Reason Maintenance and the Ramsey Test ⋆

Haythem O. Ismail

German University in Cairo
Department of Computer Science
haythem.ismail@guc.edu.eg

Abstract. The Ramsey test provides an intuitive link between conditionals and belief revision. How easy is it to incorporate a Ramsey-account of conditionals in a reason maintenance system? In this paper, it is shown that this is indeed possible, within a relevance-logical framework. In addition, it is shown that independently motivated requirements on reason maintenance systems allow us to gracefully circumvent Gärdenfors's triviality result.

1 Introduction

Frank Ramsey's so called "Ramsey test" provides an intuitive link between conditionals (sentences of the form 'If P then Q') and belief change. The test grounds the plausibility of conditionals in a process of belief change. In an often quoted excerpt from [1], Robert Stalnaker gives a procedural interpretation of the Ramsey test:

> First, add the antecedent (hypothetically) to your stock of beliefs; second, make whatever adjustments are required to maintain consistency (without modifying the hypothetical belief in the antecedent); finally, consider whether or not the consequent is then true.

This procedure is particularly appealing for researchers in artificial intelligence (AI), who have long been interested in reasoning about conditionals [2, 3, for example]. What makes it even more appealing is that, in its crucial second step, it provides a characterization of conditionals based on the (familiar to AI) concept of belief revision. The AI and philosophical literature on belief revision seem to have originated from different concerns. On one hand, the AI researchers were primarily motivated by the implementation issues of supplementing a general reasoning system with the facility to maintain consistency and revise its beliefs. This gave rise to what are known as "reason maintenance" (or "truth maintenance") systems. [4, 5, for example]. On the other hand, the philosophers were keen to uncover a tight set of *rationality postulates* that govern the principles whereby a logical theory is to be revised [6]. These two attitudes have

⋆ To appear as Haythem O. Ismail, Reason Maintenance and the Ramsey Test. In Gabriele Kern-Isberner and Christoph Beierle (eds.), Dynamics of Knowledge and Belief: Workshop at the 30th German Conference on Artificial Intelligence (KI-2007). All citations should be to the published version.

witnessed considerable convergence over the history of belief revision [7–9, for example].

To everyone's distress, however, Peter Gärdenfors [10] proved that the Ramsey test account of conditionals, together with some seemingly reasonable constraints on belief revision (three of the AGM postulates [6]), is inconsistent with a minimal set of harmless demands on a logical theory. Gärdenfors's result triggered considerable research attempting to save the intuitive interpretation of conditionals provided by the Ramsey test [11–17, for example].

In this paper, I attempt to do two things. First, I will argue that, if belief revision is interpreted in the context of an implemented reason maintenance system, Gärdenfors's triviality result is avoided. This comes as a consequence of rejecting some the AGM postulates, based on general demands on implemented reason maintenance systems that are independent of conditionals and Gärdenfors's result. Second, as a side effect, I shall outline a theory for reasoning about conditionals within an implemented knowledge representation and reasoning (KRR) system with a reason maintenance component. The discussion will primarily focus on the SNePS KRR system [18, 19] and its reason maintenance component SNeBR [5, 20, 9].

In Section 2, we review Gärdenfors's triviality theorem and, in Section 3, we examine previous attempts to rectify the damage it has wrought. Section 4 presents a reason maintenance system based on relevance logic, which is then extended to accommodate conditionals. Finally, Section 5 evaluates the system with respect to Gärdenfors's triviality result.

2 The Triviality Trap

What exactly did Gärdenfors discover? As pointed out in the introduction, he discovered that the Ramsey test is inconsistent with simple common demands on a logical system. In particular, Gärdenfors proved that the Ramsey test will introduce a contradiction into a belief set that contains none of three pair-wise contrary propositions. That such a belief set may exist is uncontroversial. Hence, the triviality result.

Nevertheless, Gärdenfors's proof is based on a mesh of background assumptions. These assumptions are primarily of two types: (i) assumptions on the belief revision process implied by the Ramsey test, and (ii) assumptions on what a "belief set" is. Attempts to circumvent the triviality result are based on dropping one or more of these assumptions. To get a deeper understanding of what exactly the problem is, and to appreciate previous approaches to solve it, we start by listing Gärdenfors's background assumptions.[1]

[1] This is not an exhaustive list. It is a list of those assumptions that are (so far) uncontroversial and/or relevant to my examination of previous work and my own proposal.

In what follows, $\mathcal{L}_0$ is a ground language of classical propositional or first-order logic including the absurd proposition $\perp$[2], $\mathcal{L}_1$ is the closure of $\mathcal{L}_0$ under the conditional connective $>$, $\mathbb{K}$ is a set of belief sets, $\mathcal{K} \in \mathbb{K}$ is a belief set, and $* : \mathbb{K} \times \mathcal{L}_1 \longrightarrow \mathbb{K}$ is a belief revision operator.

1. Assumptions on $\mathcal{K}$:
   ($A^{\mathcal{K}}1$) Belief sets are sets of sentences[3]: $\mathcal{K} \subseteq \mathcal{L}_1$.
   ($A^{\mathcal{K}}2$) Belief sets may include conditional sentences: For some $\mathcal{K} \in \mathbb{K}$, $\mathcal{K} \not\subseteq \mathcal{L}_0$.
   ($A^{\mathcal{K}}3$) Belief sets are deductively-closed: if $\phi \in \text{Cn}_0(\mathcal{K})$ then $\phi \in \mathcal{K}$.[4]
2. Assumptions on belief revision:
   (A^*1) Success: $\phi \in \mathcal{K} * \phi$.
   (A^*2) Consistency: If $\perp \in \text{Cn}_0(\mathcal{K} * \phi)$ then $\perp \in \text{Cn}_0(\{\phi\})$.
   (A^*3) Expansion 1: $\text{Cn}_0(\mathcal{K} * \phi) \subseteq \text{Cn}_0(\mathcal{K} \cup \{\phi\})$.
   (A^*4) Expansion 2: If $\neg\phi \notin \text{Cn}_0(\mathcal{K})$, then $\text{Cn}_0(\mathcal{K} \cup \{\phi\}) \subseteq \text{Cn}_0(\mathcal{K} * \phi)$.

From (A^*4), (A^*5) immediately follows:

(A^*5) Preservation: If $\neg\phi \notin \text{Cn}_0(\mathcal{K})$, then $\text{Cn}_0(\mathcal{K}) \subseteq \text{Cn}_0(\mathcal{K} * \phi)$.

