Reverse Sobel Sequences and the Dissimilarity of Antecedent Worlds[1]

David KRASSNIG — *University of Konstanz*

Abstract. We have conducted an acceptability judgement experiment to examine two hypotheses related to the felicity of reverse Sobel sequences (rSS). The first hypothesis was based upon Lewis' (2018) relevance-based variably-strict semantics: If two rSS are identical except for their respective difference in dissimilarity between their antecedent worlds, then the rSS with a higher degree of dissimilarity should be, on average, more acceptable. Our results, however, seem to support this hypothesis only weakly and appear more contradictory than supportive to the model behind it: Any kind of clear-cut dissimilarity will render an rSS felicitous, so long as its conditionals are counterfactual by nature. The second hypothesis, that rSS whose domains of quantification are entirely disjoint should be just as acceptable as regular utterances, was quasi-confirmed. Whilst a significant difference to the control items was found, these rSS differ only minimally in average acceptability and are the highest rated rSS so far. We therefore explain the slight, statistically significant decrease in acceptability as a result of the markedness of rSS structures. Finally, we provide two analyses to account for the data gathered. First, one modification of Lewis' (2018) account, where we argue that relevance may not increase closeness beyond the levels set forth by worlds similarity. Then, with the second account, we attempt to motivate the need for and the pragmatic contribution of contrastive stress with regards to rSS, using Ebert et al.'s (2008) assumption that antecedents represent their conditional's aboutness topic – thereby deriving contrastive topic with regard to the two sets of antecedent worlds.

Keywords: reverse Sobel sequences, relevance, contrastive stress, conditionals, variably-strict semantics, world closeness, contrastive topic, aboutness topic, counterfactuality

1. Introduction

A Sobel sequence (hereafter *SS*) is a sequence of conditionals whose underlying semantic structure adheres to the pattern "If ϕ then χ, but if ϕ and ψ then not χ", as seen in (1). They, and their reversals, commonly referred to as *reverse Sobel sequences* (hereafter *rSS*), as seen in (2), have played and continue to play a key role in the discussion on how to model conditionals.

(1)   If the USA threw its weapons into the sea tomorrow, there'd be war; but if all the nuclear powers threw their weapons into the sea tomorrow, there'd be peace.

(Lewis, 1973)

(2)   If all the nuclear powers threw their weapons into the sea tomorrow, there'd be peace; # but if the USA threw its weapons into the sea tomorrow, there'd be war.   (Heim, 1994)

There are two main lines of thought to the current debate: Variably-strict (Stalnaker, 1968; Lewis, 1973) and dynamic strict analyses (von Fintel, 2001; Gillies, 2007). The former predict SS to be universally felicitous, which runs counter to Heim's infelicitous example, provided in

---

(2). The dynamic strict accounts, on the other hand, predict, by design, all rSS to be infelicitous. However, as Moss (2012) pointed out, not all rSS are inherently infelicitous:

(3)   If kangaroos had no tails and they used crutches, they would not topple over. But if kangaroos had no tails, they would topple over.
                                        (adapted from Lewis (1973: p. 1,9) by Lewis (2018: p. 487))

(4)   *(Holding up a dry match, with no water around)* If I had struck this match and it had been soaked, it would not have lit. But if I had struck this match, it would have lit.
                                        (adapted from Stalnaker (1968: p. 106) by Lewis (2018: p. 487))

(5)   *(Said to someone who had just been completely alone by a frozen lake)* If you had walked on the thin ice while being supported by someone on the shore, the ice wouldn't have broken. But, of course, if you had walked on the thin ice, the ice would have broken.   (adapted from Bennett (2003: p. 166) by Lewis (2018: p. 488))

This would render the variably-strict accounts too lenient and the dynamic strict accounts too strict. As such, a recent trend in semantics and pragmatics has been to take the more lenient variably-strict semantics and combine them with additional exclusionary mechanisms that render specific rSS either infelicitous or inconsistent (Moss, 2012; Klecha, 2015; Lewis, 2018).

In one such case, Klecha (2014, 2015) argued that there are actually two types of SS: True SS (where ϕ and ψ in the antecedent are causally unrelated) and Lewis sequences (where the antecedent's propositions are causally related). He argues that only reverse Lewis sequences are universally infelicitous and provides an imprecision-based explanation for their unidirectionality. True Sobel sequences, on the other hand, are unequivocal (i.e. they are "a single pointful piece of discourse", as stated by Edgington (1995)) and, in principle, reversible. As such, we mostly concern ourselves with reverse True Sobel sequences for the remainder of this paper.

In another one such recent innovation, Lewis (2018) sought to explain the infelicity of rSS by turning the world-ordering for conditionals into a dynamic process: She argues that world closeness is determined by a function of both similarity and relevance (Lewis, 2018), rather than purely by similarity as previously believed (e.g. Lewis, 1973). In her framework, further elucidated in § 1.1, the possible infelicity of rSS comes down to (i) the perceived relevance of the $\phi \wedge \psi$-worlds and (ii) the degree of dissimilarity between $\phi \wedge \psi$-worlds and ϕ-worlds.

This paper has two main goals: First, to empirically test Lewis' (2018) prediction that rSS with similar antecedent worlds are less likely to be felicitous than rSS with dissimilar antecedent worlds. Second, to empirically test whether rSS are considered felicitous if their sets of worlds belong to different domains of quantification (live possibility worlds vs counterfactual worlds), to see the acceptability of the most likely felicitous type of rSS under both schools of thought.

This paper is structured as follows: In § 1.1, we give a summary of Lewis (2018). In § 2, we go through the materials and method for the conducted experiment. In § 3, we relay the experiment's results. Then, in § 4, we interpret and discuss the results shown in the previous section and how they relate to the first goal / hypothesis in § 4.1 and the second goal / hypothesis in § 4.2. We enhance the collected data with additional introspective observations in § 4.3 and sketch out two mutually compatible accounts for the gathered findings in § 4.4.1 and § 4.4.2. Finally, in § 5, we provide our conclusion and give thoughts for future research.

1.1. Lewis (2018)

Lewis (2018) argues that that the effect of the first conditional on the context is pragmatic by nature, as hypothesised by Moss (2012), but also that this pragmatic effect has a semantic influence on the interpretation of the second conditional, as assumed by von Fintel (2001) and Gillies (2007). In her account, first proposed to solve the problem of counterfactual skepticism[2] (Lewis, 2016), said pragmatic effect affects the world closeness ordering post-utterance and, as such, has a possible effect on the semantic evaluation of subsequent conditionals. This is done by turning closeness into a function of similarity and world relevance:

(6)   For all contexts c, $\phi\square\rightarrow\psi$ is true at w in c iff all the closest ϕ-worlds to w are ψ-worlds, where closeness is a function of both similarity and relevance.

