Learning From Conditionals: Judy Benjamin's Other Problems

Abhaya Nayak,[1] **Maurice Pagnucco,**[1] **Norman Y. Foo** [1] and **Pavlos Peppas**[2]

Abstract. The Judy Benjamin problem, a problem in probability kinematics, deals with the updating of belief functions where the new evidence itself gives the new *conditional* probability of some sentence. Similar problems arise in deterministic (as opposed to probabilistic) epistemology. We examine one such problem and propose a solution.

1 Introduction: The Judy Benjamin Problem

In Bayesian epistemology, when an epistemic agent receives evidence e, she revises her prior probability function P (representing her current belief state) to her posterior probability function (representing her revised belief state) P_e which is given by

$$P_e(h) = P(h|e) = \begin{cases} \frac{P(h \wedge e)}{P(e)} & \text{if } 0 < P(e) \\ \text{undefined} & \text{otherwise} \end{cases}$$

for any sentence h. In this orthodox Bayesian approach, the evidence e is deemed to be certain and incorrigible. The function P_e and all its future embodiments (via conditionalization) continue to assign probability 1 to e. This updating process may be viewed as follows. The agent's prior is P. For some reason or other, the agent decides to change her belief state. The only thing the agent knows about her new belief state P' is that $P'(e) = 1$. In that case, the rational way to complete P' is to equate it with P_e.

Jeffrey's account of Probability Kinematics [7] generalizes the orthodox Bayesian account to deal with corrigible and uncertain evidence. In its general form, the agent learns of the new probability function $P_{\{e_i\}}$ that $P_{\{e_i\}}(e_i) = p_i$ where $\{e_1, \ldots, e_i, \ldots, e_n\}$ is an n-celled partition over worlds and $\Sigma_{i=1}^{n} p_i = 1$. According to Jeffrey's probability kinematics then, $P_{\{e_i\}}$ is to be completed as follows: the new probability $P_{\{e_i\}}(h)$ of any sentence/proposition h is given by $\Sigma_{i=1}^{n} P_{e_i}(h) \cdot p_i$. Of special interest is the case $n = 2$: let the partition in this case be $\{e, \neg e\}$. If we let the new probability $P'(e)$ of e be p, then $P'(\neg e) = 1 - p$. The new probability $P'(h)$ of any sentence h is given by: $P'(h) = P_e(h) \cdot P'(e) + P_{\neg e}(h) \cdot P'(\neg e)$. This is the version of probability kinematics best known in the literature.

Despite the generality of probability kinematics, the nature of the evidence is fairly well constrained: it has to be such that it can be viewed as a probability function, albeit defined over a coarser event space (relative to the prior). The evidence does not always fit into this procrustean bed in an obvious manner. For instance, the evidence may be:

> All I know about about my new probability function P' is that: $P'(h|e) = p$.

[1] Knowledge Systems Group, Basser Dept. of Computer Science, Madsen Building, F09, University of Sydney, NSW 2006, Australia. Email: {abhaya | morri | norman}@cs.su.oz.au.
[2] School of MCPE, Macquarie University, North Ryde, NSW 2109, Australia. Email: pavlos@macadam.mpce.mq.edu.au.

ECAI 96. *12th European Conference on Artificial Intelligence*
Edited by W. Wahlster
Published in 1996 by John Wiley & Sons, Ltd.

In this case it follows that $P'(\neg h|e) = 1 - p$. So, obviously $P'(h) : P'(\neg h) = p : (1 - p)$. Surely there are many ways to go from P to P' such that $P'(h) : P'(\neg h) = p : (1 - p)$. The Judy Benjamin Problem is the problem of choosing the right revision from P to P' in this case. Bas van Fraassen (with colleagues) [14, 15] has shown that natural constraints are not sufficiently strong to eliminate enough of these possibilities in order to determine a unique solution.

The nomenclature of this problem is due to its dramatization by van Fraassen. We quote from [16, p. 342] at some length, since we will use variations of the scenario later.

> [The Judy Benjamin Problem] derives from the movie *Private Benjamin*, in which Goldie Hawn, playing the title character, joins the army. She and her platoon, participating in war games on the side of the 'Blue Army', are dropped in the wilderness, to scout the opposition ('Red Army'). They are soon lost. Leaving the movie script now, suppose the area to be divided in two halves, Blue and Red territory, while each territory is divided into Headquarters Company area and Second Company Area. They were dropped more or less at the center, and feel it is equally likely that they are now located in one area as in another. [See the map in Figure (1).] They have some difficulty contacting their own HQ by radio, but finally succeed and describe what they can see around them. After a while, the officer at the HQ radios: 'I can't be sure where you are. If you are in Red Territory, the odds are 3:1 that you are in HQ Coy. area. . . ' At this point the radio gives out. As in the movie, the platoon goes on to capture enemy headquarters.
> We must now consider how Judy Benjamin should adjust her opinions, if she accepts this radio message as the correct and sole constraint to impose. . . .

