Counterfactual theories of knowledge and the notion of actuality

Jan Heylen

Abstract

The central question of this article is how to combine counterfactual theories of knowledge with the notion of actuality. It is argued that the straightforward combination of these two elements leads to problems, viz. the problem of collapsing knowledge and the problem of missing knowledge. In other words, there is overgeneration of knowledge and there is undergeneration of knowledge. The combination of these problems cannot be solved by appealing to methods by which beliefs are formed. An alternative solution is put forward. The key is to rethink the closeness relation that is at the heart of counterfactual theories of knowledge.

Keywords: Knowledge; Counterfactual belief; Safety; Sensitivity; Adherence; Belief; Actuality

1 Introduction

In one form or another, Sosa (1999), Williamson (2000) and Pritchard (2005) uphold the claim that safety is a necessary condition on knowledge. Similarly, Heller (1989, 1999) and Nozick (1981) stand behind the claim that sensitivity is a necessary condition on knowledge. The simplest version of the safety condition is the following: a subject S knows that p only if there is no close possible world in which S falsely believes that p. Safely believing that p is supposed to be a stronger condition than truly believing that p: p has not only to be true at the world of evaluation but at all nearby worlds in which p is believed. The simplest version of the sensitivity condition is the following: a subject S knows that p only if at the closest worlds at which p is not true, S does not believe that p. Nozick also defends a third condition on knowledge, called the adherence condition. This condition says that a subject S knows that p only if at the closest worlds at which p is true, S believes that p.

The central question of this article is: what happens if one combines the counterfactual theories of knowledge with the notion of actuality? This question is inspired by a couple of observations. First, *any* theory of knowledge has to be able to deal with the notion of actuality. This is because one needs the notion of actuality to express certain truths that can be the object of knowledge, e.g. 'if Max ate less, he would be thinner than he actually is' (Lewis, 1970, p. 185) or 'it might have been that everyone actually happy was sad' (Hughes and Cresswell, 1996, p. 350). Second, especially any *counterfactual* theory of knowledge needs to be able to cope with the notion of actuality. After all, the entire point of counterfactual theories is to look at circumstances that are not factual or actual.

Before one can discuss the central question, one needs to familiarize oneself with current thinking about the semantics of actuality. This is the subject of Section 2. Additionally, one has to look into the combination of the notions of belief and actuality. This is the topic of Section 3. If one straightforwardly combines Kripke models for belief with standard models for actuality, one runs into two problems. A first result is that the belief that something is actually true entails that it is actually true. This is the problem of *unwanted factivity*. A second result is that it is not guaranteed that one believes that, if p is actually the case, then p is the case, and vice versa. This is the problem of *missing belief*. Related observations were

made by Rabinowicz and Segerberg (1994), and I will argue that transposing their proposed modifications to belief and actuality models solves the two noted problems.

After having secured the foundation, I will turn in Section 4 to the straightforward combination of Holliday (2012a) models for counterfactual belief, e.g. safe or sensitive belief, with standard models for actuality, with the proviso that the lessons learnt earlier have to be applied here as well. Again, two problems surface. A first result is that something being actually true and the belief that it is actually true entails that one knows or safely believes that it is actually true. This is the problem of *collapsing knowledge*. I will argue that the problem is not a special case of the problem posed by counterfactual knowledge of necessary truths and that it cannot be solved by appeal to methods of belief formation. A second result is that it is not guaranteed that one knows that, if p is actually the case, then p is the case, and vice versa. This is the problem of *missing knowledge*. It will be argued that this problem also cannot be solved by reference to methods. I will proceed by showing that a move similar to the one used earlier solves the two mentioned problems.

Throughout the article I will take care to check that the proposed solutions are not costly in the sense that they do not force proponents of counterfactual theories of knowledge to accept or reject principles that do not contain the notion of actuality and that they were previously disposed to reject or accept. In addition, I will reflect on the semantical contribution of actuality operators in doxastic or epistemic contexts.

2 Actuality

Kaplan (1977, p. 545) formulated the philosophically canonical semantical theory about actuality. The key clause is that, given a model and a context c (including an agent c_a, time c_t, position c_p and world c_w), 'actually ϕ', denoted as $@\phi$, is true with respect to a time t and a world w if and only if ϕ is true with respect to the time t and the world of that context, c_w. When there is no need for an entire context, it is common among philosophers to model actuality with the help of a distinguished world among the other worlds, viz. the unique actual world in the model (Davies and Humberstone, 1980; Fara and Williamson, 2005).

Let us consider the language of sentential logic with a necessity operator and an actuality operator, $\mathscr{L}_{\Box,@}$. If ψ is a well-formed formula of $\mathscr{L}_{\Box,@}$, then so are $\Box\psi$ and $@\psi$. The interpretation of $\mathscr{L}_{\Box,@}$ is fixed by the following definitions.

Definition 1 (Simple actuality model). A simple actuality model is a tuple

$$\langle W, w_0, R_M, V \rangle,$$

with W a non-empty set of worlds, $w_0 \in W$, R_M a two-place relation on W, V a function from proposition letters and elements of W to truth-values.

Definition 2 (Truth at a world in simple actuality models). The relation

$$\mathscr{M}, w \models \phi$$

is defined inductively as follows:

1. if $\phi = @\psi$, then $\mathscr{M}, w \models \phi$ iff $\mathscr{M}, w_0 \models \psi$;

2. all other clauses are as expected.

The above kind of models may be simple, but they are quite natural if one already makes use of distinguished worlds in models for modal logic for the purpose of defining *truth* in a model (Kripke, 1963):

$$\mathscr{M} \models \phi \quad \Leftrightarrow \quad \mathscr{M}, w_0 \models \phi.$$

Alternatively, one can define truth in a model as follows (Hughes and Cresswell, 1996, p. 39):

$$\mathscr{M} \models \phi \quad \Leftrightarrow \quad \forall w \in W : \mathscr{M}, w \models \phi.$$

Both conceptions of truth at a model come with intensionally different conceptions of validity.[1]

The previous semantical analysis can and has been generalised by philosophical logicians (Hughes and Cresswell, 1996, p. 350-355). The basic idea is to no longer define truth at a *single* world in the model but to define truth at a *world pair* in the model. One world in the pair functions as the *evaluation* world and the other world in the pair functions as the *reference* world. The latter is relevant for and only for the truth of formulas with actuality operators.

Definition 3 (Actuality model). An actuality model is a tuple $\langle W, R_M, V \rangle$, with W a non-empty set of worlds, R_M a two-place relation on W, V a function from proposition letters and worlds to truth-values.

Definition 4 (Truth in actuality models). The relation

$$\mathscr{M}, \langle w, w' \rangle \models \phi$$

is defined inductively as follows:

1. if $\phi = p$, then $\mathscr{M}, \langle w, w' \rangle \models \phi$ iff $V(p, w) = 1$;

2. if $\phi = \Box \psi$, then $\mathscr{M}, \langle w, w' \rangle \models \phi$ iff $\mathscr{M}, \langle v, w' \rangle \models \psi$ for all $v \in W$ such that $wR_M v$;

3. if $\phi = @\psi$, then $\mathscr{M}, \langle w, w' \rangle \models \phi$ iff $\mathscr{M}, \langle w', w' \rangle \models \psi$;

4. all other clauses are as expected.

As is clear from Definition 4, the first world in a world pair is the evaluation world, while the second world in the world pair is the reference world. It can easily be seen that, for every actuality model $\mathscr{M}$ of the type described in Definition 1, there exists an actuality model $\mathscr{M}^*$ of the type described in Definition 3 such that, for all worlds w in W and all formulas ϕ, $\mathscr{M}, w \models \phi$ iff $\mathscr{M}^*, \langle w, w_0 \rangle \models \phi$.

As before, there are two ways one could define truth in actuality models and, correspondingly, two ways of defining validity and satisfiability.

Definition 5 (General validity and satisfiability). For every formula ϕ, ϕ is generally valid ($\models \phi$) iff for every model $\mathscr{M}$ and for all $w, w' \in W$, $\mathscr{M}, \langle w, w' \rangle \models \phi$, and ϕ is generally satisfiable iff $\not\models \neg\phi$.

Definition 6 (Weak validity and satisfiability). For every formula ϕ, ϕ is weakly valid ($\models_{weak} \phi$) iff for every model $\mathscr{M}$ and for all $w \in W$,

$$\mathscr{M}, \langle w, w \rangle \models \phi,$$

and ϕ is weakly satisfiable iff $\not\models_{weak} \neg\phi$.

Example 1. The following formula is weakly valid but not generally valid:

$$@\phi \leftrightarrow \phi.$$

For dialectical purposes, I will try to make use of general validity for theorems about what is *valid* and of weak validity for theorems about what is *invalid* and, finally, of weak satisfiability for theorems about what is *satisfiable*. For every generally valid formula is also weakly valid (but not vice versa) and every weakly invalid formula is also generally invalid (but not vice versa) and, finally, every weakly satisfiable formula is also generally satisfiable (but not vice versa). This way the results are strongest.

---

[1]For a philosophical discussion of two competing ways of defining truth in a model of modal logic, see (Zalta, 1988), (Hanson, 2006) and (Nelson and Zalta, 2012).

3 Belief and actuality

In this section the straightforward combination of actuality models and Kripke models for belief will be shown to result in the problem of unwanted factivity and the problem of missing belief. Based on work by Rabinowicz and Segerberg (1994), an alternative combination will be proposed. This new combination avoids both problems. Finally, I will discuss something that may be seen as a remnant of the problem of missing belief.

3.1 The straightforward combination and two problems

Let us consider the language of sentential logic with a belief operator, $\mathscr{L}_B$. If ψ is a well-formed formula, then so is $B\psi$. The standard models for belief are Kripke models.[2]

Definition 7 (Kripke models for belief). A Kripke model for doxastic propositional logic is a tuple $\langle W, R_D, V \rangle$, with W a non-empty set, R_D a binary serial relation on W, and V a function from proposition letters and elements of W to truth-values.