(adapted from Lewis, 2016: p. 292)

In her system, similarity provides the basic layout for the world closeness ordering. This layout is then modified by the differing values of relevance we ascribe to each possible world. If a world is ascribed high relevance, it is moved correspondingly closer to the evaluation world. If the world in question is ascribed substandard levels of relevance, on the other hand, it is pushed further away from it. As such, even though possible worlds have similarity-based anchors in the world ordering, there is a certain degree of bidirectional mobility given to any possible world.

The relevance of worlds is mostly manipulated by conversational context and discourse. Discourse participants can thus actively, but limitedly, shape the world ordering: "They can indirectly affect what is (ir)relevant by changing the conversational purposes, by, for example, raising the standards of precision, making something salient, raising a new question under discussion, or refusing to accommodate a shift in conversational purpose." (Lewis, 2018: p. 500).

Regarding this, the raising to salience is of special import to rSS: Since discourse participants must take the antecedent of a conditional seriously,[3] in order to evaluate the counterfactual, the possibility of the antecedent is thereby automatically raised to salience (Lewis, 2018). This, in turn, may raise the relevance of the antecedent worlds and thereby their place in the world ordering. Concerning infelicitous rSS, such as the one in (2), this equates to the $\phi \wedge \psi$-worlds being pushed towards the evaluation world such that they are counted amongst the closest ϕ-worlds, rendering the utterance of the subsequent ϕ-conditional into a contradiction:
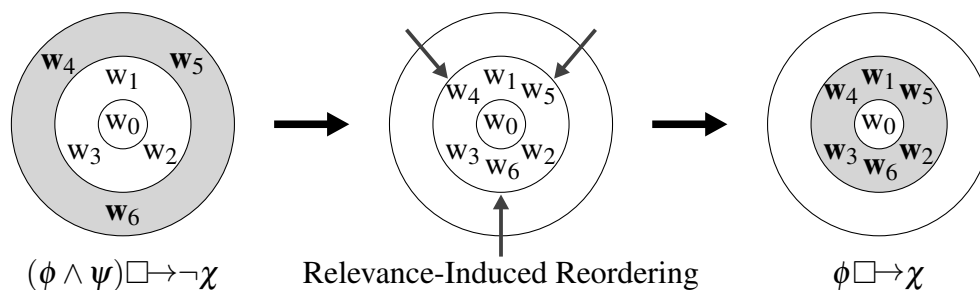

$(\phi \wedge \psi)\square\rightarrow\neg\chi$ Relevance-Induced Reordering $\phi\square\rightarrow\chi$

Figure 1: World ordering and selection for infelicitous rSS according to Lewis (2018).

[2]This refers to the worry that nearly all would-counterfactuals (that we take to be true) are false. We refer to Lewis (2016) for details due to reasons of space and a lack of direct relevance to the topic at hand.

[3]As Lewis (2018: p. 500) notes, this is done in various ways: e.g. by existence presuppositions (von Fintel, 2001), entertainability presuppositions (Gillies, 2007), or the pragmatic raising to salience of its possibility (Moss, 2012).

However, if the $\phi \wedge \psi$-worlds of the first conditional are too dissimilar to the ϕ-worlds of the second conditional, then the former are not move close enough to count amongst the latter. In (5), for example, it was specified that the person in question was very much alone by the frozen lake. When talking about whether or not that person would have broken through the ice, had they walked upon it, the possibility of a person spontaneously appearing as if out of thin air is simply not relevant enough to most discourse participants. As such, no relevance-induced world shifting takes place in (5) or in any of the other felicitous rSS.[4] As such, the ϕ-conditional does not quantify over $\phi \wedge \psi$-worlds, leading to a consistent sequence of conditionals.

$$(\phi \wedge \psi)\square\!\!\rightarrow\neg\chi \qquad \text{No Relevance-Induced Reordering} \qquad \phi\square\!\!\rightarrow\chi$$
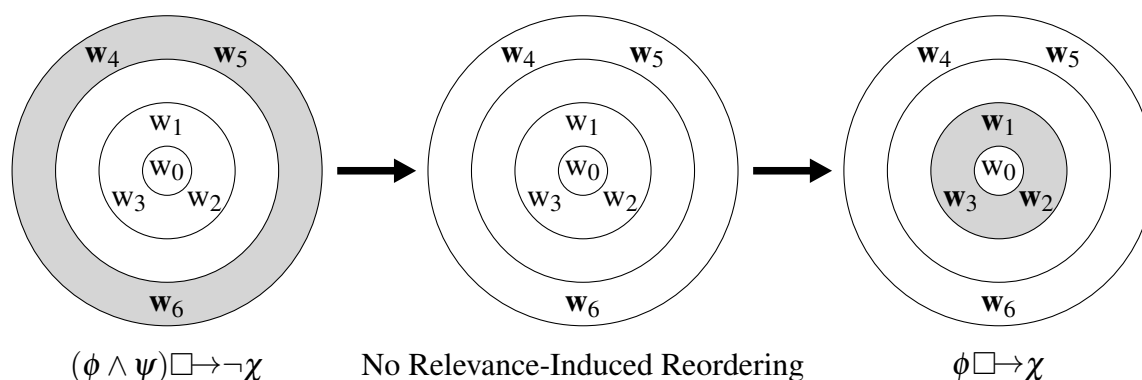
Figure 2: World orderings and selections of consistent rSS according to Lewis (2018).

Note that Lewis (2018) does not promote some specific measure by which to tell when two sets of worlds are far enough apart from one another, as this may vary from person to person or context to context. The instability concerning the felicity judgments of rSS is therefore also predicted by this account, as its sensitivity to discourse relevance grants the discourse participants some leeway in their semantic evaluation of the conditionals: "Hearing things at one moment as felicitous (consistent) and the next as infelicitous (inconsistent), or vice versa, is an expected feature of a phenomenon involving context sensitivity." (Lewis, 2018: p. 502)

As an added side-benefit, using Lewis' (2018) analysis, Klecha's (2015) observations regarding Lewis sequences and True SS can neatly be accounted for under a single phenomenon, rather than having to consider them as separate phenomena with a coincidentally equal surface structure with two separate explanations. For more details on this, we refer to Krassnig (2017).

2. Material & Methods

To accomplish our two goals, testing Lewis' hypothesis concerning the effect of world similarity on infelicity and testing whether or not disjoint antecedent worlds rSS are as acceptable as regular sentences, we have formulated two hypotheses which shaped our experiment's design:

1. If two rSS are the same except for the degree of similarity between their conditionals' antecedent worlds, then the rSS whose degree of similarity is more disparate should be considered more acceptable on average.

(hereafter the *dissimilar worlds hypothesis*)

---

[4]One could also argue that some relevance-induced restructuring does take place, but that its effects are not felt, because the $\phi \wedge \psi$ worlds have fallen short of the intended target. As this would make no difference to never having moved at all, at least to the cases within this paper, we ignore this possibility for sake of simplicity.