With these background assumptions, Gärdenfors [10] states the Ramsey test as follows.

(RT) $\phi > \psi \in \mathcal{K}$ if and only if $\psi \in \mathcal{K} * \phi$.

For ease of reference, let us break (RT) into two conditionals:

($RT1$) If $\phi > \psi \in \mathcal{K}$ then $\psi \in \mathcal{K} * \phi$.
($RT2$) If $\psi \in \mathcal{K} * \phi$ then $\phi > \psi \in \mathcal{K}$.

To simplify the proof of the triviality theorem, Gärdenfors first proves the following crucial lemma: The "monotonicity criterion".

(M) For all $\mathcal{K}, \mathcal{K}' \in \mathbb{K}$ and all $\phi \in \mathcal{L}_1$, if $\mathcal{K} \subseteq \mathcal{K}'$ then $\mathcal{K} * \phi \subseteq \mathcal{K}' * \phi$.

Proof.

1. $\mathcal{K} \subseteq \mathcal{K}'$	(Assumption)
2. $\psi \in \mathcal{K} * \phi$	(Assumption)
3. $\phi > \psi \in \mathcal{K}$	(2, ($RT2$))
4. $\phi > \psi \in \mathcal{K}'$	(1, 3)
5. $\psi \in \mathcal{K}' * \phi$	(4, ($RT1$))
6. $\mathcal{K} * \phi \subseteq \mathcal{K}' * \phi$	(2, 5)
7. (M)	(1, 6)

Before we present Gärdenfors's triviality result, we define the notion of non-triviality of a belief revision system.

[2] It should be noted, however, that the triviality result was shown to be valid for a more general class of monotonic [21] and non-monotonic [22] logics.

[3] Gärdenfors [10] uses the term "proposition" instead of "sentence". Yet, he states that "belief sets are just theories in the standard logical sense" [10, p. 83].

[4] Cn_0 is classical (deductive) logical consequence. We may assume a natural deduction system, although only modus ponens and the deduction theorem are needed.

(NT) A belief revision system $\langle \mathcal{L}_1, \mathbb{K}, * \rangle$ is non-trivial if, for some $\mathcal{K} \in \mathbb{K}$ and $A, B, C \in \mathcal{L}_1$,
1. $\{\neg(A \wedge B), \neg(A \wedge C), \neg(B \wedge C)\} \subseteq \mathrm{Cn}_0(\mathcal{K})$; and
2. $\neg A \notin \mathrm{Cn}_0(\mathcal{K})$, $\neg B \notin \mathrm{Cn}_0(\mathcal{K})$, and $\neg C \notin \mathrm{Cn}_0(\mathcal{K})$.

Theorem 1. *There is no non-trivial belief revision system that satisfies* $(A^{\mathcal{K}}1)$–$(A^{\mathcal{K}}3)$*,* (A^*1)–(A^*4)*, and RT.*

Proof.

1. $\neg A \notin \mathcal{K}$	(NT)
2. $\mathcal{K} * A = \mathrm{Cn}_0(\mathcal{K} \cup \{A\})$	$(1, A^{\mathcal{K}}3, A^*3, A^*4)$
3. $(B \vee C) \in (\mathcal{K} * A) * (B \vee C)$	(A^*1)
4. $\neg(B \vee C) \notin (\mathcal{K} * A) * (B \vee C)$	$(3,\ A^*2)$
5. $\neg C \notin (\mathcal{K} * A) * (B \vee C)$	(Assumption)
6. $\neg(A \vee B) \notin \mathcal{K}$	(NT)
7. $\mathcal{K} * (A \vee B) = \mathrm{Cn}_0(\mathcal{K} \cup \{A \vee B\})$	$(6, A^{\mathcal{K}}3, A^*3, A^*4)$
8. $\mathcal{K} * (A \vee B) \subseteq \mathcal{K} * A$	$(2{,}7{,}A^{\mathcal{K}}3, \mathrm{Cn}_0)$
9. $(\mathcal{K} * (A \vee B)) * (B \vee C) \subseteq (\mathcal{K} * A) * (B \vee C)$	$(8{,}M)$
10. $\neg C \notin (\mathcal{K} * (A \vee B)) * (B \vee C)$	(5,9)
11. $\neg(B \vee C) \notin \mathcal{K} * (A \vee B)$	$(7, NT, \mathrm{Cn}_0)$
12. $(\mathcal{K} * (A \vee B)) * (B \vee C) = \mathrm{Cn}_0(\mathcal{K} * (A \vee B) \cup \{B \vee C\})$	$(11, A^{\mathcal{K}}3, A^*3, A^*4)$
13. $(\mathcal{K} * (A \vee B)) * (B \vee C) = \mathrm{Cn}_0(\mathcal{K} \cup \{A \vee B, B \vee C\})$	$(7, 12)$
14. $(\mathcal{K} * (A \vee B)) * (B \vee C) = \mathrm{Cn}_0(\mathcal{K} \cup \{B\})$	$(13, NT)$
15. $\neg C \in \mathrm{Cn}_0(\mathcal{K} \cup \{B\}$	(NT)
16. $\neg C \in (\mathcal{K} * (A \vee B)) * (B \vee C)$	$(14, 15)$
17. $\perp$	$(10, 16)$
18. $(\neg C \notin (\mathcal{K} * A) * (B \vee C)) \supset \perp$	$(5, 17, \mathrm{Cn}_0)$
19. $(\neg B \notin (\mathcal{K} * A) * (B \vee C)) \supset \perp$	(Similarly, 5–17)
20. $\perp$	$(4, 18, 19)$

Having displayed the proof in detail, we can carefully analyze the different loopholes proposed in the literature. Each of the proposed loopholes identifies one or more of the background assumptions and/or the Ramsey test as the culprit.

3 Loopholes

We shall consider six proposals [12–17], each identifying a different set of background assumptions as the culprit, or a different way out of the triviality trap. Except for [14], all proposals attempt to invalidate the use of (M) in the proof. They all (*pace* Gärdenfors in [10, 11]) preserve the Ramsey test (some version of it, for that matter), and choose to reject M based on the background assumptions. Table 1 lists the culprits identified by each of the six proposals.

