Nina Haslinger* and Viola Schmitt*

What embedded counterfactuals tell us about the semantics of attitudes

https://doi.org/10.1515/lingvan-2021-0032
Received March 5, 2021; accepted August 19, 2021; published online July 15, 2022

Abstract: We discuss German examples where counterfactuals restricting an epistemic modal are embedded under *glauben* 'believe'. Such sentences raise a puzzle for the analysis of counterfactuals, modals, and belief attributions within possible-worlds semantics. Their truth conditions suggest that the modal's domain is determined exclusively by the subject's belief state, but evaluating the counterfactual separately at each of the subject's doxastic alternatives does not yield the correct quantificational domain: the domain ends up being determined by the facts of each particular world, which include propositions the subject does not believe. We therefore revise the semantics of counterfactuals: counterfactuals still rely on an ordering among worlds that can be derived from a premise set (Kratzer, Angelika. 1978. *Semantik der Rede: Kontexttheorie – Modalwörter – Konditionalsätze* (Monographien Linguistik und Kommunikationswissenschaft 38). Königstein: Scriptor, 2012 [1981]a. The notional category of modality. In *Modals and conditionals* (Oxford studies in theoretical linguistics 36), 27–69. Oxford: Oxford University Press), but rather than uniquely characterizing a world, this premise set can be compatible with multiple worlds. In belief contexts, the attitude subject's belief state as a whole determines the relevant ordering. This, in turn, motivates a revision of the semantics of *believe*: following Yalcin's work on epistemic modals (Yalcin, Seth. 2007. Epistemic modals. *Mind* 116. 983–1026), we submit that evaluation indices are complex, consisting of a world and an ordering among worlds. Counterfactuals are sensitive to the ordering component of an index. Attitude verbs shift both components, relativizing the ordering to the attitude subject.

Keywords: attitude predicates; counterfactuals; epistemic modality; ordering semantics; possible-worlds semantics

1 Introduction

This squib discusses counterfactual conditionals constraining existential modals in a belief context. Being native speakers of German, we focus on examples like (1) with the modal *möglicherweise* 'possibly'. Example (2) provides a scenario where (1) is judged true.[1]

(1)	*Anna glaubt,*	*dass*	*sie*	*möglicherweise*	*einen*	*Kaffeekocher*	*bekommen*	
	Anna believe.PRS.IND.3SG	that	she	possibly	a	coffeemaker	receive.PTCP	
	hätte,	*wenn*	*das*	*Paket*	*nicht*	*gestohlen*	*worden*	*wäre.*
	AUX.PST.SBJV.3SG	if	the	package	not	steal.PTCP	AUX.PTCP	AUX.PST.SBJV.3SG

'Anna believes that she would possibly have received a coffeemaker if the package had not been stolen.'

[1] Examples are glossed following the Leipzig Glossing Rules. Abbreviations used: 3 = third person, AUX = auxiliary, IMP = imperative, IND = indicative, INF = infinitive, PART = particle, PRS = present, PST = past, PTCP = participle, REFL = reflexive pronoun, SBJV = subjunctive, SG = singular.

***Corresponding authors: Nina Haslinger**, Georg-August-Universität Göttingen, Göttingen, Germany, E-mail: nina.haslinger@uni-goettingen.de. https://orcid.org/0000-0003-1749-6738; and **ViolaSchmitt**, Humboldt-Universität zu Berlin, Berlin, Germany, E-mail: viola.schmitt@hu-berlin.de

(2) SCENARIO: Last week, Anna intended to order two things online: a toaster and a coffeemaker. But ultimately, she ordered only one thing and forgot which one. This week, she was notified that a package had been delivered. When she went to pick it up, she learned it was stolen. She has no idea whether the package contained a toaster or a coffeemaker. It actually contained a toaster.

(1) **true**

Given this scenario, it is plausible to ascribe an epistemic or doxastic flavor to *möglicherweise*.[2] The modal seems to be evaluated relative to Anna's beliefs about what would have happened if the package had not been stolen, rather than the actual facts determining the most likely scenario: given the facts in scenario (2) where the package actually contains a toaster, a world where Anna receives a coffeemaker would be quite far-fetched, but this seems to be irrelevant to the truth value of (1), which suggests that the counterfactual in (1) is evaluated relative to Anna's doxastic state.

We will show that such data have unexpected consequences for the analysis of counterfactuals and of belief attributions within possible-worlds semantics (see e.g. Lewis 1973; Kratzer 1978, 2012 [1981]b for the former; Hintikka 1969 for the latter). They reveal a conflict between two widespread assumptions: first, that the quantificational domain of an epistemic modal in a belief context does not vary between the subject's doxastic alternatives, and second, that counterfactuals quantify over a set of "closest" worlds in which the antecedent holds, where "closeness" is determined by the facts of the evaluation world.