2. If the domains of quantification of an rSS are entirely disjoint, there should be no difference in acceptability between them and regular sentences (i.e. the control items).

(hereafter the *disjoint domain hypothesis*)

This section is structured as follows: In § 2.1, we explain how we designed the materials that made up our experiment. Then, in § 2.2, we explicate the method used for our experiment.

2.1. Materials

In order to test the dissimilar worlds hypothesis, we created five experimental items which contain an rSS each and were presented with two different contexts: Differing as little as possible, they either indicate that the $\phi \wedge \psi$-worlds and ϕ-worlds are very similar to one another or very dissimilar to one another. These respectively represent the SIMILAR and DISSIMILAR conditions. From a syntactic point of view, all conditionals are future-less-vivids and, deviating from the classical examples, contain the auxiliary verb *did* in front of the ϕ-antecedent's main verb. The auxiliary verb was inserted in order to ensure that the ϕ-conditional's antecedent is not a syntactic subset of the $\phi \wedge \psi$-conditional's antecedent. This was necessary, since Klecha (2015: p. 135) posited that the ϕ-antecedent being a syntactic subset of the previous conditional's antecedent would automatically render the rSS infelicitous, as some form of contrastive stress in the second conditional is required. The attentive reader may have noticed that we already encountered a counterexample to this restriction in (4). However, preliminary testing with native speakers of English prior to the experiment has shown (i) that the auxiliary verb *had* appears obligatorily stressed in this rSS, even though it occurs in either sequence conditional, and that (ii) this rSS is rendered less acceptable by turning it into the future-less-vivid conditional in (7). Introducing either *did* or *were to* then improved the acceptability for the same native speakers of English. We then decided on using *did* for the sake of simplicity. For more on this, we refer to § 4.3, where we further explore our intuitions regarding contrastive stress.[5]

(7)   *(Holding up a dry match, with no water around)* If I struck this match and it was soaked, it would not light. ? But if I struck this match, it would light.

In total, this led us to the experimental pattern in (8), which, in turn, led us to our experimental items, as seen in (9). The remainder of the experimental target items may be found here: http://bit.ly/2OyZgSNhttp://bit.ly/2OyZgSN

(8)   *Context text common to either condition*
        (i) Context text that sets the $\phi \wedge \psi$-worlds and ϕ-worlds as similar to one another.
        (ii) Context text that sets the $\phi \wedge \psi$-worlds and ϕ-worlds as dissimilar to one another.
        S: If $\phi \wedge \psi$, (then) $\neg\chi$; but if did-ϕ, (then) χ.

(9)   *Andy's friend Michael plans to go to a cabin in the woods during the semester break. He is not convinced, however, that it will be a fun trip, because he is terrified of storms and he would have no way out if one took place.*
        (i) His trip would be during an average month of the year and Andy thinks that a storm would be possible but not too likely.

---

[5]Please note that, from here on out, we refer to intuitions gathered from native speakers of English as *our intuitions* for the sake of simplicity and brevity. Whilst we share most of these intuitions, being a non-native speaker of English, we have either confirmed or first learned of these intuitions by asking native speakers of English.

> *(ii) His trip would be during the driest month of the year and Andy thinks that a storm would be impossible or at the very least extremely unlikely.*
>
> **A:** If you went to the cabin and a storm came, your trip would be horrible; but if you *did* go to the cabin, it would be a pretty good trip.

In order to test the disjoint domain hypothesis, we furthermore created five experimental items that also contain an rSS each, but which were presented with only one context. These represent the DISJOINT condition. Aside from the context, and a subsequent sentence explicating the purpose of the rSS, these items are identical in shape to the previously shown experimental items. However, there is one difference which pertains to the hypothesis tested: Whilst the ϕ-conditionals remains a standard future-less-vivid conditional, the $\phi \wedge \psi$ conditional is counterfactual by nature, though it retains the same overt tense structure as the other conditionals. This shift from counterfactual to live possibilities is motivated by the fact that both variably-strict and dynamic strict conditional models would generally assume that conditionals about live possibilities would not take counterfactual worlds into account. As such, either approach should predict felicity. See below for an example item. The remaining experimental target items may be found here: http://bit.ly/2OyZgSNhttp://bit.ly/2OyZgSN.

(10) *Alex and her friend Steve enter a construction site. Steve doesn't wear his helmet, but carries it around in his hand. This annoys Alex, since it's a dangerous site.*

  Alex: If some construction material fell on your head right now and you wore a helmet, you would probably survive the incident; but if some construction material *did* fall on your head right now, you would certainly die. So, wear your goddamn helmet.

We also created 20 control items, representing the CONTROL condition, that acted as fillers. These consisted of regular SS, generic conditional sequences, and rSS containing an exhaustifying lexical item (e.g. *only*). These items were also displayed with an appropriate context.

The total number of conditions were distributed across three lists in a Latin square design:

List	Item Order
I	A B C A B C A B C A B C A B C
II	B A B C A B C A B C A B C A B
III	A C A B C A B C A B C A B C A

Table 1: Latin square condition lists assigned to participants, the condition SIMILAR represented as A, DISSIMILAR as B, and DISJOINT as C.

List	Item order
I	F **A** F **B** F F **C** F **A** F **B** F **C** F **A** F F **B** F **C** F **A** F **B** F **C** F F **A** F **B** F F **C** F
II	F **C** F **A** F F **B** F **C** F **A** F **B** F **C** F F **A** F **B** F **C** F **A** F **B** F F **C** F **A** F F **B** F
III	F **B** F **C** F F **A** F **B** F **C** F **A** F **B** F F **C** F **A** F **B** F **C** F **A** F F **B** F **C** F F **A** F

Table 2: Latin square condition lists assigned to participants, target conditions in boldface and fillers represented as F, the SIMILAR condition as A, DISSIMILAR as B, and DISJOINT as C.

The fillers, i.e. control items, were then added to these lists such that no target condition is directly followed by another target condition, yielding the final lists shown in Table 2.

Each participant was randomly assigned to one of the three item lists. Concerning the actual test items themselves, these were also randomly assigned for each participant to each pre-placed condition slot in the randomly assigned item list, whilst the fillers were fixed in place across all lists. As such, each participant had a partially randomised item list specific to them.

There were a total of 48 participants, of which 41 remained after all exclusions. The criteria for exclusion are elaborated upon in § 2.2.

2.2. Method

The experiment itself was an acceptability rating study. Prior to participating in the study, the participants were asked to provide information concerning their age, gender, native language, whether they were raised bilingually, and other languages spoken. They were then asked to sign an agreement, that their anonymised data may be used, statistically processed, and published in its processed, anonymised form. This was followed by an explanation of what they had to do.

The participants were then individually shown and asked to rate the acceptability of each test items on a Likert scale from 1 (=*sounds very natural*) to 5 (=*sounds very unnatural*). Upon having rated a test item, the participants had to click a button to continue to the next item. The participants were unable to go back and re-rate an earlier test item.