Rott [12] argues convincingly that (M) is indeed true, but only trivially so. Informally, once we admit conditionals into belief sets (as per $(A^{\mathcal{K}}2)$), no belief set can be a proper subset of another. From this general result, the invalidity

	$A^{\mathcal{K}}1$	$A^{\mathcal{K}}2$	$A^{\mathcal{K}}3$	A^*1	A^*2	A^*3	A^*4
Rott [12]						×	×
Hansson [13]			×			×	×
Arló Costa and Levi [14]		×					
Lindström and Rabinowicz [15]	×						
Grahne [16]							×
Giordano et àl [17]						×	×

Table 1. A comparison of six proposals to escape Gärdenfors's triviality result. The crosses indicate the culprits identified by each proposal.

of (A^*3) and (A^*4) (and hence (A^*5)) follows. The result is proved by Hansson [13], making use of the following property of proper subsets.

(*PSS*) $\mathcal{K} \in \mathbb{K}$ has a proper subset if and only if there are $\mathcal{K}_1 \subseteq \mathcal{K}$ and $\mathcal{K}_2 \subseteq \mathcal{K}$ such that $\mathcal{K}_1 \not\subseteq \mathcal{K}_2$ and $\mathcal{K}_2 \not\subseteq \mathcal{K}_1$.

To informally illustrate Hansson's proof, I will refer to an example based on one due to Darwiche and Pearl [23].

Example 1. A murder occurs. John and Mary are the prime suspects. Detective 1 finds evidence incriminating John ($\mathcal{K}_1 = \mathrm{Cn}_0(\{J\})$). Detective 2, on the other hand, finds evidence incriminating Mary ($\mathcal{K}_2 = \mathrm{Cn}_0(\{M\})$). Both detectives report to their supervisor ($\mathcal{K} = \mathrm{Cn}_0(\{J, M\})$). As long as we only consider sentences in $\mathcal{L}_0$, then it is clear that $\mathcal{K}_1, \mathcal{K}_2$ and $\mathcal{K}$ satisfy (*PSS*). However, intuitively, $\neg(J \wedge M) > (J \wedge \neg M) \in \mathcal{K}_1$.[5] Similarly, $\neg(J \wedge M) > (\neg J \wedge M) \in \mathcal{K}_2$. But these two conditionals are contradictory; they cannot both be in $\mathcal{K}$.

The above example illustrates the fundamental difficulty in finding three sets satisfying (*PSS*): incomplete information licences the belief in conditionals that lose their support in a more informed belief state. Inspecting Table 1, [12, 13, 17] take issue with (A^*3) and (A^*4). Most probably [12, 13, 17] would identify the main glitch in the triviality proof with step 7.[6] Though they agree on the culprit, each of these authors proposes a different way out of the triviality trap. Rott [12] informally considers employing a non-monotonic logic instead of Cn_0. Hansson [13] presents a detailed theory founded on belief base revision. (Hence,

[5] Hansson [13, p. 529] stresses that this last conclusion is based on "prephilosophical" intuitions, and (crucially) not on the Ramsey test. However, I find the acceptance of this prephilosophical intuition together with the Ramsey test (Hansson's position) a bit strange. For this certainly commits us to accepting certain properties of belief revision. In particular, we are thus committed to (A^*4)—the very assumption that Hansson rejects.

[6] Of course, by rejecting (A^*3) and (A^*4), any step that is licensed by either is flawed. Nevertheless, those authors seem to base most of their informal arguments on the invalidity of step 7.

his rejection of $(A^{\mathcal{K}}2)$.)[7] Giordano et àl [17] restrict $(A^{*}3)$ and $(A^{*}4)$ to the maximal $\mathcal{L}_0$-subset of $\mathcal{K}$.

Grahne [16] would probably identify step 12 as the main glitch. Grahne's rejection of $(A^{*}4)$ is based, not on admitting conditionals into belief states (as per $(A^{\mathcal{K}}2)$), but on his very interpretation of the belief change operator appropriate for the interpretation of conditionals as per the Ramsey test. Instead of the classical AGM belief revision operator [6] , Grahne opts for Katsuno and Mendelzon's belief *update* operator [24]. Belief revision is appropriate for a change in belief signaled by acquiring information about a static world. Belief update, on the other hand, is needed when the change in belief is necessary due to a change in the world. When revising $\mathcal{K} * (A \vee B)$ with $B \vee C$ in step 12, we assume the world has not changed. That is, $A \vee B$ is still true. Thus, B immediately follows, since adding $B \vee C$ just gives us more specific information about which of the contraries A and B is indeed true. On the other hand, when updating $\mathcal{K} * (A \vee B)$ with $B \vee C$, we assume that the world has changed, and, thus, cannot assume that $A \vee B$ is till true. To our best knowledge, only $B \vee C$ is certain, but not B.

Arló Costa and Levi [14] (following a hard-line position of Levi's [25]) reject $A^{\mathcal{K}}2$. They argue that conditionals do not qualify as members of belief sets (or as truth-value bearers), but as representations of an agent's dispositions to change their beliefs. From the premise that this is Ramsey's own position, Gärdenfors's (RT) is disqualified as a formal rendering of the Ramsey test. Their proposal is to adopt a stratified theory, where there is a clear distinction between the belief set $\mathcal{K}$ (a subset of $\mathcal{L}_0$) and the set of sentences, $s(\mathcal{K}) \subseteq \mathcal{L}_1$, supported by $\mathcal{K}$. Only the latter may include conditionals. A stratified version of the Ramsey test may then be stated:

(SRT) For all $\phi, \psi \in \mathcal{L}_0$, $\phi > \psi \in s(\mathcal{K})$ if and only if $\psi \in \mathcal{K} * \phi$.

Although a stratified version of (M) may also be derived (with the antecedent being $s(\mathcal{K}) \subseteq s(\mathcal{K}'))$, the proof of the triviality result will be blocked at step 7, since $\mathcal{K} \subseteq \mathcal{K}'$ does not entail $s(\mathcal{K}) \subseteq s(\mathcal{K}')$.

Lindström and Rabinowicz hold yet another position [15]. They identify $A^{\mathcal{K}}1$ as the sole culprit, and, instead of taking $\mathcal{K}$ to be a set of sentences, they assume it is a set of *propositions*. How does this assumption get us out of the triviality trap? The assumption by itself may not help if sentences and propositions stand in one-to-one correspondence. This is exactly what Lindström and Rabinowicz reject. They argue that conditionals are *context-sensitive*: the same conditional may express different propositions in different contexts.[8] On their account, a context is simply a belief set.[9] By adopting a context-sensitive version of the Ramsey test, the proof of (M) is blocked and the triviality result is avoided.