In (1), *möglicherweise* is restricted by a counterfactual and simultaneously occurs in a belief context. The second assumption then predicts that the domain of the modal should vary across the subject's doxastic alternatives, while the first assumption predicts that it should not. We will argue that to derive plausible truth conditions for (1), the second assumption should be weakened: the orderings between worlds that counterfactuals are sensitive to are not based on similarity to a particular world, but on "closeness" to an information state that may be equally compatible with multiple worlds. In other words, the orderings do not have to satisfy *centering* in the sense of Lewis (1973); they do not necessarily have a unique minimum. This permits us to interpret *möglicherweise* in (1) with respect to an ordering that represents Anna's belief state as a whole, rather than a particular belief world. In addition to this weakening of the standard semantics of counterfactuals, the puzzle also motivates a revision of Hintikka's (1969) attitude semantics.

Our conclusion – that worlds are the wrong kind of parameter for the similarity relation underlying counterfactuals – is not completely new. We find related claims in the literature. First, Arregui (2008) uses data involving tense in counterfactuals to argue that worlds are not the right level of granularity, but defends a notion of similarity relativized to situations, not information states. Second, Schulz (2007) argues that counterfactuals sometimes involve a global notion of belief revision, which means they must be interpreted relative to an epistemic state. However, she adopts this type of analysis only for so-called epistemic counterfactuals, whereas we argue it should apply to all counterfactuals to account for their behavior in embedded contexts. Given space limitations, we leave a detailed comparison of our argument with these previous proposals to future work.

2 The problem

Epistemic modals in belief contexts usually range over the attitude subject's doxastic alternatives. However, (1) introduces an additional complication: assuming Kratzer's (1978, 2012 [1981]a) restrictor approach to conditionals, the domain of *möglicherweise* appears to be restricted by the counterfactual. This raises a puzzle for the traditional view of counterfactuals, on which they quantify over a set of worlds determined by the facts of the evaluation world together with the antecedent proposition. If so, a modal embedded under *believe*, as in (1), may have a different domain in each world *w* among the attitude subject's doxastic alternatives, depending on the facts of *w*. This variability leads to incorrect truth conditions for sentences like (1).

[2] For an introduction to German modals and the different flavors available to modal adverbs, see Kratzer (1978, 2012 [1981]a).

2.1 Standard assumptions about modals in belief contexts

We first consider non-counterfactual epistemic modals embedded under attitudes. Yalcin (2007) observes that conjunctions of an epistemically modalized sentence with the negation of the corresponding non-modalized sentence are contradictory – even in embedded contexts. This observation extends to the embedding of *möglicherweise* under *glauben* 'believe', as in (3).[3]

(3) #*Anna glaubt, dass sie möglicherweise einen Kaffeekocher bekommen wird*
Anna believe.PRS.IND.3SG that she possibly a coffeemaker receive.INF will.IND.3SG
und dass sie keinen Kaffeekocher bekommen wird.
and that she no coffeemaker receive.INF will.IND.3SG
'Anna believes that she will possibly receive a coffeemaker and that she will not receive a coffeemaker.'

Following Hintikka (1969), we take *believe* to quantify universally over the subject *x*'s doxastic alternatives, as determined by an accessibility relation DOX(*x*). Without further assumptions about DOX(*x*), we get the truth conditions in (4), according to which Anna believes she will not receive a coffeemaker, but also believes that she considers it possible that she will receive a coffeemaker.

(4) $[\![(3)]\!](w_0) = 1 \text{ iff } \forall w[w \in \textsc{dox}(\textbf{Anna})(w_0) \rightarrow \neg\exists y[\textbf{coffeemaker}(w)(y) \wedge \textbf{receive}(w)(y)(\textbf{Anna})] \wedge$
$\exists w'[w' \in \textsc{dox}(\textbf{Anna})(w) \wedge \exists y[\textbf{coffeemaker}(w')(y) \wedge \textbf{receive}(w')(y)(\textbf{Anna})]]]$

Example (3) would correctly come out as contradictory if *möglicherweise* under a doxastic construal had to be interpreted as ranging over exactly those worlds that 'believe' quantifies over, or a subset thereof. This can be achieved by blocking the possibility that a subject might be unaware of a belief she holds, or equivalently, requiring DOX(*x*) to be transitive:

(5) For any worlds w, w' such that $w' \in \textsc{dox}(x)(w)$, $\textsc{dox}(x)(w') \subseteq \textsc{dox}(x)(w)$.

If *möglicherweise* receives a doxastic construal, (5) guarantees that for any *w* among Anna's doxastic alternatives in w_0, *möglicherweise* when evaluated at *w* only ranges over worlds that are themselves among her doxastic alternatives in w_0. The truth conditions in (4) then become contradictory: while the first embedded conjunct states that Anna will not receive a coffeemaker in any of her doxastic alternatives, the second conjunct entails that she will receive a coffeemaker in at least one of them.

In sum, (3) suggests that *möglicherweise* in a belief context usually quantifies over a subset of the belief subject's doxastic alternatives.[4] But this assumption cannot literally extend to counterfactual cases like (1): in scenario (2), the package was stolen in each of Anna's doxastic alternatives, thus the worlds over which *möglicherweise* quantifies cannot be among these alternatives. So what happens in counterfactuals?

