Questions of identity[1]

Jelle Gerbrandy

Introduction

This article is about the interpretation of quantifiers in epistemic contexts, and the closely related notion of what it means for two objects in different possible worlds to be identical.

I will give arguments for several claims. The first claim is that the definition of identity between objects across epistemically possible worlds as being given by the identity of those objects themselves leads to empirical as well as conceptual problems. Recent examples of such a view on identity are the theory of questions of Groenendijk and Stokhof (1997) and the semantics for quantified modal logic of Groenendijk et al. (1996). Secondly, I will argue that these problems can be circumvented by taking trans-world identity to be a notion that is external to the possible worlds themselves (by some sort of counterpart theory, or by a set of individual concepts). This raises the old question what exactly this notion of identity is. I will argue by way of examples that there is no single notion of trans-world identity, but that the way we speak about objects depends on contextual factors.

In this paper, I will concentrate mostly on the analysis of questions and of 'knowing who,' but the points raised here are also relevant to issues related to *de re* knowledge and belief and the interaction of epistemic 'might' and quantifiers. I will address these and other issues in a more extended version of this paper.

Possibilities and information states.

Possible world semantics has turned out to be a useful tool for the analysis of intensional expressions, e.g. of sentences about necessity, knowledge or belief, and of questions. In this article, I will concentrate on the use of possible words semantics to interpret epistemic modal operators, and I will focus in particular on the semantics of 'knowing who' and of questions.

To interpret a language, possible worlds need to have a certain minimal structure: we need to know in each possible world what the predicates, constants and other lexical items of the language refer to. For the purposes of this paper, in which the object language is that of predicate logic, this means that possible worlds should at least contain as much structure as a first-order predicate logical model:

Definition 1 (possibilities and states)

A (epistemic) possibility w for a language $\mathcal{L}$ is a (first-order) model (D_w, I_w), where D_w is a set (called the domain of w), and I_w is an interpretation function that assigns to each non-logical constant of $\mathcal{L}$ an interpretation.

An epistemic state is a set of possibilities. □

[1]ILLC/Department of Philosophy, Amsterdam, gerbrand@illc.uva.nl. I would like to thank Maria Aloni, Henk Zeevat, Jeroen Groenendijk, Paul Harrenstein and Paul Dekker for discussion and comments. None of them wholly agree with the contents of this paper. Many of the ideas in this paper can be found in the literature on the subject and I have tried to give credit wherever possible, but I am not sure whether I have succeeded.

The information (knowledge, beliefs) of an agent are modeled by an information state. An information state contains all and only those possibilities that are compatible with what that agent knows; it consists of those models that, given the information of the agent in question, may picture the world correctly. If σ is the information state of some agent a that represents the *knowledge* of a, then we can assume that there is a possibility $w \in \sigma$ that represents 'the real world.'

The analysis of questions and of 'knowing who' that I will concentrate on is that of Groenendijk and Stokhof (1984, 1997). In their analysis, the meaning of a question is given by the set of its complete answers; answers are identified with propositions, i.e. sets of possibilities; and a complete answer to the question 'Who VP?' is represented by the set of all worlds in which the extension of VP is the same. In other words, a complete answer to the question 'Who VP?' is given by a specification of the set of objects that have the property expressed by the VP. A sentence of the form 'a knows who VP' is true just in case a knows what the complete true answer to the question 'Who VP?' is, i.e. a knows who VP just in case the set of objects with the property expressed by VP in each of a's epistemic alternatives is the same as the set of those objects that have the property VP in 'the real world.'

This analysis presupposes a notion of identity between objects in different possibilities: to know if the extension of VP is 'the same' in two possibilities, we need to know what it means for two objects in different possibilities to be identical.

Identity of objects.

There are at least two ways in which identity across possibilities can be modeled. The most simple view, which is exemplified by Kripke's work on necessity and reference, simply says that two objects are identical iff they are the same object. This is the notion of identity that Groenendijk and Stokhof use in their work.

Let us use the notation $\sigma \models \phi$ for the statement that ϕ is accepted (believed, known) in information state σ, and write $\sigma \models ?xPx$ when it is known in σ who P is. Rewriting definitions using this format 'Knowing who is P' is defined as knowing exactly which objects have the property P:

- $\sigma \models ?xPx$ iff for all w and v in σ, $I_w(P) = I_v(P)$.