[7] A belief base $\mathcal{B} \subseteq \mathcal{K}$ is a set of beliefs such that $\mathrm{Cn}_0(\mathcal{B}) = \mathcal{K}$.

[8] It should be noted that the same position was skeptically considered by Gärdenfors himself [10, p. 91].

[9] Here, I am using the term "belief set" to refer to sets of propositions. Lindström and Rabinowicz [15] use the term "belief state", instead; they reserve "belief set" for sets of sentences.

Lindström and Rabinowicz's rendering of the Ramsey test could be presented as follows, where the semantics of $<$ depends on the context $\mathcal{K}$.

(CRT) $\phi >_{\mathcal{K}} \psi \in \mathcal{K}$ if and only if $\psi \in \mathcal{K} * \phi$.

The proof of (M) blocks in step 5, which cannot be proved since $>_{\mathcal{K}'}$ is needed in place of $>_{\mathcal{K}}$ in step 4.

In what follows, I will present an assumption-based reason maintenance system. The details of the system are based on assumptions that differ fundamentally from some of those underlying the triviality result. As it turns out, these assumptions, which are independently motivated by issues of rational agency and computational complexity, allow us to gracefully escape the triviality trap when the system is extended to accommodate conditionals.

4 Reason Maintenance and Conditionals

4.1 General Requirements on Reason Maintenance

Unlike belief revision theories, reason maintenance systems are required to take into account issues of bounded computational resources and availability. These issues motivate the following three requirements on reason maintenance systems.

RM1. Belief sets are not closed under logical consequence.
RM2. Paradoxes of implication are not tolerated.
RM3. Implicit inconsistencies are tolerated.

The motivation for **RM1** is clear; no realistic computational (or rational) reasoning system can be logically closed. While we may talk about the closure of a belief set to facilitate the analysis of its potential theorems, the belief set itself must be finite, and as small as possible for that matter. Clearly, **RM1** is at odds with $(A^{\mathcal{K}}3)$.

RM2 is particularly required to block the derivation of arbitrary sentences from contradictions. From the point of view of rational agency, it is clear that agents (notably humans) can accommodate contradictory beliefs without committing to logical absurdity. From the point of view of computational reasoning systems, a system should provide useful, sound inferences even in the presence of contradictions.

RM3 is probably the least obvious. However, once **RM2** is accepted, it is clear that the harmful effects of contradictions can be isolated. In addition, inconsistencies are only tolerated as long as they are only implicit, once a contradiction is explicitly derived (that is, added to the belief state), then consolidation is triggered.

In what follows, a reason maintenance system that satisfies the above requirements will be presented. The system is based on [5] and [20, 9]. It is implemented as SNeBR, the belief revision component of the SNePS knowledge representation and reasoning system [18, 19].

4.2 The Case of $\mathcal{L}_0$

First, let us consider a reason maintenance system for the language $\mathcal{L}_0$. It is clear that the classical Cn_0 does not observe **RM2**. SNePS logic is a version of Anderson and Belnap's *relevance logic* [26, 27]. A full exposition of relevance logic is not needed (and not possible) here. Suffice it to say, that relevance logic does observe **RM2**, and that it achieves this by keeping track of the history of derivations. (Thus, we seem to have an independent motivation for recording derivation traces, which is required by assumption-based reason maintenance.) In what follows, Cn_R denotes relevance logic consequence.

Definition 1 *A* ***support set*** *of a sentence* $\phi \in \mathcal{L}_0$ *is a set* $s \subseteq \mathcal{L}_0$ *such that* $\phi \in \mathrm{Cn}_R(s)$. s *is* ***minimal*** *if, for every* $s' \subset s, \phi \notin \mathrm{Cn}_R(s')$.

The reader should note that minimal support sets of a sentence ϕ are Hansson's *ϕ-kernels* [28].

Definition 2 *A* ***belief state*** $\mathcal{S}$ *is a quadruple* $\langle \mathcal{K}, \mathcal{B}, \sigma, \preccurlyeq \rangle$*, where:*

1. $\mathcal{K} \subseteq \mathcal{L}_0$ *is a* ***belief set****.*
2. $\mathcal{B} \subseteq \mathcal{K}$*, with* $\mathcal{K} \subseteq \mathrm{Cn}_R(\mathcal{B})$*, is a* ***belief base****. If* $\phi \in \mathcal{B}$*, then* ϕ *is a* ***base belief****.*
3. $\sigma : \mathcal{K} \longrightarrow 2^{2^{\mathcal{B}}}$ *is a* ***support function****, where each* $s \in \sigma(\phi)$ *is a minimal support set of* ϕ*. If* $\phi \in \mathcal{B}$*, then* $\{\phi\} \in \sigma(\phi)$*.*
4. $\preccurlyeq \subseteq \mathcal{B} \times \mathcal{B}$ *is a total pre-order on base beliefs.*[10]

On the intuitive interpretation of the above definition, base beliefs are beliefs that have independent standing. For example, they are the result of perception or interaction with another agent (possibly a human operator/user). Crucially, they are not in the belief state based *solely* on inference. The belief set $\mathcal{K}$ is not closed under Cn_R; it represents the set of sentences that are either base beliefs or that were *actually derived* from base beliefs.[11] This is in contrast to the logically-closed $\mathrm{Cn}_R(\mathcal{K})$ which is the set of sentences *derivable* from base beliefs.

The set $\sigma(\phi)$ is the family of minimal support sets that were actually used, or discovered, to derive ϕ. $\mathcal{B}$ may include minimal support sets of ϕ that are, nevertheless, not in $\sigma(\phi)$, if they are not yet discovered to derive ϕ. The total pre-order $\preccurlyeq$ represents a preference ordering over base beliefs. This ordering will be used when belief revision requires sacrificing a base belief; the least preferred will be the victim. I will refrain from making any commitments about the origins of this ordering. In particular, unlike standard epistemic entrenchment relations [29, for example], I am not assuming any logical basis for preference. For the purpose of this paper, the ordering is just given.[12]

---

[10] A total pre-order is a complete, reflexive, and transitive binary relation.

[11] Thus, in time, a belief state can evolve into a *different* belief state that share the same base.