2.2 Some previous assumptions about counterfactuals with epistemic modals

We will start with two prima facie plausible assumptions taken from the literature on counterfactuals. First, counterfactuals quantify over a domain of worlds selected in a particular way. Most implementations of this domain-selection mechanism are based either on ordering semantics (Lewis 1973) or on premise semantics (Kratzer 1978). On both approaches, the consequent of a counterfactual is evaluated in those worlds that count as "closest" to the evaluation world among the worlds verifying the antecedent.[5] What counts as "closest" is

[3] Yalcin (2007) suggests that his observations concerning the unacceptability of sentences like *Suppose it is raining and it might not be raining* should extend to embedding under *believe*, but does not give examples.

[4] Yalcin's more general approach to this phenomenon does not rely on a transitive accessibility relation. See Section 3.

[5] To simplify the discussion, we ignore scenarios in which there are no closest worlds; see Lewis (1973, 1981).

determined by the facts and generalizations holding in the evaluation world; following Kratzer (1978, 1991, 2012 [1981]a), these can be modeled as a set of propositions that jointly uniquely characterize the world in question.[6]

The second assumption is perhaps less standard: following the work of Kratzer (1978, 1991, 2012 [1981]a) on conditionals in general, we take counterfactuals with an existential modal like *möglicherweise* to provide an existential quantifier over a set of "closest" worlds verifying the antecedent. The antecedent of a counterfactual then serves to restrict an existential quantifier over worlds. We take the semantic predictions of this analysis for non-embedded counterfactuals to be superior to an account on which the counterfactual involves a covert universal modal that scopally interacts with *möglicherweise*: If *möglicherweise* were to take scope in the consequent of a universally quantified counterfactual, the paraphrase for (6a) would be (6b). However, (6b) requires Anna's doxastic state in each of the closest worlds in which the package was not stolen to be compatible with her receiving a coffeemaker. This is implausible: if the closest antecedent-worlds included worlds in which she ordered and received a toaster before the utterance time, then in these worlds she would not be uncertain about what she ordered.[7]

(6) a. *Wenn das Paket nicht gestohlen worden wäre, hätte Anna*
if the package not steal.PTCP AUX.PTCP AUX.PST.SBJV.3SG AUX.PST.SBJV.3SG Anna
möglicherweise einen Kaffeekocher bekommen.
possibly a coffeemaker receive.PTCP
'If the package had not been stolen, Anna might have received a coffeemaker.'

b. $\forall w[w \in \text{\textsc{closest}}(w_0)(\lambda w'.\text{the package was not stolen in } w') \rightarrow \exists w''[w'' \in \text{\textsc{dox}}(\text{Anna})(w)$
$\wedge \text{ Anna receives a coffeemaker in } w'']]$

Related arguments (which we omit due to space limitations) can be adduced against an analysis on which *möglicherweise* outscopes a counterfactual with a universal modal.

Combined, our assumptions yield the following simplified truth conditions for counterfactuals with *möglicherweise*. We leave the exact nature of the relation CLOSEST unspecified, but assume that it appeals to the facts and generalizations holding in w_0.

(7) $[[\textit{möglicherweise}[\textit{wenn}\ \text{p}]\text{q}]](w_0) = 1 \text{ iff } \exists w[w \in \text{\textsc{closest}}\ (w_0)([[\text{p}]]) \wedge [[\text{q}]](w) = 1]$

2.3 The puzzle

To see why this semantics for counterfactuals clashes with our earlier observations about *möglicherweise* in belief contexts, let us try to interpret (1) by combining (7) with the standard possible-worlds semantics for *glauben* 'believe', giving us (8).

(8) $[[(1)]](w_0) = 1 \text{ iff } \forall w[w \in \text{\textsc{dox}}(\text{Anna})(w_0) \rightarrow \exists w'\ [w' \in \text{\textsc{closest}}(w)([[\text{p}]]) \wedge [[\text{q}]](w') = 1]]$
where $[[\text{p}]] = (\lambda w.\text{the package was not stolen in } w)$ and $[[\text{q}]] = (\lambda w.\text{Anna received a coffeemaker in } w)$

Here, a counterfactual with the semantics in (7) is evaluated separately in each of Anna's doxastic alternatives: (8) says that each w among these alternatives is such that in one of the closest worlds to w in which the package was not stolen, Anna received a coffeemaker. But these truth conditions are empirically inadequate: in scenario (2), Anna has no idea whether the stolen package contained a toaster or a coffeemaker. So there is a

[6] These propositions probably differ in their status; see Lewis (1979) for discussion of temporal asymmetries among facts and Kaufmann (2005), Schulz (2007), Kaufmann (2013), and others for analyses in which causal generalizations and contingent facts do not have equal weight. Our claim below – that the relevant set of propositions does not have to determine a unique world – is independent of the question whether further subdivisions within this set are needed. However, it is not obvious how to integrate the causal network formalism employed by Kaufmann and Schulz into our analysis – a question we must leave to future work.

