

Changing the Common Ground

Jelle Gerbrandy*

1 Introduction

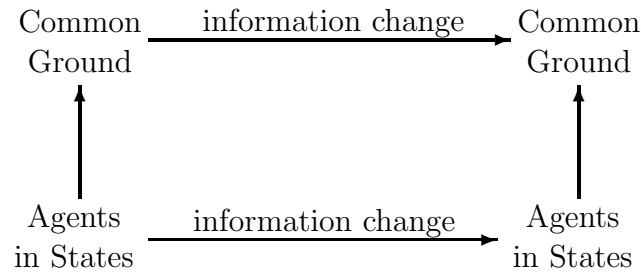
Often, in formal models of dialogue, the notion of a ‘common ground’ plays an important role: a body of public information which changes during the course of a conversation and is used to keep track of what has happened in the conversation, delimits the range of possible further utterances and influences the interpretation of those utterances. The reader need only skim many of the other papers in this volume to see that the idea is very much alive.

How exactly this common ground should be characterized is not agreed upon. To give some early examples: Lewis (1979) uses the metaphor of a ‘conversational scoreboard’ on which the relevant information about the ‘moves’ in the dialogue game are noted. Stalnaker (1978) speaks about ‘presuppositions’ as that ‘what is taken by the speaker to be the common ground of the participants in the conversation, what is treated as their common knowledge or mutual knowledge.’ Hamblin (1971) uses the metaphor of a ‘commitment slate.’ Yet other writers identify the common ground with ‘that what is mutually believed.’ Clark and Marshall (1981), for example, argue that it is necessary for a successful use of a definite description that it should be mutual knowledge what the definite refers to.

One can model changes in the common ground in one of two ways. In the first kind of model, a representation of the common ground is taken as primary, and the effect of utterances in a dialogue is modeled by showing how the utterance affects the common ground. The second approach starts out with the belief states of the dialogue participants considered separately. The effect of an utterance can then be modeled by the effect it has on the belief states of each of the separate agents. Since the contents of the common ground depend on the belief states of the participants, the effects on the states of the participants will also imply a change in the common ground; in that sense such models subsume a theory of how the common ground is changed.

Consider the following diagram:

*. Department of Philosophy, University of Amsterdam. Parts of this paper were written at the CSLI of Stanford University; the Netherlands Organization for Scientific Research (NWO) is gratefully thanked for sponsoring my visit there. I would also like to thank Henk Zeevat, Paul Dekker and Marco Hollenberg for their useful comments.



At the bottom corners of the picture, there are agents which have certain information: each of them is in a certain information state. The arrow labeled ‘information change’ on the lower end of the picture represents the change in these states that is the effect of an utterance by one of the agents: this describes the second kind of model of the information change in dialogue. The first kind of model, where the effect of an utterance seen as a direct change of the common ground, is represented by the top part of the picture.

Of course, a piece of information can only play a role in dialogue if at least one of the participants in a dialogue is aware of that information. That makes it natural to assume that the common ground is determined by the information states of the agents: the vertical arrows represent some way of extracting the common ground from the agents’ states.

The leading question in this paper is whether first taking the common ground in a certain model w of the states of the agents, and then changing the common ground according to some specified way, gives us the same result as first changing the model w , and then seeing what the common ground is in the result. Or, in other words: does the diagram above commute?

To make sense of this question, we need to be more precise about the filling in of the parameters: the kind of representation we use for states of agents and for the common ground, how these two are related, and what information change consists of. Of course, the answer to our question depends for a great deal on how we choose to fill in these parameters.

In the next two sections I will study the diagram using a classical possible worlds framework. In the first of these two sections (which is section 2) I discuss the relation between mutual belief and the common ground and discuss several definitions of mutual belief.

In the section after that, I show how a given function that describes information change can be ‘lifted’ to an operator that models ‘mutual information change.’ With this formal machinery, we have the tools to instantiate the informal picture above. We will look at operations of belief change such as expansion, contraction and revision. The main conclusions are negative: the diagram generally does not commute, not even for a relatively simple notion of belief change such as expansion. When considering weaker properties than commutativity, expansion fares fairly well, but I will argue that revision and contraction (and any kind of belief change operation that has certain minimal properties in common with these two) have properties that are incompatible with the assumption that the diagram behaves in a reasonable

way.

In section 4, I will briefly study the same questions in a more general framework. The results will be similar to those of the preceding sections. The main purpose of this section is to show that the negative results hold for any kind of model that has certain minimal properties in common with the possible worlds approach.

The paper ends with a section entitled ‘conclusions.’

2 Mutual belief and the common ground.

2.1 Possibilities

The standard way of modeling information of several agents in a possible worlds framework is by using Kripke models. Here, the information of an agent a in a world w is modeled by the set of worlds v that are accessible from w : those are the worlds that the agent a considers possible in w . We will adopt the same approach towards information, but use a different kind of model to implement it. The reason for not using Kripke semantics is that the possibilities defined below make it much more easy to define notions of information change.

Definition 2.1 Possibilities.

Let \mathcal{A} be a finite set of agents and \mathcal{P} a finite set of atomic sentences.

- Any function w on $\mathcal{A} \cup \mathcal{P}$ that assigns to each atomic sentence $p \in \mathcal{P}$ a truth value $w(p) \in \{0, 1\}$ and to each agent $a \in \mathcal{A}$ an information state $w(a)$ is a possibility.
- Any set of possibilities is an information state.

The intuition behind this definition is the following. We want a model of the world: atomic sentences are either true or false, and agents have certain information. Possibilities as defined here give us exactly that: to each atomic sentence, they assign a truth-value, and to each agent, they assign an object called an ‘information state.’ Secondly, we model the state of an agent in a traditional way: by the set of models of the world that are consistent with that agent’s information. So, an information state will be modeled as a set of possibilities.

Assuming that our set of atomic sentences contains just a single sentence *It rains*, and we have only one agent, called Francisco, an example of a possibility is a function w such that $w(\textit{It rains}) = 1$ and $w(\textit{Francisco}) = \emptyset$.¹ So, in this rather depressing possibility it is raining, and Francisco’s beliefs are not consistent: there is no possibility compatible with his beliefs.