[12] For future investigation, we may consider the possibility of moving $\preccurlyeq$ into the object language.

Belief revision in this system is a distant variant of Hansson's *semi-revision* [29]—a non-prioritized belief revision operator where success is not guaranteed. We first need to define a notion of *relevant expansion*.

Definition 3 *Let $\mathcal{S} = \langle \mathcal{K}, \mathcal{B}, \sigma, \preccurlyeq \rangle$ be a belief state. The* ***relevant expansion*** *of $\mathcal{S}$ with $\phi \in \mathcal{L}_0$ is a belief state $\mathcal{S} + \phi = \langle \mathcal{K}_{+\phi}, \mathcal{B}_{+\phi}, \sigma_{+\phi}, \preccurlyeq_{+\phi} \rangle$, with the following properties:*

(A^+1) *Success: $\mathcal{B}_{+\phi} = \mathcal{B} \cup \{\phi\}$.*
(A^+2) *Inclusion: $\mathcal{K} \subseteq \mathcal{K}_{+\phi}$.*
(A^+3) *Relevance: If $\psi \in \mathcal{K}_{+\phi} \backslash \mathcal{K}$, and $s \in \sigma_{+\phi}(\psi)$, then there is $s' \in \sigma_{+\phi}(\phi)$ such that $s' \subseteq s$.*
(A^+4) *Support update: If $\psi \in \mathcal{K}$ and $s \in \sigma_{+\phi}(\psi)$, then either $s \in \sigma(\psi)$ or there is $s' \in \sigma_{+\phi}(\phi)$ such that $s' \subseteq s$.*
(A^+5) *Order preservation: $\preccurlyeq_{+\phi}$ is the smallest total pre-order on $\mathcal{B}_{+\phi}$ satisfying*
  1. *$\preccurlyeq \subseteq \preccurlyeq_{+\phi}$ and*
  2. *for every $\psi \in \mathcal{B}$, either $\phi \preccurlyeq \psi$ or $\psi \preccurlyeq \phi$.*

Relevant expansion is simply assertion with forward inference. The belief state resulting from relevant expansion by ϕ will include ϕ and anything that follows from it. That all newly derived sentences indeed follow from ϕ is guaranteed by (A^+3), provided that ϕ was not derived in $\mathcal{K}$. In addition, old sentences may acquire new support only as a result of discovered derivations from ϕ ((A^+4)). It should be noted that, given certain constraints on Cn_R, the set $\mathcal{K}_{+\phi}$ is finite (provided that $\mathcal{K}$ is). (A^+5) makes the simplifying assumption that adding ϕ does not disturb the preference relations already established; ϕ simply gets added in some appropriate position in the $\preccurlyeq$-induced chain of equivalence classes.

Definition 4 *Let $\mathcal{S} = \langle \mathcal{K}, \mathcal{B}, \sigma, \preccurlyeq \rangle$ be a belief state. The* ***relevant revision*** *of $\mathcal{S}$ with $\phi \in \mathcal{L}_0$ is a belief state $\mathcal{S} \dot{+} \phi = \langle \mathcal{K}_{\dot{+}\phi}, \mathcal{B}_{\dot{+}\phi}, \sigma_{\dot{+}\phi}, \preccurlyeq_{\dot{+}\phi} \rangle$, with the following properties:*

($A^{\dot{+}}1$) *Base inclusion: $\mathcal{B}_{\dot{+}\phi} \subseteq \mathcal{B}_{+\phi}$.*
($A^{\dot{+}}2$) *Inclusion: $\mathcal{K}_{\dot{+}\phi} \subseteq \mathcal{K}_{+\phi}$.*
($A^{\dot{+}}3$) *Lumping: $\psi \in \mathcal{K}_{+\phi} \backslash \mathcal{K}_{\dot{+}\phi}$ if and only if, for every $s \in \sigma_{+\phi}(\psi)$, $s \not\subseteq \mathcal{B}_{\dot{+}\phi}$.*
($A^{\dot{+}}4$) *Preferential core-retainment: $\psi \in \mathcal{B}_{+\phi} \backslash \mathcal{B}_{\dot{+}\phi}$ if and only if $\bot \in \mathcal{K}_{+\phi}$ and $\psi \in \{x \mid \exists s \in \sigma_{+\phi}(\bot), x \in s, \text{ and } \forall y \in s, x \preccurlyeq_{+\phi} y\}$.*
($A^{\dot{+}}5$) *Support update: If $\psi \in \mathcal{K}_{\dot{+}\phi}$, then $\sigma_{\dot{+}\phi}(\psi)$ is the largest subset of $\sigma_{+\phi}(\psi)$ restricted to $\mathcal{B}_{\dot{+}\phi}$.*
($A^{\dot{+}}6$) *Order preservation: $\preccurlyeq_{\dot{+}\phi}$ is the largest subset of $\preccurlyeq_{+\phi}$ restricted to $\mathcal{B}_{\dot{+}\phi}$.*

Thus, relevant revision is assertion with forward inference followed by consolidation [29]. As a result of consolidation, some base beliefs might be retracted in case relevant expansion with ϕ results in a contradiction.[13] ($A^{\dot{+}}1$) captures this

[13] Technically, the contradiction need not be supported by ϕ. However, in practice, a reason maintenance system should not tolerate explicit contradiction (SNeBR does not). Thus, prior to relevant revision with ϕ, we may assume that no explicit contradiction was around, and, thus, that ϕ is somehow responsible for discovering/introducing a contradiction.

intuition. Since belief sets are not the logical closures of their bases, $(A^{\dot{+}}2)$ does not necessarily follow from $(A^{\dot{+}}1)$. It is needed to indicate that relevant revision does not result in derivations that are not accounted for by relevant expansion. $(A^{\dot{+}}3)$ makes sure that only sentence that are still supported are believable.[14]

$(A^{\dot{+}}4)$ guarantees that base beliefs that are evicted to retain (explicit) consistency indeed must be evicted. In addition, if a choice is possible, base beliefs that are least preferred are chosen for eviction. Note that, according to the above definition, this selection strategy is *skeptical*; that is, if multiple least preferred beliefs exist, all are evicted. This strategy, however, is only adopted here to simplify the exposition, and nothing relevant depends on it.