[7] Evaluating *möglicherweise* at w_0 will not help as it would make the counterfactual vacuous.

doxastic alternative w_t in which it contained a toaster. The truth conditions in (8) then entail that in at least one of the closest worlds to w_t in which the package was not stolen, Anna would nonetheless have received a coffeemaker. But if the CLOSEST relation is grounded in the facts of a particular world, how would such a coffeemaker world end up in the set of closest worlds for w_t?

While there might be worlds in which Anna ordered a toaster, but it was erroneously replaced with a coffeemaker before the package got stolen, such worlds will be very far-fetched from the perspective of w_t and should thus not be in CLOSEST(w_t)([[p]]). This is confirmed by the observation that (1) is false in a scenario where Anna is certain she only ordered a toaster. In fact, the need to exclude such exceptional worlds is exactly what motivates use of the CLOSEST relation, as opposed to a strict-conditional analysis of counterfactuals, to begin with. Thus, we cannot attribute the acceptability of (1) in scenario (2) to the availability of such worlds.

Rather, the root of the problem seems to be that the CLOSEST relation, when evaluated in a world w among Anna's doxastic alternatives, is based on a full set of propositions that uniquely characterize w, regardless of whether Anna actually believes these propositions. In Section 3, we will revise this assumption. But first, we address another potential solution: could we not assume, contra Hintikka (1969), that the quantificational force of attitude verbs is not universal[8] and weaken the analysis of belief so that it does not require us to consider any worlds in which the package contains a toaster? We cannot argue against all potential versions of weak belief here, but there is one reason why this is probably not the right explanation for the acceptability of (1) in scenario (2): when *möglicherweise* is replaced with the stronger epistemic modal *bestimmt* 'definitely', the resulting sentence (9) is no longer true in the scenario. If we could ignore the "toaster worlds" among the doxastic alternatives, so that the package contains a coffeemaker in all the doxastic alternatives under consideration, (9) and (1) should be equally acceptable.

(9) *Anna glaubt, dass sie bestimmt einen Kaffeekocher bekommen hätte,*
Anna believe.PRS.IND.3SG that she definitely a coffeemaker receive.PTCP AUX.PST.SBJV.3SG
wenn das Paket nicht gestohlen worden wäre.
if the package not steal.PTCP AUX.PTCP AUX.PST.SBJV.3SG
'Anna believes that she would definitely have received a coffeemaker if the package had not been stolen.'
false in (2)

We therefore turn to our preferred option – revising the role of the CLOSEST relation.

3 Proposal

The basic idea is that, for counterfactuals embedded under *believe*, the relevant set of closest worlds making the antecedent true is not determined on the basis of the facts of a particular world. Rather, it is based on the attitude subject's doxastic state as a whole. This doxastic state provides an ordering among worlds which must satisfy *weak centering* in the sense of Lewis (1973), but not centering; that is, it must have minimal elements, but these do not have to be unique.[9]

[8] Hawthorne et al. (2016) provide the most explicit proposal of this type that we are aware of (but see also Lassiter 2017 for a similar intuition): they suggest that a sentence *a believes p* is true iff *a*'s (subjective) confidence in *p* is larger than some contextual threshold. See Koev (2019) for a critical assessment.

[9] Strictly speaking, weak centering in Lewis's sense also requires the evaluation world that determines the choice of ordering to be one of the minimal elements in the ordering. Since we take the ordering to be determined by the matrix evaluation world w_0, this typically will not hold, as w_0 will not generally be compatible with subjects' belief states in w_0. However, there is a sense in which our belief states satisfy the spirit (although not the letter) of Lewis's definition: below, we will assume that *believe* quantifies over complex indices $\langle w, \preceq \rangle$ such that the set of $\preceq$-minimal worlds always includes w.

3.1 The role of information states

To spell this out, we first return to (8) and consider what the CLOSEST relation would have to look like to correctly predict that (1) is true in scenario (2). Let w_t be one of Anna's doxastic alternatives in which her package contained a toaster, and w_c an alternative where it contained a coffeemaker. If the counterfactual is evaluated separately in each of w_t and w_c, we get the right truth conditions only if CLOSEST(w_t)([[p]]) and CLOSEST(w_c)([[p]]) each contain worlds in which Anna actually received a coffeemaker. We are then ultimately giving up the intuition that the similarity relation accessed by counterfactual uses of *möglicherweise* is rooted in the facts and generalizations holding in a particular world. If all the facts of w_t are taken into account – including the fact that the package contained a toaster – a world in which Anna received a coffeemaker just cannot be maximally similar to w_t.

However, there is a natural way of characterizing the truth conditions of (1) that preserves the idea that counterfactuals involve similarity relations. Consider the notion of a similarity-based ordering $\preceq_w$ from Lewis (1973), where $w' \preceq_w w''$ means that w' is at least as similar to w as w'' is. If these orderings are based on the facts and generalizations holding in a world w, they must be *centered*: there must be a unique minimal element, w itself.