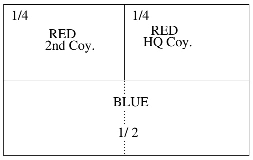

Figure 1. The Judy Benjamin Problem

The Judy Benjamin Problem points out a limitation of probability kinematics. Now, probability kinematics deals with belief change in probabilistic epistemology. There are analogous accounts of belief change in deterministic epistemology. In particular, Spohn's account of Ordinal Conditionalization Functions [13] is the analogue of probability kinematics in deterministic epistemology when an ordinal measure is imposed on the beliefs of an agent. Similarly, the account of entrenchment kinematics [8, 9, 10, 11] is its analogue in deterministic epistemology when a relational measure is imposed on the beliefs of an agent. One would expect that these accounts of belief change will face problems analogous to the original Judy Benjamin Problem. In this paper we will formulate two Judy Benjamin like problems against the backdrop of entrenchment kinematics. Due to space limitations we will examine only one of these problems, with the understanding that the other problem is amenable to analogous treatment. Unlike van Fraassen, we will argue that these problems admit intuitive solutions.

2 Two Judy Benjamin Like Problems

Orthodox conditionalization $P(\cdot|e)$ of the probability function $P(\cdot)$ is undefined if $P(e) = 0$. The Bayesian framework is not designed to deal with evidence that is currently judged to be impossible. Accordingly, Jeffrey Conditionalization is undefined for evidence $\{e_i\}$, with new probabilities $\{p_i\}$ if the old probability $P(e_i)$ for some i is 0 (but its new probability $p_i \neq 0$). Hence the Bayesian framework is not designed to address the corresponding Judy Benjamin problem, namely, when the new evidence is $P'(h|e) = p \neq 0$ but $P(e) = 0$.

However, the AGM paradigm of belief change [1, 3], does not suffer from this limitation. The AGM paradigm allows two ways of learning – belief expansion and belief revision. Expansion is the analogue of Bayesian updating: this operation determines the new belief state when the current belief state is expanded by some information that is consistent with the current beliefs. In belief revision proper, on the other hand, an agent revises her belief corpus with respect to some information that she hitherto considered impossible. Hence the new evidence, $P'(h|e) = p \neq 0$ while $P(e) = 0$, although not a serious problem for the Bayesian epistemology, has a non-probabilistic analogue which is a possible threat to the AGM paradigm. The problem in question is captured by the following scenario:

Scenario 1

Judy Benjamin's platoon is lost in the wilderness. She believes that she is in the Red territory, but she is not sure whether she is in the 2nd Coy. area or the HQ area. She calls her own HQ and gives a description of her surroundings. She is told, "You are in the 2nd Coy. area, given that you are in the Blue territory". Then radio contact is lost. What does Judy Benjamin learn?

Beside the problem with learning what is currently judged to be impossible, the Bayesian framework suffers from another serious limitation. It does not provide the agent with an opportunity to unlearn what has been fully learnt. If a proposition has probability 1, then this value cannot reduce by conditionalization. This problem also migrates to probability kinematics. Hence the corresponding Judy Benjamin Problem, namely, when the evidence is $P'(h|e) = p \neq 1$ but $P(e) = 1$, need not be addressed in a Bayesian framework either.

Again, the AGM paradigm provides a potent framework to address the issue in question. Apart from the two learning operations – expansion and revision – the AGM system also employs an unlearning operation, namely belief contraction. In the case of non-vacuous belief contraction the agent unlearns what she had fully learnt. Accordingly, the AGM paradigm should also address the following problem:

Scenario 2

Judy Benjamin's platoon is lost in the wilderness. She believes that she is in the Blue territory, but she is not sure whether she is in the 2nd Coy. area or the HQ area. She calls her own HQ and gives a description of her surroundings. She is told, "Well, if we don't assume that you are in the Blue territory, then you are in the 2nd Coy. area". Then radio contact is lost. What does Judy Benjamin learn?

The two scenarios outlined above raise the question of revising a given belief state by a (possibly counterfactual) conditional. If we allow the Ramsey-test account of conditionals [3, ch. 7], this problem boils down to rationally moving from the current belief state S to a new belief state S' such that if the agent were to learn the antecedent (of the conditional representing the evidence) in S', she would be obliged to believe the consequent. Hence, the solution will involve an account of iterated belief change. Since the original AGM account of belief change does not concern iterated belief change, we will employ entrenchment kinematics which is a generalization of the AGM paradigm. As we pointed out earlier, we will confine the discussion to the first scenario only. This discussion is deferred until we give an outline of entrenchment kinematics in the next section.