Unfortunately, simply using standard set-theory as our background theory will not give us enough different possibilities to model everything we want to model. For example, there is no object in the ZFC set-theoretical universe that corresponds with a possibility w in which w itself is consistent with Francisco’s information.

1. In the following, we will often leave the precise structure of \mathcal{A} and \mathcal{P} implicit where it is not likely to lead to confusion.

This is why we use non-well-founded set theory, as it is developed in Aczel (1988). In this theory, the axiom of foundation is left out of the ZFC -theory, leaving us with an axiom system standardly denoted by ZFC^- . Instead, Aczel adds a new axiom, the axiom of anti-foundation, which for our purposes can be expressed as follows: “For each world in a Kripke model there is a unique possibility that is bisimilar to that world.”²

This axiom guarantees us that we have enough possibilities to do epistemic logic: for every bisimulation class of Kripke models, there exists a corresponding possibility. If we want to use these models as a semantics for a modal language –and we do– this means that for every two Kripke models that are distinguishable in infinitary modal logic (modal logic with arbitrary conjunction added, a very strong language), there are corresponding possibilities that can be similarly distinguished. A fortiori, this holds for the weaker language we will use in our paper, which is a finitary modal language with an operator added for common belief.

Possibilities are studied in more detail in Gerbrandy (1997) and Gerbrandy and Groeneveld (1997). Possibilities are very similar to the states in the model for transition systems developed by Aczel (1988). Finally, the work in Barwise and Moss (1996) on using modal logic to describe non-well-founded sets is related to the present approach to epistemic logic.

Truth of classical modal sentences in a possibility can be defined in a way analogous to the definition of truth for Kripke models.

Definition 2.2 Let w be a possibility.

$$\begin{aligned}
w \models p & \text{ iff } w(p) = 1 \\
w \models \phi \wedge \psi & \text{ iff } w \models \phi \text{ and } w \models \psi \\
w \models \neg\phi & \text{ iff } w \not\models \phi \\
w \models \Box_a\phi & \text{ iff for all } v \in w(a) : v \models \phi
\end{aligned}$$

There are two kinds of possibilities that we will be particularly interested in: the possibilities in which agents have introspective information (where they know exactly what information state they are in), and the possibilities in which their information is not only introspective, but also true. Introspection holds in those possibilities w such that $w(a)$, the state of a in w , contains only possibilities v in which a gets assigned the information state she is actually in, i.e. such that $v(a) = w(a)$. Moreover, this property is also assumed to hold for each $v \in w(a)$. Truthfulness corresponds

2. The idea is that a Kripke model and a possibility are bisimilar just in case that for all practical purposes they have the same structure. Formally, a relation Z is a bisimulation iff its domain consists of worlds in Kripke models (i.e. pairs (K, x) where K is a Kripke model $(W, (R_a)_{a \in \mathcal{A}}, V)$, and x a world in K) and its range consists of possibilities. Moreover, if $(K, x)Zw$, then for each $p \in \mathcal{P}$, $V(x)(p) = 1$ iff $w(p) = 1$, and for each y such that $xR_a y$, there is a $v \in w(a)$ such that $(K, y)Zv$, and for each $v \in w(a)$, there is a y in K such that $xR_a y$ and $(K, y)Zv$. We say that (K, x) and w are bisimilar iff there is a bisimulation Z such that $(K, x)Zw$.

to the property that for each a , $w \in w(a)$: a considers the ‘real world’ possible. We will call possibilities that are both introspective and truthful ‘S5-possibilities.’³

2.2 Mutual Belief

A modest part of the discourse on logic is concerned with the relation between mutual belief and beliefs of separate agents. By definition, a sentence ϕ is mutually believed iff each participant believes that ϕ to be the case, each participant believes that all other participants believe ϕ to be the case, etcetera, *ad infinitum*.

As I remarked in the introduction, one way of seeing the common ground is as seeing it as that which is mutually believed. This is the idea we will adopt in this section, so some comparison between this view and other views on the common ground are in order. First, I will argue that mutual belief can be seen as a stronger notion than that of the common ground – everything in the common ground must be (or can be taken to be) mutually believed – but that it depends on the view one adopts towards the common ground whether the converse holds, i.e. whether all mutual beliefs are in the common ground.

When seeing the common ground as the ‘conversational scoreboard,’ or as a ‘commitment slate,’ one can argue that whatever is on that scoreboard is independent of the beliefs of the separate agents. Lies, for example, will be added to the scoreboard in the same way as honestly believed utterances are (since the liar is committed to them in the same way as he is committed to honest utterances). Clearly then, there may be sentences on the scoreboard that are not believed, let alone mutually believed.

I think this point is valid, but it does not necessarily imply that the concept of mutual belief is irrelevant to the concept of the common ground when it is seen as a commitment slate. If we want a useful model of the common ground, it should also apply to conversations in which the participants try not to mislead. Given our little problem that is concerned with belief change resulting from utterances in dialogue, we can restrict the study of it to changes in the common ground that arise from honest utterances alone. In other words, we may assume that the participants *really* follow Grice’s maxim of quality, and are really cooperative. Within this restriction on the kind of dialogue studied, it will never happen that sentences appear on the commitment slate or the conversational scoreboard that the participants are not committed to.

In any case, we will assume in the rest of this paper that the information in the common ground is in fact believed to be true by each of the participants.

A property of the common ground that, as far as I know, is shared in each model of the common ground that has anything to say about higher-order information (beliefs about beliefs and such) is that the common ground is in some sense ‘publicly accessible’: each of the participants knows what information is in the common ground. Given that whatever is in the common ground is believed by everybody,

3. More formally, we take the class of introspective possibilities to be the largest class \mathcal{I} such that for each $w \in \mathcal{I}$ and $v \in w(a)$: $v(a) = w(a)$ and $v \in \mathcal{I}$. The class of S5-possibilities is the largest class included in \mathcal{I} such that for each possibility w in that class, $w \in w(a)$ for each a .

the public accessibility of the common ground implies that everybody believes that everybody believes the information in the common ground. We can repeat this argument to get arbitrary iterations of ‘everybody believes ...’