Definition 8 (Truth at a world in Kripke models for belief). The relation

$$\mathscr{M}, w \models \phi$$

is defined inductively as follows:

1. if $\phi = B\psi$, then $\mathscr{M}, w \models \phi$ iff $\mathscr{M}, w' \models \psi$ for all w' such that $wR_D w'$;

2. all other clauses are as expected.

Let us now consider the language of sentential logic with a necessity operator, an actuality operator and a belief operator, $\mathscr{L}_{\Box, @, B}$. One can straightforwardly combine Kripke models for belief with actuality models.

Definition 9 (Simple belief and actuality models). A simple belief and actuality model is a tuple

$$\langle W, R_D, R_M, V \rangle,$$

with W a non-empty set of worlds, R_D a serial two-place relation on W, R_M a two-place relation on W, and V a function from proposition letters and worlds to truth-values.

Definition 10 (Truth in a simple belief and actuality model). The relation

$$\mathscr{M}, \langle w, w' \rangle \models \phi$$

is defined inductively as follows:

1. if $\phi = B\psi$, then $\mathscr{M}, \langle w, w' \rangle \models \phi$ iff $\mathscr{M}, \langle v, w' \rangle \models \psi$ for all v such that $wR_D v$;

2. the other clauses are as expected.

The simple belief and actuality model $\langle W, R_D, R_M, V \rangle$ can thus be seen as the result of merging an actuality model $\langle W, R_M, V \rangle$ with a Kripke model for belief $\langle W, R_D, V \rangle$.

The resulting models have philosophically undesirable consequences. It is almost universally acknowledged that beliefs are not in general factive. In the Kripke models the doxastic accessibility relation is not stipulated to be reflexive, which would turn belief into a factive notion. However, the straightforward combination of Kripke models for belief and actuality models results in a wide category of beliefs being factive after all.

---

[2]The classic source of possible world semantics for epistemic and doxastic logic is (Hintikka, 1962).

Theorem 1 (Unwanted factivity). $\models B@\phi \rightarrow @\phi$.

Proof. Suppose that $\mathcal{M}, \langle w, w' \rangle \models B@\phi$. Then for all v such that $wR_D v$, it is the case that $\mathcal{M}, \langle v, w' \rangle \models @\phi$. Consequently, $\mathcal{M}, \langle w', w' \rangle \models \phi$. Hence, $\mathcal{M}, \langle w, w' \rangle \models @\phi$. $\qquad\square$

So, any student who believes that actually the answer to the third exam question is such-and-such is correct in his or her belief that actually the answer to the third exam question is such-and-such. This does not sit well with the experience of exam graders.

A related problem was noted by Rabinowicz and Segerberg (1994). They studied the straightforward combination of Kripke models for *knowledge* (K) and *simple* actuality models in the context of Edgington (1985)'s reformulation of the verificationist claim in the light of the Church-Fitch (1963) paradox.[3] In Kripke models for knowledge the two-place serial relation R_D on W is replaced with R_E, a two-place reflexive relation on W. What Rabinowicz and Segerberg proved is that the combination yields omniscience of actual truths.

Theorem 2. $\models @\phi \rightarrow K@\phi$.

Proof. See (Rabinowicz and Segerberg, 1994, p. 128-129). $\qquad\square$

It can easily be checked that their result also holds for the combination of Kripke models for knowledge and the general actuality models in the sense of Definitions 3 and 4.

Let us briefly compare the problem of unwanted factivity and the problem of omniscience of actual truths. First, factivity of belief is unwanted, whereas the corresponding factivity of knowledge ($\models K@\phi \rightarrow @\phi$) *is* wanted. Second, note that I could also have proven that all actual truths are believed to be actual truths ($\models @\phi \rightarrow B@\phi$). So, it would seem that the problem of omniscience of actual truths is a more telling symptom than the problem of unwanted factivity. Yet, there is also a sense in which the problem of unwanted factivity is more revealing. After all, if one wants to avoid the problem of logical omniscience or the problem of belief in all logical truths, then a good option is to introduce explicit belief sets (Fagin and Halpern, 1988), i.e. one believes ϕ at a world only if ϕ belongs to the explicit belief set at that world. Such a move would block not only the derivation of logical omniscience but also the derivation of the belief in all actual truths. Still, it would be powerless against the problem of unwanted factivity.

Both the problem of unwanted factivity and the problem of belief in all actual truths are but symptoms of a root cause: the doxastic (or epistemic) accessibility relation R_D (or R_E) leaves the reference world, i.e. the second world of the world pair, fixed. As will become clear in the next subsection, if one somehow allows the reference world to change when considering doxastic (or epistemic) possibilities, the problem will disappear.

A second problem is that some beliefs go missing. On the one hand, for every generally valid formula, it is a general logical truth that it is believed. On the other hand, for *no* weakly but not generally valid formula (which does not contain the belief operator) is it a weak logical truth that it is believed. In particular, it is not a weak logical truth that it is believed that something is actually true if and only if true.

Theorem 3 (Missing belief - particular case). $\not\models_{weak} B(@\phi \leftrightarrow \phi)$.

Proof. Consider a model $\mathcal{M}$ with the following characteristics:

1. $W = \{w_1, w_2\}$;

2. $R_D = \{\langle w_2, w_1 \rangle, \langle w_1, w_1 \rangle\}$;

3. $V(p, w_1) = 1$ and $V(p, w_2) = 0$.

---

[3]This is not entirely correct: Rabinowicz and Segerberg (1994, p. 127-128) include in the models also a set of propositions that is closed under certain operations. These sets of propositions are not relevant for our current purposes, so we will not discuss them.

Then $\mathcal{M}, \langle w_2, w_2 \rangle \not\models B(@p \leftrightarrow p)$. $\qquad\square$

Theorem 4 (Missing belief - general case). $\not\models_{weak} B\phi$ *for any* $\phi \in \mathcal{L}_{\Box,@}$ *such that* $\models_{weak} \phi$ *but* $\not\models \phi$.

Proof. Suppose that $\models_{weak} \phi$ but $\not\models \phi$. Furthermore, $\phi \in \mathcal{L}_{\Box,@}$. Then there is a $\mathcal{M}$ and $w_1, w_2 \in W$ such that $\mathcal{M}, \langle w_1, w_2 \rangle \not\models \phi$. Now define $\mathcal{M}'$ as follows:

1. $W^{\mathcal{M}'} = W^{\mathcal{M}}$;

2. $R_D^{\mathcal{M}'} = R_D^{\mathcal{M}} \cup \{\langle w_2, w_1 \rangle\}$;

3. $R_M^{\mathcal{M}'} = R_M^{\mathcal{M}}$;

4. $V^{\mathcal{M}'}(p, w) = V^{\mathcal{M}}(p, w)$ for all $w \in W^{\mathcal{M}}$ and for all sentence letters p.

One can then easily prove by induction on the complexity of $\psi \in \mathcal{L}_{\Box,@}$ that, for all $w, v \in W^{\mathcal{M}}$:

$$\mathcal{M}, \langle w, v \rangle \models \psi \quad \Leftrightarrow \quad \mathcal{M}', \langle w, v \rangle \models \psi.$$

It follows that $\mathcal{M}', \langle w_1, w_2 \rangle \not\models \phi$. Given the definition of $R_D^{\mathcal{M}'}$, it is a consequence that $\mathcal{M}', \langle w_2, w_2 \rangle \not\models B\phi$. $\qquad\square$

On the one hand, one may not want the result that, for any generally valid formula, it is a general logical truth that it is believed. On the other hand, one may also not want the result that, for any weakly but not generally valid formula (which does not contain the belief operator), it is *not* a weak logical truth that it is believed. After all, there is something odd about not believing that, if it actually rains, then it rains, and vice versa. The first issue can be dealt with by introducing explicit belief sets, for instance. However, this is a powerless move against the second problem. Explicit belief sets can only shrink belief not expand it.

Both the problem of unwanted factivity and the problem of missing belief call for a solution. This will be provided in the next subsection.

3.2 A better combination

Rabinowicz and Segerberg (1994, Sect. 3) propose to redefine the epistemic accessibility relation R_E: it is no longer a relation between worlds but between *pairs* of worlds, which they call *states*. Their motivation for doing so (apart from solving the problem) is the following:

> What is the intuitive interpretation of the epistemic accessibility relation between states? When we consider a state $\langle w, v \rangle$, the set of states to which $\langle w, v \rangle$ bears the epistemic relation E delimits the range of knowledge obtaining in $\langle w, v \rangle$. When this set is smaller, the range of knowledge is larger – the epistemic uncertainty decreases. Now, in a given state, there are *two sources* of epistemic uncertainty. Uncertainty concerns in part the reference-point of the state, in part the world that constitutes the point of perspective. Our knowledge of the world that is referred to (= described) and of the world that constitutes the perspective is more or less limited. The relation $[R_E]$ is meant to model both types of uncertainty. This is why, when $\langle w, v \rangle [R_E] \langle w', v' \rangle$, we must allow not only that v' may differ from v but also that w' and w may differ. (Rabinowicz and Segerberg, 1994, p. 132)

They add a 'coincidence condition' on the newly defined epistemic accessibility relation: if at a state the evaluation world is the same as the reference world, then every epistemically accessible state is one in which the evaluation world is the same as the reference world. A state in which the evaluation world is the same as the reference world is called 'self-centered'.

When Rabinowicz and Segerberg (1994)'s solution is transposed to belief, the two problems noted in Subsection 3.1 are solved. Let us first for the sake of definiteness define the new models.

Definition 11 (Belief and actuality models)**.** A belief and actuality model is a tuple $\langle W, R_D, R_M, V \rangle$, with W, R_M and V as in simple belief and actuality models and with R_D a two-place serial relation on $W \times W$ such that, if $\langle w, w' \rangle R_D \langle v, v' \rangle$ and $w = w'$, then $v = v'$.

Definition 12 (Truth in belief and actuality models)**.** The relation $\mathscr{M}, \langle w, w' \rangle \models \phi$ is defined inductively as follows:

1. if $\phi = B\psi$, then $\mathscr{M}, \langle w, w' \rangle \models \phi$ iff $\mathscr{M}, \langle v, v' \rangle \models \psi$ for every v, v' such that

$$\langle w, w' \rangle R_D \langle v, v' \rangle;$$

2. the other clauses are as expected.