So, $?xPx$ is accepted in a state σ just in case the extension of the predicate P is the same set of objects in each possibility in σ. If we assume that σ is a knowledge state, then a (representation of) the real world will be among the possibilities in σ. This assumption garantees that if A is the set of objects that have the property P in 'the real world', then it is known in σ who have the property P just in case in each of the possibilities in a's information state, it holds that A is indeed the set of objects that have the property P. For example, if Superman is the object d, then Lois Lane knows who Superman is just in case d is Superman in each of the possibilities in her epistemic state.

Although taking identity across possibilities as 'given' with the objects that are in the domain of those possibilities is appealing in its simplicity, this view is not without its problems.

Consider for example an information state σ that contains all and only isomorphic copies of one single model in which there is exactly one person who is the murderer. This is a state in which one knows all the properties of the murderer (all information necessary

to find him, arrest him and convict him):[1] typically a case in which one knows who the murderer is. The definitions above predict exactly the opposite: σ does not contain any information about who the murderer is (it could be any object in the domain).

Another example of a situation in which the definitions give counterintuitive predictions is a state σ that contains all models in which the object d is the unique object that is the murderer. This is a state in which one knows none of the properties of the murderer (except that he or she is the murderer): as far as the information in σ goes, the murderer could be a man, woman or child, he or she could be called Frederick Pluto Bulsara or Madonna Louise Veronica Ciccone. This seems to be a typical case of not knowing who the murderer is. Again, the definitions above predict the opposite: this state is one in which one knows who the murderer is, and in which there is just a single person who might be the murderer (namely d).

These examples show that the definition of 'knowing who VP' as 'knowing which objects have the property VP' is not correct as it stands. For the more philosophically oriented among us, this is not very surprising. The objects in the domains of the possibilities are completely independent of the properties they might have in those possibilities: for any object d and any property, there is a model in which d has that property.[2] The notion of an object that is independent of the properties it might have is reminiscent of Kant's notion of a *Ding an sich*, and philosophers from Kant to Husserl to Hintikka have argued that knowledge about the objects that 'underlie our experience' is unattainable for mere humans.[3] The definition says that it is exactly this kind of knowledge that is needed to know who someone is.

Identity between objects.

Elegant though it may be, the object-oriented view on 'knowing who' runs into both empirical and conceptual problems. At the same time, the analysis of 'knowing who VP' as knowing which objects have the property VP is intuitively appealing. Moreover, Groenendijk and Stokhof's analysis of 'knowing who' is only a small part of what is probably the most sophisticated theory of the semantics of questions that exists today, and it would be unfortunate if we had to discard this whole theory on the basis of examples such as the above.

One way to avoid the problems mentioned and at the same time keep the spirit of Groenendijk and Stokhof's analysis intact, is to re-interpret identity across possibilities. Instead of modeling identity across possibilities by taking it to be given by the objects that constitute the domains of the possibilities, we assume that some equivalence relation R over the objects in different possibilities is given, and say that an object d in one possibility is identical to an object d' in another possibility just in case dRd'.[4] I will refer to such an

[1] Presuming that the models are sufficiently rich to contain such information.

[2] To avoid confusion: Kripke's arguments do not apply here. His arguments apply to counterfactual situations and 'metaphysical' necessity, where it does seem to make sense to say that the domains of the possible worlds consists of the objects that exist in the 'real world'. Such a view will not work in epistemic contexts: consider Quine's example of Ralph, who has seen the Reverend Ortcutt on different occasions (on the beach and at a party), and believes that the man he saw on the beach is different from the man he saw at the party. There is no way to model this situation, and at the same time represent the man on the beach and the man at the party (which happen to be the same person) by the same object in the information state that models Ralph's beliefs.

[3] For a more knowledgeable discussion, see, for example, the collection Dreyfus (1982) and in particular Smith (1983).

[4] To be more precise, R should be a relation between pairs (w, d), where w is a possibility and d is an

equivalence relation as an 'individuation scheme.'

Relative to an individuation scheme R, we can interpret 'knowing who' in the following way:[5]

- $\sigma \models_R ?xPx$ iff for all $w, v \in \sigma$, $I_w(P)$ is R-identical to $I_v(P)$.