So, under the assumption that the common ground is mutually accessible, and that all information contained in it is believed, it follows that all information in the common ground is mutually believed: there are at least as many things mutually believed as there are in the common ground.

The answer to the question whether all mutual beliefs are in the common ground depends on the view one takes of that common ground. If the common ground is seen as a kind of conversational scoreboard, or as only containing the information that the dialogue participants are committed to by utterances actually made, the answer will be ‘no’: surely many facts that are not explicitly stated in the dialogue can be taken to be mutual beliefs, such as the fact that the participants speak a certain language, that the speaker has an enormously big red nose, that there is a vase of flowers on the table between them, etcetera.

Other theories include all such information in the common ground. Clark and Marshall (1981), for example, argue that for a correct interpretation of definites such as the vase on this table,’ facts such as that there is a vase on the table between the participants should be mutual belief. If one defines the common ground as ‘all information that should be accessible for the dialogue participants so that their dialogue works,’ the mutual belief that there is a vase on the table should be in the common ground as well.

2.3 Formal notions of mutual belief

Apart from the definition above, there have been many other definitions and characterizations of mutual belief. Jon Barwise (1989), in an article in one of the books on situation theory, compares three characterizations of the concept of common knowledge⁴, and concludes that in situation theory, all three can be distinguished. In our format, these three notions can be, roughly, represented as follows:

The iterated approach is just a straightforward rewriting of the informal definition above.⁵

$$w \models C^{iter} \phi \quad \text{iff} \quad w \models \Box_{a_1} \dots \Box_{a_n} \phi$$

for each sequence $a_1 \dots a_n$ of agents.

The fixed point approach is based on the intuition that the mutual belief of a formula ϕ is a property (that is, a set) P of possibilities that holds of a possibility w

4. The term common knowledge and mutual belief are used interchangeably in the literature.

5. The reason to use a ‘ C ’ to denote mutual belief is because the operator we will define is essentially the $C_{\mathcal{A}}$ -operator of Fagin, Halpern, Moses, Vardi (1995). The reason they use this symbol is because they use the term ‘common knowledge’ instead of ‘mutual belief.’ Fagin et al. also study the logic of this operator, and provide a completeness theorem.

just in case it holds in w that a knows that ϕ is the case, and each possibility that is in the information state of a also has the property P .

If we were to denote this property by ' $\models C^{fix}\phi$ ', then this condition is formally expressed by the following equivalence:

$$w \models C^{fix}\phi \quad \text{iff} \quad \forall a \forall v \in w(a) : v \models \phi \text{ and } v \models C^{fix}\phi$$

This does not uniquely identify a property though. For reasons explained in Barwise's article, we let $\models C^{fix}\phi$ be the *largest* property that satisfies the equation above.⁶

According to the 'shared situation' approach, a sentence ϕ is mutually believed just in case there is a situation σ in which (1) ϕ holds, and (2) the situation σ implies, or gives reason enough to assume, that each of the agents knows (or believes) that the situation σ in fact obtains, and (3) each of the agents does believe that σ obtains. This kind of definition has been proposed by Lewis (1969), Schiffer (1972) and Clark and Marshall (1981).

A typical example is a situation of sitting around a table on which stands a vase of flowers: such a situation would give each of the agents enough reason to assume that the fact that there is a vase of flowers on the table is mutual belief. Another typical example is the utterance of a sentence followed by an acknowledgment of the hearer: this situation would be reason enough to assume that the fact that the utterance is made is now mutually believed.⁷

If we identify a situation with a set of possibilities –'all maximal extensions' of that situation, if one wants, or 'all possibilities in which that situation obtains'– we can transpose Barwise's analysis in our framework and define:

$$w \models C^{share}\phi \quad \text{iff} \quad \text{there is a set of possibilities } \sigma \text{ such that:}$$

- (1) $v \in \sigma \Rightarrow v \models \phi$
- (2) $v \in \sigma \Rightarrow v(a) \subseteq \sigma$ for each a
- (3) $w(a) \subseteq \sigma$ for each a

If we compare the three definitions, it turns out that all three are equivalent.⁸

Fact 2.3 For each possibility w :

$$w \models C^{iter}\phi \Leftrightarrow w \models C^{fix}\phi \Leftrightarrow w \models C^{share}\phi$$

6. Of course, the fact that such a largest set exists needs proof. We will omit it, just as we omit the motivation for choosing the largest property instead of, e.g. the smallest.

7. Note that such knowledge is not meant to be infallible in any way. The negative results of Halpern and Moses (1991) in the context of message-passing systems shows that if one reads the 'knowledge' in 'common knowledge' in the strong sense as implying truth, it can never happen that any non-trivial information becomes common knowledge; at least not under the quite reasonable assumptions that message passing takes time, and is never completely reliable.

8. Fagin et al. (1995) contains a proof of the equivalence of the iterated and the fixed point accounts

proof:

[From the iterated account to the fixed points] Assume $w \models C^{iter}\phi$. It is not hard to see that for each a and $v \in w(a)$: $v \models \phi$ and $v \models C^{iter}\phi$. Since we have defined $\models C^{fix}\phi$ as the largest set with exactly this property, it follows that $w \models C^{fix}\phi$.

[From fixed points to shared situations] Consider the set $\sigma = \{v \mid v \models \phi \text{ and } v \models C^{fix}\phi\}$. Then $v \in \sigma$ implies that $v(a) \subseteq \sigma$ by definition of C^{fix} , and clearly, $v \models \phi$ for each $v \in \sigma$. Assume $w \models C^{fix}\phi$. Then clearly, $w(a) \subseteq \sigma$, so $w \models C^{share}\phi$.

[From shared situations to the iterated approach] Assume $w \models C^{share}\phi$. We need to show that $w \models \Box_{a_1} \dots \Box_{a_n}\phi$ for each sequence $a_1 \dots a_n$ of agents. That this holds is easily proven by an induction on n . \square

So, in a classical possible worlds framework (the definitions can be easily reformulated to apply to Kripke models, and the equivalence results will continue to hold) the three different characterizations of common knowledge collapse.