3 Entrenchment Kinematics: A Quick Review

The AGM account of belief revision uses a relational measure, namely that of epistemic entrenchment, over the body of beliefs to determine what current beliefs to give up if (and when) the accepted evidence conflicts with the current beliefs. However it does not provide any guideline as to the entrenchment of beliefs in the revised belief state. Hence the AGM paradigm proper cannot guide iterated belief change.

Entrenchment kinematics, a generalization of the AGM account, alleviates this problem. The account of entrenchment kinematics offered in [10, 11] is closer in spirit to the AGM account in that the evidence is represented as a single sentence. On the other hand, the account offered in [8, 9] is more general: the evidence is represented here as a prioritised body of sentences, the prioritization relation in question satisfying the same constraints as an entrenchment relation. It is this latter (more general) framework that concerns us here.

The constraints on an entrenchment relation allow one to extract from it the body of beliefs (respectively evidence) it prioritizes. The general account of entrenchment kinematics takes advantage of this fact and represents the current belief state, the input information (evidence) and the resultant belief state as epistemic entrenchment relations. Since the result of a revision is again an entrenchment relation, iterated belief change is easily handled in this framework.

The entrenchment relation used in entrenchment kinematics (EE relation) is a slight variant of the standard entrenchment relation (SEE relation) used in the AGM literature [3, 4]. Let us assume a finite propositional language $\mathcal{L}$ with the Boolean connectives $\neg, \vee, \wedge, \rightarrow$ and $\leftrightarrow$. $\top$ (Truth) and $\bot$ (Falsum) are two distinguished members of $\mathcal{L}$. We also assume that its logic is governed by a standard logical consequence relation $\vdash$ satisfying supraclassicality, compactness and the deduction theorem. (Occasionally we will use the logical consequence operation denoted by Cn and defined as: $Cn(X) = \{y|X \vdash y\}$.) An EE relation $\preceq$ is defined as any relation over $\mathcal{L}$ that satisfies the following constraints:

$(EE1)$	If $x \preceq y$ and $y \preceq z$ then $x \preceq z$	(transitivity)
$(EE2)$	If $x \vdash y$ then $x \preceq y$	(dominance)
$(EE3)$	For any x and y, $x \preceq x \wedge y$ or $y \preceq x \wedge y$	(conjunctiveness)
$(EE4)$	Given that $\bot \prec z$ for some z, if $y \preceq x$ for all y, then $\vdash x$	(maximality).

We abbreviate $(x \preceq y) \wedge (y \preceq x)$ as $x \equiv y$ and $(x \preceq y) \wedge (y \not\preceq x)$ as $x \prec y$. An EE relation may represent the belief state of an agent or the evidence that she has decided to accept. If an EE relation represents a belief state, then the belief set (the beliefs of the agent, usually denoted by K) can be obtained as the epistemic content of this relation:

Definition 1 Epistemic Content

$$EC(\preceq) = \begin{cases} \{x|\bot \prec x\} & \textit{if } \bot \textit{ is not } \preceq\textit{-maximal} \\ K_\bot & \textit{otherwise} \end{cases}$$

Accordingly, $\mathcal{L} \setminus EC(\preceq)$, represents the non-beliefs of the agent. In this context, the sentence $x \preceq y$ may be interpreted as saying that it is not easier to give up the belief y than the belief x. On the other hand, when an EE relation represents the evidence, its epistemic content represents the set of evidential sentences.

In the AGM framework, the revision of a belief set K (with accompanying EE relation $\preceq_K$) by the sentential evidence x is determined by the following result [3, 4]:[3]

(EE$*$) $\quad y \in K^*_x$ iff either $x \rightarrow \neg y \prec_K x \rightarrow y$ or $x \vdash y$.

Entrenchment kinematics may be viewed as offering a generalization of this construction (for details, see [9, § 5.1]). Given a prior EE relation $\preceq_K$ and an evidential EE relation $\preceq_E$, the posterior EE relation $\preceq_{K*E}$ is defined as follows:

Definition 2 *For any sentences x and y of a finite language $\mathcal{L}$,*

$$x \preceq_{K*E} y \text{ iff} \begin{cases} \text{either} & x \prec_E y \\ \text{or} & x \equiv_E y \text{ and } (x^\#_E \rightarrow x) \preceq_K (x^\#_E \rightarrow y) \end{cases}$$

where, for any sentence x, the sentence $x^\#_E$ represents the sentence $\bigwedge(\{x' : x \prec_E x'\})$, with the proviso that $\bigwedge(\emptyset) = \top$. The intuitive motivation behind entrenchment kinematics is primarily semantic. This arises from the fusion of Spohn's account of Ordinal Conditional Functions [13] and Grove's modelling of belief change via systems of spheres [5]. We refer the reader to [9] for details. Non-semantic motivations for this construction can be found in [8].