As a simple corollary, it follows from $(A^{\dot{+}}3)$ and $(A^{\dot{+}}4)$ that the resulting belief state is not known to be inconsistent:

$(A^{\dot{+}}7)$ Contradiction ignorance: $\bot \notin \mathcal{K}_{\dot{+}\phi}$

4.3 The Case of $\mathcal{L}_1$

To extend the reason maintenance system presented above to $\mathcal{L}_1$, a number of superficial alterations of the definitions are needed. The important point, however, is to devise an extension of Cn_R that accommodates conditionals. Following [26], I am assuming a natural deduction system. Adding the connective $>$ to the language, we need two inference rules—one for elimination and one for introduction. First, a piece of notation.

Definition 5 *Let $\mathcal{S} = \langle \mathcal{K}, \mathcal{B}, \sigma, \preccurlyeq \rangle$ be a belief state. The* ***hypothetical expansion*** *of $\mathcal{S}$ with $\phi \in \mathcal{L}_1$ is a belief state $\mathcal{S} \mp \phi = \langle \mathcal{K}_{\mp\phi}, \mathcal{B}_{\mp\phi}, \sigma_{\mp\phi}, \preccurlyeq_{\mp\phi} \rangle$ where*

1. *$\mathcal{B}_{\mp\phi} = \mathcal{B} \cup \{\phi\}$;*
2. *$\sigma_{\mp\phi}(\phi) = \{\{\phi\}\}$; and*
3. *for every $\psi \in \mathcal{B}$, $\psi \preccurlyeq_{\mp\phi} \phi$ and $\phi \not\preccurlyeq_{\mp\phi} \psi$*

Hypothetical expansion (re)introduces ϕ into the belief state with independent standing as a most preferred belief. It is similar to the *do* operator of Pearl [31] in that it detaches ϕ from its derivational history (causal history, in the case Pearl). We now define the elimination and introduction rules for $>$ as follows.

$(> E)$ If $\phi, \phi > \psi \in \mathcal{K}$, then ψ may be added to $\mathcal{K}$, with $\sigma(\psi) = \{s_\phi \cup s_> \mid \langle s_\phi, s_> \rangle \in \sigma(\phi) \times \sigma(\phi > \psi)\}$.

$(> I)$ If $\psi \in (\mathcal{K}_{\mp\phi})_{\dot{+}\phi}$, then $\phi > \psi$ may be added to $\mathcal{K}$ provided that $\sigma(\phi > \psi) = \{s \backslash \{\phi\} \mid s \in (\sigma_{\mp\phi})_{\dot{+}\phi}(\psi) \backslash \sigma_{\mp\phi}(\psi)\}$ is not empty.

The elimination rule $(> E)$ is a direct extension of Anderson and Belnap's rule for the elimination of material implication [26]. The introduction rule $(> I)$ is also an extension of Anderson and Belnap's rule for the introduction of material implication. The extension in this case is, by no means, direct though.

[14] By "lumping", I'm referring to the lumping operation of Kratzer [30], whereby certain propositions either stay together or go together.

$(> I)$ is actually the right-to-left direction of the Ramsey test $(RT2)$, within the context of relevance logic. In simple English, the rule describes a procedure whereby one may decide whether to believe in the conditional $\phi > \psi$:

1. Hypothetically expand the belief state with ϕ.
2. Perform relevant forward inference to derive all sentences that could be derived from ϕ.
3. Consolidate the resulting belief state, giving ϕ highest preference.
4. If ψ is in the resulting belief state, accept $\phi > \psi$.

In addition to deciding on whether to accept $\phi > \psi$, we also compute its support sets along the way. The relevance of this derivation is guaranteed by two measures. The first is the hypothetical expansion step. The reason why we need this is that we need to make sure that any derivation of ψ following relevant expansion with ϕ follows from ϕ itself, not merely from its supports. The second is the procedure used to compute $\sigma(\phi > \psi)$: We only consider support sets that were added as a result of relevant expansion with ϕ. This eliminates cases where a conditional is only accepted as a result of its consequent being already in the belief set. The final removal of ϕ from the sets of supports is inherited from Anderson and Belnap's rule for material implication introduction.

By adding $(> E)$ and $(> I)$ to our repertoire of inference rules, we define an extension $\text{Cn}_{R>}$ of Cn_R for relevant conditional consequence. All the definitions of Section 4.2 may now be extended to $\mathcal{L}_0$ by replacing each occurrence of $\mathcal{L}_0$ by $\mathcal{L}_1$, and each occurrence of Cn_R by $\text{Cn}_{R>}$. This may be considered overly permissive by many scholars. For, now, we allow two things that are traditionally not allowed.

We allow, *pace* Hansson [13], (i) belief bases to include conditional sentences and (ii) belief states to be revised with conditional sentences. The justification for this is the same: we view conditionals to possibly have independent standing. For a rational agent or knowledge representation and reasoning system, this is actually reasonable. The following example of a "useful counterfactual" is due to Costello and McCarthy [2, p. 1].

(1) If another car had come over the hill when you passed that car, there would have been a head-on collision

A natural context in which (1) may be uttered is one in which the speaker is teaching someone an important safety rule of car driving. Most probably, the person receiving this utterance does not have enough experience to have concluded it themselves. The person can learn from (1), though. And they can do that without ever trying to achieve its antecedents. Thus, it seems that in similar cases, the only reasonable thing to assume is that the conditional is a base belief that probably could not have been derived without external help.

5 The Trap Reentered

What can we say now about the Ramsey test and Gärdenfors's triviality result? First, as $(> I)$ indicates, only one direction of the Ramsey test is adopted; we use it as a rule of introduction for conditionals. A direct reversal of $(> I)$ would yield.

$(RRT1)$ If $\phi > \psi$ may be added to $\mathcal{K}$ with
$\sigma(\phi > \psi) = \{s \backslash \{\phi\} \mid s \in (\sigma_{\mp\phi})_{\dot{+}\phi}(\psi) \backslash \sigma_{\mp\phi}(\psi)\}$, then $\psi \in (\mathcal{K}_{\mp\phi})_{\dot{+}\phi}$ and $\{s \backslash \{\phi\} \mid s \in (\sigma_{\mp\phi})_{\dot{+}\phi}(\psi) \backslash \sigma_{\mp\phi}(\psi)\}$ is not empty.