The problem of unwanted factivity disappears when we redefine the doxastic accessbility relation as a relation between pairs of worlds.

Theorem 5. $\not\models_{weak} B@\phi \to @\phi$.

Proof. Consider a model with the following characteristics:

1. $W = \{w_1, w_2\}$;

2. $R_D = \{\langle \langle w_1, w_1 \rangle, \langle w_2, w_2 \rangle \rangle, \langle \langle w_2, w_2 \rangle, \langle w_2, w_2 \rangle \rangle, \langle \langle w_1, w_2 \rangle, \langle w_1, w_2 \rangle \rangle, \langle \langle w_2, w_1 \rangle, \langle w_2, w_1 \rangle \rangle\}$;

3. $V(p, w_1) = 0$ and $V(p, w_2) = 1$

Then $\mathscr{M}, \langle w_1, w_1 \rangle \models B@p$ but $\mathscr{M}, \langle w_1, w_1 \rangle \not\models @p$. $\square$

The problem raised by Rabinowicz and Segerberg (1994) is solved in essentially the same way, but I leave that to the reader.

The problem of missing belief vanishes by imposing the coincidence condition on the doxastic accessibility relation.

Theorem 6.

$$\models_{weak} \phi \quad \Rightarrow \quad \models_{weak} B\phi,$$

and in particular,

$$\models_{weak} B(@\phi \leftrightarrow \phi).$$

Proof. See Rabinowicz and Segerberg (1994, p. 132). (Replace K with B in their argument.) $\square$

Both of the problems that showed up as a result of straightforwardly combining Kripke models for belief with models for actuality have been solved. Moreover, the solution proposed in this section is in a sense conservative.

Theorem 7. *For every* $\phi \in \mathscr{L}_{\square, B}$, *if there is an ordinary Kripke model* $\mathscr{M}$ *and a world w such that* $\mathscr{M}, w \models \phi$, *then there is a model* $\mathscr{M}'$ *in the sense of Definition 11 and a pair of worlds* $\langle v, v' \rangle$ *such that* $\mathscr{M}', \langle v, v' \rangle \models \phi$, *and vice versa.*

Proof. For the left-to-right direction, let $\mathscr{M}'$ be identical to $\mathscr{M}$, except that $R_D^{\mathscr{M}'}$ is defined as follows:

$$R_D^{\mathscr{M}'} = \{\langle \langle w, u \rangle, \langle v, v \rangle \rangle \mid \langle w, v \rangle \in R_D^{\mathscr{M}} \& u \in W^{\mathscr{M}}\}.$$

It can be checked that $R_D^{\mathcal{M}'}$ is a two-place serial relation on $W^{\mathcal{M}'} \times W^{\mathcal{M}'}$ that trivially satisfies the co-incidence relation. It is easy to see that $wR_D^{\mathcal{M}}v$ if and only if $\langle w,w \rangle R_D^{\mathcal{M}'} \langle v,v \rangle$. One can then prove by induction on the complexity of formulas that $\mathcal{M},w \models \phi$ if and only if $\mathcal{M},\langle w,w \rangle \models \phi$.

For the right-to-left direction, define W as follows:

$$W = \{w_{v,v'} \mid \langle v,v' \rangle \in W^{\mathcal{M}'} \times W^{\mathcal{M}'}\}.$$

Next, stipulate that $w_{v,v'} R_D^{\mathcal{M}'} w_{u,u'}$ if and only if $\langle v,v' \rangle R_D' \langle u,u' \rangle$. Furthermore, stipulate that $w_{v,v'} R_M^{\mathcal{M}'} w_{u,v'}$ if and only if $vR_M^{\mathcal{M}}u$. Finally, let $V^{\mathcal{M}'} = V^{\mathcal{M}}$. Then one can prove by induction on the complexity of formulas that $\mathcal{M}',\langle v,v' \rangle \models \phi$ if and only if $\mathcal{M},w_{v,v'} \models \phi$. $\qquad\square$

To conclude, the class of validities formulated in $\mathscr{L}_{\Box,B}$ remains the same, regardless of whether one adopts Kripke-style models or Rabinowicz and Segerberg (1994)-style models for belief.

3.3 More missing belief?

One of the two problems that has been solved is the problem of missing belief. Relative to definitions 9 and 10, one gets $\not\models_{weak} B(@\phi \leftrightarrow \phi)$ (Theorem 3) and, more generally, $\not\models_{weak} B\phi$, for any $\phi \in \mathscr{L}_{\Box,@}$ such that $\models_{weak} \phi$ but $\not\models \phi$ (Theorem 4). Relative to definitions 11 and 12, one gets that $\models_{weak} B(@\phi \leftrightarrow \phi)$ and, more generally, $\models_{weak} B\phi$, for any ϕ such that $\models_{weak} \phi$ (Theorem 6). Interestingly, one does *not* get that

$$\models B(@\phi \leftrightarrow \phi).$$

To establish the latter claim, note that in the model used in the proof of Theorem 5 the following holds: $\mathcal{M},\langle w_1,w_2 \rangle \not\models B(@p \leftrightarrow p)$. In order to get $\models B(@\phi \leftrightarrow \phi)$, it is sufficient to impose a stronger coincidence condition on the doxastic accessibility relation of models of the type described in Definition 11: if $\langle w,w' \rangle R_D \langle v,v' \rangle$, then $v = v'$. If $\models_{weak} \phi$, then ϕ is by definition true in all self-centered states in all models. Moreover, only self-centered states are accessible, because of the stronger coincidence condition. Therefore, from any given state in a model only states where ϕ is true are doxastically accessible. Hence, by definition, ϕ is believed in the given state.

Rabinowicz and Segerberg (1994) do not discuss the aforementioned possibility, but Holliday and Perry (2014, Sect. 4.5) advocate the stronger coincidence condition on the doxastic accessibility relation. However, it should be noted that they do this in the context of a discussion about models for a certain type of language that does not include the actuality operator. In fact, Holliday and Perry (2014) go one step further: if $\langle w,w' \rangle R_D \langle v,v' \rangle$, then $v = v'$ *and* $\langle w,u \rangle R_D \langle v,v' \rangle$ for all $u \in W$. Note the connection between their stipulation on the one hand and the definition of $R_D^{\mathcal{M}'}$ in the proof of the left-to-right direction of Theorem 7. The extra condition is also relevant for the right-to-left direction of a conservativeness theorem similar to Theorem 7. Let the Kripke model $\mathcal{M}'$ be in every respect identical to model $\mathcal{M}$, except that $wR_D^{\mathcal{M}'}v$ iff $\langle w,w \rangle R_D^{\mathcal{M}} \langle v,v \rangle$. Then one can prove that, for any $\phi \in \mathscr{L}_{\Box,B}$, $\mathcal{M}',w \models \phi$ iff $\mathcal{M},\langle w,w \rangle \models \phi$. Given the extra condition, the latter implies that, for any $\phi \in \mathscr{L}_{\Box,B}$ and for any $w' \in W$, $\mathcal{M}',w \models \phi$ iff $\mathcal{M},\langle w,w' \rangle \models \phi$. So, with the extra condition in place, the move from Kripke models to Rabinowicz and Segerberg (1994)-style models and back again is very straightforward.

Let us return to the main question by first turning our attention to the counterpart condition on the *epistemic* accessibility relation (i.e., if $\langle w,w' \rangle R_E \langle v,v' \rangle$, then $v = v'$). The crucial issue seems to be whether or not one accepts the general validity of the principle of the factivity of knowledge. Let K be the knowledge operator. If ψ is a well-formed formula, then so is $K\psi$. Then the question is whether or not one accepts the following:

$$\models K\phi \rightarrow \phi.$$

This corresponds to the question whether one should stipulate that the epistemic accessibility relation is reflexive or not, i.e. whether or not $\langle w,v \rangle R_E \langle w,v \rangle$ for all $w,v \in W$. Note that the issue is not the *weak*

validity of the principle of the factivity of knowledge. Indeed, all self-centered states can and should be epistemically accessible from themselves: if $w = w'$, then $\langle w, w' \rangle R_E \langle w, w' \rangle$. So, we still have

$$\models_{weak} K\phi \to \phi.$$

The stronger coincidence condition on R_E is not compatible with the reflexivity condition on R_E. Suppose that $w \neq w'$. Then the stronger condition implies that it is *not* the case that $\langle w, w' \rangle R_E \langle w, w' \rangle$. Yet this is required by the reflexivity condition. One might be willing to forego the general validity of the principle of the factivity of knowledge in order to retain the general validity of $K(\phi \leftrightarrow @\phi)$.

Let us have a look at the extent of the failure of the general validity of the principle of the factivity of knowledge. First, stipulate that, if $\langle w, w' \rangle R_E \langle v, v' \rangle$, then $v = v'$ *and* $\langle w, u \rangle R_E \langle v, v' \rangle$ for all $u \in W$, and if $w = w'$, then $\langle w, w' \rangle R_E \langle w, w' \rangle$. The first conditional is the epistemic counterpart of the *full* condition on the doxastic accessibility relation that is put forward by Holliday and Perry (2014). The second conditional is the condition needed for the *weak* validity of the principle of the factivity of knowledge. Then it follows that $\langle w, w' \rangle R_E \langle w, w \rangle$. Given the full (Holliday and Perry, 2014)-condition on the doxastic *and* the epistemic accessibility relations, one can prove that, for all $\phi \in \mathscr{L}_{\Box, B, K}$, for all models $\mathscr{M}$ and worlds w_i, w_j, $\mathscr{M}, \langle w_i, w_j \rangle \models \phi$ iff $\mathscr{M}, \langle w_i, w_i \rangle \models \phi$. The proof is left to the reader. Consequently, $\models K\phi \to \phi$ for all $\phi \in \mathscr{L}_{\Box, B, K}$. The situation changes when one adds the actuality operator.