4 Proposed Solution

Let us now examine the first scenario against the backdrop of entrenchment kinematics. We will use the following abbreviations:

x: Judy's platoon is in the Blue territory
$\neg x$: Judy's platoon is in the Red territory
y: Judy's platoon is in the 2nd Coy. area
$\neg y$: Judy's platoon is in the HQ Coy. area

First we address this problem from the point of view of epistemic entrenchment. Later on we examine it from a semantic angle.

We will use Rott's [12] convention to cogently represent entrenchment relations. An EE relation $\preceq$ will be represented as $\bot \prec X_1 \prec X_2 \prec \ldots \prec X_n \prec \top$ where each X_j is a set of sentences and for each $x_{i<j} \in X_i$ it holds that $x_i \notin Cn(\bigcup_{k=j}^n X_k)$. This represents the following EE relation: $x \prec y$ iff there exists X_m such that $y \in Cn(\bigcup_{i=m}^n X_i)$ but $x \notin Cn(\bigcup_{i=m}^n X_i)$. When X_i is a singleton set we drop the brackets if it does not cause any confusion.

Entrenchment Approach

Judy's current belief set (relevantly pruned) may be equated with $Cn(\{\neg x\})$. Furthermore, since she is not sure whether she is in the 2nd Coy. area or the HQ area, we can assume that the beliefs $\neg x \vee y$ and $\neg x \vee \neg y$ are equally entrenched. Hence her current belief state may be represented by the entrenchment relation $\bot \prec \neg x \prec \top$. This is actually the coarsest possible EE relation with epistemic content $Cn(\{\neg x\})$ that allows both $\neg x \vee y$ and $\neg x \vee \neg y$ to be equally entrenched. However the crux of the problem is the representation of the evidence, to which we now turn.

The sole evidence that Judy accepts is that she is in the 2nd Coy. area, given that she is in the Blue territory. This may be captured by the conditional $x \hookrightarrow y$, meaning that if she were to learn that x, then she would believe that y. Note that she is not being told of a fact about her current belief state; she is not being told that if she were to accommodate x into her current belief state, she would believe y. On the contrary, she is being told that she should revise her belief state in a rational fashion such that if, in this new belief state she were to learn x, she would believe y. In short, assuming that her current belief set is K and $*$ her favorite belief revision operation, unless already $y \in K^*_x$, she should revise K to K' in a rational manner such that $y \in K'^*_x$. Let us denote the EE relation that would accompany K' by $\preceq'$ (and its strict part by $\prec'$). Then it follows from (EE$*$) that $y \in K'^*_x$ iff either (i) $x \rightarrow \neg y \prec' x \rightarrow y$ or (ii) $x \vdash y$. In other words, the evidence $x \hookrightarrow y$ should be represented in such a manner that as a result of accepting it, either (i) $x \rightarrow \neg y$ becomes less entrenched than $x \rightarrow y$ or (ii) $x \vdash y$ comes to hold. However, whether $x \vdash y$ holds or not is a logical fact and has nothing to do with Judy's beliefs. Hence the only relevant constraint here is (i). Assume that Judy revises her prior EE relation $\preceq$ to the posterior $\preceq'$ via entrenchment kinematics (and represents $x \hookrightarrow y$ as an evidential EE relation $\preceq_e$). Applying Definition 2 to constraint (i) we conclude that either (a) $x \rightarrow \neg y \prec_e x \rightarrow y$ or (b) both $x \rightarrow \neg y \equiv_e x \rightarrow y$ and $(x \rightarrow \neg y)^\#_e \rightarrow (x \rightarrow \neg y) \prec (x \rightarrow \neg y)^\#_e \rightarrow (x \rightarrow y)$. The following argument shows that (b) is not a viable option, so (a) holds. Note that (b) refers to both the prior EE relation and the evidential EE relation. Considering that Judy's prior is given by the EE relation $\bot \prec \neg x \prec \top$, she currently considers all nontautological consequences of $\neg x$ to be equally entrenched, and places them just below the logical theorems. Hence, in order that (b) might hold, it must be the case that $\vdash (x \rightarrow \neg y)^\#_e \rightarrow (x \rightarrow y)$, i.e., $(x \rightarrow \neg y)^\#_e \vdash x \rightarrow y$. But this conflicts with the assumption that $x \rightarrow \neg y \equiv_e x \rightarrow y$ given the construction of $(x \rightarrow \neg y)^\#_e$. So the only constraint on $\preceq_e$ is that $x \rightarrow \neg y \prec_e x \rightarrow y$.