This does not look right. For what does it mean to say that $\phi > \psi$ *may be* added to $\mathcal{K}$? Note that it does not mean that $\phi > \psi$ is *in* $\mathcal{K}$, for $\mathcal{K}$ is not logically closed. The primary way, within our system, we can make sure that $\phi > \psi$ *may be* added to $\mathcal{K}$ is probably by using $(> I)$. But, then, $(RRT1)$ will not be very useful; it is only telling us something that we already know about $\mathrm{Cn}_{R>}$.

So maybe we can simply replace "may be added to $\mathcal{K}$" with "is in $\mathcal{K}$". But, then, the first antecedent $(\phi > \psi \in \mathcal{K})$ would imply the second consequent $(\sigma(\phi > \psi) \neq \varnothing)$, and the second antecedent (the definition of $\sigma(\phi > \psi)$) would imply the first consequent $(\psi \in (\mathcal{K}_{\mp\phi})_{\dot{+}\phi})$. Again, this version of *RRT1* does not seem useful.

Finally, we can simply drop all mention of supports and state that $\phi > \psi \in \mathcal{K}$ if and only if $\psi \in (\mathcal{K}_{\mp\phi})_{\dot{+}\phi}$. This is far from being the converse of $(> I)$. In addition, I do not see how useful it may be, at least from the point of view of rational agency or a KRR system. If anything, it may save us some time by caching one of the results of hypothetically revising with ϕ (we can also easily reconstruct the supports of ψ in the resulting belief state). This, however, is not unproblematic. For even assuming $\phi > \psi \in \mathcal{K}$, it need not be the case that $\psi \in (\mathcal{K}_{\mp\phi})_{\dot{+}\phi}$. This may happen, for example, if deriving ψ using $(> E)$ may result in a contradiction. Considering the following variant of Example 1 may clarify this point (where $\phi = M$ and $\psi = \neg J$).

Example 2. The supervisor believes (probably by default) that only one person committed the murder: $\{J > \neg M, M > \neg J\}$. Detective 1 reports evidence incriminating John. This results in adding the beliefs J and $\neg M$ to the supervisor's belief set. Now detective 2 reports evidence incriminating Mary. At this point, the supervisor needs to do some consolidation. Assuming that the evidence provided by both detectives is highly reliable, the supervisor would disbelieve $J > \neg M$, rendering $\neg M$ no longer supported. In addition, note that $M > \neg J$ also needs to be removed, to block the derivation of $\neg J$. Thus, when faced with strong evidence to the contrary, the supervisor gives up the belief in a single murderer.

Now, even if we assume some reasonable converse of $(> I)$, the triviality proof will not go through. Any step in the proof that depends on closure $(A^{\mathcal{K}}3)$, success (A^*1), or consistency (A^*2) will be invalid. In addition, since revisions are now based on $\mathrm{Cn}_{R>}$, step 8 is obviously invalid. For, even if we admit closure, $\mathrm{Cn}_{R>}(\mathcal{K} \cup \{A \vee B\}) \nsubseteq \mathrm{Cn}_{R>}(\mathcal{K} \cup \{A\})$ (since $A \vee B \notin \mathrm{Cn}_{R>}(\{A\})$).

In addition, Hansson [13, p. 531–532] argued that even if two belief sets are identical, base-revising them may yield different results. Thus any step in the proof that relies on the equality of revising two identical belief sets will be invalid (for example, step 13).

Example 3 [13, p. 531–532]. Let $\mathcal{S}_1$ be a belief state with $\mathcal{B}_1 = \{p\}$. Consider, $\mathcal{S}_2 = \mathcal{S}_1 \dot{+} q$ and $\mathcal{S}_3 = \mathcal{S}_1 \dot{+} p \Leftrightarrow q$. Clearly, $\mathcal{B}_2 = \{p, q\}$, $\mathcal{B}_3 = \{p, p \Leftrightarrow q\}$, and $\mathrm{Cn}_{R>}(\mathcal{B}_2) = \mathrm{Cn}_{R>}(\mathcal{B}_3)$. In particular, if $p \Leftrightarrow q \in \mathcal{K}_2$ then $\sigma_2(p \Leftrightarrow q) = \{\{p, q\}\}$. similarly, if $q \in \mathcal{K}_3$, then $\sigma_3(q) = \{\{p, p \Leftrightarrow q\}\}$. Now consider $\mathcal{S}_2 \dot{+} \neg p$ and $\mathcal{S}_3 \dot{+} \neg p$, assuming $p \preccurlyeq \neg p$ in both $\mathcal{S}_2$ and $\mathcal{S}_3$. Given Definition 4, q is in the first belief set and $\neg q$ is derivable in the second.

Finally, similar to [15], the proof of (M) will be blocked due to the context-sensitivity of conditionals. In our system, context-sensitivity is defined by the dependence of $> I$ (and its presumed converse) on sets of supports. Even with closure, success, and consistency reinstated, similar to [15], the proof of (M) will be blocked at step 5:

1. $\mathcal{K} \subseteq \mathcal{K}'$ (Assumption)
2. $\psi \in (\mathcal{K}_{\mp\phi})_{\dot{+}\phi}$ and $\{s \backslash \{\phi\} \mid s \in (\sigma_{\mp\phi})_{\dot{+}\phi}(\psi) \backslash \sigma_{\mp\phi}(\psi)\} \neq \varnothing$ (Assumption)
3. $\phi > \psi \in \mathcal{K}$
with $\sigma(\phi > \psi) = \{s \backslash \{\phi\} \mid s \in (\sigma_{\mp\phi})_{\dot{+}\phi}(\psi) \backslash \sigma_{\mp\phi}(\psi)\}$ (2, $(> I)$)
4. $\phi > \psi \in \mathcal{K}'$
with $\sigma'(\phi > \psi) = \{s \backslash \{\phi\} \mid s \in (\sigma_{\mp\phi})_{\dot{+}\phi}(\psi) \backslash \sigma_{\mp\phi}(\psi)\}$ (1, 3)

At this point we are stuck; we cannot prove $\psi \in (\mathcal{K}'_{\mp\phi})_{\dot{+}\phi}$ since this requires σ', and not σ, in the definition of the support of $\phi > \psi$. For instance, in Example 2, it is clear that the set of supports of $M > \neg J$ in $\mathcal{K} = \{J > \neg M, M > \neg J\}$ is different from that in $\mathcal{K}' = \mathcal{K} \cup \{J\}$, where in the former it is simply $\{\{M > \neg J\}\}$ and in the latter it is the empty sets.