I am not sure whether this result should be seen as a positive or a negative one. In contrast to the analysis above, on Barwise's analysis of the three definitions in situation theory, all three definitions turn out to give different situation-theoretic notions of common knowledge. But the differences between the three kinds of definitions come up only at the transfinite level; restricting Barwise's analysis to models in which agents believe only finitely logically independent facts (a natural assumption to make on any agent), the three notions collapse also in situation theory. This makes it, at least to me, very hard to see how the distinctions between the three characterizations correspond to pre-situation-theoretic distinctions. To put it bluntly: it seems that situation-theory is making trouble where there was no trouble to be found.

Whatever the conclusion is, the fact that the three different characterizations come down to the same semantical characterization in our framework makes the choice between the definitions meaningless: we can take either one.

We will represent the common ground in a possibility w by an information state that contains exactly the information that is mutual belief. This information state contains all and only possibilities v for which it holds that one of the agents considers v possible (in w), or that one of the agents considers it possible that one of the agents considers v possible, etcetera. We let the notation $C(w)$ stand for this set of possibilities.

Definition 2.4 The common ground between the agents in a possibility w , $C(w)$, is the set of all possibilities v such that there is a sequence of possibilities and agents $w_0, a_0, w_1 \dots a_n, w_{n+1}$ such that $w_0 = w$, $w_{i+1} \in w_i(a_i)$ for each $i \leq n$, and $w_{n+1} = v$.

It turns out that this characterization is consistent with what we said previously: a sentence is accepted in the state $C(w)$ ⁹ exactly when it is common knowledge in w :

9. We say that a sentence ϕ is accepted in a state σ just in case ϕ is true in each possibility in σ

Fact 2.5 $C(w) \models \phi$ iff $w \models C\phi$.

Before going back to our diagram, I would like to make some remarks about $C(w)$.

First, note that in $C(w)$, we have lost information about w : in general, there are w and v different from each other such that $C(w) = C(v)$. This also holds within the class of introspective possibilities. In particular, we cannot see from $C(w)$ alone where its possibilities ‘come from’: there is no way of telling from the structure of $C(w)$ whether some $v \in C(w)$ is there because some a thought it possible, or because some a thought some b considered it possible. We will return to this observation later.

Another remark concerns the complexity of $C(w)$: it contains possibilities in which information of agents is represented, the information they have about each other’s information, etcetera. Often, in models of dialogue, the common ground is not taken to be that complex at all: sometimes it contains only world-information (information that can be expressed by non-modal sentences), and in general, higher-order information (information about information) is only represented up to some very restricted finite depth. Also this point will be taken up in the next section, where we really start proving things about our diagram.

We end this section by noting some formal properties of common grounds, and comparing the common grounds introduced here with those of Zeevat (this volume).

Consider the following operation on sets of possibilities that collects all worlds considered possible in one of the possibilities of that set. We call the operation E .¹⁰

Definition 2.6 $E(\sigma) = \bigcup \{w(a) \mid w \in \sigma, a \in \mathcal{A}\}$.

If σ is a singleton set $\{w\}$, we will write $E(w)$ for $E(\{w\})$.

Fact 2.7 $C(w)$ is the smallest set σ containing $E(w)$ such that $E(\sigma) \subseteq \sigma$.

We end this section by comparing our common grounds to those of Zeevat (this volume). In his article, an information state σ is a common ground just in case it has the following property:

$$\sigma = E(\sigma)$$

Let’s call this property the ‘Zeevat property.’ It turns out that many, but not all, possibilities have a common ground with the Zeevat property:

Fact 2.8 It holds that $C(w)$ has the Zeevat property iff $E(w) \subseteq E(C(w))$

That $C(w) = E(C(w))$ is not a very strong property of common grounds. For example, it is implied by introspection:

Fact 2.9 If w is introspective, then $E(w) \subseteq E(C(w))$.

10. The ‘E’ is from ‘everyone.’ The reason for this is that just as $C(w) \models \phi$ iff $w \models C\phi$, so it holds that $E(\{w\}) \models \phi$ iff $w \models \Box_a \phi$ for each a , i.e. just in case ‘everyone knows ϕ .’

Which means that each $C(w)$ belonging to an introspective possibility has the Zeevat property.

3 Changing the common ground

Suppose we are given an operator F over information states that expresses some sort of information change. What I have in mind is an operator such as ‘expand with p ’ or ‘revise with ϕ ,’ (Alchourrón, Gärdenfors, Makinson, 1985) or the update functions from update semantics (Veltman, 1996). The first question that I will try to answer here is what it means for a group of agents to apply such a function together; the second is how such functions behave in the diagram.

3.1 Multi-agent Expansion

We will start our discussion with one special and relatively simple case: that of expansion. Expansion with a certain sentence means simply adding the information contained in that sentence to the information you already have: in our case, that means discarding all possibilities in which the sentence is false:¹¹

Definition 3.1 If σ is an information state, then $\sigma + \phi = \{v \in \sigma \mid w \models \phi\}$.

This definition may be familiar from update semantics, and if one takes classical logic as the ‘base logic’ in the work on belief revision (e.g. Alchourrón et al., 1985), this is essentially the definition of expansion used there. To keep things from getting too complicated, we will restrict our language to non-modal sentences in the following.

We are looking for a definition of ‘mutual update’ on the level of possibilities that corresponds with a change in the common ground. Consider the following definition, in which the notation $+^* \phi$ stands for a mutual expansion with ϕ :

Definition 3.2 $w +^* \phi = v$ iff $w[\mathcal{A}]v$ and for each $a \in \mathcal{A}$, $v(a) = \{v +^* \phi \mid v \in w(a) + \phi\}$.

In this definition, the notation $w[\mathcal{A}]v$ stands for the fact that w and v differ at most in the states they assign to agents in \mathcal{A} : w and v assign the same truth-values to the atomic sentences. For later use, we define the mutual update of an information state as $\sigma +^* \phi = \{v +^* \phi \mid v \in \sigma\}$. We will use this operation to change the common ground: it corresponds with learning that ϕ , and learning that all agents have learned that ϕ .