The earlier definition is a special case of this one, where R identifies objects just in case they are the same object, i.e. R is 'real identity.' But, as we have seen, such an identity relation does not give the right predictions about what is known in which state. The first example that I gave above seems to show that a minimal condition on an individuation scheme is that it should identify two objects in different possibilities if they have exactly the same properties. Relative to such an individuation scheme, it holds that if σ is a state that contains any number of isomorphic copies of one single model, it is always known who VP is, which is exactly what we would expect. After all, in such a state, all the properties of the things that VP are known.

Switching Schemes.

There has been a lot of debate on the question what the right notion of trans-world identity is. I believe that with respect to epistemic possibilities, this question is not a very good one to ask. Asking what *the* identity relation is presupposes that there is only one such a relation, one single way to identify objects. I will argue that this assumption leads to wrong predictions about the semantics of wh-complements and questions. There are other examples that I cannot discuss here for lack of space – examples concerning the interaction of epistemic 'might' with quantifiers, *de re* readings of epistemic modals, and the analysis of anaphora in dynamic semantics (cf. Aloni in this volume). These examples seem to provide further evidence that there is not one single way of identifying objects across possibilities in epistemic states.

This leads to the conclusion that the quantification into modal contexts presupposes a given 'mode of individuation', which may vary in different contexts.[6]

It has often been observed (for example, in Ginzburg (1996), Boër and Lycan (1986), Carlson (1983)) that both the correctness of an answer to a question and the truth of sentences of the form 'a knows who VP' are highly context dependent: what constitutes a good answer to a question depends on the goals and interests of the questioner; and whether someone can be said to know who VP depends on contextual factors in a similar way. For example, depending on the context, all of the following sentences may express a complete answer to the question 'Who is Bill Clinton?', as well as the information that a needs so that the sentence 'a knows who Bill Clinton is' is true.

---

object in the domain of w. If we assume that any two possibilities have disjoint domains, we can let R be a relation between objects *simpliciter*. I will do this in the following.

[5]This definition, although it looks quite different, is in fact very close to Hintikka's analysis of 'knowing who.' Hintikka's analysis only applies to sentences of the form 'a knows who NP is', where NP is a definite description (i.e. an expression that identifies a unique object in each possibility, such as a proper name or an NP of the form 'the N'). Hintikka uses individual concepts as his method of identifying objects across possible worlds, but given Hintikka's assumptions on such a set, this is just a special case of our notion of an identification scheme. The only essential difference between this analysis and that of Hintikka is that he assumes that 'knowing who NP is' presupposes the existence of an object satisfying the description NP.

[6]This conclusion is not new. For example, Hintikka (1969), Kaplan (1979) and Kraut (1983) argue (or, at least, claim) that quantifiers range over a set of individual concepts that varies depending on the context.

(1) Bill Clinton is the president of the United States.

(2) That is Bill Clinton (pointing to him).

(3) Bill Clinton was born on august 19, 1946.

The first sentence would be a complete answer to the question who Bill Clinton is in the context of, for example, a highschool exam. If a student has the information expressed by (1), his teacher would have enough reason to say that this student knows who Bill Clinton is.[7] This does not mean, however, that the student would be able to recognize Clinton. At a fund-raising party at which Bill Clinton is present, the question who Bill Clinton is not satisfactory answered with (1); pointing him out to the questioner would be a better answer. In situations like these, knowing who Bill Clinton is, is knowing which of the people present is Bill Clinton. In a completely different situation, say at a conference of astrologers, his job or looks may be considered completely irrelevant. Instead, his date of birth would be considered much more important. In such a situation, sentence (3) would be considered a good answer to the question who Bill Clinton is, and many an astrologer would agree that he does not know who Bill Clinton is unless he knows his date of birth.

In general, it seems that when we ascribe knowledge to people by using a circumlocution of the form 'a knows who ... is,' there will be different kind of information that support such a judgment. For some purposes, knowing someone's name will be enough evidence for knowing who that person is (and stating his name a good answer to the question who he is), for other purposes, knowing where he is and what he looks like may be the kind of information needed, in other cases, it may be important what his birthdate is, etcetera.

