

THE ROLE OF BELIEF DYNAMICS FOR KNOWLEDGE

DANIEL SCHOCH

ABSTRACT. It has been emphasised that Belief Dynamics plays a vital role within the theory of knowledge. Previous attempts at defining this connection have left a gap between both fields. We provide a framework in which knowledge can be defined from AGM belief dynamics, and vice versa. Our main theorem states that certain Kripkean knowledge operators can be represented by (local) Belief Revision functions such that knowledge is understood as a kind of undefeated justification, thus coming close to Lehrer's defeasibility analysis.

The AGM conditionals involved induce orderings of possible worlds, which can be regarded as degrees of plausibility. This interpretation helps to understand the specific way of knowledge attribution in Gettier-like scenarios.

This formalism, however, is adequate only for AGM belief states. In scenarios where the subject is unaware of defeated defeaters (Tom Grabit), the representation theorem cannot be applied. But unawareness definitely transcends the scope of classic belief change theory. Here again, problems of knowledge correspond to problems of belief change theory.

1. INTRODUCTION

1.1. The Gap Between Belief and Knowledge Theory. There are two fields of research in epistemology, which hardly meet, although it is widely assumed that they are intertwined [Rott 2002], [Spohn 2001]. On the one hand, we have a host of highly elaborated formal theories about the statics and dynamics of doxastic states, of which we will only mention two. First, there is the most prominent AGM theory of belief revision and, second, there is Spohn's theory of ranking functions [Spohn 1988], which became famous in the community of computational sciences as an alternative to Bayesian Nets. On the other hand, we have the theory of knowledge, which is the main subject of traditional epistemological research, especially in the aftermath of Gettier's famous argument. There hardly seem to be points of contact between these two disciplines, although the theory of knowledge refers to some sort of (true and justified) belief, which should under certain constraints qualify as knowledge. Consequently, Plantinga dismisses belief theories for the lack of relevance to the problem of knowledge [Plantinga 1993]. As Spohn points out, a problem also occurs on the side of belief theorists, who erroneously assume that a "pure" theory of doxastic states can be formulated without reference to justification [Spohn 2001]. In contrast, Spohn expresses his hope that a theory of justification

Date: July 2nd, 2006.

Key words and phrases. Epistemology, Knowledge, Gettier, Internalism, Belief Dynamics, AGM, Plausibility, Unawareness.

I owe thanks to Ulrich Nortmann and Thomas Grundmann for helpful comments on earlier versions of this paper. I thank the anonymous referee for her or his efforts.

This paper is in final form and no version of it will be submitted for publication elsewhere.

is implicitly contained within some sort of belief change theory and that further research on the latter would reveal that connection.

The aforementioned gap between belief and knowledge occurs also on the side of knowledge theorists. The standard analysis stating that knowledge is justified true belief plus x is being increasingly doubted. Williamson considers knowledge as a primary term which is not reducible to belief and justification [Williamson 2000]. Here the standard view that belief is the basic concept by which knowledge has to be defined is turned upside down. Indeed, there are several observations supporting the thesis that a sufficiently sophisticated concept of belief is on a par with knowledge. For example, in almost all non-probabilistic frameworks, belief states are considered deductively closed. If deduction is one, albeit very basic relation of justification, this should nevertheless support Spohn's thesis that what belief theorists call belief already incorporates structures of justification. Moreover, within the AGM theory a kind of conditional belief relation can be defined, which has recently been interpreted as an abductive conditional. Alternatively, Spohn's theory of ranking functions is essentially connected with a subjective causal relation, and was even originally intended to model such a conditional. Thus, also higher structures of justification seem to be implicitly incorporated in the dynamic of doxastic states.

If Spohn's thesis is true and a sufficiently rich, dynamic belief structure already contains justification, would it be going too far to claim that it also implicitly contains knowledge (if only some veridical dimension was added)? The acceptance of such a thesis depends on whether the alleged theory of knowledge can deal with Gettier-like problems. It is the aim of this paper to confirm this, both by giving examples and theoretical reasons. Moreover, we will also defend the converse case: Given a suitable concept of Knowledge, a dynamic belief structure can be derived from it.

From the side of epistemology, Lehrer was the first to try to formulate a general theory of knowledge which focuses on how the subject is inclined to change her or his belief. Belief theorists followed in his footsteps, but found it difficult to express his concepts in terms of belief revision. Rott finds Lehrer's basic concept of undefeated justification too strong [Rott 2002, p. 14f.]. Actually, this problem is connected (exactly) with Lehrer's own "Tom Grabit" example, thereby also spelling trouble for our own theory. Spohn, when trying to adopt Lehrer's theory for his own belief change theory, also arrives at a no-go theorem [Spohn 2002]. His negative result is especially interesting because it can be carried over from the formalism of ranking functions to the AGM framework. He uses a special belief change operation, called contraction, to remove every false belief from the subject's doxastic state.

Unfortunately, in the special case investigated, knowledge reduces merely to true belief.¹

There have been attempts from the side of computer science to bring knowledge and belief dynamics together. Halpern and Friedman have investigated a hybrid system containing disjoint representations for knowledge and plausibility.² Their system, however, deals with a completely different form of knowledge, which is not the one philosophers are targeting. As Rott rightly proclaims, philosophers are not allowed to confuse epistemic and doxastic concepts, however close these may be.

1.2. Belief Dynamics: An Example. Let us consider a motivating example of how belief change theory might be relevant to knowledge. Before we go into the formal details, we will only mention that, according to basic belief change theories, the doxastic state of a person can be represented by a ranking of possible worlds. The ordinal information contained in such a ranking can, without loss of generality, be represented by an ordering called plausibility measure. The current belief state of the subject is usually formed by the worlds of maximal plausibility such that a proposition is contained in the belief state if and only if it is true in each of the most plausible worlds. The rest of the ordinal ranking is needed to define the dynamics of the belief state, as we will see later.

Now assume a person is answering a multiple-choice test where more than one answer can be selected. Any combination of answers could be considered as a complete description, and thus, the possible marks on the answer sheets could be identified with (small) possible worlds. The examinee marks the first answer in figure 1. After being told that it is partially incorrect, she or he revises to the second answer. Again, the answer turns out to be unsatisfactory. Finally, the examinee's third answer turns out correct.

Which knowledge can we attribute to the examinee? Obviously, it depends on whether her or his choice is guided by thoughtful reasoning or rather by wishful thinking or mere guessing. If all seven possible solution outcomes were equally likely to be chosen, we could deny (non-trivial) knowledge. For item a , the subject has never changed her or his opinion; a necessary condition for attributing knowledge of a should be that the subject would be surprised if a was wrong. In all three

¹Expressed in terms of subsets $K \subseteq W$ of possible worlds ("belief cores"), the contraction operation Spohn considers as a candidate for knowledge is $K \div \{w_0\}$, where $\{w_0\}$ stands for the negation of the conjunction of all propositions valid in the actual world w_0 . This enforces all beliefs to be true. Unfortunately, it turns out that $K \div \{w_0\} = K \cup \{w_0\}$, thus knowledge reduces merely to true belief.

The same result could be found in AGM theory for partial meet contraction: It is easy to see that both full meet and maxichoice contraction evaluate to the above result, and therefore any partial meet contraction must do so, too. In both cases, the origin of the failure to produce non-trivial knowledge is the controversial recovery postulate, which holds both for ranking functions [Spohn 1999, p. 8] and partial meet contraction [Gärdenfors 1992, Theorem 7].