This definition is circular, but it does in fact define a unique function over possibilities.¹² The idea behind the definition is this: one of the participants a lear-

11. Often in the work on belief revision and expansion, information is represented by sets of sentences closed under some ‘base logic.’ We use classical possible worlds. However, if we assume this base logic to contain classical logic, the two modes of representation are equivalent.

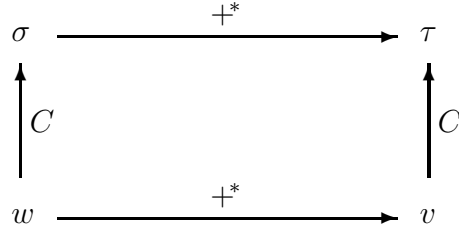
12. More precisely, the definition can be read as defining a system of equations (where objects of the form ‘ $w +^* \phi$ ’ are seen as the indeterminates), which has a unique solution by the axioms of

ning that all participants have expanded with ϕ is the same thing as a learning ϕ herself, and moreover, changing each of the possibilities in her resulting state to the effect that the participants have mutually learned that ϕ . The proof of fact 3.4 contains an example.

One way of viewing the operation $+^*$ is that it models a certain fact becoming common knowledge:

Fact 3.3 $w +^* \phi \models C\phi$.

Now that we have given all parts of our diagram a formal interpretation, we can redraw it:



This diagram commutes just in case $C(w +^* \phi) = C(w) +^* \phi$. It turns out that this is not the case:

Fact 3.4 There are w and ϕ such that:

$$C(w +^* \phi) \neq C(w) +^* \phi$$

There is even an S5-counterexample.

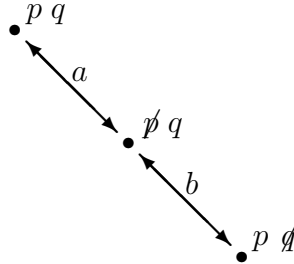
proof: Consider the three possibilities given by the following equations:

$$w_0(p) = 1, w_0(q) = 1, w_0(a) = \{w_0, w_1\}, w_0(b) = \{w_0\}.$$

$$w_1(p) = 0, w_1(q) = 1, w_1(a) = \{w_0, w_1\}, w_1(b) = \{w_1, w_2\}.$$

$$w_2(p) = 1, w_2(q) = 0, w_2(a) = \{w_2\}, w_2(b) = \{w_1, w_2\}.$$

We can draw this model as follows:



In this picture, the topmost dot represents w_0 , the middle represents w_1 , and the lowest dot is w_2 . I have not drawn reflexive arrows: but there should be both a and b -arrows going from each world to itself.

non-well-founded set theory. In Gerbrandy and Groeneveld (to appear), a proof is given.

Consider $w_0 +^* p$. The state $w_0(a)$ contains two worlds, w_0 itself, in which p is true, and w_1 , in which p is false, so $w_0(a) + p = \{w_0\}$. Applying the definition of $+^*$, this gives us that $(w_0 +^* p)(a) = \{w_0 +^* p\}$.

The state $w_0(b)$ contains only w_0 itself, so $(w_0 +^* p)(b) = \{w_0 +^* p\}$.

This means that $w_0 +^* p$ is that possibility in which both p and q are true, and each agent is fully informed about the world. We could draw the possibility by a single dot with a reflexive a and b -arrow. The common ground in $w_0 +^* p$ consists of just a single world: $C(w_0 +^* p) = \{w_0 +^* p\}$.

Consider now the common ground in w_0 : $C(w_0) = \{w_0, w_1, w_2\}$. That means that $C(w_0) +^* p = \{w_0 +^* p, w_2 +^* p\}$. Since in $w_0 +^* p$, q is true, and in $w_2 +^* p$, q is false, $w_2 +^* p$ is different from $w_0 +^* p$, and hence, $C(w_0) +^* p$ is different from $C(w_0 +^* p)$. \square

I could have chosen a counterexample that is less complex, but w_0 is the most simple example I could find that both is S5 and in which the update with p makes sense as the effect of an utterance in a dialogue between b and a . The possibility w_0 has the property $w_0 \models \diamond_a \neg p$ and $w_0 \models \Box_b p$, which makes $w_0 +^* p$ a candidate for the result of b uttering ‘ p ’ in w_0 . Since b believes p in w_0 , she is in a position to say that p , and a has no reason to disagree: he considers p possible.

If one inspects w_0 in the proof of fact 3.4, one sees that a believes that b considers a $\neg q$ -world possible only if $\neg p$ is the case. So, in a sense, the fact that $\neg q$ is a possibility in the common ground in w_0 depends on the fact that a considers $\neg p$ possible. In fact, a and b both know this, although it is not common ground that they both do. This is the reason that when one considers the mutual expansion with p , $\neg q$ disappears from the mutual beliefs, but that this is not reflected in the p -expansion of the common ground.

We get the same kind of result when we choose to represent the common ground in a less detailed way – containing only information about atomic sentences, for example.

Definition 3.5 σ contains less world-information than τ , $\sigma \preceq \tau$ iff for all $v \in \tau$ there is a $w \in \sigma$ such that $w[\mathcal{A}]v$

σ and τ are atomically equivalent, $\sigma \approx \tau$ iff $\sigma \preceq \tau$ and $\tau \preceq \sigma$.

Commutativity modulo atomic equivalence (which is in essence the same thing as representing the common ground as a state containing only information about atomic sentences), fails in the same way as it did before:

Fact 3.6 $C(w +^* \phi) \not\approx C(w) +^* \phi$

proof: Use the same counterexample as before. \square

As I remarked above, mutual belief may be too strong a notion to use it as a model for the common ground. So one may view the property that the expansion of the common ground will never give you any results that are not also mutually

believed in the mutually expanded possibility as a minimal correctness condition. In that respect $+^*$ behaves correctly:

Fact 3.7 $C(w +^* \phi) \subseteq C(w) +^* \phi$.