Participants were excluded if they (i) failed to rate 75% of all control items three or better, (ii) provided an age below 18 or above 65, (iii) consistently rated the experimental items in a systematic fashion (e.g. 1-5-1-5, 1-2-3-4-5, etc.), (iv) were either non-native speakers of English or were raised bilingually, or (v) did not complete the entire experiment run.

The experiment was conducted entirely online, and the participants were gathered from various social network sites. Participation was entirely voluntary and not financially compensated.

3. Results

From the results, it was clear that the control items were rated the most acceptable, followed by the DISJOINT rSS, the DISSIMILAR rSS, and finally by the SIMILAR rSS:

Condition	Average Acceptability	Variance
CONTROL Condition	1.48	0.45
DISJOINT Condition	2.15	0.65
DISSIMILAR Condition	3.41	1.54
SIMILAR Condition	4.4	0.43

Table 3: Average acceptability of each condition, ranked from highest to lowest

We analysed the obtained data with a one-factor ANOVA, which showed a statistically significant difference between groups as determined by one-way ANOVA ($F(3, 1021) = 643.1$, $p < .01$). Post hoc conducted one-tailed t-tests between each condition showed that there

is a significant difference between each condition: Comparing the CONTROL condition to the DISJOINT condition, we obtained $t(204) = -8.8, p < 0.05$, to the DISSIMILAR condition, $t(204) = -19.3, p < 0.05$, and to the SIMILAR condition, $t(204) = -45.1, p < 0.05$. Comparing the SIMILAR condition to the DISSIMILAR condition, we obtain $t(204) = 10.3, p < 0.05$, and to the DISJOINT condition, $t(204) = 30.7, p < 0.05$. Comparing the DISSIMILAR condition to the DISJOINT condition, we obtain $t(204) = 12.2, p < 0.05$.

The comparatively high variance of the DISSIMILAR condition led us to take a closer look at its raw data: Intuitively, it seemed that some participants consistently rated these items far lower than other participants. We therefore opted to test our intuition with a k-means clustering analysis, to see if our participants could be divided into more than one population. The cluster analysis of the participants' results showed that the data could be split into two distinct populations, with relatively high values of confidence. The first population cluster, Cluster 1, consists of 26 participants (i.e. 63% of all participants), and the second population cluster, Cluster 2, consists of 15 participants (i.e. 37% of all participants).
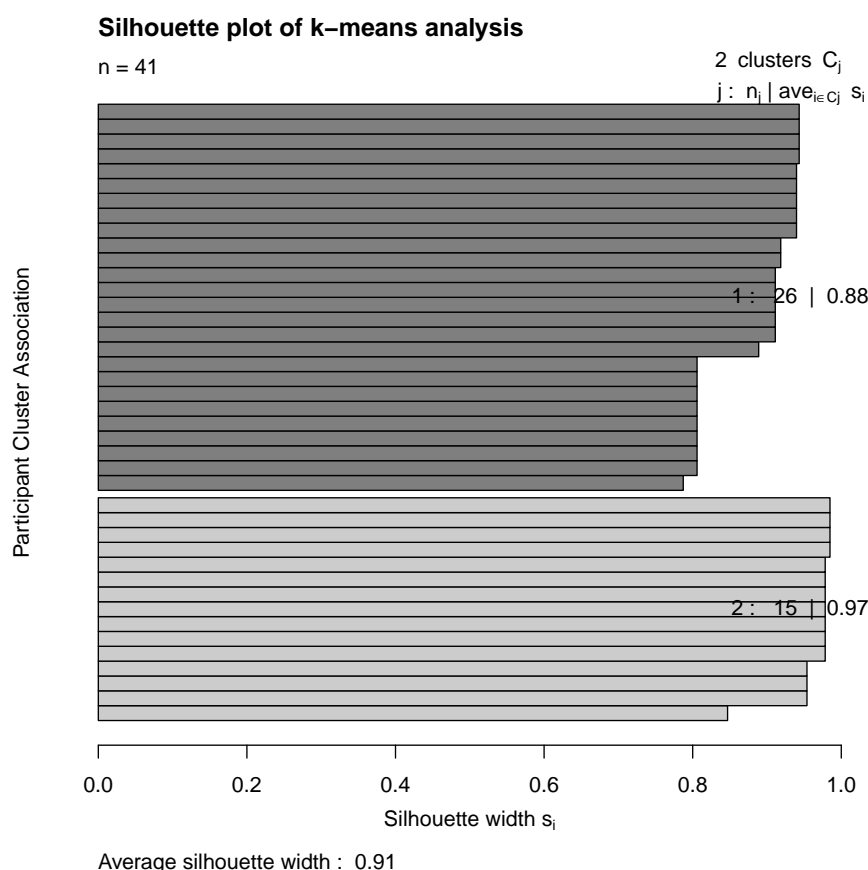

Figure 3: Silhouette plot of k-means analysis of participants by acceptability ratings

Subsequent one-tailed t-tests showed that there are no differences for each condition between population clusters, except for the DISSIMILAR condition with $t(69) = 10.5, p < 0.05$. For the CONTROL condition, we obtain $t(139) = -0.6, p > 0.1$, for the SIMILAR condition, $t(69) = 0.7, p > 0.1$, and for the DISJOINT condition, $t(69) = 0.3, p > 0.1$.

For the DISSIMILAR condition, the variance and acceptability is greatly reduced for Cluster 2:

Condition	Average Acceptability		Variance	
	Cluster 1	Cluster 2	Cluster 1	Cluster 2
CONTROL Condition	1.50	1.44	0.47	0.41
DISJOINT Condition	2.13	2.17	0.65	0.67
DISSIMILAR Condition	2.87	4.46	1.21	0.51
SIMILAR Condition	4.38	4.44	0.45	0.40

Table 4: Average acceptability of each condition, divided by population cluster

One-tailed t-tests showed that, for Cluster 1, each condition is significantly different to every other condition. Comparing the CONTROL condition to the DISJOINT condition, we obtained $t(134) = -7.2, p < 0.05$, to the DISSIMILAR condition, $t(134) = -12.4, p < 0.05$, and to the SIMILAR condition, $t(134) = -35.6, p < 0.05$. Comparing the SIMILAR condition to the DISSIMILAR condition, we obtain $t(134) = 13.8, p < 0.05$, and to the DISJOINT condition, $t(134) = 23.6, p < 0.05$. Comparing the DISSIMILAR condition to the DISJOINT condition, we obtain $t(134) = 6.4, p < 0.05$.