²A combination of plausibility orderings with a Kripke framework for knowledge has already been performed by [Halpern 1994]. In Halpern and Friedman's approach, however, the Kripke accessibility relation and the plausibility ordering are different, while they are identical here. This goal conflicts with the authors' CONS principle, which would imply that the subject has to consider all possible worlds which are less plausible than the actual world as impossible. Moreover, their use of the philosophically untenable S5 principle for knowledge, $\neg Kp \rightarrow K\neg Kp$, might suggest that the concept of knowledge Halpern and Friedman have in mind is not the one philosophers are using, which is different from justified true belief. Nevertheless, their other axioms, including REF, SDP, UNIF and RANK are satisfied within our framework.

	a	b	c
Solution 1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Solution 2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Solution 3	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

FIGURE 1. Questionnaire

cases, the answers b or c (or both) are selected. If the examinee has chosen this sequence for good reasons, she or he should be attributed the knowledge, i.e., that b or c (or both) is correct. Had a $\neg b \wedge \neg c$ -solution been ranked by the examinee at least as plausible as the third solution, then no knowledge with respect to b and c could have been attributed.

This example demonstrates how important the *dynamics* of belief is for attribution of knowledge. Without asking the subject to revise her or his belief, we would not have been able to attribute knowledge of the proposition $b \vee c$, since the examinee mixed up the markings for b and c and was wrong on both items, but not on the adjunction. Moreover, we cannot be sure about the attribution of knowledge with respect to item a , unless we have assured ourselves that, given some true information, the examinee would not have given up her or his belief that a . In our case, in all subsequent belief revisions, the item a is chosen. Thus, we are in a position to attribute knowledge that a .

The main intuitive lesson of this paper is that belief dynamics reflects what happens in Gettier-like stories. Here we interpret the different cases in figure 1 as possible worlds. To illustrate this in case of Gettier's own example, let b stand for 'Mr. Nogot drives a certain car' and c for 'Mr. Havit drives such a car.' For a subject having only evidences for b but not for c , the b -worlds are then rationally preferred over all $\neg b$ -worlds. If, in the actual world, $\neg b \wedge c$ holds, and thus $b \vee c$ is justified as true belief, this does still not constitute knowledge since the $\neg b \wedge \neg c$ -scenario ranks equal to the actual world. Thus, no knowledge could be attributed.

Alternatively, the Romeo example is closer to our questionnaire scenario. Let b stand for "Mr. Romeo arrived with the 11 a.m. flight" and c be the proposition "Mr. Romeo arrived with the 10 a.m. flight." If I meet him leaving the airport at 11.30 a.m., I will believe that b . But Mr. Romeo, as the name suggests, has a romantic secret such that b is false and, surprisingly, c is true. This corresponds to the ranking in figure 1 with world 2 removed (of course, he could not have taken *both* flights). Thus, given that these were the only flights from the city where he was seen earlier that morning, there is independent support for $b \vee c$ (more precisely, for the disjunction). This additional support is reflected by world 3 being ranked above everything except world 1. Consequently, the proposition 'either b or c ' is commonly attributed as knowledge, even though the justification for b has been misleading.

1.3. Outline of the Argument. The main thesis of this article is that the connection between epistemic and doxastic states is so close that one has good reasons

to assume that knowledge *supervenes* on a sufficiently rich belief state attribution, together with the information whether a proposition is true or false. We propose a very simple definition of knowledge, which only in a small detail differs from a proposal independently advanced by Klein and Hilpinen [Klein 1971] [Hilpinen 1971]. We find an axiomatic approach and derive a representation theorem. Although Lehrer has good reasons - which we accept - to claim the approach by Klein and Hilpinen is incompatible with one of his examples, our investigation shows that, in fact, the problem lies not only in their definition of knowledge, but rather in some hidden assumptions on the belief states.

We derive two representation theorems, one for a global belief state independent of the world, and a local one. The global version allows representation of any knowledge operator in terms of a belief change conditional satisfying all AGM conditions up to one (and the latter only locally). The local approach deals with a special case of knowledge only, which, among others, presumes the well-known KK-Principle.

Lehrer has introduced the principle of undefeated justification as a condition for knowledge [Lehrer 2000]. The basic idea is that true information should not lead the subject to defeat her or his original belief. In contrast to the common opinion, in our definition 1 a defeater has to be *disbelieved* in order to constitute a counterargument for some alleged belief. Condition (ii) is introduced in order to deal with cases where the subject is *unaware of a defeater, which is itself defeated* by another circumstance. The problem with unaware defeaters is that they do constitute a valid counterargument only in case they are not themselves defeated, a condition hard to formulate axiomatically, and impossible to model in AGM belief revision theory. Alternatively, a disbelieved defeater always constitutes a reason against belief attribution, whether it is indeed a valid argument or only subjectively believed to be one. The converse relation that failure of knowledge implies the existence of a *disbelieved* defeater is less obvious, but is provable as a corollary of the local representation theorem under the aforementioned conditions on a knowledge operator.

We propose a much simpler definition of undefeated justification than Lehrer, using a two-place operator to describe both conditional and unconditional justification in the sense of belief revision. More precisely, our definition of knowledge reads as follows:

Definition 1. *Let the subject be justified in some proposition p . A proposition q is a **disbelieved defeater** for p , if and only if*

- (i) *Proposition q is true,*
- (ii) *The subject is justified in $\neg q$, and*
- (iii) *The subject is no more justified in p conditional to q .*

Definition 2. *A subject is **undefeatedly justified** in accepting that p , if and only if the subject is justified in p , and there is no disbelieved defeater for p .*

Definition 3. *A subject **rationally knows** that p if and only if she or he is undefeatedly justified in accepting that p .*

The attribute "rational" comes from rational choice theory where it is often used as a shortcut to Richter rationality [Richter 1971]. Our representation theorems will justify this terminology. Similar to Lehrer's definition, there is no truth clause in ours, either, since it can be proved that rational knowledge implies truth. Observe

that the claim is not that all knowledge is rational knowledge in our sense - possibly, rational knowledge could be a definitional part of knowledge. We merely claim that knowledge is rational knowledge if and only if the subject is in an AGM state of belief, or, at least some relevant properties of her or his dynamic belief state can be captured by AGM theory. Then, as we will argue, defeaters cannot be ignored by the subject.

2. KNOWLEDGE FOR AGM BELIEF STATES

2.1. Non-Monotonic Reasoning in AGM. AGM belief theory can be represented in several equivalent formulations. Here we adopt the notion of the non-monotonic conditional which can be defined in terms of an ordering \preceq on the set W of possible worlds [Katsuno 1988, section 5]. The ordering \preceq itself stems from a representation theorem on belief revision. In semantic terms, it could easily be defined by

$$(2.1) \quad p \mid\sim q \Leftrightarrow \max_R [p] \subseteq [q],$$

where

$$(2.2) \quad \max_{\preceq} A = \{w \in A \mid \forall v \in A : v \preceq w\}.$$

The conditional holds independently of the actual world. Obviously, whenever q follows logically from p , the conditional $p \mid\sim q$ holds. If p is unconditionally believed, $\mid\sim p$, then $p \mid\sim q$ is being interpreted as an open conditional; and from the postulates ($\mid\sim 1$) and ($\mid\sim 3$) below, it follows that q is then also believed. Whenever p is disbelieved, the conditional $p \mid\sim q$ expresses a counterfactual of the form "if p was the case then q would be" [Gärdenfors 1992, p. 24]. It has been pointed out that $p \mid\sim q$ is a defeasible conditional, well suited for non-monotonic reasoning such as "the grass is wet" which abductively follows from "it rains," but not from "it rains and the grass is covered" [Boutilier/Becher 1995].