Consider a model $\mathscr{M}$ with only two worlds, w_1 and w_2, with $V(p, w_1) = 0$, $V(p, w_2) = 1$, with R_E the relation that relates every state $\langle w_i, w_j \rangle$ to $\langle w_i, w_i \rangle$ and only that state, and with R_M the universal relation. Note that, if $\langle w_i, w_j \rangle R_E \langle w_k, w_l \rangle$, then $w_k = w_l$ and $\langle w_i, w_m \rangle R_E \langle w_k, w_l \rangle$ for all $w_m \in W$. In other words, the full condition put forward by Holliday and Perry (2014) is satisfied. Then $\mathscr{M}, \langle w_1, w_2 \rangle \models K(p \leftrightarrow @p) \wedge (\neg p \wedge @p)$, and

$$\mathscr{M}, \langle w_2, w_2 \rangle \models \Diamond \left(K(p \leftrightarrow @p) \wedge (\neg p \wedge @p) \right).$$

So, an instance of $K\phi \wedge \neg\phi$ is generally satisfiable and an instance of $\Diamond(K\phi \wedge \neg\phi)$ is weakly satisfiable. Failure of factivity is not confined to 'mixed' formulas, i.e. formulas that have subformulas that are not within the scope of an actuality operator (e.g., p) and subformulas that are within the scope of an actuality operator (e.g., $@p$). Indeed, $\mathscr{M}, \langle w_2, w_1 \rangle \models K@p \wedge \neg@p$. So, $K@p \wedge \neg@p$ is generally satisfiable and $\Diamond(K@p \wedge \neg@p)$ is weakly satisfiable.

To sum up, one has to make a choice between the stronger coincidence condition on the epistemic accessibility relation on the one hand and the reflexivity condition on that same relation on the other hand, since they cannot be at the same time satisfied. The reflexivity condition corresponds to the general validity of the factivity of knowledge principle. As we have seen, it is possible to retain the stronger coincidence condition and reject the *general* validity of the *unrestricted* factivity principle without sacrificing, first, the *weak* validity of the factivity principle and, second, the general validity of the restriction of the factivity principle to formulas without the actuality operator. However,

$$\Diamond \left(K(p \leftrightarrow @p) \wedge (\neg p \wedge @p) \right)$$

and

$$\Diamond \left(K@p \wedge \neg@p \right)$$

are weakly satisfiable. The other option is to retain the reflexivity condition and to reject the stronger coincidence condition. As we have seen in Subsection 3.2, one can do this without rejecting the *weak* validity of $K(@\phi \leftrightarrow \phi)$. For that one only needs the weaker coincidence condition of Definition 11.

If one retains the reflexivity condition and rejects the stronger coincidence condition on the *epistemic* accessibility relation, then one has reason to reject the stronger coincidence condition on the *doxastic* accessibility relation as well. Suppose that one adheres to the general validity of the principle that knowledge (materially) implies belief ($\models K\phi \to B\phi$). In the models under consideration this boils down to stipulating that, if $\langle w, w' \rangle R_D \langle v, v' \rangle$, then $\langle w, w' \rangle R_E \langle v, v' \rangle$. Furthermore, assume that one rejects the stronger

coincidence condition on the *epistemic* accessibility relation. If the epistemic accessibility relation does not satisfy the stronger coincidence condition, then it is generally satisfiable that $K(\neg p \wedge @p)$. To see this, consider a model $\mathscr{M}$ with only two worlds, w_1 and w_2, with $V(p,w_1) = 0$, $V(p,w_2) = 1$, and with R_E the relation that relates every state to itself and only to itself. Then $\mathscr{M}, \langle w_1, w_2 \rangle \models K(\neg p \wedge @p)$. If knowledge implies belief, it follows that it is also generally satisfiable that $B(\neg p \wedge @p)$. But if the doxastic accessibility relation satisfies the stronger coincidence condition, then it is generally valid that $\neg B(\neg p \wedge @p)$. Contradiction. So, on the plausible assumption that it is generally valid that knowledge implies belief, giving up the stronger coincidence condition on the *epistemic* accessibility relation forces one to give up the stronger coincidence condition on the *doxastic* accessibility relation.

In summary, the question arose whether the problem of missing belief has only been solved to some extent. In Subsection 3.2 the weak validity of $B(@\phi \leftrightarrow \phi)$ has been restored, but not its general validity. The first has been accomplished by laying down a coincidence condition for the doxastic accessibility condition. The second can be accomplished by stipulating a stronger coincidence condition. However, if one accepts the general validity of the principle that knowledge (materially) implies belief, and if one accepts the general validity of the factivity of knowledge principle, then one cannot go down that route. Rather than exploring the options, I would like to point to a crucial presupposition in the above argument, namely that knowledge can be adequately modelled using accessibility relations only. It is because of this presupposition that it was easy to prove that, relative to models without the stronger coincidence condition on the epistemic accessibility relation, it is generally satisfiable that $K(\neg p \wedge @p)$. In the next section different kinds of models for knowledge will be used. We will then return to this issue.

4 Counterfactual belief and actuality

In this section the straightforward combination of actuality models and Holliday (2012a) models for counterfactual belief, e.g. safe and sensitive belief, will be shown to result in the problem of collapsing knowledge and the problem of missing knowledge. The type of solution used in Subsection 3.2 will be used to solve the two new problems as well. Finally, I will discuss something that may be seen as a remnant of the problem of missing knowledge.

4.1 The straightforward combination and two problems

Let us consider the language of sentential logic with a necessity operator, an actuality operator, a belief operator and a knowledge operator, $\mathscr{L}_{\Box, @, B, K}$. Holliday (2012a) provides a landmark study of the formal properties of counterfactual accounts of knowledge. In what follows I will make use of Holliday models of counterfactual belief.

Definition 13 (Simple counterfactual belief model)**.** A simple counterfactual belief model $\mathscr{M}$ is a tuple $\langle W, R_D, \leq, V \rangle$, with W a non-empty set, R_D a binary serial relation on W, $\leq$ assigns to each $w \in W$ a binary, reflexive, transitive, well-founded relation $\leq_w$ on some $W_w \subseteq W$ such that $w \in W_w$ and for all $w' \in W_w$, $w \leq_w w'$, and V is a function from proposition letters and elements of W to truth-values.

Definition 14 (Closest S-worlds)**.** Let $S \subseteq W$. Then $Min_{\leq_w}(S) =_{df} \{v \in S \cap W_w \mid \neg \exists u \in S : u <_w v\}$, with $u <_w v$ iff $u \leq_w v$ but not $v \leq_w u$.

Definition 15 (Extension and complement)**.** Let $[\![\phi]\!]_x^{\mathscr{M}} =_{df} \{w \in W \mid \mathscr{M}, w \models_x \phi\}$ and let $\overline{[\![\phi]\!]}_x^{\mathscr{M}}$ be its complement in W.

Definition 16 (Truth in simple counterfactual belief models)**.** The relation

$$\mathscr{M}, w \models_x \phi$$

(with $x \in \{h, n, s\}$) is defined inductively as follows:

1. if $\phi = B\psi$, then $\mathscr{M}, w \models_x \phi$ iff $\mathscr{M}, w' \models_x \psi$ for all w' such that wR_Dw';

2. if $\phi = K\psi$, then $\mathscr{M}, w \models_h \phi$ iff $\mathscr{M}, w \models_h B\psi$ and

   (sensitivity) for all $w' \in Min_{\leq_w}\left(\overline{\llbracket \psi \rrbracket}_h^{\mathscr{M}}\right)$, $\mathscr{M}, w' \not\models_h B\psi$;

3. if $\phi = K\psi$, then $\mathscr{M}, w \models_n \phi$ iff $\mathscr{M}, w \models_n B\psi$ and

   (sensitivity) for all $w' \in Min_{\leq_w}\left(\overline{\llbracket \psi \rrbracket}_n^{\mathscr{M}}\right)$, $\mathscr{M}, w' \not\models_n B\psi$;

   (adherence) for all $w' \in Min_{\leq_w}\left(\llbracket \psi \rrbracket_n^{\mathscr{M}}\right)$, $\mathscr{M}, w' \models_n B\psi$ or, equivalently,
       for all $w' \in Min_{\leq_w}(W)$, $\mathscr{M}, w' \models_n \psi \to B\psi$;

4. if $\phi = K\psi$, then $\mathscr{M}, w \models_s \phi$ iff $\mathscr{M}, w \models_s B\psi$ and

   (safety) for all $w' \in Min_{\leq_w}\left(\llbracket B\psi \rrbracket_s^{\mathscr{M}}\right)$, $\mathscr{M}, w' \models_s \psi$ or, equivalently,
       for all $w' \in Min_{\leq_w}(W)$, $\mathscr{M}, w' \models_s B\psi \to \psi$;

5. all other clauses are as expected.

The indices h, n, s refer to Heller (1989, 1999), Nozick (1981) and Sosa (1999) respectively. The index x will be used as a variable that can take the values h, n, s. Heller claims that sensitivity is a necessary condition for knowledge. Nozick makes the same claim about sensitivity *and* adherence and Sosa does the same for safety. They respectively claim that sensitivity, sensitivity and adherence, or safety are necessary conditions for knowledge. Neither Heller (1999) nor Sosa (1999) thinks that sensitivity or safety is *sufficient* for knowledge. In what follows, I will take care to indicate which direction is needed to prove what.

One can straightforwardly combine Holliday models for counterfactual belief with actuality models. In doing so we apply the lesson learnt in Section 3.

Definition 17 (Simple counterfactual belief and actuality model). A simple counterfactual belief and actuality model $\mathscr{M}$ is a tuple $\langle W, R_D, R_M \leq, V \rangle$, with W, $\leq$ and V as in simple counterfactual belief models and with R_D and R_M as in belief and actuality models.

Definition 18 (Extension and complement). Let

$$\llbracket \phi \rrbracket_x^{\mathscr{M}, w'} =_{df} \{w \in W \mid \mathscr{M}, \langle w, w' \rangle \models_x \phi\}$$

and let $\overline{\llbracket \phi \rrbracket}_x^{\mathscr{M}, w'}$ be its complement in W.