Now, if $x \vdash y$ then the evidence would be quite trivial. So let us grant that $x \nvdash y$. From the general constraints on EE relations, it follows then that $x \rightarrow y \prec_e \top$. So the total constraints on $\preceq_e$ are given by $x \rightarrow \neg y \prec_e x \rightarrow y \prec_e \top$. Note, this constraint is equivalent to $\neg x \prec_e x \rightarrow y \prec_e \top$.

In order to obtain a full entrenchment relation from these constraints, we will use the principle of "coarsest relevant partitioning" (see [14]). An EE relation $\preceq$ partitions the language into equivalence classes under $\preceq$. The condition $\neg x \prec_e x \rightarrow y \prec_e \top$ implies that $x \rightarrow y$ is in the epistemic content of $\preceq_e$. But we are under no compulsion to include $\neg x$ in the epistemic content of $\preceq_e$. The following two EE relations are the coarsest among those satisfying the constraint $\neg x \prec_e x \rightarrow y \prec_e \top$:

(1) $\bot \prec_e x \rightarrow y \prec_e \top$, $\qquad$ (2) $\bot \prec_e x \wedge y \prec_e \top$.

Relation (1) is obtained by essentially declaring that a sentence is in the epistemic content of $\prec_e$ just in case it is a consequence of $x \rightarrow y$ whereas relation (2) is obtained by declaring that a sentence is not in the epistemic content of $\prec_e$ when $x \rightarrow \neg y$ is a consequence of it.

Assuming that we should restrict our choice of representing the evidential EE relation to (1) and (2), how does one determine which is the right choice? This can be done only by investigating the result of adopting these respective representations. Of these two, the second one is stronger since $x \wedge y \vdash x \rightarrow y$. So it might prove beneficial to examine whether under some circumstances it is appropriate to adopt (2) as the right representation of the evidential EE relation. If there are no such circumstances, then (1) is always the right representation.

A crucial consequence of adopting (2) as the representation in question is that the agent comes to believe $x \wedge y$ after accepting this

[3] This actually follows from their account of belief contraction given the Levi Identity. The definition we are alluding to is : $y \in K^-_x$ iff both $y \in K$ and either $x \prec_K x \vee y$ or $\vdash x$.

conditional input. The sole circumstances in which we would expect the evidence "if x is (were) the case, then y is (would be) the case" to warrant belief in both x and y seems to be when the agent already believes that x.[4] Hence, a case can be made for the following principle:

Represent $x \hookrightarrow y$ as
(1) $\bot \prec_e x \rightarrow y \prec_e \top$, when $x \notin K$, and
(2) $\bot \prec_e x \wedge y \prec_e \top$, when $x \in K$.

Another completion of the constraint $\neg x \prec_e x \rightarrow y \prec_e \top$ is:
(3) $\bot \prec_e \neg x \prec_e x \rightarrow y \prec_e \top$.
In fact, (3) might look a more natural completion of the constraint in question than either of (1) and (2). However, this is not one of the "coarsest partitions" since it introduces one more equivalence class of sentences than the former two representations. Furthermore, a necessary consequence of adopting this representation of the evidence is that $\neg x$ becomes a belief as a result of accepting the conditional evidence in question. As we pointed out earlier, that may not be bad if the agent gives enough credence to conversational implicature. The matter of conversational implicature, however, is a matter of pragmatics, not of semantics. We will assume that the agent might use this representation only when, due to contextual pressure, the conversational implicature overrides the dictates of rationality (namely, minimal mutilation).

Judy's current belief state is the EE relation $\bot \prec \neg x \prec \top$. Since she believes $\neg x$, her evidence is (1) $\bot \prec_e x \rightarrow y \prec_e \top$. If we now follow the entrenchment revision process outlined in section 2 above, we will get as the result the EE relation $\bot \prec \neg x \prec x \rightarrow y \prec \top$. The epistemic content of this resultant belief state is $Cn(\neg x)$ – exactly the epistemic content of the earlier state. So, this evidence does not actually result in changing the beliefs *per se*. The relative entrenchment among beliefs change however. As a result of accepting evidence $x \hookrightarrow y$, Judy now believes the sentence $\neg x \vee y$ more firmly than the sentence $\neg x \vee \neg y$; she believed both these sentences equally firmly before accepting the evidence $x \hookrightarrow y$.