Let us now think of Anna's belief state in w_0 as a consistent set S of propositions whose intersection is the set DOX(Anna)(w_0).[10] Worlds could then be partially ordered by a relation $\preceq_{\text{Anna}, w_0}$, where $w \preceq_{\text{Anna}, w_0} w'$ iff w comes at least as close as w' to satisfying all the propositions in S. The exact definition of this ordering is irrelevant for our purposes (but see in particular Lewis 1981; Kaufmann 2005, 2013; Schulz 2007). But importantly, in Lewis's (1973) terminology, it will be weakly centered, but not centered, that is, it may have multiple minimal elements: a world w will be a minimal element if and only if w satisfies all the propositions in S. So Anna's doxastic alternatives are the minimal elements of S.

We submit that, when a counterfactual with a modal like *möglicherweise* is evaluated in the context of Anna's beliefs in w_0, the modal's quantificational domain does not vary across doxastic alternatives, but is determined by the ordering $\preceq_{\text{Anna}, w_0}$ which encodes Anna's doxastic state as a whole.[11] The domain will consist of the minimal worlds in $\preceq_{\text{Anna}, w_0}$ that satisfy the antecedent. The resulting truth conditions for (1) are paraphrased in (10a), where the CLOSEST relation is defined as in (10b). Note that, since the modal's domain remains constant across doxastic alternatives, the universal quantifier contributed by *believe* is vacuous.

(10) a. $[[(1)]](w_0) = 1$ iff $\forall w[w \in \text{DOX}(\text{Anna})(w_0) \rightarrow \exists w'[w' \in \text{CLOSEST}(\preceq_{\text{Anna}, w_0})([[\text{p}]]) \wedge [[\text{q}]](w') = 1]]$
where $[[\text{p}]] = (\lambda w.\text{the package was not stolen in } w)$ and
$[[\text{q}]] = (\lambda w.\text{Anna received a coffeemaker in } w)$

b. For a partial ordering $\preceq$ among worlds and a proposition p,
$\text{CLOSEST}(\preceq)(p) = \{w \in \text{dom}(\preceq) \mid p(w) = 1 \wedge \neg \exists w'[w' \prec w \wedge p(w') = 1]\}$

How does this solve our original puzzle? Consider a world w_0 described by scenario (2). Among Anna's doxastic alternatives, there are "toaster worlds" w_t and "coffeemaker worlds" w_c. Further, there is a world $w_{t'}$ that differs from w_t only in that the package was not stolen; in particular, the package in $w_{t'}$ contains a toaster. Similarly, there is a "coffeemaker world" $w_{c'}$ minimally different from w_c in which the package was not stolen. The new CLOSEST relation in (10b) has us consider the lowest-ranked worlds in $\preceq_{\text{Anna}, w_0}$ where the package was not stolen.

[10] As discussed by Kratzer (1978, 2012 [1981]b) for non-embedded conditionals, there will be many such sets that pick out the same doxastic alternatives. The problem of choosing between multiple premise sets arises in all versions of premise semantics.

[11] That belief states can be related to similarity orderings of the kind needed to interpret counterfactuals is not a new idea in the context of belief revision. A reviewer mentions Harper (1976), which we did not have access to at the time of writing; another reviewer points out a connection with Grove (1988), who uses Lewis's notion of a weakly centered similarity ordering to model belief revision, and, also in the context of belief revision, Gärdenfors (1988). Given these connections, let us clarify that we do not analyze counterfactuals in general as involving belief revision or as sensitive to belief states regardless of the context they occur in. While counterfactuals embedded under overt belief predicates are sensitive to belief states, we do not assume this for counterfactuals embedded under other attitudes.

Since Anna has no idea whether there was a toaster or a coffeemaker in the package, we can assume that neither of $w_{c'}$ and $w_{t'}$ ranks lower than the other. This corresponds to the intuition that $w_{c'}$ and $w_{t'}$ come equally close to satisfying all of Anna's beliefs. So (1) is correctly predicted true in scenario (2).

In sum, we propose to account for the weak semantics of existential counterfactuals in belief contexts by weakening the centering assumption adopted in Lewis's (1973) ordering semantics, or equivalently adopting a weaker version of premise semantics where the premises do not jointly characterize a unique world: The orderings we use to interpret counterfactuals are based on an information state that may leave certain issues undetermined. Two worlds that differ only on one of these undetermined issues may then be incomparable, whereas the standard semantics of counterfactuals would require us to order one world below the other, depending on how the issue is resolved in the evaluation world.