Definition 19 (Truth in simple counterfactual belief and actuality models). The relation

$$\mathscr{M}, \langle w, w' \rangle \models_x \phi$$

is defined inductively as follows:

1. if $\phi = B\psi$, then $\mathscr{M}, \langle w, w' \rangle \models_x \phi$ iff $\mathscr{M}, \langle v, v' \rangle \models_x \psi$ for all v, v' such that

$$\langle w, w' \rangle R_D \langle v, v' \rangle;$$

2. if $\phi = K\psi$, then $\mathscr{M}, \langle w, w' \rangle \models_h \phi$ iff $\mathscr{M}, \langle w, w' \rangle \models_h B\psi$ and

   (sensitivity) for all $v \in Min_{\leq_w}\left(\overline{\llbracket \psi \rrbracket}_h^{\mathscr{M}, w'}\right)$, $\mathscr{M}, \langle v, w' \rangle \not\models_h B\psi$;

11

3. if $\phi = K\psi$, then $\mathscr{M}, \langle w, w' \rangle \models_n \phi$ iff $\mathscr{M}, \langle w, w' \rangle \models_n B\psi$ and

   (sensitivity) for all $v \in Min_{\leq_w} \left(\overline{[\![\psi]\!]_n^{\mathscr{M}, w'}} \right)$, $\mathscr{M}, \langle v, w' \rangle \not\models_n B\psi$;

   (adherence) for all $v \in Min_{\leq_w} \left([\![\psi]\!]_n^{\mathscr{M}, w'} \right)$, $\mathscr{M}, \langle v, w' \rangle \models_n B\psi$ or, equivalently,

   for all $v \in Min_{\leq_w} (W)$, $\mathscr{M}, \langle v, w' \rangle \models_n \psi \to B\psi$;

4. if $\phi = K\psi$, then $\mathscr{M}, \langle w, w' \rangle \models_s \phi$ iff $\mathscr{M}, \langle w, w' \rangle \models_s B\psi$ and

   (safety) for all $v \in Min_{\leq_w} \left([\![B\psi]\!]_s^{\mathscr{M}, w'} \right)$, $\mathscr{M}, \langle v, w' \rangle \models_s \psi$ or, equivalently,

   for all $v \in Min_{\leq_w} (W)$, $\mathscr{M}, \langle v, w' \rangle \models_s B\psi \to \psi$;

5. the other clauses are as expected.

Two problems emerge. The first problem is one of overgeneration: a certain kind of knowledge comes on the cheap. The second problem is one of undergeneration: a certain kind of knowledge is lost.

Knowledge of actual truths comes at low cost. If something is actually the case and you believe it to be actually the case, then you know it to be actually the case. This collapses the notion of knowledge of actual truths into the notions of actual truth and belief in actual truth.

Theorem 8. *For any ϕ, $\models_{h,s} (@\phi \wedge B@\phi) \to K@\phi$.*

Proof. Suppose that $\mathscr{M}, \langle w, w' \rangle \models_{h,s} @\phi \wedge B@\phi$. Then $\mathscr{M}, \langle w', w' \rangle \models_{h,s} \phi$. Hence, for all $v \in Min_{\leq_w} (W)$, it is the case that $\mathscr{M}, \langle v, w' \rangle \models_s @\phi$. So, the safety condition is satisfied. Moreover,

$$ Min_{\leq_w} \left(\overline{[\![@\phi]\!]_h^{\mathscr{M}, w'}} \right) = \emptyset, $$

so that the sensitivity condition is also satisfied. $\square$

The above result does not extend to Nozick's theory, if in the latter the clause for belief is suitably modified, as in Subsection 3.2. This is left as an exercise to the reader.

It may be objected that the proof of Theorem 8 makes use of sensitivity or safety as *sufficient* conditions for knowledge. There are putative counterexamples to the sufficiency of the safety condition – see (Lackey, 2006, p. 288) and (Pritchard, 2012, p. 260). But even if one reads the K-operator not as a knowledge operator but as a sensitive belief or safe belief operator Theorem 8 is problematic in that the true belief of the student that actually the answer to the third exam question is such-and-such entails that the student sensitively believes or safely believes that actually the answer to the third exam question is such-and-such.

It is noteworthy that the problem is not due to the fact that actual truths are necessarily actual truths ($\models @\phi \to \Box @\phi$). In the relevant models it is *not* in general the case that worlds close to a given world are modally accessible from that world. For that reason the models don't support

$$ (\Box\phi \wedge B\phi) \to K\phi. $$

Of course, adding the stipulation that the worlds close to a given world are modally accessible from that world may be reasonable. However, the fact that the problem surfaces even without that stipulation is telling. Again, the root cause seems to be that the reference world remains fixed when moving to close worlds.

It is worth emphasizing that, unlike the problem made explicit in Theorem 2, adding explicit belief sets would not make a difference. In addition, it should be remarked that the problem does not disappear

when safety, sensitivity and adherence are relativized to methods or, more precisely, if safety, sensitivity and adherence are formulated in terms of beliefs based on methods. Epistemologists have defended this idea for independent reasons – see e.g. (Nozick, 1981) and (Luper-Foy, 1984). Holliday (2012b, p. 98-104) gives a formal rendering of this idea, but the details do not matter. Let us denote 'belief that actually ϕ via method b' as $B(@\phi, b)$. It is easy to check that, if one systematically replaces $B@\phi$ by $B(@\phi, b)$ in Theorem 8 and its proof, the proof still goes through. The reason is that, as long as moving to a close world does not change the reference world, $@\phi$ will remain true and, hence, the safety and sensitivity conditions are trivially satisfied. The consequent of the safety condition is always true (relative to the given reference world) and, therefore, the safety condition as well. Moreover, there is not going to be a closest world at which $@\phi$ is false (relative to the given reference world), so the sensitivity condition is trivially satisfied. The semantics of $B(\phi, b)$ does not need to do any work here.

By redefining the doxastic accessibility relation, it is a weak logical truth that one *believes* that something is actually the case if and only if it is the case. Notwithstanding this it is *not* a weak logical truth that one *knows* that something is actually the case if and only if it is the case.

Theorem 9 (Missing knowledge - particular case 1). $\not\models_{weak,x} K(@\phi \leftrightarrow \phi)$.

Proof. Consider a model $\mathcal{M}$ with the following characteristics:

1. $W = \{w_1, w_2\}$;

2. $R_D = \{\langle\langle w_2, w_1, \rangle, \langle w_1, w_1\rangle\rangle\} \cup \{\langle\langle w_i, w_j\rangle, \langle w_i, w_j\rangle\rangle \mid \neg(i = 2 \wedge j = 1) \wedge w_i, w_j \in W\}$;

3. $V(p, w_1) = 1; V(p, w_2) = 0$;

4. $Min_{\leq_{w_1}}(W) = \{w_1, w_2\}$.

Then $\mathcal{M}, \langle w_1, w_1\rangle \not\models_s K(@p \leftrightarrow p)$, since $\mathcal{M}, \langle w_2, w_1\rangle \models_x B(@p \leftrightarrow p)$ but

$$\mathcal{M}, \langle w_2, w_1\rangle \not\models_s @p \leftrightarrow p.$$

Furthermore, note that $Min_{\leq_{w_1}}\left(\overline{[\![@p \leftrightarrow p]\!]_{h,n}^{\mathcal{M}, w_1}}\right) = \{w_2\}$. So,

$$\mathcal{M}, \langle w_1, w_1\rangle \not\models_{h,n} K(@p \leftrightarrow p).$$

$\square$

Although I do not have a proof for the general case, I do have a proof for the fact that infinitely many cases of weak but not generally valid formulas are not known at every self-centered world pair.

Theorem 10 (Missing knowledge - particular case 2).

$$\not\models_{weak,x} KB^n(@\phi \leftrightarrow \phi),$$

with B^n standing for n iterations of the belief operator B for any $n \in \mathbb{N}$.

Proof. Consider a model $\mathcal{M}$ with the following characteristics:

1. $W = \{w_1, w_2, w_3\}$;

2. $R_D = \{\langle\langle w_3, w_1\rangle, \langle w_2, w_1\rangle\rangle, \langle\langle w_2, w_1, \rangle, \langle w_1, w_1\rangle\rangle\} \cup \{\langle\langle w_i, w_j\rangle, \langle w_i, w_j\rangle\rangle \mid \neg((i = 2 \vee i = 3) \wedge j = 1) \wedge w_i, w_j \in W\}$;

3. $V(p, w_1) = 1; V(p, w_2) = 0$;

13

4. $Min_{\leq_{w_1}}(W) = \{w_1, w_3\}$.

It is then easy to prove that $\mathcal{M}, \langle w_1, w_1 \rangle \not\models_x KB(@p \leftrightarrow p)$. To prove the same for $KBB(@p \leftrightarrow p)$ we can use a structurally similar model with four worlds. In general, one can use a structurally similar model with $n+2$ worlds for $KB^n(@p \leftrightarrow p)$. $\qquad\square$

On the one hand, one may not want the result that, for any generally valid formula, it is a general logical truth that it is known. On the other hand, one may also not want the result that, for infinitely many weakly but not generally valid formulas, it is *not* a weak logical truth that they are known. After all, there is something odd about not knowing that, if it actually rains, then it rains, and vice versa. The first issue can be dealt with by introducing explicit belief sets for instance. However, this is a powerless move against the second problem. Note that the problem requires only safety or sensitivity as *necessary* conditions for knowledge.

As before, adding explicit belief sets does nothing to solve the problem. Relativizing beliefs to methods is a more interesting move. A bit more detail is needed to discuss it properly. Holliday (2012b, p. 99) adds to the simple counterfactual belief models a function $\mathfrak{B}$ that assigns to each combination of a formula ϕ and a method term b a subset of W (subject to certain condition that do not matter for this discussion). The clause for belief then becomes the following:

- $\mathcal{M}, w \models_x B(\phi, b)$ iff $w \in \mathfrak{B}(\phi, b)$ and $\mathcal{M}, w' \models_x \phi$ for all w' such that wR_Dw'.