It turns out that the class of possibilities for which the diagram commutes, modulo \approx , coincides exactly with the class of possibilities where $E(w) = C(w)$: the class of possibilities in which ‘everybody believes ϕ ’ implies ‘it is mutually believed that ϕ :’

Fact 3.8 $E(w) \approx C(w)$ iff $C(w +^* \phi) \approx C(w) +^* \phi$ for all ϕ .

Because the language we are considering is not very rich in expressive power, we cannot prove a result corresponding to the fact above with \approx interchanged with real identity. We do have the following result:

Fact 3.9 If $C(w) = E(w)$, then $C(w +^* \phi) = C(w) +^* \phi$.

One can see the fact that the diagram commutes only for possibilities in which everybody’s belief is mutual belief as a kind of diagnosis of the problem: in $C(w)$, the difference between first-order and higher-order knowledge has disappeared, and the only possibilities in which this does not lead to loss of information are those in which this distinction was not made in the first place.

3.2 Other Operators

The trick above, lifting an operator like expansion to a different operator corresponding to a mutual application can easily be generalized: we simply copy the definition for mutual expansion and apply it to an arbitrary function on information states.

Definition 3.10 Let F be any operation on information states. F^* is the following function over possibilities:

$$F^*(w) = v \text{ iff } v[\mathcal{B}]w \text{ and } v(a) = \{F^*(u) \mid u \in F(w(a))\}$$

Lifting an operation F to F^* has the following effect: each of the agents applies the operation F to his or her own information state, and then updates the possibilities in the resulting state with F^* . Just as in the case of mutual expansion, we will omit the proof of the existence and uniqueness of the function F^* .

I will show that assuming that $F^*(C(w)) \preceq C(F^*(w))$ for each w is inconsistent with assuming that F satisfies the postulates for either contraction or revision, together with the assumption that F is flat:

Definition 3.11 F is flat iff for all s and t : if $s \approx t$, then $F(s) \approx F(t)$.

An operator is flat just in case when the results of applying F to any two states that contain the same world-information will result in states that contain the same world-information too. I think that in general this assumption is not warranted, but when one assumes that F is meant to describe change in world-information only, the assumption is reasonable: if F expresses change in world-information only, its effects on the world-information should depend on world-information only.

It turns out when F is flat, monotony of F over \preceq is a necessary condition for the property that $F^*(C(w)) \preceq C(F^*(w))$

Definition 3.12 An update operator F is monotone over an ordering \preceq iff it holds that $\sigma \preceq \tau$ implies that $F(\sigma) \preceq F(\tau)$. We say that F is propositionally monotone iff it is monotone over \preceq .

Fact 3.13 If F is flat, and for each w , $F^*(C(w)) \preceq C(F^*(w))$, then F is monotone over \preceq .

proof: Assume F is flat, and $F^*(C(w)) \preceq C(F^*(w))$ for each w . Take any σ and τ such that $\tau \preceq \sigma$. Take any $v \in \tau$ (assuming that τ is not empty: in that case, $\tau = \sigma$, and we are finished), and define an S5-possibility w as follows:

$$w(p) = v(p) \text{ for each } p \in \mathcal{P}.$$

$$w(a) = \{v \mid \exists u \in \sigma : u \approx v \text{ and } v(c) = w(a) \text{ for all } c \in \mathcal{B}\} \cup \{w\}$$

$$w(b) = \{v \mid \exists u \in \tau : u \approx v \text{ and } v(c) = w(b) \text{ for all } c \in \mathcal{B}\} \cup \{w\}$$

Since $w(a)$, and each information state occurring anywhere in $w(a)$, is atomically equivalent to σ , and $w(b)$ and each information state occurring in $w(b)$ is atomically equivalent to τ , it follows that $C(w) \approx \sigma \cup \tau \approx \tau$. Since F is flat, it follows that $F(\tau) = F(C(w))$, and hence that $F^*(\tau) \approx F^*(C(w))$.

By assumption, we know that $F^*(C(w)) \preceq C(F^*(w))$.

By definition of C , $F^*(w)(a) \subseteq C(F^*(w))$, so $C(F^*(w)) \preceq F^*(w)(a)$. By definition of F^* , $F^*(w)(a) = F^*(w(a))$ and F^* , $F^*(w(a)) \approx F(w(a))$. Since we have defined w in such a way that $w(a) \approx \sigma$, we have by flatness that $F(w(a)) \approx F(\sigma)$.

Putting all of this together, we get that $F(\tau) \preceq F(\sigma)$. Since we chose σ and τ arbitrarily, we may conclude that F is monotone over \preceq . \square

This result is interesting, because the notions of revision and contraction are not propositionally monotone. Consider for example the following two postulates that have been proposed as conditions on any contraction function:

K-3 If $\sigma \not\approx \phi$, then $\sigma - \phi = \sigma$ (vacuity)

K-4 If $\not\approx \phi$, then $\sigma - \phi \not\approx \phi$ (success)

Fact 3.14 If p is atomic, then any function $-p$ that satisfies (K-3) and (K-4) is not propositionally monotone.

The proofs of this fact and the following are similar to those given in section 4.

To show that revision functions are not monotone over \preceq either, we need the following two postulates:

K*3 If $\sigma \not\models \neg\phi$, then $\sigma * \phi = \sigma + \phi$.

K*4 If $\not\models \neg\phi$, then $\sigma * \phi \neq \emptyset$

Fact 3.15 If p is atomic, and $*p$ satisfies to (K*3) and (K*4) then $*p$ is not monotone.

4 A general framework.

I have shown how to formalize our informal picture in possible worlds semantics. In this section, I will try to assume as little as possible about the structure of information states, the common ground, or the relation between simple and mutual updates, and try to see which assumptions were really essential for the result to go through.