In AGM theory, the conditional $\mid\sim$ has been identified with the belief change operator for a fixed belief set such that $p \mid\sim q$ reads that q is believed after revision with p . Thus, $p \mid\sim q$ is just another way to write the more common notion $q \in K * p$, where the AGM term K represents a fixed set of initial beliefs, not to be mixed up with the knowledge operator. Observe that $\mid\sim$ is different for each such initial belief set and each belief change operator $K*$. It has been shown that the conditional/revision operator satisfies the famous eight AGM postulates if and only if a complete ordering \preceq exists on the set W of possible worlds such that the conditional could be represented by (2.1). In order to avoid confusion between the belief set named K and the knowledge operator K , we will formulate the AGM postulates directly in terms of the conditional, which is an equivalent way of doing it.

The same conditional has been extensively studied in the literature under the name of abductive inference.³ Here, $p \mid\sim q$ reads that q follows abductively from p . The ordering \preceq is interpreted as a measure of plausibility of the world. According

³For an overview containing interesting epistemological applications, see [Boutilier/Becher 1995]. The abductive conditional is denoted by \Rightarrow , which we need for the implication in metalanguage. There the syntactic definition of the abductive conditional can also be found. The plausibility order is revised to an implausibility order to match the conventions in computer science literature. We leave the question open whether Boutilier's claim is justified that this conditional models the abductive inference to the best explanation.

to this notion, q follows abductively from p if and only if the most plausible p -worlds are all q -worlds. We endorse this reading since it clarifies the function of the ordering \preceq : Worlds are ordered with respect to a degree of plausibility, and (conditional) belief is formed by the most plausible worlds. The origin of the ordering is commonly left open. Belief theorists and computer scientists take this ordering to be completely subjective in nature. Epistemologists, however, who are fond of explicating internalistic knowledge, should insist on a more rationalistic interpretation where the ordering is based on good reasons. A suitable approach could be found in coherence theory [Bartelborth 1996] [Bartelborth 1999], where orderings are considered which measure the degree of constraint fulfilment given by general inference relations [Schoch 2000].

Definition 4. A binary operator $|\sim$ is called a (non-monotonic) **AGM conditional** if and only if the following axioms hold

($|\sim 1$) If $p |\sim q_i$ for all $q_i \in A$ and $A \vdash r$, then $p |\sim r$. (Closure)

($|\sim 2$) $p |\sim p$ (Reflexivity)

($|\sim 5$) If $p |\sim \perp$, then $p \vdash \perp$. (Consistency Preservation)

($|\sim 6$) If $p \leftrightarrow q$ is a logical truth, and $p |\sim r$, then $q |\sim r$. (Left Logical Equivalence)

($|\sim 7$) If $p \wedge q |\sim r$, then $p |\sim q \rightarrow r$. (Conditionalisation)

($|\sim 8$) If $p |\approx \neg q$ and $p |\sim r$, then $p \wedge q |\sim r$. (Rational Monotonicity)

Remark 1. An AGM conditional also satisfies the weaker conditions

($|\sim 3$) If $p |\sim q$, then $\top |\sim p \rightarrow q$. (Weak Conditionalisation)

($|\sim 4$) If $\top |\approx \neg p$ and $\top |\sim p \rightarrow q$, then $p |\sim q$. (Weak Rational Monotony)

Here ($|\sim 3$) is simply ($|\sim 7$) with $p \equiv \top$, the tautological proposition. Condition ($|\sim 4$) can be seen as a special case of ($|\sim 8$) by letting $p \equiv \top$ and $r \equiv q \rightarrow s$, and applying ($|\sim 1$) and ($|\sim 2$). In the AGM notation, both axioms are logically redundant and are named only for notational convenience because they correspond to similar axioms for the belief revision operator. Furtheron we will write $|\sim p$ as a shortcut for $\top |\sim p$. AGM conditionals are well known to be representable in the form (2.1), where R is a complete ordering on W . However, here we need the following, more detailed result:

Lemma 1. Let R be an arbitrary relation on W . Then, by (2.1), a conditional is declared satisfying ($|\sim 1$), ($|\sim 2$), ($|\sim 6$), and ($|\sim 7$). If R is transitive, then ($|\sim 8$) holds. If, restricted on a finite subset $V \subseteq W$, R is a complete ordering, then ($|\sim 5$) holds for every p with $[p] \subseteq V$.

Proof. ($|\sim 1$): Let $p |\sim q_i$ for all $q_i \in A$ and $A \vdash r$. Thus, $\max_R [p] \subseteq \bigcap_i [q_i] \subseteq [r]$, or $p |\sim r$.

($|\sim 2$): Obviously, $\max_R [p] \subseteq [p]$.

($|\sim 6$): If p and q are logically equivalent, then $[p] = [q]$, and thus $p |\sim r$ implies $q |\sim r$.

($|\sim 7$): Let $p \wedge q |\sim r$. For $w \in \max_R [p]$, if $w \in [q]$ then also $w \in \max_R [p \wedge q] \subseteq [r]$. Thus, $w \in [q \rightarrow r]$ and we arrive at $\max_R [p] \subseteq [q \rightarrow r]$, which is what had to be shown.

($|\sim 8$): Assume that R is transitive. Assume further $p |\approx \neg q$ and $p |\sim r$. In other words, $\max_R [p] \not\subseteq [\neg q]$ and $\max_R [p] \subseteq [r]$. From the first statement we find that there is a $z \in \max_R [p]$ with $z \in [p] \cap [q]$. Thus, for all $v \in [p]$, vRz . Our

goal is to show that $\max_R [p \wedge q] \subseteq [r]$. For every $w \in \max_R [p \wedge q]$, we have zRw . Thus, for all $v \in [p]$, $vRz \wedge zRw$, and, by transitivity, vRw . We finally obtain $w \in \max_R [p] \subseteq [r]$, or $\max_R [p \wedge q] \subseteq [r]$, or $p \wedge q \mid \sim r$.

($\mid \sim 5$): Assume that R is a complete order on $[p]$. Let $p \mid \sim \perp$, or $\max_R [p] \subseteq \emptyset$. Since R is a complete order on a finite set $[p]$, $\max_R [p] \neq \emptyset$ if and only if $[p] \neq \emptyset$. Thus, $[p] = \emptyset$, or $p \vdash \perp$. ■

2.2. The Knowledge Operator. We represent knowledge by the standard truth condition for a Kripke operator,

$$(2.3) \quad Kp \text{ is true in } w_0 \Leftrightarrow \forall w : w_0Rw \Rightarrow p \text{ is true in } w.$$

Here R denotes the Kripkean accessibility relation R , which is defined on the set W of possible worlds. More formally, a Kripkean operator K is generated by a relation R on W , also represented as a Kripke frame $\langle R, W \rangle$, by

$$(\text{Kripke}) \quad w_0 \in [Kp] \Leftrightarrow \forall w : w_0Rw \Rightarrow w \in [p].$$

Here, for each sentence p , $[p]$ denotes the set of p -worlds such that $w \in [p]$ iff p is true in w . We will use double arrows ' \Rightarrow ' and ' \Leftrightarrow ' for metalanguage implication and equivalence, respectively, and single arrows ' \rightarrow ' and ' \leftrightarrow ' for the corresponding junctors in the object language, which contains the operator K .