3.2 An implementation based on Yalcin's (2007) domain semantics

Given our proposal that the ordering relation exploited by counterfactuals is not always grounded in the facts of a particular world, the universal modal quantification introduced by 'believe' in (10a) is vacuous. This aspect of our paraphrase is crucial to account for the coffeemaker puzzle, but it raises a compositionality problem: if the counterfactual is evaluated relative to an ordering relation that depends on the evaluation world w_0, and 'believe' shifts the world parameter to worlds other than w_0 in the usual way, how does the counterfactual get access to this ordering?[12]

We submit that some expressions, including counterfactuals, are sensitive to an information state that acts as a separate parameter of the semantic interpretation function, independently of the world parameter. The truth value of sentences containing such expressions will in general not be determined by the evaluation world. To implement this idea without committing to premise semantics, we model information states as partial orderings among worlds. For instance, a belief state will be a partial ordering with the subject's doxastic alternatives as minimal elements; the interpretation of counterfactuals relative to this state will depend on the

[12] A reviewer suggests this problem could be solved by appealing to the introspection properties of belief: Natural language semantics seems to treat subjects as having full knowledge of their own belief states (but see Hintikka 1962). We could therefore assume that in each of Anna's doxastic alternatives relative to w_0, her belief state must be the same as in w_0. Extending this idea to counterfactual beliefs would mean that $\preceq_{\text{Anna},w_0} = \preceq_{\text{Anna},w}$ must hold for each doxastic alternative w. Each of the doxastic alternatives would thus make Anna's belief state in w_0 accessible. However, there is an empirical reason to prefer our grammatical approach: The domain of epistemic modals is not always determined by a belief state; under other attitudes, the relevant information state is determined by the embedding attitude (Yalcin 2007). For instance, if an epistemic modal is embedded under verbs meaning 'suppose' or 'imagine', it seems to quantify over worlds compatible with the imagined situation, not over worlds compatible with the subject's beliefs in the imagined situation. This is illustrated by the minimal pair in (i), adapted from Yalcin (2007): (ia) demonstrates that one can consistently suppose that some proposition *p* is true and that one considers $\neg p$ epistemically possible. Yet, this reading cannot be expressed using *might* (ib). The unacceptability of (ib) follows immediately if the embedding attitude shifts the domain of *might* to a set of worlds compatible with the subject's supposition, as sketched for *believe* in the main text.

(i) a. *Suppose Max is lying to you and you do not believe he is lying to you.*
b. *#Suppose Max is lying to you and he might not be lying to you.*

As (ii) shows, German *möglicherweise* behaves like *might* in (i) when embedded under *sich vorstellen* 'imagine, suppose'.

(ii)	*#Stell*	*dir*	*vor,*	*der*	*Max*	*würde*	*dich*	*belügen*	*und*	*er*	*würde*	*dich*	*möglicherweise*	*nicht*
	imagine.IMP	REFL	PART	the	Max	AUX.SBJV.PRS.3SG	you	lie.to	and	he	AUX.SBJV.PRS.3SG	you	possibly	not
	belügen.													
	lie.to													

'Suppose/imagine Max is lying to you and he is possibly not lying to you.'

Generally speaking, such data suggest that we need to model a dependency between *möglicherweise* and the embedding attitude, regardless of the introspective properties of that attitude. (Iterating *suppose* or German *vorstellen* does not give rise to the same introspection behavior as iterating *believe/glauben* either.) That being said, embedded *möglicherweise* is restricted in ways that are not well understood.

arrangement of the non-minimal worlds in the ordering. This provides us with an analysis of *believe* that extends to cases like (1): *believe* manipulates both the world parameter and the information-state parameter, shifting the latter to the ordering representing the subject's belief state.

This idea is a slight adaptation of Yalcin's (2007) proposal about epistemic modals in attitude contexts (see Section 2.1). Yalcin (2007) argues that the indices passed to the interpretation function are pairs consisting of a world and what he calls an *information parameter*, which he represents as an unordered set of worlds. This has two consequences: First, the denotations of expressions can be sensitive to different components of complex indices; for example, epistemic modals only consider the information parameter. Second, the semantics of expressions can shift or quantify over the components of complex indices individually. Yalcin proposes that attitude verbs shift both components: they quantify over different potential values of the world parameter, but also shift the information parameter (in the case of *believe*, to the subject's epistemic state).

We adopt the general structure of this proposal, but generalize Yalcin's notion of an information parameter: Since this parameter will provide the quantificational domain for counterfactuals, unordered sets of worlds will be insufficient. Instead, we define indices – the elements of D_s – as ordered pairs $i = \langle w_i, \preceq_i \rangle$, with w_i a world and $\preceq_i$ a partial ordering among worlds that represents an information state.

Given this richer notion of indices, we will now develop an analysis of (1). For expressions like (11a), which contain no modals, counterfactuals or similar expressions, our semantics does not differ significantly from the traditional view. The intension of (11a) is a function from complex indices to truth values, but this function is insensitive to the information-state component of the index.

(11) a. *The package wasn't stolen.*

b. $[\![(11a)]\!] = \lambda i.$the package was not stolen in w_i.

Next, we consider counterfactuals with existential modals. Without going into their internal composition, we assume following Kratzer (1978, 1991, 2012 [1981]b) that they quantify existentially over the minimal worlds in the relevant ordering which make the antecedent true. However, this ordering is now determined by the information state $\preceq_i$ of the evaluation index *i*. (In the unembedded case, this information state could arguably be determined by the common ground [see Yalcin 2007 for discussion], so that the ordering's minimal elements would constitute the context set.) Example (12) shows that the counterfactual selects the minimal elements relative to $\preceq_i$ among the worlds making the antecedent true. The consequent is evaluated at indices consisting of one of these worlds and $\preceq_i$ itself. Hence, the counterfactual quantifies over a domain determined by the information-state component of its index and completely ignores the world component.