A first observation is that the definition of $\mathfrak{B}$ has to be reformulated in terms of *world pairs* rather than single worlds. There is no point in stipulating that:

- $\mathcal{M}, \langle w, w' \rangle \models_x B(\phi, b)$ iff $w \in \mathfrak{B}(\phi, b)$ and $\mathcal{M}, \langle v, v' \rangle \models_x \phi$ for all $\langle v, v' \rangle$ such that $\langle w, w' \rangle R_D \langle v, v' \rangle$.

For consider the proof of Theorem 9. If one wants to block the proof by stipulating that

$$w_2 \notin \mathfrak{B}(@p \leftrightarrow p, b),$$

then

$$\mathcal{M}, \langle w_2, w_1 \rangle \not\models_x B(@p \leftrightarrow p, b)$$

but also $\mathcal{M}, \langle w_2, w_2 \rangle \not\models_x B(@p \leftrightarrow p, b)$. Consequently,

$$\mathcal{M}, \langle w_2, w_2 \rangle \not\models_x K(@p \leftrightarrow p).$$

But let us suppose that $\mathfrak{B}$ is redefined for world pairs rather than single worlds. Even this would not suffice. One would need to impose certain conditions that exclude, say, $\langle w_2, w_1 \rangle \in \mathfrak{B}(@p \leftrightarrow p, b)$ in a revised proof of Theorem 9. Let us suppose for the sake of the argument that one can come up with such conditions. It will not matter. For consider a variation on the model used in the proof of Theorem 9. Add a world w_3 and let $V(p, w_3) = V(p, w_1)$ and stipulate that $\langle w_3, w_1 \rangle R_D \langle w_2, w_1 \rangle$ and that $w_3 \leq_{w_1} w_1$. It follows that $\mathcal{M}, \langle w_3, w_1 \rangle \models_x @p \leftrightarrow p$, while $\mathcal{M}, \langle w_3, w_1 \rangle \not\models_x B(@p \leftrightarrow p, b)$. This is completely independent from conditions on $\mathfrak{B}$, since it only involves the doxastically accessible worlds. But then $\mathcal{M}, \langle w_1, w_1 \rangle \not\models_n K(@p \leftrightarrow p)$. So, the conclusion still holds for the Nozick account, which includes the adherence condition. This is significant, because the Heller and Sosa accounts succumbed to the problem of collapsing knowledge, whereas the Nozick account escaped that problem. Even if one could manage to save the Heller and Sosa accounts from the problem of missing knowledge by appealing to methods, this is not going to save the Nozick account. So, in the end all three accounts face at least one major problem.

Both the problem of collapsing knowledge and the problem of missing knowledge require a solution. This will be provided in the next subsection.

14

4.2 A better combination

In the case of the problems of unwanted factivity and missing belief the source was traced back to the fact that the reference world remained fixed even when doxastic alternatives were considered. My diagnosis of the problems of collapsing knowledge and missing knowledge is analogical: the source of those problems is that the reference world remains fixed even when close worlds are considered. The solution is also analogous to the one introduced in Subsection 3.2. The key idea is to redefine the closeness relation not as a relation between worlds but between pairs of worlds. In addition, one also needs to add a kind of coincidence condition.

Definition 20 (Counterfactual belief and actuality model)**.** A counterfactual belief and actuality model $\mathscr{M}$ is a tuple $\langle W, R_D, R_M \leq, V \rangle$, with W, R_D, R_M and V as in a simple safety and actuality model and with $\leq$ a function that assigns to each $\langle w, w' \rangle \in W \times W$ a binary, reflexive, transitive, well-founded relation $\leq_{w,w'}$ on some $W_{w,w'} \subseteq W \times W$ such that:

1. $\langle w, w' \rangle \in W_{w,w'}$ and for all $\langle v, v' \rangle \in W_{w,w'}$, $\langle w, w' \rangle \leq_{w,w'} \langle v, v' \rangle$;

2. if $\langle v, v' \rangle \leq_{w,w'} \langle u, u' \rangle$ and $w = w'$, then $u = u'$.

Definition 21 (Closest world pairs)**.** Let $S \subseteq W \times W$. Then

$$Min_{\leq_{w,w'}}(S) =_{df} \{\langle v, v' \rangle \in S \cap W_{w,w'} \mid \neg \exists \langle u, u' \rangle \in S : \langle u, u' \rangle <_{w,w'} \langle v, v' \rangle\},$$

with $\langle u, u' \rangle <_{w,w'} \langle v, v' \rangle$ iff $\langle u, u' \rangle \leq_{w,w'} \langle v, v' \rangle$ but not $\langle v, v' \rangle \leq_{w,w'} \langle u, u' \rangle$.

Definition 22 (Extension and complement)**.** Let $[\![\phi]\!]_x^{\mathscr{M}} =_{df} \{\langle w, w' \rangle \in W \times W \mid \mathscr{M}, \langle w, w' \rangle \models_x \phi\}$ and let $\overline{[\![\phi]\!]}_x^{\mathscr{M}}$ be its complement in $W \times W$.

Definition 23 (Truth in a counterfactual belief and actuality model)**.** The relation

$$\mathscr{M}, \langle w, w' \rangle \models \phi$$

is defined inductively as follows:

1. if $\phi = K\psi$, then $\mathscr{M}, \langle w, w' \rangle \models_h \phi$ iff $\mathscr{M}, \langle w, w' \rangle \models_h B\psi$ and

   (sensitivity) for all $\langle v, v' \rangle \in Min_{\leq_{w,w'}}\left(\overline{[\![\psi]\!]}_h^{\mathscr{M}}\right)$, $\mathscr{M}, \langle v, v' \rangle \not\models_h B\psi$;

2. if $\phi = K\psi$, then $\mathscr{M}, \langle w, w' \rangle \models_n \phi$ iff $\mathscr{M}, \langle w, w' \rangle \models_n B\psi$ and

   (sensitivity) for all $\langle v, v' \rangle \in Min_{\leq_{w,w'}}\left(\overline{[\![\psi]\!]}_n^{\mathscr{M}}\right)$, $\mathscr{M}, \langle v, v' \rangle \not\models_n B\psi$;

   (adherence) for all $\langle v, v' \rangle \in Min_{\leq_{w,w'}}\left([\![\psi]\!]_n^{\mathscr{M}}\right)$, $\mathscr{M}, \langle v, v' \rangle \models_n B\psi$ or, equivalently,

   for all $\langle v, v' \rangle \in Min_{\leq_w}(W)$, $\mathscr{M}, \langle v, v' \rangle \models_s \psi \rightarrow B\psi$;

3. if $\phi = K\psi$, then $\mathscr{M}, \langle w, w' \rangle \models_s \phi$ iff $\mathscr{M}, \langle w, w' \rangle \models_s B\psi$ and

   (safety) for all $\langle v, v' \rangle \in Min_{\leq_{w,w'}}\left([\![B\psi]\!]_s^{\mathscr{M}}\right)$, $\mathscr{M}, \langle v, v' \rangle \models_s \psi$ or, equivalently,

   for all $\langle v, v' \rangle \in Min_{\leq_{w,w'}}(W)$, $\mathscr{M}, \langle v, v' \rangle \models_s B\psi \rightarrow \psi$;

4. the other clauses are as expected.

By redefining the relation of closeness as a relation between pairs of worlds, the problem of collapsing knowledge goes away.

Theorem 11. $\not\models_{weak,x} (@\phi \wedge B@\phi) \rightarrow K@\phi$.

Proof. Consider a model with the following characteristics:

1. $W = \{w_1, w_2, w_3\}$;

2. $R_D = \{\langle\langle w_1,w_1\rangle,\langle w_2,w_2\rangle\rangle, \langle\langle w_3,w_3\rangle,\langle w_2,w_2\rangle\rangle, \langle\langle w_2,w_2\rangle,\langle w_2,w_2\rangle\rangle\} \cup$
   $\{\langle\langle w_i,w_j\rangle,\langle w_i,w_i\rangle\rangle \mid i \neq j \wedge w_i, w_j \in W\}$;

3. $Min_{\leq_{w_1,w_1}} = \{\langle w_1,w_1\rangle, \langle w_3,w_3\rangle\}$;

4. $V(p,w_1) = V(p,w_2) = 1$ and $V(p,w_3) = 0$.

Then $\mathscr{M}, \langle w_1,w_1\rangle \models_x @p \wedge B@p$. But $\mathscr{M}, \langle w_3,w_3\rangle \models_x \neg @p \wedge B@p$. Therefore,

$$\mathscr{M}, \langle w_1,w_1\rangle \not\models_x K@p.$$

$\square$

By adding the coincidence condition on the new closeness relation while keeping the coincidence condition on the doxastic accessibility relation, the problem of missing knowledge goes away.

Theorem 12.
$$\models_{weak,x} \phi \quad \Rightarrow \quad \models_{weak,x} K\phi$$

and, in particular
$$\models_{weak,x} K(@\phi \leftrightarrow \phi).$$

Proof. Note that by the coincidence condition, for all $\langle v,v'\rangle$ such that $\langle w,w\rangle \leq_{w,w} \langle v,v'\rangle$, it is the case that $v = v'$ and therefore $\mathscr{M}, \langle v,v'\rangle \models_x \phi$, if $\models_{weak} \phi$. So, the safety and sensitivity conditions are satisfied. Moreover, for every $\langle u,u'\rangle$ such that $\langle w,w\rangle R_D \langle u,u'\rangle$, it is the case that $u = u'$. Therefore, $\mathscr{M}, \langle w,w\rangle \models_x B\phi$, given that $\models_{weak} \phi$. So, the belief and adherence conditions are satisfied as well.

$\square$

Both of the problems that showed up as a result of straightforwardly combining Holliday models for counterfactual belief with models for actuality have been solved. Again, the solution is conservative, since models in the sense of Definition 20 are modally equivalent with respect to $\mathscr{L}_{\square,B,K}$ to Holliday models.

