The condition that each model defines a unique class of situations modeled by it boils down to saying that a Kripke model has a unique decoration:

(2) Each Kripke model has a unique decoration.

From (1) and (2), it follows that if w and v are bisimilar, then $\delta(w) = \delta(v)$. This takes some work.

We define the canonical model $\mathcal{S} = (S, \rightarrow, V)$ as follows. S is the class of classes of situations that are modeled by some Kripke model, i.e. it $S = \{\delta(w) \mid w \text{ is a Kripke model, and } \delta \text{ is its decoration }\}$. We set $\delta(w) \rightarrow \delta(v)$ iff $w \rightarrow v$, and $V(\delta(w)) = V(w)$, for each w and v. This is well-defined because δ satisfies the conditions given by (1).

As S is a class, the canonical model is, strictly speaking, not a Kripke model. But we can easily extend our notion of bisimulation to apply the the canonical model as well.

Our first claim is that any model is bisimilar to its decoration in the canonical model, i.e. that each Kw is bisimilar to $\mathcal{S}\delta(w)$. This follows from the observation that the relation $w \sim s$ iff $\delta(w) = s$ is a bisimilarity relation.

Our second claim is that the canonical model has the property that no different worlds in it are bisimilar (a model with this property is called strongly bisimulation-extensional).

To see this, let s and s' be bisimilar worlds in the canonical model. We show that s and s' are in fact the same. Define a Kripke model $\mathcal{K} = (W, \rightarrow, V)$ by letting W be the cartesian product of the set that includes s and all worlds that can be reached by $\rightarrow$-transitions from s, with the set that contains s' and all worlds that are reachable from s'. We are now back in the set-theoretical universe: W is a proper set. In the new model, we let $(x, y) \rightarrow (x', y')$ iff $x \rightarrow x'$ and $y \rightarrow y'$ in the canonical model, and let $V(x, y) = V(x)$ (note that by bisimilary, $V(x) = V(y)$).

We now define projection functions δ_0 and δ_1 by setting $\delta_0(x, y) = x$ and set $\delta_1(x, y) = y$ for each x and y. By the way that the model is constructed, δ_0 and δ_1 are decorations of $\mathcal{K}$, so by (2), they must be the same. In particular, this means that $\delta_0(s, s') = \delta_1(s, s')$, and thus, $s = s'$.

To prove our statement, suppose w and v are bisimilar. Then $\delta(w)$ and $\delta(v)$ are bisimilar as well, and therefore, $\delta(w) = \delta(v)$. □

Proposition 2 There is an extensional equivalence relation that is preserved under quotients, but that is not equal to bisimulation.

proof:

Let $\cong$ be the relation of isomorphy, and let $\simeq$ be the relation that holds between Kw and $K'w'$ iff $\{Kv \mid w \rightarrow v\} \cong \{K'v' \mid w' \rightarrow v'\}$ (the two models are "isomorphic in Finsler's sense" in the terminology of Aczel (1988)).

Let $(\cdot)^{\simeq}$ is the operation of taking the quotient of a model under isomorphy, let $(Kw)^*$ be the the fixed point of $(\cdot)^{\simeq}$ applied to Kw relative to isomorphy. More formally we define, for each ordinal α, the model $(Kw)^\alpha$:

- $(Kw)^0 = Kw$
- $(Kw)^{\alpha+1} = ((Kw)^\alpha)^{\simeq}$

- if α is a limit ordinal, $(Kw)^\alpha = (W, \to, V, w^\alpha)$, where:
  $x^\alpha = \{Kx' \mid (Kx')^\beta \cong (Kx)^\beta \text{ for some } \beta < \alpha\}$
  $W = \{x^\alpha \mid x \text{ is a world in } K\}$
  $(Kx)^\alpha \to (Kx')^\alpha$ iff there is an $\beta < \alpha$ such that $(Kx)^\beta \to (Kx')^\beta$
  $V(x^\alpha) = V(x)$

We now let $(Kw)^* = (Kw)^\alpha$ for the smallest α such that $(Kw)^\alpha = (Kw)^{\alpha+1}$, which is easily seen to exist. We slightly abuse notation, and use x^* or $(Kx)^*$ to refer to the worlds x^α in $(Kw)^*$.

Note that the function $(\cdot)^*$ takes a model to its unique (modulo isomorphism) strongly $\simeq$-extensional representation– in $(Kw)^*$, there is no two $\simeq$-equivalent worlds.

The example we are looking for is the relation $\equiv$ that is defined as: $Kw \equiv K'w'$ iff $(Kw)^*$ is isomorphic to $(K'w')^*$.

We claim that $\equiv$ is extensional and preserved under quotients, and that it is stricly weaker than bisimulation.

In the proof, we will omit reference to the model; the variables w, w' and x will refer to worlds in K, and v, v' and y are worlds in K', and we will write w^* for $(Kw)^*$ (etc.). To simplify, we will omit the part of the proof that refers to the valuations V.

1) $\equiv$ is extensional.

For suppose $\{w' \mid w \to w'\} \equiv \{v' \mid v \to v'\}$. Then for each w' with $w \to w'$ there is a v' with $v \to v'$ and w'^* isomorphic to v'^*. It follows immediately that $\{w'^* \mid w \to w'\} \cong \{v'^* \mid v \to v'\}$; this means that $w^* \simeq v^*$. Since w^* and v^* are $\simeq$-extensional, it holds that w^* and v^* are isomorphic, and therefore, $w \equiv v$.

For the other direction, assume $w \equiv v$, and let $w \to w'$. As w^* is isomorphic to v^*, there is some v'^* such that $w' * \cong v'^*$ and $v^* \to v'^*$. This means that there is some v' such that $w' \equiv v'$ and $v \to v'$.

2) $\equiv$ is preserved under quotients, i.e. $w^\equiv \equiv w$.

We show that $w^\equiv \cong w^*$. Let R be the relation given by $x^\equiv R x^*$. By definition, $x \equiv y$ iff $x^* \cong y^*$, and since x^* and y^* occur in the same model, it holds that $x^* \cong y^*$ iff $x^* = y^*$. So R is one-one. It is not difficult to see that R is a bisimilarity relation as well, and so it is an isomorphism.

We need to show that $(w^\equiv)^* \cong w^*$, which follows from the previous statement and the fact that $(w^*)^* = w^*$

3) $\equiv$ is strictly weaker than bisimulation. Consider the model K with worlds $\{w_1, w_2\}$ with $w_1 \to w_2 \to w_2 \to w_1$. Then Kw_1 and Kw_2 are bisimilar, but it does not hold that $Kw_1 \equiv Kw_2$ □