Fortunately, we already have the tools available to model these different sorts of knowledge ascription. The only assumption we need to give up is that there is one fixed individuation scheme relative to which the quantifier $?x$ is interpreted. Relative to an individuation scheme R that identifies people by their position in society (R is a relation that holds between two objects in different worlds just in case they have the same 'position in society'; in particular, if d is the president of the United States in some possibility, and d' is the president in another possibility, then dRd'), it holds that you know who Bill Clinton is if you know that he is the president, i.e.[8]

$$\sigma \models_R ?x(x = \mathsf{Bill\ Clinton}) \text{ iff } \sigma \models (1)$$

Relative to an individuation scheme R that identifies people just in case they are born on the same day, it holds that

$$\sigma \models_R ?x(x = \mathsf{Bill\ Clinton}) \text{ iff } \sigma \models (3)$$

As an extra argument for the claim that we really cannot make do with a single individuation scheme, imagine a small room in which two people sit: a butler and a gardener. One of them is called John, the other Mary, but you do not know whether John is the butler or the gardener. The butler has committed a horrible crime; the gardener had nothing to do with it. So, you know who has done it: the butler.

[7] This example excludes the view that (1) is only a *partial* answer, because in that case, the student would not have a complete answer to the question, and thus, in the present analysis, he does not know who Bill Clinton is.

[8] Given that (1) is in fact true, and σ is a state that models the knowledge of some person, i.e. that the information reprsented in σ is in true.

On the other hand, you do not know if John is the butler or not, and so you don't know whether he has done it. The same holds for Mary. So, it seems you do not know who has done it after all: it could be either John or Mary.

We seem to have a situation here in which, in one respect, you know who has done it (namely the butler) and from another viewpoint, you don't (it could be either John or Mary). This situation can be accounted for by switching identification schemes: relative to an R in which the butler and the gardener are identified, it holds that you accept $?x(\mathsf{done\ it}(x))$, relative to a scheme in which John and Mary are identified, it holds that you do not accept that $?x(\mathsf{done\ it}(x))$.

Conclusion

Identity between objects in possibilities should not be seen as given by the objects themselves: such a view is both conceptually and empirically questionable. Instead, identity across possibilities should be taken as a non-primitive notion. This raises the question what this identity relation is. I have argued, by giving examples from wh-constructions, that the notion of trans-world identity should be seen as a context dependent notion.

I will end this paper with some general questions that I have not addressed. One issue is whether an individuation scheme should be modeled as an equivalence relation (as I have done here), or as a set of individual concepts, or even as a set of salient properties. A second question is what (if any) the general constraints on individuation schemes are (I have mentioned that such schemes should always identify objects in case they have exactly the same properties. But one could also argue that any object should be identified with at most one other object in any possibility). Given that individuation schemes are context-dependent, there is the question of where this parameter should be put. It could be a variable in the logical form of a sentence (and if so, is it introduced by the quantifiers or by the modal operators?) or a parameter in the context itself (which implies that all quantifiers and wh-terms are interpreted relative to the same individuation scheme).

References

Maria Aloni (1997) *Quantification in dynamic semantics*. This volume.

Steven E. Boër and William G. Lycan (1986) *Knowing who*. The MIT Press.

Lauri Carlson (1983) *Dialogue games*. Reidel, Dordrecht.

Hubert L. Dreyfus, ed. (1982) *Husserl, intentionality and cognitive science*. The MIT Press.

Jonathan Ginzburg (1996) *Interrogatives*. In Shalom Lappin, ed., *Handbook of contemporary semantic theory*, pages 1385–422. Blackwell, Oxford.

Jeroen Groenendijk and Martin Stokhof (1984) *Studies on the semantics of questions and the pragmatics of answers*. Jurriaans BV, Amsterdam.

Jeroen Groenendijk and Martin Stokhof (1997) *Questions*. In Johan van Benthem and Alice ter Meulen, eds., *Handbook of logic and language*, pages 1055–1124. Elsevier.

Jeroen Groenendijk, Martin Stokhof and Frank Veltman (1996) *Coreference and modality*. In Shalom Lappin, ed., *Handbook of contemporary semantic theory*, pages 179–213. Blackwell, Oxford.

Jaakko Hintikka (1969) *Semantics for propositional attitudes*. In J. W. Davis, D. J. Hockney and W. K.Wilson, eds., *Philosophical Logic*, pages 21–45.

David Kaplan (1979) *Transworld Heir Lines*. In Michael J. Loux, ed., *The possible and the actual*, pages 88–109. Cornell University Press, Ithaka and London.

Robert Kraut (1983) *There are no de dicto attitudes*. Synthese, 54:275–294.

David Woodruff Smith (1983) *Kantifying in*. Synthese, 54:261–273.