For Cluster 2, the same one-tailed t-tests showed that each condition is significantly different to every other condition with the exception of the one-tailed t-test between SIMILARITY and DISSIMILARITY, with $t(69) = -0.12, p > 0.1$. Comparing the CONTROL condition to the DISJOINT condition, we obtained $t(69) = -5.1, p < 0.05$, to the DISSIMILAR condition, $t(69) = -23.6, p < 0.05$, and to the SIMILAR condition, $t(69) = -27.7, p < 0.05$. Comparing the SIMILAR to the DISJOINT condition, we obtained $t(69) = 20.1, p < 0.05$. Comparing the DISSIMILAR condition to the DISJOINT condition, we obtain $t(69) = 16.2, p < 0.05$.
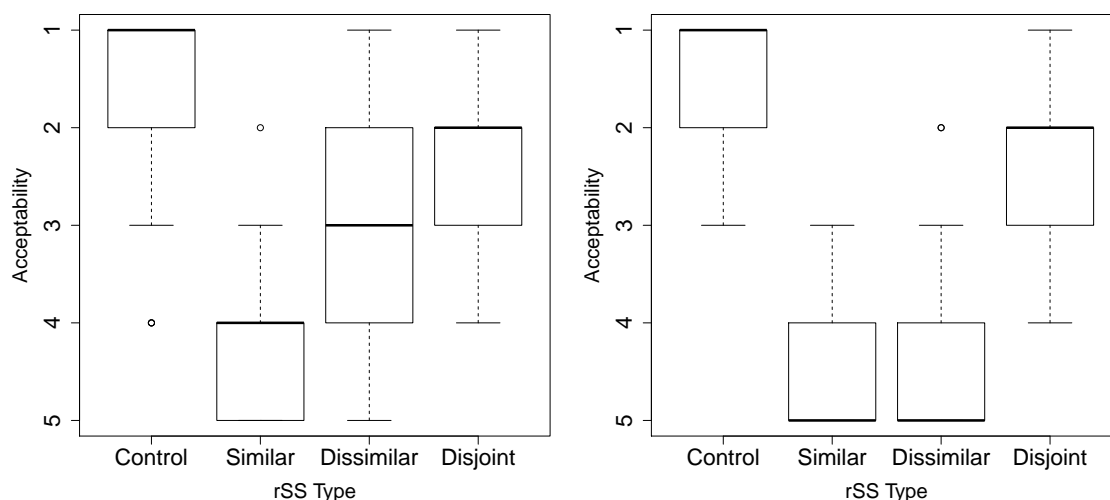

Figure 4: Boxplot for Cluster 1 (left) and Cluster 2 (right).

4. Discussion

Our experiment set out to test two independent hypotheses, repeated below.

1. If two rSS are the same except for the degree of similarity between their conditionals' antecedent worlds, then the rSS whose degree of similarity is more disparate should be considered more acceptable on average.

<div align="right">(the dissimilar worlds hypothesis)</div>

2. If the domains of quantification of an rSS are entirely disjoint, there should be no difference in acceptability between them and regular sentences (i.e. the control items).

<div align="right">(the disjoint domain hypothesis)</div>

4.1. Dissimilar Worlds Hypothesis

For the dissimilar worlds hypothesis, the experiment yielded somewhat contradictory results. In the previous section, it was shown that, for the undivided participant population, the SIMILAR condition is significantly different from the DISSIMILAR condition and that the DISSIMILAR condition yields higher values of acceptability than the SIMILAR condition by approximately one point of acceptability on average. As such, the hypothesis was technically confirmed, though the difference in acceptability was somewhat smaller and the variance of the DISSIMILAR condition much higher than anticipated. However, the k-means cluster analysis has shown that there are actually two distinct population clusters within our group of participants. Cluster 1 continued to rate the DISSIMILAR condition significantly higher in acceptability than the SIMILAR condition, now by approximately 1.5 points, but Cluster 2 appears to make no distinction between the two conditions whatsoever. Not only that, but the variance for the DISSIMILAR condition in the first population cluster is still very high ($\sigma^2 = 1.21$) and the actual distribution of acceptability judgements disconcertingly even across the board, as seen in Figure 4. As such, it seems that the participants in the first population cluster were unsure of what to do with these rSS, rather than considering them a simple improvement on the SIMILAR condition's rSS.

Both findings are unexpected by or contradictory to Lewis' (2018) account, though to different degrees. If considered only on its own, Cluster 2 would directly falsify the dissimilar worlds hypothesis. The variance of Cluster 1 would suggest that dissimilarity – whilst clearly having a positive impact on acceptability in some cases – is not the sole deciding factor (aside from relevance) behind acceptability.

We must consider whether these two findings may be explained by Lewis' (2018) model in its current state. Concerning Cluster 2, there are two explanations apparent to us: First, as the world closeness is an interaction between similarity and relevance, it might be the case that the participants of this population cluster consistently ascribe enough relevance to the $\phi \wedge \psi$-worlds s.t. they are always moved to be amongst the closest ϕ-worlds irrespective of dissimilarity. The second possibility would be that they interpret the DISSIMILAR $\phi \wedge \psi$-worlds as more similar than intended. The latter option would indicate that there might be an error in the experiment's design; more specifically, in how the DISSIMILAR condition items were created. The former option would introduce the question why these participants would consistently go through the trouble of rearranging their world ordering – even though most people would not, given the disparity in similarity – if this leads to a contradictory reading. Both by principle of economy

and charitability, it would be more suitable to leave the $\phi \wedge \psi$-worlds in their original place in the world ordering, given the vast distance they would have to cross to count amongst the closest ϕ-worlds. Concerning the variance of the first population cluster, we have a similar option to argue in favour of Lewis' (2018) model: If we were to assume that the $\phi \wedge \psi$-worlds of the DISSIMILAR condition are regarded as more similar than intended, then the participants might be more inclined to provide them with lower acceptability ratings – the higher ratings would then be an act of charitable interpretation on their part. This would however raise the question why their charitability – a successful strategy – is only intermittently employed.

Excluding, for the sake of argument, the possibility of there being an inherent flaw in the design of the DISSIMILAR rSS and assuming that Cluster 2 is not that anti-charitable, we would argue that our results, whilst weakly supporting the dissimilar worlds hypothesis, are more contradictory to than supportive of Lewis' (2018) model. Rather, the data would suggest to us that there is another main factor behind the acceptability of rSS, as further explored in § 4.3.

4.2. Disjoint Domain Hypothesis

Concerning the disjoint domain hypothesis, things are a bit more clear: Since there is a significant difference between the CONTROL condition and the DISJOINT condition, and the latter is less acceptable on average, the null hypothesis has been falsified. However, the DISJOINT rSS are, on average, only 0.67 points less acceptable than the control items. They are also far more acceptable than the other types of rSS. We reckon that this slight – though significant – degradation in acceptability may be chalked up to the markedness of rSS in general: The markedness of going from a specific case to a more general case that then seemingly contradicts the specific case, even if only on the surface. Generally speaking, in language, the reverse appears far more common (e.g. precisification). In the same line of reasoning, it appears, to us at least, quite difficult to find a natural occurrence of rSS within any given corpus – written or spoken.