Semantic Analysis

Let us now examine the problem from the semantic view point. The agent's current belief state is represented by a Grove system of spheres (SOS) [5] i.e., a set of worlds ordered by a connected, transitive and reflexive relation (total preorder) $\sqsubseteq$. If one views the set $\{\omega' | \omega' \sqsubseteq \omega\}$ for some world ω as a sphere, then $\sqsubseteq$ generates a nested system of spheres. That an SOS semantically represents an EE relation is clear from the following connection: $x \preceq y$, for contingent $x,\ y \in \mathcal{L}$, just in case $S_{\neg x} \subseteq S_{\neg y}$ where S_α, for any sentence α, is the smallest sphere intersecting the set of α-worlds $[\alpha] = \{\omega | \alpha \text{ holds in } \omega\}$ [5, 3]. Nayak [9] offers an account of how a given SOS can be revised by an evidential SOS. Let us say that the initial belief state is represented by an SOS $\mathcal{S}_1$, the evidence $x \hookrightarrow y$ by the SOS $\mathcal{S}_e$ and the result by $\mathcal{S}_2$. All we know about $\mathcal{S}_2$ is that if it was to be revised by x, then y would be believed. This fact may be represented by the constraint $S_{x \wedge y} \subset S_{x \wedge \neg y}$ on $\mathcal{S}_2$. (This constraint on $\mathcal{S}_2$ is equivalent to $x \rightarrow \neg y \prec x \rightarrow y$ and can be achieved by imposing the constraint $S_{x \wedge y} \subset S_{x \wedge \neg y}$ on $\mathcal{S}_e$.) This would be the case exactly when there is at least one $x \wedge y$ world (strictly) closer than every $x \wedge \neg y$ world. By relevant coarsening, we assume that all worlds in $[x \wedge y]$ are closer to the center than any world in $[x \wedge \neg y]$. The relevant partition on the space of worlds seems to be this: $\{[x \wedge y], [x \wedge \neg y], [\neg x \wedge y], [\neg x \wedge \neg y]\}$. Furthermore, since all we know is that worlds in $[x \wedge y]$ are closer to the center than the worlds in $[x \wedge \neg y]$, there is actually not enough reason to separate the worlds in $[\neg x \wedge y]$ from the worlds in $[\neg x \wedge \neg y]$; the worlds in $[\neg x]$ may be viewed as constituting a single cell in the partition. But even this three celled partition,

$(0')$ $\{[x \wedge y], [x \wedge \neg y], [\neg x]\}$

is not the *coarsest* relevant partition; any of the following two-celled partitions will do as long as the constraint in question is satisfied:

$(1')$ $\{[x \wedge y], [\neg x \vee \neg y]\}$, $\qquad$ $(2')$ $\{[\neg x \vee y], [x \wedge \neg y]\}$.

Partition $(1')$ is obtained from $(0')$ by merging the two cells $[x \wedge \neg y]$ and $[\neg x]$ together; on the other hand, $(2')$ is obtained from $(0')$ by merging the two cells $[x \wedge y]$ and $[\neg x]$ together. Since we are constrained to viewing the worlds in $[x \wedge y]$ to be closer to the center than the worlds in $[\neg x]$, it follows then that, if option $(1')$ is chosen, the worlds in $[x \wedge y]$ are viewed to be closer to the center than the worlds in $[\neg x \vee \neg y]$. Similarly, in case $(2')$, the worlds in $[\neg x \vee y]$ are viewed to be closer to the center than the worlds in $[x \wedge \neg y]$. Accordingly, the two SOS's that are plausible representations of the evidence $x \hookrightarrow y$ are the ones shown in Figure 2.

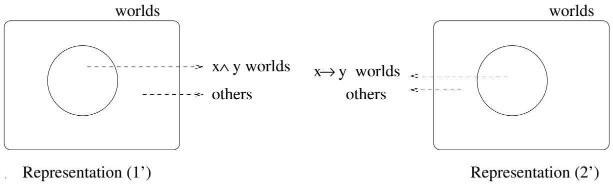

Figure 2. Coarsest representation of the evidence $x \hookrightarrow y$

It is easily verified that SOS $(1')$ and $(2')$ correspond, respectively, to EE relations (1) and (2). So, given our earlier discussion, $(1')$ is the appropriate evidential SOS when the agent already believes that x; otherwise $(2')$ is the appropriate evidential SOS. In particular, $2'$ is the correct evidential representation here.

However, one may find the ground on which we decided to merge the worlds in $[\neg x]$ with others (thus generating $(1')$ and $(2')$) rather tenuous, and want to reserve a separate cell for $[\neg x]$ itself. In that case we have a three celled partition and, given our constraint, there are three possibilities, as pictured in Figure 3.