Theorem 13. *For every ϕ belonging to $\mathscr{L}_{\square,B,K}$, if there is an ordinary Holliday model $\mathscr{M}$ and a world w such that $\mathscr{M}, w \models_x \phi$, then there is a model $\mathscr{M}'$ in the sense of Definition 20 and a world pair $\langle v,v'\rangle$ such that $\mathscr{M}', \langle v,v'\rangle \models_x \phi$, and vice versa.*

Proof. For the left-to-right direction, define $\leq_{w,w'}^{\mathscr{M}'}$ as follows:

1. if $w \neq w'$, then
$$\leq_{w,w'}^{\mathscr{M}'} = \{\langle\langle v,w'\rangle,\langle u,w'\rangle\rangle \mid \langle v,u\rangle \in \leq_w^{\mathscr{M}}\};$$

2. if $w = w'$, then
$$\leq_{w,w'}^{\mathscr{M}'} = \{\langle\langle v,v\rangle,\langle u,u\rangle\rangle \mid \langle v,u\rangle \in \leq_w^{\mathscr{M}}\}.$$

The rest is as in Subsection 3.2. For the right-to-left direction, define $\leq_{w_{v,v'}}^{\mathscr{M}}$ as follows:

$$\leq_{w_{v,v'}}^{\mathscr{M}} = \{\langle w_{u,u'}, w_{s,s'}\rangle \mid \langle\langle u,u'\rangle,\langle s,s'\rangle\rangle \in \leq_{v,v'}^{\mathscr{M}'}\}.$$

The rest is as in the proof of Theorem 7.

$\square$

So, the class of validities formulated in $\mathcal{L}_{\Box,B,K}$ remains the same, regardless of whether one adopts Holliday (2012a)-style models or the style models of Definition 20 for belief.

Theorem 13 is interesting for another reason as well. In the (Holliday, 2012a)-models the closeness relation holds between single worlds. According to Lewis (1973, p. 91) '[i]t is vague – very vague – but in a well-understood way'. The natural question to ask then is whether we understand the reconceptualized closeness relation, viz. a closeness relation holding between world pairs, including an evaluation world and a reference world. It is here that the proof of Theorem 13 is relevant. Note that the proof is constructive. Given any (Holliday, 2012a)-model with a well-understood closeness relation between single worlds, one can construct a unique model with a closeness relation between world pairs, where the latter closeness relation is completely determined by the former closeness relation. Moreover, given any model with a closeness relation between world pairs, one can construct a unique model with a closeness relation between single worlds, where the latter is completely determined by the former. So, if we understand closeness between single worlds, then we can understand closeness between pairs of worlds by the given construction.

4.3 More missing knowledge?

One of the two problems that has been solved is the problem of missing knowledge. Relative to definitions 17-19, one gets $\not\models_{weak,x} K(@\phi \leftrightarrow \phi)$ (Theorem 9) and more knowledge of the kind is missing in the same way (Theorem 10). Relative to definitions 20-23, one gets that $\models_{weak,x} K(@\phi \leftrightarrow \phi)$ and, more generally, $\models_{weak,x} K\phi$, for any ϕ such that $\models_{weak,x} \phi$ (Theorem 12). Interestingly, one does *not* get that

$$\models_x K(@\phi \leftrightarrow \phi).$$

In the model used in the proof of Theorem 11 the following holds: $\mathcal{M}, \langle w_2, w_3 \rangle \models_x p \wedge \neg @p$. Since sensitive or safe belief entails true belief, one gets that

$$\mathcal{M}, \langle w_2, w_3 \rangle \models_x \neg K(@p \leftrightarrow p).$$

It is noteworthy that in this model the doxastic accessiblity relation satisfies the stronger coincidence conditions, viz. if $\langle w, w' \rangle R_D \langle v, v' \rangle$, then $v = v'$. So, even though the stronger coincidence condition on the doxastic accessibility relation is sufficient for $\models B(@\phi \leftrightarrow \phi)$, it is not sufficient for $\models K(@\phi \leftrightarrow \phi)$.

One might contemplate stipulating a stronger coincidence condition on the closeness relation as well, viz. if $\langle v, v' \rangle \leq_{w,w'} \langle u, u' \rangle$, then $u = u'$. However, if $w \neq w'$, the stronger coincidence condition on the closeness relation conflicts with the stipulations that $\leq_{w,w'}$ is a *reflexive* relation on $W_{w,w'}$ and that $\langle w, w' \rangle \in W_{w,w'}$.

A possible solution consists in dropping $\langle w, w' \rangle$ from $W_{w,w'}$. Then $\leq$ can be redefined as follows: $\leq$ is a function that assigns to each $\langle w, w' \rangle \in W \times W$ a binary, reflexive, transitive relation $\leq_{w,w'}$ on some $W_{w,w'} \subseteq W \times W$ such that:

1. $\langle w, w \rangle \in W_{w,w'}$ and for all $\langle v, v' \rangle \in W_{w,w'}$, $\langle w, w \rangle \leq_{w,w'} \langle v, v' \rangle$;

2. if $\langle v, v' \rangle \leq_{w,w'} \langle u, u' \rangle$, then $u = u'$.

In effect, $W_{w,w'}$ now only contains self-centered states. In combination with the stronger coincidence condition on the doxastic accessibility relation, the foregoing is sufficient for obtaining the general validity of $K(@\phi \leftrightarrow \phi)$. The argument is left to the reader.

Recall that Holliday and Perry (2014) did not only stipulate the stronger coincidence condition on the doxastic accessibility relation, they also added the following extra condition: if $\langle w, w' \rangle R_D \langle v, v' \rangle$, then $\langle w, u \rangle R_D \langle v, v' \rangle$ for all $u \in W$. Inspired by this extra condition, one can stipulate that, if $\langle v, v \rangle \in W_{w,w'}$, then $\langle v, v \rangle \in W_{w,u}$ for all $u \in W$. Then $W_{w,w'} = W_{w,u}$. Note that this would make the left-to-right direction

of a conservativeness theorem similar to Theorem 13 easier, because there would not be any difference between $\leq_{w,w'}$ when $w \neq w'$ and when $w = w'$ any more. Moreover, only item 2 of the definition of $\leq_{w,w'}^{\mathscr{M}'}$ should be retained, since it makes all *relata* self-centered states. Furthermore, the right-to-left direction would also be easier: let $\mathscr{M}'$ be identical to $\mathscr{M}$ in every respect, except that $\leq_w^{\mathscr{M}'} = \{\langle v,u\rangle \mid \langle \langle v,v\rangle, \langle u,u\rangle\rangle \in \leq_{w,w}^{\mathscr{M}}\}$. With the extra condition in place, the move from Holliday models to models of the new type is very straightforward.

So, the above combination of stipulations is sufficient for making $K(@\phi \leftrightarrow \phi)$ *generally* valid, while at the same time keeping the closeness relation reflexive. Moreover, one can prove a conservativeness theorem in a very straightforward way. However, a consequence of the above solution is that the general validity of the principle of the factivity of knowledge is no longer guaranteed, because it is no longer the case that $\langle w, w'\rangle \in W_{w,w'}$. Let us have a quick glance at the extent of the failure of the principle of factivity.

First of all, the *weak* validity of the principle of the factivity of knowledge is upheld, since $\langle w,w\rangle \in W_{w,w}$. Second, knowledge of formulas that do not contain the actuality operator is also factive, if the *full* (Holliday and Perry, 2014)-condition is satisfied by the doxastic accessibility relation and if the analogous condition is satisfied by the closeness relation. Suppose that $\mathscr{M}, \langle w_1, w_2\rangle \models_x B\phi$. If $\langle w_1, w_2\rangle R_D \langle v,v\rangle$, then $\langle w_1, u\rangle R_D \langle v,v\rangle$ for all $u \in W$ and in particular w_1. Hence, $\mathscr{M}, \langle w_1, w_1\rangle \models_x B\phi$. Since $\langle w_1, w_1\rangle \in W_{w_1, w_1}$, it follows that $\mathscr{M}, \langle w_1, w_1\rangle \models_x \phi$. Suppose that, if $\langle v,v\rangle \in W_{w,w'}$, then $\langle v,v\rangle \in W_{w,u}$ for all $u \in W$. Then one can prove that, for all $\phi \in \mathscr{L}_{\Box, B, K}$, $\mathscr{M}, \langle w_i, w_j\rangle \models \phi$ iff $\mathscr{M}, \langle w_i, w_i\rangle \models \phi$. Therefore, $\mathscr{M}, \langle w_1, w_2\rangle \models_x \phi$.

Once more, the situation changes when the actuality operator is added. Consider a model with the following characteristics: $W = \{w_1, w_2\}$, $V(p, w_1) = 1$, $V(p, w_2) = 0$, $R_M = W \times W$,

$$R_D = \{\langle \langle w_i, w_j\rangle, \langle w_i, w_i\rangle\rangle \mid i, j \in \{1, 2\}\},$$

$\leq_{w_i, w_j} = \{\langle w_i, w_i\rangle\}$. In this model the doxastic accessibility relation and the closeness relation are of the type proposed above. Note that $\mathscr{M}, \langle w_1, w_2\rangle \models K(@p \leftrightarrow p) \wedge K@p$. But it is also the case that $\mathscr{M}, \langle w_1, w_2\rangle \models p \wedge \neg @p$.

One might think that the reason why the general validity of the factivity principle has to go has to do with the *specific* way in which the general validity of $K(@\phi \leftrightarrow \phi)$ has been backed up. But the tension between those two principles does not depend on most of the specifics of the models. Suppose that

$$\models K(@\phi \leftrightarrow \phi)$$

and that

$$\models K(@\phi \leftrightarrow \phi) \rightarrow (@\phi \leftrightarrow \phi).$$

Then

$$\models @\phi \leftrightarrow \phi.$$

But general validity is closed under necessitation. So, we get that

$$\models \Box(@\phi \leftrightarrow \phi).$$

Therefore,

$$\models \Box\Box @\phi \leftrightarrow \Box\phi.$$

But

$$\models \Box @\phi \leftrightarrow @\phi,$$

so

$$\models @\phi \leftrightarrow \Box\phi.$$

In other words, actual truths are equivalent to necessary truths. We end up with a collapse. So, the conflict between the general validity of the factivity principle on the one hand and the general validity of $K(@\phi \leftrightarrow \phi)$ on the other hand is not really tied to the specifics of the models.