We start with some minimal assumptions that we need for representing agents with certain information. First of all, assume there is a set of agents \mathcal{A} , and a set of (information-)states \mathcal{S} that those agents may be in. We will assume \mathcal{A} to be finite, lets say $\mathcal{A} = \{1, \dots, n\}$, and we will use s_i as variables over states that agent i is ‘in’. (States may be represented by sets of possible worlds, by sets of sentences, discourse representation structures, databases, situation-theoretical objects, anything that suits your fancy.) We also assume that there is a transitive and reflexive relation \preceq on $\mathcal{S} \times \mathcal{S}$, similar to the one we defined in the previous section. The idea is that this relation expresses ‘containing less information about the world’, i.e. it is a measure of information that disregards information about the epistemic states of the agents. We will write that $s \approx s'$ iff $s \preceq s'$ and $s' \preceq s$.

We want to be able to talk about agents having certain information in common, so we need a notion of agents being in certain states together. The simplest way to do this is by representing such a situation by a sequence $\bar{s} = \langle s_0 \dots s_n \rangle$.

Another thing that we assume is that there is some function that extracts the common ground in a situation \bar{s} , and we assume that the common ground can be represented by the same kind of object that represent the states of the agents, i.e. we have a function C on situation \bar{s} , such that $C(\bar{s})$ is a state from \mathcal{S} . The following assumption can be seen as a minimal assumption on the function C :

common ground $C(\langle s_0 \dots s_n \rangle) \preceq s_i$ for any $i \leq n$.

We assume that the common ground in a situation contains less information than each of the agents has in that situation. I don’t think this is a controversial assumption in any way.

Now take an operation $F : \mathcal{S} \mapsto \mathcal{S}$ and a corresponding notion of a mutual application of this function $F^* : \bar{\mathcal{S}} \mapsto \bar{\mathcal{S}}$ that operates on sequences of states. I will propose a number of assumptions on these functions (all of which were assumed in the previous sections) which together are strong enough to give results similar to those we got in the previous section.

distributivity If $F^*(\langle s_0 \dots s_n \rangle) = \langle t_0 \dots t_n \rangle$, then $F(s_i) \approx t_i$ for all $i \leq n$.

To accept this postulate, keep in mind that \preceq orders states with respect to world-information only. What the assumption says is that if the agents in \mathcal{A} mutually perform the operation F , then their higher-order information may change in all kinds of ways, but the changes in the information they have about the world will be the same as when each of the agents would have applied the operation ‘on her own.’

We need a third assumption to guarantee that we have enough states to work with:

fullness We assume that for every two states s and t such that $s \preceq t$, there is a situation \bar{s} that contains a state t such that $t \approx t'$, and which is such that $C(\bar{s}) \approx s$

This is not a very strong assumption, I believe, but it may help to unravel the definition a little. Fullness says that for any two states s and t such that s contains less world-information than t , there is a situation \bar{s} such that the world-information that is mutually known in \bar{s} is the same as that contained in s , while one of the agents in \bar{s} has the same world-information that is contained in t .¹³

The last assumption we make is the same as we did before:

flatness If $s \approx t$, then $F(s) \approx F(t)$.

Given these four assumptions, we can prove that if our diagram commutes, then F must be monotone over \preceq . In fact, we prove something slightly stronger, corresponding to fact 3.13, namely that monotony is a necessary condition for $F(C(\bar{s})) \preceq C(F^*(\bar{s}))$:

Fact 4.1 Assume that the four properties formulated above hold. Then it also holds that if $F(C(\bar{s})) \preceq C(F^*(\bar{s}))$, then F is monotone over \preceq .

proof: Take any s and t such that $t \preceq s$. Since \mathcal{S} is full, we can find $\bar{s} = \langle s_0 \dots s_n \rangle$ such that $s \approx s_i$ for some $i \leq n$ and $C(\bar{s}) \approx t$. Since F is flat, $F(t) \approx F(C(\bar{s}))$. By assumption, $F(C(\bar{s})) \preceq C(F^*(\bar{s}))$.

Let $F^*(\bar{s}) = \langle t_0 \dots t_n \rangle$. By the assumption on the common ground, $C(F^*(\bar{s})) \preceq t_i$. By distributivity, $t_i \approx F(s_i)$, and using flatness again, we have that $F(s_i) \approx F(s)$.

Since we assumed that \preceq is transitive, we can combine these observations and conclude that $F(t) \preceq F(s)$.¹⁴ \square

Since none of the operations considered above was originally defined to be applied to such abstract objects as the states introduced above, we still need to show that this abstract result applies to contraction or revision functions. Of course, we will not be able to prove anything about the original notions of expansion, contrac-

13. Also for this assumption it is important to note that \preceq pertains to world information only. Assuming this, I can see no reason for this assumption to fail in any of the representational frameworks that I know of. The proof of fact 3.13 contains a construction of such a state in a possible worlds model.

14. I have skipped over matters pertaining to the possible partiality of the function F . If one defines monotony as a property that need only hold for values on which F is defined, the proof will work just as well.

tion and revision. Instead, I will reformulate some postulates yet again (in general slightly weakening them), and then prove how failure of monotony follows from them.

To show that contraction functions are not monotone over \preceq , we need to reformulate the postulates for contraction in such a way that they apply to states in general. And for doing that, we need to extend our ontology: we need a language and a relation of \models of ‘acceptance’ between \mathcal{S} and this language. Think of $s \models \phi$ as meaning that ϕ is accepted in state s , that the information that ϕ is subsumed by the information in s , or that the information that ϕ is contained in s .

Consider the following postulates for contraction:

K’-2 $s - \phi \preceq s$.

K’-3 If $s \not\models \phi$, then $s - \phi = s$. (vacuity)

K’-4 If ϕ is not a tautology, then $\sigma - \phi \not\models \phi$. (success)

The original formulation of (K’-2) uses a stronger notion of than \preceq , so the present formulation may be seen as a weaker version. (K’-3) is exactly the same as the original definition. (K’-4) introduces the notion of a ‘tautology’: this may be taken as a primitive notion, or it may be taken as defined as ‘being accepted in each state’ or as ‘being accepted in the minimal state.’

To show a function $-\phi$ satisfying these three postulates is not monotone over ϕ , if ϕ is not a tautology, we need to be sure that (\mathcal{S}, \preceq) contains enough structure.