The existence of such a Kripke frame is equivalent to the following two principles:

$$(\text{Tautology}) \quad [p] = W \Rightarrow w \in [Kp],$$

$$(\text{Closure}) \quad Kp \wedge K(p \rightarrow q) \rightarrow Kq.$$

A minimum requirement is that knowledge implies truth,

$$(\text{Truth}) \quad Kp \rightarrow p.$$

Definition 5. A *knowledge operator* K is an operator satisfying (Tautology), (Closure) and (Truth).

Another axiom often suggested for knowledge is the so-called KK-principle. Many authors, among them Chisholm, have proposed that a decisive criterion for epistemological internalism is that knowledge implies knowledge of knowledge. This can easily be made clear. Assume that in some world Kp is true, but not KKp . Then there must be a proposition q with $Kp \vdash q$ and $\neg Kq$, since otherwise KKp would hold. Observe that there might not be a logically weakest q with this property, but we have a rather weak proposition in mind. Since q is not known, this could not be the strictest possible form of internalism. The KK-principle reads

$$(\text{KK}) \quad Kp \rightarrow KKp.$$

The KK-principle is equivalent to the transitivity of the accessibility relation R . The reflexivity of R is given by the Truth principle. Thus, according to the common view, internalistic knowledge can be represented by an S4 operator where the accessibility relation R is a (partial) weak order. We do not enter the debate whether the KK-Principle should hold in general for an operator representing knowledge. Rather, our aim is to point out its specific role within our link of knowledge to belief revision theory. The result arrived at below seems to indicate that a failure of the KK-Principle is not due to an epistemological misrepresentation of the metaphysical connection of knowledge to the world, but rather to the fact that the subject is not in an AGM-like state of belief.

Another rather strong condition has to be imposed to obtain our representation theorem. In order to let (2.2) be well defined for certain subsets of W , we require the accessibility relation R of the knowledge operator to be weakly connected,

$$(2.4) \quad wRx \wedge wRy \rightarrow xRy \vee x = y \vee yRx.$$

For a knowledge operator, R is always a reflexive relation so that the former condition simplifies to

$$(2.5) \quad wRx \wedge wRy \rightarrow xRy \vee yRx.$$

The general characterisation of (2.4) for a Kripke operator J with accessibility relation R is the rather complicated formula

$$(2.6) \quad J(p \wedge Jp \rightarrow q) \vee J(q \wedge Jq \rightarrow p).$$

The modal logical system for a knowledge operator satisfying (KK) and (2.4) is called S4.3, and is well studied in the literature [Hughes / Cresswell, 128-130]. For a knowledge operator K satisfying the KK-Principle, (2.6) can be written as⁴

$$(LD) \quad K(Kp \rightarrow Kq) \vee K(Kq \rightarrow Kp).$$

We call this the **local dependency** condition (LD). It can be understood only in terms of the AGM belief theory. Here, for the unconditional belief $Jp := |\sim p$, (2.6) can easily be derived. According to Spohn's thesis, the belief structure includes justification such that J has to be understood as justified acceptability. But within our theory, knowledge should be determined by the justification structure and truth alone. Thus, if justification satisfies local dependence, knowledge should do so, too. We call a knowledge operator which satisfies both conditions rational because, as we will see later, it can be represented by a belief structure which is expressed by rational choice theory in the ordinary economic sense [Richter 1971].

Definition 6. A *rational knowledge operator* is a knowledge operator K satisfying the KK-Principle and the LD-Axiom.

Note that this definition, and especially condition (LD), is compatible with epistemically independent propositions.

2.3. The Global Representation Theorem. We are now in a position to explicate rational knowledge (undefeated justification) in terms of belief change theory. As mentioned before, this can best be described as the stability of belief under true information. Indeed, we are able to prove that this condition defines a knowledge operator in the sense of our definition 5.

Definition 7. Let $|\sim$ be an AGM conditional in the sense of definition 4 describing the belief state of a subject in world w_0 . Then the subject has **rational knowledge** that p (denoted by Kp) in w_0 , if and only if the subject is justified in p ($|\sim p$) in w_0 , and, for all propositions q , if q is disbelieved ($|\sim \neg q$) and true in w_0 , then conditional to q , the subject is still justified in p ($q|\sim p$) in world w_0 .

⁴For a knowledge operator, since $Kp \rightarrow p$, (2.6) for operator K is equivalent to

$$K(Kp \rightarrow q) \vee K(Kq \rightarrow p).$$

Instanciating Kp for p and Kq for q , we can derive

$$K(KKp \rightarrow q) \vee K(KKq \rightarrow p).$$

From this, according to the KK-Principle, it follows (LD). Conversely, from (LD) we obtain the first formula above by applying the truth condition $Kp \rightarrow p$.

Lemma 2. *Let $|\sim$ be an AGM conditional in the sense of definition 4. Then, by the above definition, a knowledge operator K is defined.*

Proof. We first prove that, given $|\sim$ be an AGM conditional, the above definition (2.7)

$w_0 \in [Kp] \Leftrightarrow w_0 \in [|\sim p] \wedge \text{For all propositions } q : w_0 \in [q] \wedge w_0 \in [|\sim \neg q] \Rightarrow w_0 \in [q |\sim p]$
is equivalent to

$$(2.8) \quad w_0 \in [Kp] \Leftrightarrow \text{For all propositions } q : w_0 \in [q] \Rightarrow w_0 \in [q |\sim p]$$

Only direction " \Rightarrow " has to be shown. Let $w_0 \in [Kp]$ in the sense of (2.7) and assume $w_0 \in [q]$ for some q . We find $w_0 \in [|\sim p]$, and, by condition ($|\sim 1$), moreover $w_0 \in [|\sim q \rightarrow p]$. If $w_0 \in [|\sim \neg q]$ holds, then by (2.7) we are finished. Conversely, if $w_0 \in [|\sim \neg q]$ holds, then by condition ($|\sim 4$), since $|\sim q \rightarrow p$ and $|\sim \neg q$ holds in w_0 , also $q |\sim p$ is valid in w_0 , which is what has to be shown.

To show (Tautology), assume that p is a tautology, $[p] = W$, and let $w_0 \in W$ be a possible world. We have to show that $w_0 \in [Kp]$. Let q be a proposition with $w_0 \in [q]$. Since p is a tautology, we have $q |\sim p$ according to ($|\sim 1$). Thus, the definiens is satisfied.

For (Closure), assume that Kp and $K(p \rightarrow q)$ hold in $w_0 \in W$. Thus, for all propositions r with $w_0 \in [r]$ we have $r |\sim p$ and $r |\sim p \rightarrow q$. Then, by ($|\sim 1$), since $\{p, p \rightarrow q\} \vdash q$, we obtain $r |\sim q$ for every propositions r with $w_0 \in [r]$. Thus, Kq holds in $w_0 \in W$.