Williamson (2000, p. 39, 271), one of the most influential defenders of the safety condition on knowledge, explicitly endorses that, for any proposition p, *necessarily*, if p is known, then p is true. This is a consequence of his theory that knowledge is the most general factive stative attitude. Moreover, Williamson (2000, p. 291, 292) is willing to apply the principle of the necessity of the factivity of knowledge in contexts in which the actuality operator occurs. The relevance for the issue at hand is that it follows from Williamson's commitment that he has to reject the general validity of $K(@\phi \leftrightarrow \phi)$. It is an application of the necessity of factivity principle that

$$\Box(K(@\phi \leftrightarrow \phi) \to (@\phi \leftrightarrow \phi)).$$

If one accepts the general validity of $K(@\phi \leftrightarrow \phi)$, one also has to accept that

$$\Box K(@\phi \leftrightarrow \phi).$$

As above, this will quickly lead to the equivalence of actual truths with necessary truths. So, Williamson should reject the general validity of $K(@\phi \leftrightarrow \phi)$. Of course, this leaves open that other defenders of counterfactual conditions on knowledge make another choice or that readers may choose differently.

As we have seen above, the stronger coincidence condition on the doxastic accessibility relation did not suffice for obtaining the general validity of $K(@\phi \leftrightarrow \phi)$. For this reason a stronger coincidence condition on the closeness relation was added. Those who would reject the general validity of $K(@\phi \leftrightarrow \phi)$ can do this by dropping just one of these stronger coincidence conditions. In particular, one could drop the stronger coincidence condition on the closeness relation while retaining the stronger coincidence condition on the doxastic accessibility relation. This option implies that $B(@\phi \leftrightarrow \phi)$ is generally valid. Since the general validity of the principle that knowledge (materially) implies belief is guaranteed by counterfactual belief models of knowledge, $K(\neg\phi \land @\phi)$ is *not* generally satisfiable.

In conclusion, one has to make a choice between the general validity of

$$K(@\phi \leftrightarrow \phi)$$

on the one hand and the general validity of the factivity of knowledge principle on the other hand, if one wants to avoid the equivalence of necessary truths with actual truths. If one chooses the first, one can still retain the weak validity of the factivity principle and the restriction of the factivity principle to formulas that do not contain the actuality operator. At least one of the champions of the safety condition on knowledge endorses the necessity of the factivity of knowledge, which also conflicts with the general validity of $K(@\phi \leftrightarrow \phi)$. By moving from Kripke-style models of knowledge to Holliday-style models of knowledge one can have

$$\models B(@\phi \leftrightarrow \phi),$$
$$\models K\phi \to \phi,$$

and

$$\models K\phi \to B\phi.$$

This is in contradistinction to the situation at the end of Subsection 3.3. It means we don't have to make a choice between three prima facie plausible principles. But in that case one does have to accept that

$$\not\models K(@\phi \leftrightarrow \phi),$$

although one will retain that

$$\models \neg K\neg(@\phi \leftrightarrow \phi).$$

The latter is important, because it was the satisfiability of $K(\neg p \land @p)$ that drove a wedge between the general validity of the principle that knowledge (materially) implies belief, the general validity of the factivity of knowledge principle, and the general validity of $B(@\phi \leftrightarrow \phi)$.

19

5 Conclusion

The straightforward combination of Kripke models for belief with actuality models leads to the problem of unwanted factivity: $\models B@\phi \to @\phi$ (Theorem 1). Moreover, it also yields the problem of missing belief (Theorem 3): $\not\models_{weak} B(@\phi \leftrightarrow \phi)$. In fact, for no weak validity of $\mathscr{L}_{\Box,@}$ is it weakly valid that it is believed (Theorem 4), despite the fact that for every general validity it is generally valid that it is believed.

A modification proposed by Rabinowicz and Segerberg (1994) was accepted as a solution to the noted problem. The solution consists of two parts. First, the doxastic accessibility relation is no longer a relation between single worlds but a relation between pairs of worlds, one of which is the evaluation world and one of which is the reference world. This solves the problem of unwanted factivity (Theorem 5). Second, a coincidence condition is imposed: only self-centered world pairs, in which the reference world is identical to the evaluation world, are doxastically accessible from other self-centered world pairs. This solves the problem of missing belief (Theorem 6). The proposed solution is conservative with respect to validities that do not contain the actuality operator (Theorem 7).

Even with the proposed modifications, there still is some missing belief:

$$\not\models B(@\phi \leftrightarrow \phi).$$

Strengthening the coincidence condition can fix this, but there is a choice to be made. One would have to give up $\models K\phi \to \phi$ or $\models K\phi \to B\phi$ (see Subsection 3.3). An important presupposition of the argument is that knowledge can be modelled with an accessibility relation only. In the second part of the article we have looked at different kinds of models for knowledge.

The straightforward combination of Holliday models for knowledge with actuality models (incorporating the earlier solution for the problem of unwanted factivity) leads to the problem of collapsing knowledge (Theorem 8): $\models_{h,s} (@\phi \wedge B@\phi) \to K@\phi$. The indices refer to the theories of Heller (1989, 1999) and Sosa (1999). I have argued that the problem is not a special case of the problem of knowledge of necessary truths. A second problem is the problem of missing knowledge (9): $\not\models_{weak,x} K(@\phi \leftrightarrow \phi)$. (See also Theorem 10.) The index refers to the fact that it not only affects the theories of Heller (1989, 1999) and Sosa (1999) but also the theory of Nozick (1981).

The combination of the two problems cannot be solved by appeal to methods for belief formation. I have proposed a solution similar to the one used before. First, the closeness relation is no longer a relation between single worlds but between world pairs. This solves the problem of collapsing knowledge (Theorem 11). Second, a coincidence condition was laid down: given a self-centered world pair as starting point, only self-centered world pairs are in the closeness relation to the given world pair. This solves the problem of missing knowledge (Theorem 12). It was demonstrated that the solution put forward is conservative with regard to valid formulas that are free of the actuality operator (Theorem 13). The proof of Theorem 13 turned out to be relevant for our understanding of the relation between closeness between single worlds on the one hand and closeness between world pairs on the other hand.

Even with the proposed modifications, there is still some missing knowledge:

$$\not\models K(@\phi \leftrightarrow \phi).$$

Strengthening the coincidence condition on both the doxastic accessibility relation and the closeness relation helps, but one faces a choice again. One would have to give up $\models K\phi \to \phi$. Alternatively, one could have $\models K\phi \to \phi$, $\models K\phi \to B\phi$, and $\models B(@\phi \leftrightarrow \phi)$. But then one would have to accept $\not\models K(@\phi \leftrightarrow \phi)$ as well, although one will also have $\models \neg K\neg(@\phi \leftrightarrow \phi)$. So, one can escape the conundrum at the end of Subsection 3.3 by modelling knowledge differently. See Subsection 4.3.

In this article the central question was how to combine counterfactual theories of knowledge with the notion of actuality. Answering this question involved rethinking the doxastic accessibility relation and the closeness relation as relations not between single worlds but between states or pairs of worlds,

where each pair contains an evaluation world and a reference world. It turned out to be very important to carefully consider under which conditions states are doxastically accessible from other states or close to other states.[4]

References

Davies M, Humberstone L (1980) Two notions of necessity. Philosophical Studies 38(1):1–31

Edgington D (1985) The paradox of knowability. Mind 94(376):557–568

Fagin R, Halpern JY (1988) Belief, awareness, and limited reasoning. Artificial Intelligence 34:39–76

Fara M, Williamson T (2005) Counterparts and actuality. Mind 114(453):1–30

Fitch FB (1963) A logical analysis of some value concepts. Journal of Symbolic Logic 28(2):135–142

Hanson WH (2006) Actuality, necessity, and logical truth. Philosophical Studies 130(3):437–459

Heller M (1989) Relevant alternatives. Philosophical Studies 55(1):23–40

Heller M (1999) Relevant alternatives and closure. Australasian Journal of Philosophy 77(2):196–208

Hintikka J (1962) Knowledge and Belief. Ithaca, N.Y.,Cornell University Press

Holliday WH (2012a) Epistemic closure and epistemic logic i: Relevant alternatives and subjunctivism. Journal of Philosophical Logic Forthcoming

Holliday WH (2012b) Knowing what follows: Epistemic closure and epistemic logic. PhD thesis, Stanford University

Holliday WH, Perry J (2014) Roles, rigidity, and quantification in epistemic logic. In: Baltag A, Smets S (eds) Trends in Logic, Outstanding Contributions: Johan van Benthem on Logic and Information Dynamics, Springer, pp 591–629

Hughes GE, Cresswell MJ (1996) A New Introduction to Modal Logic. Routledge, London

Kaplan D (1977) Demonstratives. In: Almog J, Perry J, Wettstein H (eds) Themes From Kaplan, Oxford University Press, pp 481–563

Kripke SA (1963) Semantical analysis of modal logic i. normal propositional calculi. Zeitschrift fur mathematische Logik und Grundlagen der Mathematik 9(56):67–96

Lackey J (2006) Review: Pritchard's epistemic luck. Philosophical Quarterly 56(223):284–289

Lewis D (1970) Anselm and actuality. Noûs 4(2):175–188

Lewis DK (1973) Counterfactuals. Blackwell Publishers

Luper-Foy S (1984) The epistemic predicament: Knowledge, nozickian tracking, and scepticism. Australasian Journal of Philosophy 62(1):26–49

Nelson M, Zalta EN (2012) A defense of contingent logical truths. Philosophical Studies 157(1):153–162

Nozick R (1981) Philosophical Explanations. Harvard University Press

Pritchard D (2005) Epistemic Luck, vol 29. Clarendon Press

Pritchard D (2012) Anti-luck virtue epistemology. Journal of Philosophy 109(3):247–279

Rabinowicz W, Segerberg K (1994) Actual truth, possible knowledge. Topoi 13(2):101–115

Sosa E (1999) How must knowledge be modally related to what is known? Philosophical Topics 26(1/2):373–384

Williamson T (2000) Knowledge and its Limits. Oxford University Press

Zalta EN (1988) Logical and analytic truths that are not necessary. Journal of Philosophy 85(2):57–74