4.3. Further Introspective Judgments

In § 4.1, we argued that the experimental results suggest that relevance and dissimilarity may not be the only important factors for the acceptability of rSS – perhaps not even the main ones. What, then, renders rSS acceptable in the rare instances when this is the case? From the results of the DISJOINT condition, we know that its ingredients are a recipe for (limited) success. Recalling the conditions for their creation from § 2.1, we know them to be (i) establishing the $\phi \wedge \psi$-conditional as counterfactual (by meaning, if not by syntax), whilst making the ϕ-conditional a regular future-less-vivid conditional and (ii) having both sequence conditionals share a common discourse goal explicitly named by a sentence following the rSS. As such, we tried to pin down what makes an rSS acceptable by systematically creating rSS with only one of these features or even neither of them whilst trying to keep the changes to a minimum.

(11)   **Counterfactual + Unified Discourse Purpose**
       (Said, over the telephone, to someone who is completely alone by a frozen lake, though 15 minutes away from the lake, there are some people who have told him that they definitely won't be coming to the lake, unless he came and told them he needed them.)
       Yes, if you walked on the thin ice right now while being supported by someone on the shore, the ice wouldn't break and you'd be fine; but if you DID walk on the thin ice

right now, the ice would break and you would die! So, if you really want to walk on it, go and fetch one of those people you told me about!

(12)   **Counterfactual + No Unified Discourse Purpose**
(Said, over the telephone, to someone who is completely alone by a far-off frozen lake with nobody even remotely in the vicinity.) Yes, if you walked on the thin ice right now while being supported by someone on the shore, the ice wouldn't break and you'd be fine; but if you DID walk on the thin ice right now, the ice would break and you would die! So, don't walk on the thin ice!

(13)   **Not Counterfactual + Unified Discourse Purpose**
(Said, over the telephone, to someone who is currently planning on going to a remote frozen lake next week. It is known that said person has not decided on going completely alone.) If you walked on the thin ice next week while being supported by someone on the shore, the ice wouldn't break and you'd be fine; but # if you DID walk on the thin ice next week, the ice would break and you would die! So, only walk on the ice if you have someone with you!

(14)   **Not Counterfactual + No Unified Discourse Purpose**
(Said, over the telephone, to someone who is currently planning on going to a frozen lake next week.) If you walked on the thin ice next week while being supported by someone on the shore, the ice wouldn't break and you'd be fine; but # if you DID walk on the thin ice next week, the ice would break and you would die! So, don't walk on the thin ice!

From this, it would appear that only counterfactuality is strictly needed for a felicitous rSS. The $\phi \land \psi$-conditional in (12) does not really support the goal of convincing the person near the lake not to walk upon the thin ice – at best, it does not detract from said goal, since it mentions a non-possibility with no direct bearing on the situation. In that sense, (12) is not unequivocal in the sense of being a "single pointful piece of discourse" (Edgington, 1995), i.e. sharing a common discourse purpose, though still unequivocal in the sense of the speaker not changing their mind or retracting their previous statement (Klecha, 2015).

The requirement of actual counterfactuality appears too strict, however: As seen below, rSS may be felicitous if the possibility of $\phi \land \psi$ has merely been epistemically excluded from occurring, though, in reality, it remains an option, no matter how unlikely:

(15)   **Epistemically Excluded Possibility**
(Said, over the telephone, to someone who is currently planning on going to a remote frozen lake next week that is known for usually nobody ever going there. It is known that said person has decided on going completely alone and is extremely adamant about it, because he wants get away from everything.) Listen, if you walked on the thin ice next week while being supported by someone on the shore, the ice wouldn't break and you'd be fine; but if you DID walk on the thin ice next week, the ice would break and you would die! So, I'm begging you to be careful and not to go on the ice!

With this, we may have identified the factor that threw off the results for the SIMILAR rSS. However, considering how inconclusive the results for this condition were, we also further examined Lewis' (2018) criterion it was based upon: The hypothesis that two differently similar

sets of worlds may intermingle due to relevance if the worlds in question are "similar enough". To this effect, we created a scenario that fulfilled the previously identified criterion of counterfactuality with only a minimal difference in similarity:

(16)   *(Said to John, who is at a remote cabin next to a frozen lake with some friends. John asked whether they wanted to go for a walk, but all of them declined, even though Mary nearly agreed.)* If Mary hadn't decided against coming with you and you had walked on the ice supported by her from the shore, the ice wouldn't have broken; but if you HAD walked on the thin ice, the ice would've broken. You made the right choice.

Considering that (16) seems felicitous and that we have found no rSS for which dissimilarity appears to play a role in felicity (short of the $\phi \wedge \psi$-worlds being so dissimilar that they ought to be considered excluded possibilities), we would argue that this criterion may be formally dropped: Any two sets of worlds appear similar enough so long as there is no counterfactuality or epistemically excluded possibility involved. These could be summed up as *perceived non-reality worlds*. Only if an rSS involves a set of perceived non-reality worlds does similarity play a role in whether or not the sequence is rendered felicitous or infelicitous.

Another point of interest is the stress on the auxiliary verb: The placement of the stress in the experimental items is easily motivated, since the ϕ-antecedent's *did* was the only lexical item not found within both antecedents. However, for counterfactual rSS, such as the one in (16), the matter is somewhat more complicated, since the stressed auxiliary verb *had* is part of either conditional's antecedent. This technically violates Klecha's (2015) constraint on rSS that the second conditional's antecedent may not be a syntactic subset of the preceding conditional, so as to allow for contrastive stress of some kind. Rather, here, the contrastive stress appears obligatorily placed upon the auxiliary verb *had*, as the same rSS is rendered less acceptable if the stress is prevented from being placed upon it – e.g. by using a contraction:

(17)   *(Said to someone who had just been completely alone by a frozen lake)* If you had walked on the thin ice while being supported by someone on the shore, the ice wouldn't have broken; ?? but if you'd walked on the thin ice, the ice would have broken.

The acceptability is somewhat increased, if the stress is shifted to the verb, but appears not to reach the same levels of acceptability as rSS where the auxiliary verb is stressed.

4.4. Possible Accounts

Here, we sketch out two possible, not necessarily mutually exclusive ways to account for both the data gathered in the experiment and the introspective data of the previous subsection.

4.4.1. A Modification of Lewis (2018)

First, we attempt to reconcile the data with Lewis' (2018) system with minimal modifications: As discussed in the previous section, the notion of the two antecedent world sets having to be *similar enough* has been somewhat discredited. As such, we would propose to eliminate this criterion so long as no appropriate example that is clearly in its favour has been found. The previous section has shown that counterfactuality or the epistemic exclusion of possibility appears to be the main driving factor behind rSS felicity.