The (Truth) principle can be shown as follows: Assume Kp holds in w_0 , we have to demonstrate that p is true in w_0 . Assume the contrary, $w_0 \notin [p]$. Then $w_0 \in [\neg p]$. By definition of K , since $w_0 \in [Kp]$ has been presupposed, $\neg p |\sim p$. But by ($|\sim 2$), also $\neg p |\sim \neg p$ holds, and by ($|\sim 1$), we conclude $\neg p |\sim p \wedge \neg p$, or $\neg p |\sim \perp$. Now, by ($|\sim 5$), $\neg p \vdash \perp$. Therefore, $[p] = W$, and $w_0 \in [p]$, or p is true in w_0 , which has to be proved. ■

Surprisingly enough, for finite languages a general representation theorem for every knowledge operator can be formulated. In such languages, possible worlds correspond to finite descriptions and are generally more or less tacitly assumed to be characterised by propositions of that language. We call a language discrete if it has such a property. One can find discreteness implicitly assumed in [Spohn 2002].

Definition 8. *A language with a set of possible worlds W is called **discrete** if and only if for every world $w \in W$ there is a proposition q with $[q] = \{w\}$.*

Theorem 1. *In a discrete language, every knowledge operator K can be represented by a conditional $|\sim$ in the form (2.8). If, moreover, the KK -principle holds for K , then $|\sim$ satisfies the AGM axioms except ($|\sim 5$), and can therefore be written in the form given by the definition 7 of rational knowledge.*

Proof. Let K be a knowledge operator with accessibility relation R . We demonstrate that (Kripke) is equivalent to (2.8), where $|\sim$ is defined by (2.1) using the accessibility relation R . That is, to show the following equivalence:

$$(\forall w : w_0 R w \Rightarrow w \in [p]) \Leftrightarrow \text{For all propositions } q : w_0 \in [q] \Rightarrow \max_R [q] \subseteq [p].$$

" \Rightarrow ": Assume that q is a proposition with $w_0 \in [q]$. Now let w be an arbitrary world with $w \in \max_R [q]$. Thus, by definition (2.2), $w_0 R w$. From the left-hand side it follows $w \in [p]$. We have arrived at $\max_R [q] \subseteq [p]$.

" \Leftarrow ": Assume the right-hand side of the equivalence and let w be a world with $w_0 R w$. Since the language is discrete, there are propositions q_1, q_2 with $[q_1] = \{w_0\}$ and $[q_2] = \{w\}$. Then, also $q \equiv q_1 \vee q_2$ is a proposition of the language, and we find $[q] = \{w_0, w\}$. Since, by assumption, $w_0 R w$, from definition (2.2) we obtain $w \in \max_R [q]$. Since also $w_0 \in [q]$ holds, by the right-hand side of the equivalence it follows $w \in [p]$, which is what has to be shown.

The rest of the proposition follows from lemma 1. ■

It remains to explain why the consistency preservation axiom ($|\sim 5$) fails to hold in general. According to lemma 1, ($|\sim 5$) holds on those areas $V \subseteq W$, where R is connex ($x R y \vee y R x$ for all $x, y \in V$). Obviously, this cannot be a general property of the accessibility relation R on a sufficiently rich set of possible worlds W . It is easy to see that R being connex would imply that for each proposition p and q , the corresponding propositions expressing their knowledge would be logically dependent,

$$(2.9) \quad (Kp \vdash Kq) \vee (Kq \vdash Kp),$$

i.e., either $Kp \rightarrow Kq$ or $Kq \rightarrow Kp$ (or both) would be a logical truth. This is much stronger than our condition (2.6). Alternatively, such a logical connection should hold under the condition that the subject's (actual and dispositional) belief is kept constant. Since, in general, the subject's state of belief is expected to vary with the actual world, there are worlds belonging to different belief systems. Such two worlds can be considered epistemologically incomparable, and are therefore not connected by the relation R .

2.4. The Local Representation Theorem. We now refine our first result in terms of a collection of rational choice functions indexed by possible worlds.

Definition 9. A *rational choice function* on W is a function $\varphi : \mathcal{L} \rightarrow \mathcal{L}$ such that for all $A, B \in \mathcal{L}$

$$(2.10) \quad \varphi(A) = \emptyset \Rightarrow A = \emptyset,$$

$$(2.11) \quad \varphi(A) \subseteq A,$$

$$(2.12) \quad \varphi(A) \cap B \neq \emptyset \Rightarrow \varphi(A \cap B) = \varphi(A) \cap B.$$

A *local rational choice function* $\{\varphi_w\}_{w \in W}$ is a collection of rational choice functions φ_w for each world $w \in W$.

Rational Choice Theory states that on a finite set W φ is a rational choice function if and only if it can be written in the form

$$(2.13) \quad \varphi(A) = \max_{\preceq} A$$

for some complete ordering \preceq on W . Equation (2.13) is the standard notation of Richter rationality [Richter 1971], which was originally proposed for *arbitrary* (not necessarily transitive) relations \preceq , while the special case of a complete ordering is obtained using Tchernov's axiom (2.12) for the choice function.

Lemma 3. A local rational choice function $\{\varphi_w\}_{w \in W}$ induces an AGM conditional by

$$(2.14) \quad w_0 \in [p \mid \sim q] \Leftrightarrow \varphi_{w_0}([p]) \subseteq [q].$$

Proof. ($|\sim 1$): Let $w_0 \in [p \mid \sim q_i]$ for all $q_i \in A$ and $A \vdash r$. Thus, $\varphi_{w_0}([p]) \subseteq \bigcap_i [q_i] \subseteq [r]$, or $w_0 \in [p \mid \sim r]$.

($|\sim 2$): By (2.11), $\varphi_{w_0}([p]) \subseteq [p]$.

($|\sim 5$): Let $w_0 \in [p \mid \sim \perp]$, or $\varphi_{w_0}([p]) \subseteq \emptyset$. Then, by (2.10), $[p] = \emptyset$, or $p \vdash \perp$.

($|\sim 6$): If $p \longleftrightarrow q$, then $[p] = [q]$, and thus $\varphi_{w_0}([p]) = \varphi_{w_0}([q])$. We conclude that $w_0 \in [p \mid \sim r]$ implies $w_0 \in [q \mid \sim r]$.

($|\sim 7$): Let $w_0 \in [p \wedge q \mid \sim r]$. For $w \in \varphi_{w_0}([p])$, if $w \in [q]$ then, by (2.12), also $w \in \varphi_{w_0}([p]) \cap [q] = \varphi_{w_0}([p \wedge q]) \subseteq [r]$. Thus, $w \in [q \rightarrow r]$ and we arrive at $\varphi_{w_0}([p]) \subseteq [q \rightarrow r]$, which is what had to be shown.

($|\sim 8$): Assume that $w_0 \notin [p \mid \sim \neg q]$ and $w_0 \in [p \mid \sim r]$. In other words, $\varphi_{w_0}([p]) \not\subseteq [\neg q]$, or $\varphi_{w_0}([p]) \cap [q] \neq \emptyset$, and $\varphi_{w_0}([p]) \subseteq [r]$. From (2.12) we infer that $\varphi_{w_0}([p \wedge q]) \subseteq \varphi_{w_0}([p]) \subseteq [r]$, or $w_0 \in [p \wedge q \mid \sim r]$. ■

Definition 10. For any set $A \subseteq W$, relation R on W , and any world $w_0 \in W$, let

$$(2.15) \quad A_{w_0}^R := \{x \in A \mid w_0 R x\},$$

denote the elements of A reachable from w_0 by R .

Remark 2. Let R be a transitive and locally connected (2.5) relation. Then, for every $A \subseteq W$, the set $A_{w_0}^R$ is totally ordered.