6 Conclusions

Based on [5, 9], I have presented a reason maintenance system with an underlying relevance logic. I have shown how such a system could be extended to accommodate a relevance-logical account of conditionals. One direction of the Ramsey test is used as a conditional-introduction inference rule. The rule defines the set of supports of the derived conditional in such a way that effects context-sensitivity as per [15]. Given independently motivated assumptions on the underlying logic, the belief state, and the belief revision operator, Gärdenfors's triviality result is avoided.

The system presented here is similar to that of [13] in its use of base revision. It is different, however, in allowing conditional sentences to be members of belief bases and candidates for expansion and revision. Compared to the system of [16], the revision-based approach presented here is a practical alternative to the update-based approach of [16]. It is my conviction that, ultimately, both revision and update are needed for reasoning about conditionals. In particular, the relation between indicative conditionals and belief revision on one hand, and subjunctive conditionals and belief update on the other hand, remains to be investigated.

References

1. Stalnaker, R.: A theory of conditionals. In Rescher, N., ed.: Studies in Logical Theory. Number 2 in American Philosophical Quarterly, Monograph Series. Basil Blackwell Publishers, Oxford (1968) 98–112
2. Costello, T., McCarthy, J.: Useful counterfactuals. Linköping Electronic Articles in Computer and Information Science **4**(12) (1999)
3. Ortiz, C.: Explanatory update theory: Applications of counterfactual reasoning to causation. Artificial Intelligence **108**(1–2) (1999) 125–178
4. Doyle, J.: A truth maintenance system. Artificial Intelligence **12**(3) (1979) 231–272
5. Martins, J., Shapiro, S.C.: A model for belief revision. Artificial Intelligence **35**(1) (1988) 25–79
6. Alchourron, C.E., Gärdenfors, P., Makinson, D.: On the logic of theory change: Partial meet contraction and revision functions. The Journal of Symbolic Logic **50**(2) (1985) 510–530
7. Nebel, B.: A knowledge level analysis of belief revision. In Brachman, R., Levesque, H., Reiter, R., eds.: Principles of Knowledge Representation and Reasoning: Proceedings of the First International Conference (KR'89), Toronto, ON, Morgan Kaufmann (1989) 301–311
8. Williams, M.A.: Transmutations of knowledge systems. In Doyle, J., Sandewall, E., Torasso, P., eds.: Principles of Knowledge Representation and Reasoning: Proceedings of the Fourth International Conference (KR'94), San Francisco, CA, Morgan Kaufmann (1994) 619–629
9. Johnson, F.L., Shapiro, S.C.: Dependency-directed reconsideration: Belief base optimization for truth maintenance systems. In: Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI-05). (2005) 313–320
10. Gärdenfors, P.: Belief revisions and the Ramsey test for conditionals. The Philosophical Review **95**(1) (1986) 81–93
11. Gärdenfors, P.: Variations on the Ramsey test: More triviality results. Studia Logica **46** (1987) 321–327
12. Rott, H.: Conditionals and theory change: Revisions, expansions, and additions. Synthese **81** (1989) 91–113
13. Hansson, S.O.: In defence of the Ramsey test. The Journal of Philosophy **89**(10) (1992) 522–540
14. Arló Costa, H., Levi, I.: Two notions of epistemic validity: Epistemic models for Ramsey's conditionals. Synthese **109** (1996) 217–262
15. Lindström, S., Rabinowicz, W.: The Ramsey test revisited. In Grocco, G., Fariñas del Cerro, L., Herzig, A., eds.: Conditionals: From philosophy to computer science. Oxford University Press, Oxford, UK (1996) 147–191
16. Grahne, G.: Updates and counterfactuals. Journal of Logic and Computation **8**(1) (1998) 87–117
17. Giordano, L., Gliozzi, V., Olivetti, N.: Belief revision and the Ramsey test: A solution. In: AI*IA 2001: Advances in Artificial Intelligence : 7th Congress of the Italian Association for Artificial Intelligence. Number 2175 in LNAI, Berlin/Heidelberg, Springer-Verlag (2001) 165–175

18. Shapiro, S.C., Rapaport, W.J.: SNePS considered as a fully intensional propositional semantic network. In Cercone, N., McCalla, G., eds.: The Knowledge Frontier. Springer-Verlag, New York (1987) 263–315
19. Shapiro, S.C.: SNePS: A logic for natural language understanding and commonsense reasoning. In Iwańska, Ł.M., Shapiro, S.C., eds.: Natural Language Processing and Knowledge Representation: Language for Knowledge and Knowledge for Language. AAAI Press/The MIT Press, Menlo Park, CA (2000) 175–195
20. Johnson, F.L., Shapiro, S.C.: Automatic belief revision in SNePS. In Baral, C., Truszczynski, M., eds.: Proceedings of the 8th International Workshop on Non-Monotonic Reasoning (NMR2000). (2000) Unpaginated, 5 pages.
21. Segerberger, K.: A note on an impossibility theorem of Gärdenfors. Noûs **23**(3) (1989) 351–354
22. Makinson, D.: The Gärdenfors impossibility theorem in non-monotonic contexts. Studia Logica **49**(1) (1990) 1–6
23. Darwiche, A., Pearl, J.: On the logic of iterated belief revision. Artificial Intelligence **89** (1997) 1–29
24. Katsuno, H., Mendelzon, A.: On the difference between updating a knowledge base and revising it. In: Principles of Knowledge Representation and Reasoning: Proceedings of the second International Conference (KR'91). (1991) 387–394
25. Levi, I.: Iteration of conditionals and the Ramsey test. Synthese **76** (1988) 49–81
26. Anderson, A., Belnap, N.: Entailment. Volume I. Princeton University Press, Princeton, NJ (1975)
27. Shapiro, S.C.: Relevance logic in computer science. In Anderson, A., Belnap, N., Dunn, M., eds.: Entailment. Volume II. Princeton University Press, Princeton, NJ (1992) 553–563
28. Hansson, S.O.: Kernel contraction. Journal of Symbolic Logic **59**(3) (1994) 845–859
29. Hansson, S.O.: A survey of non-prioritized belief revision. Erkenntnis **50**(2–3) (1999) 413–427
30. Kratzer, A.: An investigation into the lumps of thougth. Linguistics and Philosophy **12** (1989) 607–653
31. Pearl, J.: Causality. Cambridge University Press, New York (2000)