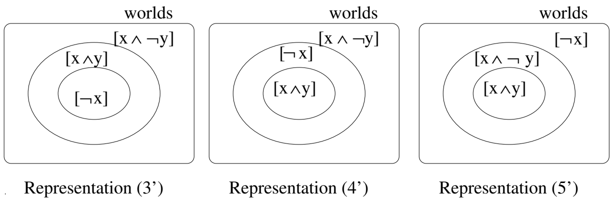

Figure 3. Representation of the evidence $x \hookrightarrow y$: not coarsest

As one would expect, $(3')$ of course corresponds to the EE relation (3), and is of pragmatic importance when the agent gives enough credence to the speaker so as to adopt its presuppositions. Although $(4')$ and $(5')$ correspond to certain EE relations, it is difficult to see under what circumstances the evidence should be so represented.

Going back to Judy Benjamin's problem, we assumed that the representation of her current beliefs in terms of EE relations is $\bot \prec \neg x \prec \top$. In terms of SOSs; this may be viewed as a 2-celled SOS, the x-worlds surrounding the $\neg x$-worlds. If we superimpose the evidential SOS $2'$ on the prior SOS and do the revision as suggested

[4] Unless, of course, the agent attaches enough weight to what the speaker conversationally implicates, given the speaker uses the counterfactual "if x were the case, then y would be the case". We will consider this case later.

in [9, § 4.1]), we get a 3-celled SOS centered (again!) on $[\neg x]$. This is pictured in Figure 4. As before, we noticed that the evidence $x \hookrightarrow y$

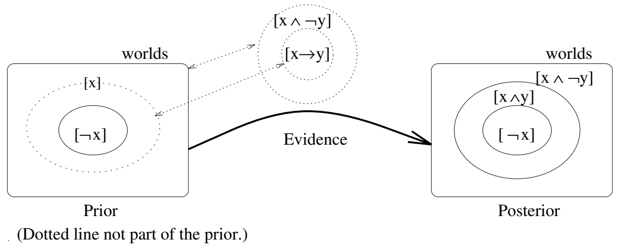

Figure 4. Revision of the prior SOS by the evidence $x \hookrightarrow y$

does not change the *beliefs* of Judy Benjamin; $[K]$ is still $[\neg x]$. However, whereas the prior does not distinguish among the outer worlds (x-worlds), the posterior assumes that the $x \wedge y$-worlds are closer than the $x \wedge \neg y$ worlds. The import of this becomes clear if we look at this fact from an EE point of view, as pointed out earlier. After accepting $x \hookrightarrow y$, the belief $x \rightarrow y$ becomes more entrenched than $x \rightarrow \neg y$ which ensures that if x were to be believed at this point, y would be believed as well.

5 Discussion

In the belief revision literature, there are two other works [2, 6] that consider revision by conditionals. Of these two, the work of Boutilier and Goldszmidt [2] lends itself to easier comparison with our approach. As in our approach, they recognise that the solution to the problem of accepting $x \hookrightarrow y$ lies in the rational shift to a possibly new belief state such that acceptance of x would end up in believing y. The semantic representation of a belief state is also identical in both approaches. Both approaches agree that the rational shift in question is based on "minimal" change. The difference lies in the interpretation of what amounts to a minimal change in the current belief state. In particular, they suggest that the change in question lies in minimally shifting, if necessary, the $\sqsubseteq$-minimal $x \wedge y$ worlds down such that they become closer to the center than the $\sqsubseteq$-minimal $x \wedge \neg y$ worlds. This ensures that if x were to be accepted, y would be believed. When applied to our example (see Figure 4), we get the same resultant belief state. However, in general, the two approaches give different results.

Here is a possible objection to our approach. Revising the same prior EE relation, respectively, by the EE relation $\bot \prec \neg x \vee y \prec \top$ as suggested here (also see [9]) and by the sentence $\neg x \vee y$ as prescribed in [10] results in identical resultant states. On the other hand, we have been pressing the point that the evidence $x \hookrightarrow y$ may be represented as the EE relation $\bot \prec \neg x \vee y \prec \top$ provided x is not already believed by the agent. Hence, a consequence of this approach is that when the agent does not believe x, the evidence $x \hookrightarrow y$ collapses to the truth functional sentence $x \rightarrow y$! But it is almost an article of faith in the literature that the conditional $\hookrightarrow$ is not truth functional.

We do not consider this objection to be very serious. Firstly, the two conditionals, $\rightarrow$ and $\hookrightarrow$ are traditionally supposed to have different belief conditions: believing $x \rightarrow y$ and believing $x \hookrightarrow y$ are two different things. But our concern here is not with beliefs *per se*, but with evidence. Besides, we are not claiming that when x is not believed, $x \rightarrow y$ and $x \hookrightarrow y$ *mean* the same thing; all we are claiming is that when x is not believed, accepting $x \hookrightarrow y$ as evidence has the very same effect as accepting $x \rightarrow y$ as the evidence.