To account for this, we propose to install a further restriction on the relevance-induced move-
ment of worlds: If a world is no longer a live possibility, then said world cannot move any
closer to the evaluation world than its original value of similarity. In other words, similarity
acts as a hard upper limit imposed upon world closeness when it comes to counterfactuals.
Intuitively, this may be explained as follows: Counterfactuals range over worlds that are, by
definition, no longer possible. Even if we put much relevance into what might have happened,
this does not make said worlds appear any more close, similar, or familiar to the actual world.
For conditionals that range over live possibilities, on the other hand, similarity may rather be
seen as a metric for expectation, and expectations are rather malleable by nature.

For epistemically excluded possibility worlds, there are two options: First, to analyse them as
counterfactual worlds. Second, to assign to them a relevance value so marginal that salience
alone has no discernible impact upon the closeness of these worlds (i.e. they are so irrelevant
that they do not move closer to the evaluation world, even if mentioned).

With this, we would already explain why the DISJOINT rSS are so acceptable and why the
DISSIMILAR rSS are not rated any higher than the SIMILAR rSS for Cluster 2: Even though
they may quantify over dissimilar and highly improbable worlds, they do not approach the
same level of epistemic exclusion as (15) does and, therefore, move close enough to count
amongst the closest ϕ-worlds. The respective results of Cluster 1 may be explained by an
assumption of charitability: Cluster 1 participants actively try to accommodate for the rSS to
be felicitous by attempting to interpret the extreme dissimilarity between the $\phi \wedge \psi$-worlds and
ϕ-worlds as a sign for the epistemic exclusion of $\phi \wedge \psi$ as a possibility, even though this is not
explicitly stated or sufficiently alluded to within the context. Dissimilarity being an unreliable
factor for this kind of accommodation then causes the results to fluctuate out of control or, in
other words, causes the participants' confusion regarding these matters.

These modifications have very little impact upon Lewis (2016, 2018). Aside from her reasoning
on the infelicity of rSS, essentially all of her points are left functionally untouched: Her analysis
for the problem of counterfactual scepticism remains as is, and Krassnig's (2017) reunification
of Lewis sequences and True Sobel sequences into a single phenomenon, using Lewis' (2018)
model and Bennett (2003) or Arregui's (2009) similarity metric, should also remain unaffected.

4.4.2. Contrastive Stress Account

The second attempt to account for the data within this paper would be to take a closer look at
contrastive stress and the role it plays in rSS – something we have so far ignored in
Lewis' (2018) account. Specifically, its effect on the auxiliary verb and why the stress is obliga-
torily placed upon it if there is a lack of better alternatives, as previously demonstrated with (15)
and (17). Optimally, any analysis that accounts for this would also account for why the simi-
larity values between $\phi \wedge \psi$ and ϕ must be clearly different[6] and why, for non-counterfactuals,
epistemic exclusion of possibility appears necessary, but simple unlikelihood does not. To
achieve this, we need to make two main assumptions that have already been argued for and

---

[6]This may appear a trivial requirement but is far from it. A broad, simple exhaustification-based account, for
example, might lead to an analysis along the lines of interpreting appropriately stressed rSS as "$\phi \wedge \psi \square\!\!\rightarrow \chi$, but
$\phi \wedge \neg\psi \square\!\!\rightarrow \neg\chi$", which would yield felicity regardless of similarity. Something that would contradict our empirical
and introspective findings and would turn even reverse Lewis sequences felicitous.

been independently motivated in the literature: (i) We assume that focus or stress on a bound pronoun and other anaphoric lexical items of this kind affects not the function meaning of the pronoun but rather the domain of its binder (Jacobson, 2004: p. 155ff.). (ii) We also assume that the antecedents of conditionals actively set the current aboutness topic (Ebert et al., 2008), not only for conditionals, but also for counterfactuals (Ebert et al., 2008: p. 139).[7]

With these assumptions in mind, we refer to the fact that tense and mood have, in one way or another, frequently been associated with the world selection process of the antecedent for the subsequent evaluation of the consequent (see, amongst others, Palmer, 1986; Iatridou, 2000; Arregui, 2009; Romero, 2014; Schulz, 2014). We posit that, lacking any available difference in syntax, the only difference in semantics – the difference in the selection of worlds – is the target for the obligatory contrastive stress. The auxiliary verb is merely stressed because it is bound by the tense and mood of the antecedent and stressing it therefore directly affects its binder's domain. This is analogous to how the focus effect of a stressed bound pronoun does not affect the pronoun itself but rather the domain of its binder (Jacobson, 2004: p. 155ff.). However, rather than assuming that this contrastively used stress yields a simple exhaustifying implicature, i.e. that the ϕ-conditional is effectively interpreted as $\phi \wedge \neg\psi$, we assume that this is a case of contrastive (aboutness) topic – since the antecedents, in general, represent the aboutness topics of their respective conditional (Ebert et al., 2008). However, for a contrastive aboutness topic to actually be contrastive, the contrastive topic needs to be necessarily different from the topic it is contrasted against. If a contrastive topic construction is used, such as some fronting construction, but the sentence in question refers to an already present aboutness topic, the acceptability of said utterance degrades greatly (see, amongst others, Reinhart, 1981; Büring, 2016). See the following classical example, adapted from Reinhart (1981: p. 63):

(18)   a.   Felix is an obnoxious guy.
          b.   Even Matilda can't stand him.

(19)   a.   Felix is an obnoxious guy.
          b.   ?As for Matilda, even she can't stand him.

The degradation in acceptability in (19) is blamed by Reinhart (1981) on the fact that (19a) sets Felix as the aboutness topic, (19b) then contrastively sets Mathilda as the current aboutness topic, yet loops back to a sentence that is still quintessentially about the previous aboutness topic: Felix. For rSS with contrastive stress on the auxiliary verb, something very similar happens. The contrastive stress (or accent) on the auxiliary verb marks the antecedent not only as a contrastive aboutness topic, but also specifies in which regard the current topic deviates from the previous one: The domain of worlds under which the consequent is to be evaluated. The regular rSS pattern could therefore be paraphrased along the lines of (19):

(20)   a.   In all the closest $\phi \wedge \psi$-worlds, $\chi = 0$ needs to be the case.
          b.   As for all the closest ϕ-worlds, $\chi = 1$ needs to be the case.

---

[7]Strictly speaking, Ebert et al. (2008) only claim that antecedents act as aboutness topics when they are fronted (i.e. they precede their consequent). In our experiment and the remainder of this paper, we have only considered and analysed such fronted antecedent conditionals. As such, the topicality of non-fronted antecedents is not of immediate concern. However, should rSS prove just as acceptable with non-fronted antecedents – given otherwise identical conditions – then this will have to be addressed. In such a case, if this account was to extend to rSS without fronted antecedents, the topicality of fronted antecedents would most likely have to be extended to non-fronted antecedents as well.