Proof. Since R is transitive, $A_{w_0}^R$ is partially ordered by R . It remains to show that R is connex on $A_{w_0}^R$. Let $x, y \in A_{w_0}^R$, then $w_0 R x$ and $w_0 R y$. By (2.5), $x R y$ or $y R x$, which is what had to be shown. ■

Lemma 4. Let R be a transitive and locally connected (2.5) relation. By

$$(2.16) \quad \varphi_{w_0}(A) := \begin{cases} \max_R A_{w_0}^R, & \text{if } A_{w_0}^R \neq \emptyset \\ A, & \text{else} \end{cases}$$

for $A \subseteq W$ and $w_0 \in W$, a local rational choice function is defined.

Proof. (2.10): Assume $\varphi_{w_0}(A) = \emptyset$. If $\varphi_{w_0}(A) = \max_R A_{w_0}^R$, then $\max_R A_{w_0}^R = \emptyset$. Since, by remark 2, $A_{w_0}^R$ is totally ordered by R , it follows $A_{w_0}^R = \emptyset$, a contradiction. Thus, $\varphi_{w_0}(A) = A = \emptyset$.

(2.11): Obviously, since $\max_R A_{w_0}^R \subseteq A_{w_0}^R \subseteq A$.

(2.12): Assume $\varphi_{w_0}(A) \cap B \neq \emptyset$. In case (i) $A_{w_0}^R = \emptyset$, also $(A \cap B)_{w_0}^R = \emptyset$, and $\varphi(A \cap B) = A \cap B = \varphi(A) \cap B$. In case (ii) $A_{w_0}^R \neq \emptyset$, we have $\varphi_{w_0}(A) = \max_R A_{w_0}^R$, thus $(\max_R A_{w_0}^R) \cap B \neq \emptyset$. By standard rational choice theory, we conclude $\max_R (A_{w_0}^R \cap B) = (\max_R A_{w_0}^R) \cap B$. Now, since $(A \cap B)_{w_0}^R = A_{w_0}^R \cap B \neq \emptyset$, we have $\varphi(A \cap B) = \max_R (A \cap B)_{w_0}^R = \max_R (A_{w_0}^R \cap B) = (\max_R A_{w_0}^R) \cap B = \varphi(A) \cap B$. ■

Having collected all our ingredients, we now need to state the main Theorem. The most important statement is that rational knowledge operators have a representation in the form given by definition 7 with an AGM conditional $|\sim$. In other words, a rational (S4.3) knowledge operator can always be represented as undefeated justification with respect to some AGM dynamic doxastic state.

Theorem 2. Let \mathcal{L} be a discrete language. Let $\{\varphi_w\}_{w \in W}$ be a local rational choice function, and $|\sim$ be the AGM-conditional derived from it in lemma 3. Then, by definition 7 of rational knowledge through conditional $|\sim$, a knowledge operator is defined. Every rational knowledge operator can be represented in this form.

Proof. From lemmata 3 and 2 it follows that definition 7 defines a knowledge operator from $|\sim$ given by (2.14). Conversely, let K be a rational knowledge operator with accessibility relation R . Thus, R is reflexive, transitive and locally connected. We demonstrate that (Kripke) is equivalent to (2.8), where $|\sim$ is defined by (2.14) using the local rational choice function (2.16) derived from the accessibility relation R . Since R is reflexive, $w_0 \in [q]$ implies that $[q]_{w_0}^R \neq \emptyset$ and thus $\varphi_{w_0}([q]) = \max_R [q]_{w_0}^R$. That is, we have to prove the following equivalence:

$$(\forall w : w_0 R w \Rightarrow w \in [p]) \Leftrightarrow \text{For all propositions } q : w_0 \in [q] \Rightarrow \max_R [q]_{w_0}^R \subseteq [p].$$

" \Rightarrow ": Assume that q is a proposition with $w_0 \in [q]$. Since R is reflexive, $w_0 \in [q]_{w_0}^R$. Now let w be an arbitrary world with $w \in \max_R [q]_{w_0}^R$. Thus, by definition (2.15), $w_0 R w$. From the left-hand side it follows $w \in [p]$. We have arrived at $\max_R [q]_{w_0}^R \subseteq [p]$.

" \Leftarrow ": Assume the right-hand side of the equivalence and let w be a world with $w_0 R w$. Since the language is discrete, there are propositions q_1, q_2 with $[q_1] = \{w_0\}$ and $[q_2] = \{w\}$. Then, also $q \equiv q_1 \vee q_2$ is a proposition of the language, and we find $[q] = \{w_0, w\}$. Since R is reflexive and, by assumption, $w_0 R w$, we find $[q]_{w_0}^R = [q]$. From definition (2.2) we obtain $w \in \max_R [q] = \max_R [q]_{w_0}^R$. Since $w_0 \in [q]$ also holds, by the right-hand side of the equivalence it follows $w \in [p]$, which is what had to be shown. ■

3. INTERPRETATION

3.1. Examples. Many philosophical attempts have been made to define knowledge by other concepts, such as belief or justification, in order to solve the puzzles Gettier and his followers have left to us. We show that such a definition is not needed. The interpretation presented in the former section is sufficient for this. For simplicity, we assume that the person's doxastic state does not change with the outer world. It is the usual, tacit assumption of "Gettierology" to keep the subject in a fixed belief state and vary the outer world to study which knowledge should be attributed to it. From our representation theorem, we find that the doxastic state is represented by a plausibility ordering, which at the same time constitutes the accessibility relation for the knowledge operator. For the first two examples, there is no harm in working under the strong (2.9) condition such that reference is made only to the global representation theorem. In the third case, the plausibility ordering comes out partial.

3.1.1. Simple Gettier. Let us start with a very simple example. The only two propositions in our possible worlds are "Nogot" for "Mr. Nogot driving a car" and "Havit" for "Mr. Havit doing so." The only evidence is in favour of Mr. Nogot driving a car, which constitutes the only belief the subject has.

Plausibility	+	○	+	○
World	w_0	w_1	w_2	w_3
Havit	T	T	F	F
Nogot	T	F	T	F

Here degrees of plausibility are denoted in rising order by $-$, \circ , $+$, $++$. The actual world is w_1 . The subject's actual belief state is characterised by the worlds of maximal plausibility, $\{w_0, w_2\}$, which corresponds to the proposition 'Nogot,' while her or his knowledge, characterised by the set of all possible worlds at least

as plausible as w_1 , $\{w_0, w_1, w_2, w_3\}$, is void. The subject does not know that one of the two persons drives a car, although this is justified true belief.

3.1.2. *Overlapping Gettier*. Overlapping evidences are known as the Romeo example in the literature. The class of these scenarios constitutes a problem for all those approaches based solely on possible world semantics over belief states, since the belief state remains the same as in the former example. We take the former case as a basis, but this time, besides the misleading evidence, there is another independent evidence that at least one of the two men is driving a car. For example, we might think of seeing a car parked at the university campus, although on that particular day and time no other seminar is taking place. The subject might be justified in believing that at least one of the two co-participants of her or his *privatissimum* seminar is driving a car. Thus, the plausibility order has changed as follows:

Plausibility	+	○	+	-
World	w_0	w_1	w_2	w_3
Havit	T	T	F	F
Nogot	T	F	T	F

Assuming that the actual world is still w_1 , while the subject's belief state remains the same, the state of knowledge has changed to "Havit or Nogot," characterised by the set $\{w_0, w_1, w_2\}$. The conditional $\neg\text{Nogot}|\sim\text{Havit}$ holds, in accordance with our finding.