(12) a. *possibly* [*if* [$_p$ *the package had not been stolen*]] [$_q$ *Anna would have received a coffeemaker*]

b. $$[\![(12a)]\!] = \lambda i.\ \exists w\Big[\big[w \in \mathrm{dom}(\preceq_i) \wedge [\![\mathrm{p}]\!]\big(\langle w, \preceq_i\rangle\big) = 1 \wedge \neg \exists w'\big[w' \prec_i w \wedge [\![\mathrm{p}]\!]\big(\langle w', \preceq_i\rangle\big) = 1\big]\big] \wedge [\![\mathrm{q}]\!]\big(\langle w, \preceq_i\rangle\big) = 1\Big]$$

The final prerequisite is a revised semantics for *believe* (13), which has two jobs: to perform a shift to an information state $\preceq_{x, w_i}$ that encodes the subject's beliefs in the evaluation world, and to quantify over indices consisting of this new information state and some doxastic alternative of the subject's. Notice that (13) appeals to the minimal elements of $\preceq_{x, w_i}$ (in a global sense, not relative to some antecedent proposition), which are the subject's doxastic alternatives.

(13) $$[\![\textit{believe}]\!] = \lambda i.\lambda p_{\langle s, t\rangle}.\lambda x_e.\ \forall w\Big[\Big[w \in \mathrm{dom}\big(\preceq_{x, w_i}\big) \wedge \neg \exists w'[w' \prec_{x, w_i} w]\Big] \rightarrow p\big(\langle w, \preceq_{x, w_i}\rangle\big)\Big]$$

For complements that do not depend on the information-state component of an index, (13) preserves Hintikka's (1969) intuition that *believe* quantifies universally over doxastic alternatives: in (14), the proposition expressed by the embedded clause is evaluated at each index consisting of Anna's belief state and a doxastic alternative.

(14) a. *Anna believes that the package wasn't stolen.*

b. $$\Big[\Big[(14a)\Big]\Big] = \lambda i.\ \forall w\Big[\Big[w \in \mathrm{dom}\Big(\preceq_{\mathrm{Anna}, w_i}\Big) \wedge \neg \exists w'\Big[w' \prec_{\mathrm{Anna}, w_i} w\Big]\Big]$$
$$\rightarrow \text{the package was not stolen in } w]$$

Embedding the counterfactual in (12a) under *believe* finally yields our motivating example (1). As (15) shows, *believe* quantifies universally over indices with the information-state component shifted to $\preceq_{\text{Anna},w_i}$ and the world component shifted to one of Anna's doxastic alternatives. The counterfactual denotation (12b) is evaluated at each of these indices. Importantly, the truth conditions of this counterfactual do not depend on the world component of its evaluation index; it simply quantifies over the minimal worlds in the respective information state at which the package was not stolen. Since all the indices *believe* quantifies over have the same information-state component, quantification over the world variable *w* in (15) is vacuous and the truth conditions reduce to a single instance of existential quantification over worlds. Thus, (15) expresses the desired truth conditions: in some of the minimal worlds relative to $\preceq_{\text{Anna},w_i}$ in which the package was not stolen, Anna receives a coffeemaker.

(15)
$$\begin{aligned}
&[[(1)]] = \lambda i.\ \forall w\left[\left[w \in \text{dom}\left(\preceq_{\text{Anna},w_i}\right) \wedge \neg \exists w' [w' \prec_{\text{Anna},w_i} w]\right] \rightarrow [[(12a)]]\left(\langle w, \preceq_{\text{Anna},w_i} \rangle\right)\right]\\
&= \lambda i.\ \forall w\left[\left[w \in \text{dom}\left(\preceq_{\text{Anna},w_i}\right) \wedge \neg \exists w'\left[w' \prec_{\text{Anna},w_i} w\right]\right]\right.\\
&\rightarrow \exists w'\left[w' \in \text{dom}\left(\preceq_{\text{Anna},w_i}\right) \wedge \text{the package was not stolen in } w' \wedge\right.\\
&\neg \exists w'' [w'' \prec_{\text{Anna},w_i} w' \wedge \text{the package was not stolen in } w''] \wedge \text{Anna received a coffeemaker in } w']]\\
&= \lambda i.\ \exists w'\left[w' \in \text{dom}\left(\preceq_{\text{Anna},w_i}\right) \wedge \text{the package was not stolen in } w' \wedge\right.\\
&\neg \exists w'' [w'' \prec_{\text{Anna},w_i} w' \wedge \text{the package was not stolen in } w''] \wedge \text{Anna received a coffeemaker in } w']
\end{aligned}$$

Let's sum up the idea behind this analysis. We argued in Section 2 that counterfactuals in belief contexts should not be sensitive to all the facts characterizing a particular doxastic alternative. They should only consider propositions actually believed by the subject. We suggested that, while counterfactuals are sensitive to an ordering among worlds, this ordering may be determined by an information state that leaves certain issues unresolved and therefore does not have to be centered, only weakly centered (Lewis 1973). Our implementation dissociates the truth conditions of counterfactuals from the evaluation world, following Yalcin's (2007) analogous claim about epistemic modals. Instead, counterfactuals (and epistemic modals) depend on a separate parameter passed to the interpretation function. We took the possible values of this parameter to be orderings among worlds, although one could also use Kratzer's (2012 [1981]b) *conversational backgrounds*. Essentially, our claim is that embedded uses of counterfactuals reveal that counterfactuals and epistemic modal constructions are more similar than they might seem: neither requires the information state used to select the domain for the modal to characterize a unique world.