The contrastive marking of the ϕ-antecedent therefore tries to introduce an entirely new aboutness topic. If the closest $\phi \wedge \psi$-worlds were just as close to the evaluation world as the closest ϕ-worlds, then this would entail that the set of the closest $\phi \wedge \psi$-worlds is a subset of the set of closest ϕ-worlds.[8] We would, therefore, not strictly introduce a new aboutness topic, but rather loop back to the old one, even though we may add a couple of worlds in the process. We believe that, in addition to the semantic contradictions that may arise, the failure to introduce an adequately contrastive aboutness topic is responsible for the unacceptability of rSS whose antecedents' world similarity values are either identical or unclear. Furthermore, if no contrastive stress / topic is found, then the ϕ-conditional may be interpreted as still belonging to the same aboutness topic as before – motivating a modal subordination reading of the ϕ-conditional, yielding a contradiction, motivating a proposal such as Klecha's (2015: p. 136). Generalising all of this, we argue that the topicality of antecedents therefore enforces, pragmatically, that a difference in similarity between the rSS' antecedent worlds needs to be clearly distinguishable.

This would explain the results of the DISJOINT condition and also account for our introspective observation that the difference in dissimilarity need only be minimal, yet clearly delineated. Yet, this still leaves open why indicatives and future-less-vivid conditionals are generally infelicitous, short of epistemically excluded possibilities. Why this is strictly the case needs to be further developed, though it is likely linked to the fact that, for live possibility worlds, the similarity ordering is, by nature, far less clearly demarcated than their counterfactual counterparts'.

For a more developed, formalised account of this nature, we would refer to Krassnig (in prep.).

5. Conclusion

We have shown that dissimilarity's role is contrary to Lewis' (2018) reasoning: Rather than two sets of worlds having to be similar enough to one another in order for rSS to become infelicitous, any clear-cut degree of dissimilarity is enough to render rSS felicitous. That is, so long as the rSS in question is either a counterfactual or treats the closest $\phi \wedge \psi$-worlds as epistemically excluded possibilities. We have also shown that contrastive stress need not necessarily require an overtly different item but may also target semantic differences – e.g. when the ϕ-antecedent's auxiliary verb of a counterfactual rSS is stressed. We have sketched out two accounts: a modification of Lewis (2018) s.t. her predictions are more in line with our results and one account that motivates why contrastive stress is necessary, why contrastive stress may fall upon the auxiliary verb (obligatorily so, lacking any other candidates), and why the dissimilarity requirement is enforced not only on a semantic level but also at the level of discourse. Concerning future research, we believe it necessary for further rSS experiments to be conducted, as there is a lot left to uncover or confirm with regards to what makes and what unmakes rSS (in)felicity. The contrastive topic account also requires further formalisation and exploration – for instance, regarding the (in-)felicity of rSS whose antecedents are not fronted. Further independent motivation for relevance interacting with world similarity to bring about world closeness would also be a worthwhile avenue, to better independently motivate such a model apart from the problem of counterfactual scepticism, rSS, and similar conditionals.

---

[8]The entailment would either hold true due to world closeness being equal to world similarity (Stalnaker, 1968; Lewis, 1973) or, carrying over the assumptions from § 4.4.1, because the $\phi \wedge \psi$-worlds would automatically get pushed to be as close to the evaluation world as they are similar to it, upon becoming salient, under our modified account of Lewis (2018) – at least insofar as counterfactuals are concerned.

References

Arregui, A. (2009). On similarity in counterfactuals. *Linguistics and Philosophy 32*(3), 245–278.

Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford: Oxford University Press.

Büring, D. (2016). (Contrastive) topic. In C. Féry and S. Ishihara (Eds.), *The Oxford handbook of information structure*, pp. 64–85. Oxford University Press.

Ebert, C., C. Endriss, and S. Hinterwimmer (2008). Topics as speech acts: An analysis of conditionals. In N. Abner and J. Bishop (Eds.), *Proceedings of the 27th West Coast Conference on Formal Linguistics*, Somerville, MA, pp. 132–140. Cascadilla Proceedings Project.

Edgington, D. (1995). On conditionals. *Mind 104*(414), 235–329.

Gillies, A. S. (2007). Counterfactual scorekeeping. *Linguistics and Philosophy 30*(3), 329–360.

Heim, I. (1994). Unpublished MIT seminar on conditionals.

Iatridou, S. (2000). The grammatical ingredients of counterfactuality. *Linguistic Inquiry 31*(2), 231–270.

Jacobson, P. (2004). Kennedy's puzzle: What I'm named or who I am. In R. Young (Ed.), *Proceedings of SALT XIV*, pp. 145–162.

Klecha, P. (2014). *Bridging the Divide: Scalarity and Modality*. Ph. D. thesis, University of Chicago.

Klecha, P. (2015). Two kinds of Sobel sequences: Precision in conditionals. In U. Steindl, T. Borer, H. Fang, A. G. Pardo, P. Guekguezian, B. Hsu, C. O'Hara, and I. C. Ouyyang (Eds.), *Proceedings of the 32nd West Coast Conference on Formal Linguistics*, Somerville, MA, pp. 131–140. Cascadilla Proceedings Project.

Krassnig, D. (2017). Sobel sequences: Relevancy or imprecision? In A. Cremers, T. van Gessel, and F. Roelofsen (Eds.), *Proceedings for the Amsterdam Colloquium 2017*, pp. 325–334.

Krassnig, D. (in preparation). *The Semantics and Pragmatics of Counterfactual Sentences at the Discourse Level: Sobel-Sequences and the Licensing of Negative Polarity Items*. Ph. D. thesis, University of Konstanz.

Lewis, D. K. (1973). *Counterfactuals*. Malden: Blackwell.

Lewis, K. (2016). Elusive counterfactuals. *Noûs 50*(2), 286–313.

Lewis, K. (2018). Counterfactual discourse in context. *Noûs 52*(3), 481–507.

Moss, S. (2012). On the pragmatics of counterfactuals. *Noûs 46*(3), 561–586.

Palmer, F. R. (1986). *Mood and Modality*. Cambridge: Cambridge University Press.

Reinhart, T. (1981). Pragmatics and linguistics: An analysis of sentence topics. *Philosophica 27*, 53–94.

Romero, M. (2014). 'Fake tense' in counterfactuals: A temporal remoteness approach. In L. Crnič and U. Sauerland (Eds.), *The Art and Craft of Semantics: A Festschrift for Irene Heim*, Volume 2, pp. 47–63. MITWPL 71.

Schulz, K. (2014). Fake tense in conditional sentences: A modal approach. *Natural Language Semantics 22*, 117–144.

Stalnaker, R. C. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in Logical Theory*, pp. 98–112. Oxford: Basil Blackwell Publishers.

von Fintel, K. (2001). Counterfactuals in a dynamic context. In M. Kenstowicz (Ed.), *Ken Hale: A Life in Language*, pp. 123–152. Cambridge: MIT Press.