The above point becomes clearer if we consider the evidence $x \hookrightarrow y$ when x itself is a belief. If we take the trouble to analyze this case as we did above for the other case, it would turn out that accepting $x \hookrightarrow y$ as evidence has the same effect as accepting $x \wedge y$. We are not conflating $\hookrightarrow$ with $\wedge$ here; it is indeed difficult to confuse $\hookrightarrow$ and $\wedge$ since the latter is not a conditional at all. However, the fact of the matter is that when x is a belief, accepting $x \hookrightarrow y$ compels us to believe y (unless one takes the counterfactual presupposition seriously enough to adopt $\neg x$ as a belief on the basis of $x \hookrightarrow y$), and the principle of conservation entails that we revise the current belief state by $x \wedge y$.

We conclude with a brief remark regarding the connection of this work with nested conditionals. The evidential sentences in these Judy Benjamin like problems may be viewed as conditionals. Thus our account gives a recipe for handling conditional evidence. Viewed from another angle, this account gives the assertability conditions for nested conditionals such as $(x \hookrightarrow y) \hookrightarrow z$. On the other hand, the account of entrenchment kinematics discussed in [10, 11] gives the assertability conditions for nested conditionals such as $x \hookrightarrow (y \hookrightarrow z)$. Thus these two accounts supplement each other in a very nice manner. When fully generalized, we will obtain a complete account of nested conditionals. This problem is currently being pursued.

Acknowledgements

The authors would like to thank Craig Boutilier and David Makinson for insightful discussions and suggestions.

REFERENCES

[1] Alchourrón, C. E., P. Gärdenfors, and D. Makinson, "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions", *Journal of Symbolic Logic,* vol. 50, pp.510-530, 1985.

[2] Boutilier, B. and M. Goldszmidt, "Revision By Conditional Beliefs", in *Proceedings of the Eleventh National Conference on Artificial Intelligence*, Washington DC, MIT Press/AAAI Press, pp. 649-654, 1993.

[3] Gärdenfors, P. *Knowledge in Flux. Modeling the Dynamics of Epistemic States,* Bradford/MIT, Cambridge (Massachussettes), 1988.

[4] Gärdenfors, P. and D. Makinson, "Revision of Knowledge Systems using Epistemic Entrenchment," in *Proceedings of the Second Conference on Theoretical Aspects of Reasoning about Knowledge,* M. Vardi ed., Los Altos, CA: Morgan Kaufmann. pp. 83-95, 1988.

[5] Grove, Adam "Two Modeling for Theory Change," *Journal of Philosophical Logic,* vol. 17, pp. 157-170, 1988.

[6] Hansson, Sven Ove "In Defense of the Ramsey Test," *Journal of Philosophy,* vol. 89, pp. 522-540, 1992.

[7] Jeffrey, R. C. *The Logic of Decision,* Chicago University Press, 1983.

[8] Nayak, A. C., P. Nelson and H. Polansky, "Belief Change as Change in Epistemic Entrenchment", *Synthese,* forthcoming.

[9] Nayak, A. C. "Iterated Belief Change Based on Epistemic Entrenchment", *Erkenntnis,* vol. 41, pp. 353-390, 1994.

[10] Nayak, A., N. Foo, M. Pagnucco and A. Sattar, "Entrenchment Kinematics 101", *Proc. of the 7th Australian Joint Conference on AI,* C. Zhang, *et al.* (eds.), World Scientific, pp. 157-164, 1994.

[11] Nayak, A., N. Foo, M. Pagnucco and A. Sattar, "Changing Conditional Beliefs Unconditionally", *Proc. of the 6th Conf. on Theoretical Aspects Of Rationality And Knowledge*, Morgan Kaufmann, pp. 119-135, 1996.

[12] Rott, Hans "A Nonmonotonic Conditional Logic for Belief Revision I", in A. Fuhrmann and M. Mourreau, eds., *The Logic of Theory Change,* Lecture Notes in AI 465, Springer Verlag, pp. 135-183, 1991.

[13] Spohn, W. "Ordinal Conditional Functions: A Dynamic Theory of Epistemic States", in W. L. Harper and B. Skyrms (eds), *Causation in Decision, Belief Change and Statistics,* vol.2, pp. 105-134, 1988.

[14] van Fraassen, B. "Discussion: A Problem for Relative Information Minimizers", *Brit. J. for the Phil. of Science*, vol. 32, pp. 375-379, 1981.

[15] van Fraassen, B. *et al.* "A Problem for Relative Information Minimizers in Probability Kinematics, Continued", *Brit. J. for the Phil. of Science*, vol. 37, pp. 453-475, 1986.

[16] van Fraassen, B. *Laws and Symmetry*, Clarendon Press, 1989.