Acknowledgments: Special thanks to Magdalena Kaufmann and Frank Sode for their detailed and insightful comments and criticism. In this contribution, we could not do justice to all of their comments for reasons of space, but we aim to address them in more detail in future work. We would also like to thank two anonymous reviewers for this volume, our audiences at the workshop "Semantics and Pragmatics of Conditional Connectives" at DGfS 2021 and the Göttingen English Linguistics Oberseminar, and an anonymous reviewer for the DGfS. All remaining errors are of course our own.
Research funding: Viola Schmitt's work was supported by the Austrian Science Fund (FWF) project P-32939 "The Typology of Cumulativity".

References

Arregui, Ana. 2008. Resolving similarity in embedded contexts. In Olivier Bonami & Patricia Cabredo (eds.), *Empirical issues in formal syntax and semantics 7: Papers from CSSP 7*, 35–52. Paris: CSSP. http://www.cssp.cnrs.fr/eiss7/ (accessed 5 March 2021).

Gärdenfors, Peter. 1988. *Knowledge in flux: Modeling the dynamics of epistemic states*. Cambridge, MA: MIT Press.

Grove, Adam. 1988. Two modellings for theory change. *Journal of Philosophical Logic* 18(2). 157–170.

Harper, William L. 1976. Ramsey test conditionals and iterated belief change (a response to Stalnaker). In William L. Harper & Clifford Alan Hooker (eds.), *Foundations of probability theory, statistical inference, and statistical theories of science*, 117–135. Dordrecht: Springer.

Hawthorne, John, Daniel Rothschild & Levi Spectre. 2016. Belief is weak. *Philosophical Studies* 173(5). 1393–1404.

Hintikka, Jaakko. 1962. *Knowledge and belief*. Ithaca: Cornell University Press.

Hintikka, Jaakko. 1969. Semantics for propositional attitudes. In *Models for modalities. Selected essays*, 87–111. Dordrecht: Reidel.

Kaufmann, Stefan. 2005. Conditional predictions. A probabilistic account. *Linguistics and Philosophy* 28. 181–231.

Kaufmann, Stefan. 2013. Causal premise semantics. *Cognitive Science* 37. 1136–1170.

Koev, Todor. 2019. Strong beliefs, weak commitments. In Maria Teresa Espinal, Elena Castroviejo, Manuel Leonetti, Louise McNally & Cristina Real-Puigdollers (eds.), *Proceedings of Sinn und Bedeutung 23*, vol. 2, 1–18. Bellaterra (Cerdanyola del Vallès): Universitat Autònoma de Barcelona. https://semanticsarchive.net/Archive/Tg3ZGI2M/Koev.pdf (accessed 5 March 2021).

Kratzer, Angelika. 1978. *Semantik der Rede: Kontexttheorie – Modalwörter – Konditionalsätze* (Monographien Linguistik und Kommunikationswissenschaft 38). Königstein: Scriptor.

Kratzer, Angelika. 1991. Conditionals. In Arnim von Stechow & Dieter Wunderlich (eds.), *Semantics: An international handbook of contemporary research*, 651–656. Berlin: de Gruyter.

Kratzer, Angelika. 2012 [1981]a. The notional category of modality. In *Modals and conditionals* (Oxford studies in theoretical linguistics 36), 27–69. Oxford: Oxford University Press.

Kratzer, Angelika. 2012 [1981]b. Partition and revision: The semantics of counterfactuals. In *Modals and conditionals* (Oxford studies in theoretical linguistics 36), 72–84. Oxford: Oxford University Press.

Lassiter, Daniel. 2017. *Graded modality: Qualitative and quantitative perspectives*. Oxford: Oxford University Press.

Lewis, David. 1973. *Counterfactuals*. Cambridge, MA: Harvard University Press.

Lewis, David. 1979. Counterfactual dependence and time's arrow. *Noûs* 13(4). 455–476.

Lewis, David. 1981. Ordering semantics and premise semantics for counterfactuals. *Journal of Philosophical Logic* 10. 217–234.

Schulz, Katrin. 2007. *Minimal models in semantics and pragmatics: Free choice, exhaustivity, and conditionals*. Amsterdam: ILLC. https://hdl.handle.net/11245/1.272471 (accessed 5 March 2021).

Yalcin, Seth. 2007. Epistemic modals. *Mind* 116. 983–1026.