3.1.3. *Undercutting an Argument*. A more general Gettier case can be formulated which does not rely on false premises. Assume that Smith walks into a room and seems to see Jones in it; he immediately forms the justified belief, "Jones is in the room." But in fact, it is not Jones that Smith saw; it was a life-size replica propped up in Jones's chair. Nevertheless, Jones is in the room; he is just hiding behind a curtain while his replica makes it seem as though he is there. So Smith's belief is not only justified but also true.

An undercutting argument should not establish knowledge. This case is more difficult to handle, compared to the foregoing ones. Assume a certain evidence e supports a proposition p . Assume further that both propositions are true in the actual world. If an undercutting proposition f is true, the argument does not contribute to plausibility. For example, assume that we have been told that p . By default, the evidence e of having received this message makes it more plausible that p . But the argument might nevertheless fail for some reason f , even though both e and p are true. In this case, we are not inclined to attribute knowledge that p .

In this example, two independent factors contribute to the (partial) order of plausibility. First, there is a positive measure of plausibility, designated by $+$, in all e -worlds. Second, another positive measure will be added to those worlds, in which the argument supporting p is valid. This is exactly the world in which p and e are true, but the undercutting factor f is false, which is assigned the highest degree of plausibility $++$. There is only one exception. If f is the only potentially undercutting factor, and if e is true and f false, we would find it extremely implausible if p was wrong. The negative plausibility should in this case outweigh the positive one stemming from e and denoted by $-$. To the rest of the worlds, a neutral degree of plausibility \circ is assigned. In figure 2, values \circ and $-$ are considered incomparable such that the resulting ordering is partial.

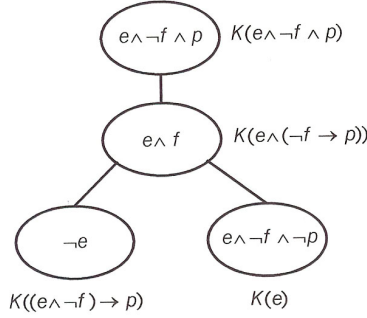


FIGURE 2. Undercutting an Argument

Plausibility	+	++	○	○	+	-	○	○
World	w_0	w_1	w_2	w_3	w_4	w_5	w_6	w_7
Proposition p	T	T	T	T	F	F	F	F
Evidence e	T	T	F	F	T	T	F	F
Undercutting Factor f	T	F	T	F	T	F	T	F

It is an artefact of the AGM theory to assume that, since the subject believes the argument to be correct, she or he would deny any undercutting factor f , even though neither f nor $\neg f$ is believed in the strict sense of the word, and she or he might even be unaware of this. If p is unconditionally acceptable, $|\sim p$, and given f this belief would be changed to the opposite, $f |\sim \neg p$, then also $|\sim \neg f$ holds, which means that $\neg f$ is acceptable (for the proof, see below). Consequently, the subject’s state of acceptance is characterised by $\{w_1\}$, or $p \wedge e \wedge \neg f$. The actual world is w_0 . Thus, knowledge is represented by $\{w_0, w_1, w_4\}$, or $e \wedge (\neg f \rightarrow p)$. In other words, the subject knows the evidence and the conditional $\neg f \rightarrow p$. Neither she or he knows that p nor that $\neg f$. This is exactly what we expect, since there is no valid argument for either proposition.

3.2. The "Tom Grabit" Case - A Non-AGM Belief State. Lehrer introduced the Tom Grabit example to distinguish his own theories from those of his predecessors, who propose that in order to have knowledge attributed to her or him, the subject’s justification should not break down if she or he was confronted with a true fact. This has been regarded as an explication of the condition that justification should not depend on false premises. We will tell Lehrer’s story in the version of [Lehrer 2000, p. 158 f.], which contains a true fact acting as a defeater for a proposition assumed to be known, which is itself undercut by another true statement. The crucial point is that the subject is totally unaware both of the defeater and the undercutting argument. Had she or he been aware of the first fact without being aware of the second, the subject would have changed her or his original belief.

Suppose I see a man, Tom Grabit, with whom I am acquainted and have seen often before, standing a few yards from me in the library. I observe him take a book off the shelf and leave the library. I am justified in accepting that Tom Grabit took a book, and, assuming he did take it, I know that he did. Imagine, however, that Tom Grabit’s father has, quite unknown to me, told someone that Tom

was not in town today, but his identical twin brother, John, who he himself often confuses with Tom, is in town at the library getting a book. Had I known that Tom's father said this, I would not have been justified in accepting that I saw Tom Grabit take the book, for if Mr. Grabit confuses Tom for John, as he says, then I might surely have done so, too. [...] The truth is that Tom's thieving ways have driven his father quite mad and caused him to form the delusion that Tom has a twin, John, who took the book from the library that was actually taken by Tom. Mr. Grabit thus protects his wish that Tom is honest. I know none of this, but I did see Tom Grabit take the book.

Although the story is in accordance with our general definitions 1 ff. or 7 of rational knowledge (formula 2.7), it does not comply with formula (2.8), which is equivalent under AGM. There is a true proposition, namely the defeater, under which the subject changes her or his original belief. The subject is unaware of this defeater, which does not constitute a counterargument to attributing knowledge within our general definition. But it conflicts with formula (2.8). Thus, (2.8) denies knowledge where it should be attributed. Now, one could argue that this example shows our formalism to be defective. But this is not the whole story: We cannot achieve a more adequate representation of knowledge solely by changing the definition of knowledge. Moreover, the representation of the doxastic state by an AGM conditional is not adequate here.

Unfortunately, as stated above, within the AGM theory, a defeater always has to be disbelieved. That means AGM does not leave room for circumstances the subject is unaware of or does not believe, which could give rise to a belief change. Let p be the proposition that Tom stole the book, and q the fact that Mr. Grabit told his story. Unconditionally, the epistemic subject accepts that p , $|\sim p$. If she or he came to believe that q , however, according to the scenario, the opposite, $\neg p$, should be acceptable, which is expressed by the conditional $q |\sim \neg p$. But then, by ($|\sim 3$), it follows $|\sim q \rightarrow \neg p$, and by ($|\sim 1$), $|\sim p \rightarrow \neg q$, and thus, again by ($|\sim 1$), $|\sim \neg q$. It turns out that the subject is justified in accepting $\neg q$. But this is a significantly different story, since now $\neg q$ is a false belief on which the justification of p depends. And this is a standard case *which is in accordance with our definition of knowledge*.

The best approximation within our formalism can be found in figure 3. Proposition p and q are as before, f stands for the fact that the twin brother is a delusion of the father. Here, in the actual world on the right side of the figure, where the fact p , the defeaters q and defeater of the defeater f are true, we find $K(p \wedge (q \rightarrow f))$. Again, the attribution of knowledge is compatible with our concept of rational knowledge stemming from a weakly connected plausibility ordering. The deviations are due to the fact that the subject is not in an AGM-like state of belief. Instead of $q |\sim \neg p$, we find the dual conditional $\neg q |\sim p$ true in the actual world. Moreover, p , $\neg q$, and f turn out acceptable. The problem with the Tom Grabit example, a case of a defeated defeater, could not be handled from the doxastic side of the problem by using standard belief revision theory.