We will assume that there are states s and t in \mathcal{S} , such that $s \not\models p$, $t \not\models p$, and for all u such that $s \preceq u$ and $t \preceq u$, $u \models p$. Moreover, we assume that there is in fact a u such that $s \preceq u$ and $t \preceq u$. (For an intuitively acceptable example, consider states s and t such that $s \models q$, $t \models q \rightarrow p$.)

We now have enough material to prove that $-p$ is not monotone over \preceq . For assume that $-p$ is monotone. We know that $s - p = s$ and $t - p = t$, by (K-3). Now take any u that contains more information than both s and t . By monotony, $u - p$ must contain more information than both s and t . But by assumption, every such state is one in which p is accepted, contradicting (K’-4).

The postulates for revision assume we have negation in our language, and that \mathcal{S} contains an inconsistent state \perp . We will assume that if s is a state such that $s \models p$ and $s \models \neg p$ for some sentence p , then $s \approx \perp$. Consider the following postulates:

K*2 $s * \phi \models \phi$.

K’*3 If $s \not\models \neg \phi$, then $s \preceq s * \phi$.

K’*4 If $\neg \phi$ is not a tautology, then $\sigma * \phi \not\approx \perp$.

The postulate (K*2) is just the original postulate. It is not hard to see that (K’*3) is a weakening of (K*3), assuming at least that $s \preceq s + \phi$. Similarly, we have weakened (K*4) to the effect that if ϕ is not a contradiction, then a revision with ϕ will not be atomically equivalent with the inconsistent state.

Let p be such that $\neg p$ is not a tautology, and assume we have states s and t such that $s \not\models \neg p$ and $t \not\models \neg p$, and for all u such that $s, t \preceq u$, $u \models \neg p$. Assume moreover that there exists such a u . (Consider, e.g., $s \models q \rightarrow \neg p$, $t \models q$, similar as

before.) Take any u such that $s, t \preceq u$. It holds, by (K'*3), that $s \preceq s * p$, and by monotony, that $s * p \preceq u * p$. Similarly, $t \preceq t * p \preceq u * p$. But then, by assumption, $u * p \models \neg p$. But according to (K*2), $u * p \models p$, from which it follows that $u \approx \perp$, which contradicts (K'*4).

5 Conclusions

In this paper, I have compared two ways to model changes in the common ground: changing a representation of the common ground directly versus seeing such changes as derived from changes in the belief states of the participants involved. It turns out that the two ways of modeling give different results for the two approaches.

The main conclusion to draw from the results of this paper is that mutual belief and common knowledge are not simply two sides of the same coin; at least not when one considers information change. This holds even for expansion.

Expanding the common ground may give results that are too weak when compared with mutual expansion, but they will not lead to additions in the common ground that are not also added when in the mutual expansion. If this discrepancy is a problem at all, I don't think it is a very serious one. Firstly because it seems that one does not seem to need all mutual beliefs to be in the common ground of a dialogue. But also because one of the reasons of using a separate representation of the common ground is that it is a less complicated way of modeling dialogue than keeping track of the states of the participants; this means losing certain information about the relations between world-information and higher-order information, but fact 3.7 shows that this is basically harmless when considering expansion.

The result that a function that is flat has to be monotone for the changes in the common ground to be mutual beliefs, and that neither contraction nor revision are monotone seems more serious. On the other hand, both notions are notorious for their indeterminacy. What the results seem to say is that if one uses a simple deterministic function to model contraction or revision of the common ground, it may be that the resulting common ground will contain information that is not mutually believed. But if one takes a more lenient view on contraction or revision, as a process that involves some more or less arbitrary decisions on what kind of information to discard, i.e. if one considers the result of a revision process as, to a certain extent, unpredictable, it will be unclear in general what exactly is in the resulting common ground, and it will be even less clear to each of the participants what is mutually believed (since the latter involves reasoning about the belief change of the other agents, and their reasoning about each other's belief change, etcetera). The negative results seem to give just another argument that revision and contraction processes are not to be modeled by deterministic functions.

References

- Aczel, P.: 1988, *Non-Well-Founded Sets*, Vol. 14 of *CSLI Lecture Notes*, CSLI Publications, Stanford
- Alchourrón, C., Makinson, D., and Gärdenfors, P.: 1985, On the logic of theory change: partial meet functions for contraction and revision., *Journal of Symbolic Logic* 50, 510–530
- Barwise, J.: 1989, On the model theory of common knowledge, in *The Situation in Logic*, pp 201–220, CSLI Lecture Notes, Stanford
- Barwise, J. and Moss, L.: 1996, *Vicious Circles*, CSLI Publications, Stanford University
- Clark, H. and Marshall, C. R.: 1981, Definite reference and mutual knowledge, in A. Joshi, B. L. Webber, and I. A. Sag (eds.), *Elements of discourse understanding*, pp 0–00, Cambridge University Press, Cambridge, U.K.
- Fagin, R., Halpern, J., Moses, Y., and Vardi, M.: 1995, *Reasoning about Knowledge*, MIT Press, Cambridge (Mass.)
- Gerbrandy, J., *Dynamic Epistemic Logic*, To appear in the proceedings of the Second Conference on Information-Theoretic Approaches to Logic, Language, and Computation
- Gerbrandy, J. and Groeneveld, W., *Reasoning about information change*, To appear in the *Journal of Logic, Language and Information*
- Halpern, J. and Moses, Y.: 1990, Knowledge and common knowledge in a distributed environment, *Journal of the Association for Computing Machinery* 37(3), 549–587
- Hamblin, C. L.: 1971, Mathematical models of dialogue, *Thoeria* 37, 130–155
- Lewis, D.: 1979, Scorekeeping in a language game, *Journal of Philosophical Logic* 8, 339–359
- Stalnaker, R.: 1978, Assertion, in P. Cole (ed.), *Pragmatics (Syntax and Semantics 9)*, pp 315–332, Academic Press, New York
- Veltman, F.: 1996, Defaults in update semantics, *Journal of Philosophical Logic* 25, 221–26
- Zeevat, H.: 1997, The common ground as a dialogue parameter, in *this volume*