3.3. Conclusion and Outlook. We have formally defined a defeasibility analysis of knowledge, proved a representation theorem and uncovered its limitation. This lies in the presumption of an AGM-like dynamic doxastic state, which is not the

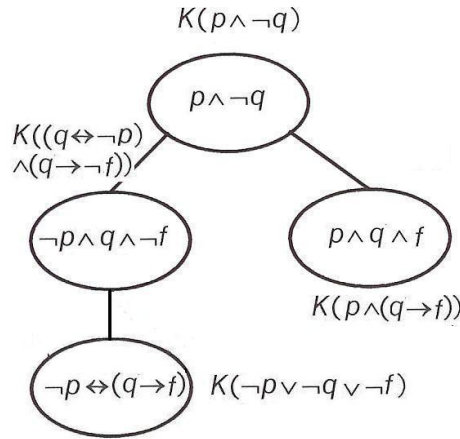


FIGURE 3. Tom Grabit

most general form of doxastic states compatible with human behaviour. Although it is not a complete theory of knowledge, the result presented here has several implications which might influence further research strategies. First, leaving our definition of rational knowledge untouched, we can read the form (2.8) used within the representation theorem as a *sufficient* condition for knowledge, although it might not be necessary. Second, one might interpret the conflict between default reasoning and AGM from an epistemological perspective: Although in many cases plausibility measures seem to be perfectly suited for modelling default reasoning [Boutilier/Becher 1995], there seem to be limitations from an epistemological standpoint.

Nevertheless, there are several reasons to consider the defeasibility analysis by rational knowledge in our sense as an interesting paradigm. First, it is the only theory we are able to comprehend completely through our representation theorems. Second, in many Gettier-like examples the theory produces exactly the results we normally expect and moreover provides a very simple fundamental mechanism as an explanation for *why* we have the sort of knowledge we have. Third, although much of criticism has been mounted against AGM in the past decades, it is still a powerful theory which links doxastic states to classic theory of economic rational choice. Together, these reasons motivate a strategy of research which retains the epistemic part of this approach and concentrates on the doxastic states with the goal of extracting plausibility measures from them in order to insert these into our definition of knowledge. This strategy is analogous to those in rational choice theory where, in case subjects fail to reveal choice functions compatible with the classic definition, one tries to discover hidden preferences in their actual behaviour as a kind of reconstructed rationality.

A fourth reason is that upholding a classic rationality paradigm for a while paves the way for interdisciplinary connections with fields like economics and neurosciences. But most of all, our simplified concept of knowledge avoids the usual *ad hoc* strategies of refining the definition of knowledge more and more until it turns out virtually unapplicable. It combines a purely internalistic approach, close to that

of Lehrer, with the basic insight, most prominently put forward by Williamson, that knowledge is not justified true belief plus x , which has become more prominent in recent years. Our representation theorem can both handle the knowledge operator as a primary concept and at the same time reduce it to plausibility orderings over possible worlds. Here lies the potential to unify epistemological and doxastic states without confusing them - which classic definitorial approaches are lacking.

A fifth and last reason is that even though subjects in general fail to be in AGM states or have rational (S4.3) knowledge, they could still be attributed rational knowledge on a *quotient space* of the possible world space, where possible worlds are partitioned into equivalence classes. This could be a direction for further research, namely to tackle the problem of the unaware defeated defeaters, quite similar to recent approaches in decision theory to represent decision under unawareness [Li 2006]. The main point here is that people are unaware of some facts because they are unable to tell some possible worlds apart. Studying the relation between unawareness belief states and (as we call them) rational states of acceptance in terms of embedding relations could be a fruitful proposal for future research.

REFERENCES

- [Bartelborth 1996] Bartelborth, Thomas, 1996, *Begründungsstrategien*, Akademie Verlag, Berlin.
- [Bartelborth 1999] Bartelborth, Thomas, 1999, "Coherence and Explanation," *Erkenntnis* **50**, 209-224.
- [Boutilier/Becher 1995] Boutilier, Craig and Becher, Verónica, 1995, "Abduction as Belief Revision," *Artificial Intelligence* **77(1)**, 43-94.
- [Gärdenfors 1992] Gärdenfors, Peter, 1992, "Belief Revision: An Introduction," in: *Belief Revision*, ed. by P. Gärdenfors, Cambridge University Press.
- [Halpern 1994] Halpern, Joseph Y. and Friedman, Nir, 1994, "A Knowledge-Based Framework for Belief Change. Part I: Foundations," in: *Theoretical Aspects of Reasoning about Knowledge*, Morgan Kaufmann.
- [Halpern 2001] Halpern, Joseph Y., 2001, "Plausibility Measures: A General Approach For Representing Uncertainty," *Proceedings of the 17th International Joint Conference on AI (IJCAI 2001)*, 1474-1483.
- [Hilpinen 1971] Hilpinen, Risto, 1971, "Knowledge and Justification," *Ajatus* **33**, 7-39.
- [Hughes / Cresswell] Hughes, Georges and Cresswell, Maxell, 1996, *A New Introduction to Modal Logic*. Routledge, London.
- [Katsuno 1988] Katsuno, H. and Mendelzon, A., 1988, "Propositional knowledge base revision and minimal change," *Artificial Intelligence*, 52, 263-294.
- [Lehrer 2000] Lehrer, Keith, 2000, *Theory of Knowledge*. Second Edition. Boulder.
- [Klein 1971] Klein, Peter, 1971, "A Proposed Definition of Propositional Knowledge," *Journal of Philosophy* **67**, 471-82.
- [Li 2006] Li, Jing, 2006, *Modeling Unawareness in Arbitrary State Space*. Working paper, University of Pennsylvania.
- [Plantinga 1993] Plantinga, A., 1993, *Warrant: The Current Debate*. New York.
- [Richter 1971] Richter, M.K., 1971, "Rational Choice," in: J.S. Chipman, M.K. Richter and H. Sonnenschein eds., *Preference, Utility and Demand*. Harcourt Brace Jovanovich, New York.
- [Rott 2002] Rott, Hans, 2002, "Lehrer's dynamic theory of knowledge," in: Erik Olsson, *The Epistemology of Keith Lehrer*, Philosophical Studies Series, Dordrecht: Kluwer.
- [Schoch 2000] Schoch, Daniel, 2000, *Explanatory Coherence*. *Synthese* **122/3**, 291-311.
- [Spohn 1988] Spohn, Wolfgang, 1988, "Ordinal Conditional Functions: A Dynamic Theory of Epistemic States," in: W.L. Harper, B. Skyrms (eds.), *Causation in Decision, Belief Change, and Statistics*, Kluwer, Dordrecht, 105-134.

- [Spohn 1999] Spohn, Wolfgang, 1999, "Ranking Functions, AGM Style," in: B. Hansson, S. Halldén, N.-E. Sahlin, W. Rabinowicz (eds.), *Internet Festschrift for Peter Gärdenfors*, Lund, s.: <http://www.lucs.lu.se/spinning/>
- [Spohn 2001] Spohn, Wolfgang, 2001, "Vier Begründungsbegriffe," in: Tomas Grundmann (Hrsg.), *Erkenntnistheorie*, Mentis, Paderborn.
- [Spohn 2002] Spohn, Wolfgang, 2002, "Lehrer Meets Ranking Theory," in: E. Olsson (ed.), *The Epistemology of Keith Lehrer*, Kluwer, Dordrecht.
- [Williamson 2000] Williamson, Timothy, 2000, *Knowledge and its Limits*. Oxford University Press.

PHILOSOPHICAL INSTITUTE, SAARLAND UNIVERSITY, D-66123 SAARBRÜCKEN/GERMANY
E-mail address: d.schoch@mx.uni-saarland.de
URL: <http://www.uni-saarland.de/philosophy/schoch>