Counterfactual Dependence and Arrow

Thomas Kroedel

Humboldt University of Berlin

Franz Huber

University of Konstanz

Penultimate version: please cite the paper in *Noûs* – Franz Huber, 07/07/2014

ABSTRACT: We argue that a semantics for counterfactual conditionals in terms of comparative overall similarity faces a formal limitation due to Arrow's impossibility theorem from social choice theory. According to Lewis's account, the truth-conditions for counterfactual conditionals are given in terms of the comparative overall similarity between possible worlds, which is in turn determined by various aspects of similarity between possible worlds. We argue that a function from aspects of similarity to overall similarity should satisfy certain plausible constraints while Arrow's impossibility theorem rules out that such a function satisfies all the constraints simultaneously. We argue that a way out of this impasse is to represent aspectual similarity in terms of ranking functions instead of representing it in a purely ordinal fashion. Further, we argue against the claim that the determination of overall similarity by aspects of similarity faces a difficulty in addition to the Arrovian limitation, namely the incommensurability of different aspects of similarity. The phenomena that have been cited as evidence for such incommensurability are best explained by ordinary vagueness.

KEYWORDS: counterfactual conditionals; comparative overall similarity; Arrow's impossibility theorem; ranking functions; incommensurability

1 Introduction

Lewis's (1973a) semantics for counterfactual conditionals has been highly popular, at least among those philosophers who believe that counterfactuals have truth-conditions.[1] According to his semantics, a counterfactual is non-vacuously true in the actual world if and only if there is a possible world where both its antecedent and its consequent are true that is closer to the actual world than any worlds where its antecedent is true while its consequent is false. In addition to this semantics, Lewis gives an account of the closeness relation that is at work in his semantics. This closeness relation is to be understood as a relation of comparative overall similarity: a possible world *w* is closer to the actual world than a possible world *v* if and only if *w* is more similar overall to the actual world than *v*. The overall similarity between worlds, Lewis holds, is a function of different aspects of similarity between worlds, just as, say, the overall similarity between people is a function of different aspects of similarity between people (1973c: 420–21). So far we do not disagree with Lewis. We shall argue, however, that the relation of comparative overall similarity cannot be a function of mere relations of comparative aspectual similarity for very general reasons. If aspects of similarity are to determine overall similarity, we have to make additional assumptions about how these aspects are measured on pain of falling prey to an application of Arrow's impossibility theorem.[2]

2 Overall Similarity

Here is a minimal model of how the truth or falsity of a counterfactual is determined that is suggested by Lewis's account (see 1973a: 50–52). A given counterfactual is asserted in a given context. This somehow determines a number of relevant aspects of similarity. How

[1] Edgington (1995, 2004) denies this.

[2] We owe the inspiration to this application of Arrow's impossibility theorem to Samir Okasha, who, in a talk at the University of Konstanz in July 2009, applied it to the aggregation of the overall virtue of scientific theories (see Okasha 2011). The applicability of the theorem to similarity relations was discovered independently by Michael Morreau (2010). Morreau draws more radical conclusions than we do, which we criticise in section 4 below.

possible worlds are ordered according to their comparative similarity to the actual world in these aspects then somehow determines how they are ordered according to their comparative overall similarity to the actual world. The latter then determines whether or not the counterfactual is true according to Lewis's truth-conditions. Let us have a closer look at how comparative aspectual similarity determines comparative overall similarity. (We shall assume, at least for the sake of argument, that there are no problems with the determination of the relevant aspects of similarity by the assertion of a given counterfactual in a given context.) Technically speaking, we are looking for a function that maps a number of relations of the form 'world *w* is at least as similar to the actual world according to aspect of similarity *i* as world *v* is' to the relation 'world *w* is at least as similar overall to the actual world as world *v* is'. Let an *aspectual similarity ordering* be a relation of the former kind and an *overall similarity ordering* a relation of the latter kind. Let a *profile* be a given number of aspectual similarity orderings in a given order.[3] Then we can state more precisely what we are looking for: a function from a profile to an overall similarity ordering.

Not any such function will do. A reasonable function from a profile to an overall similarity ordering should satisfy a number of constraints. One such constraint is that the overall similarity ordering should not overrule a unanimous judgement, by all aspects of similarity, that a certain world is more similar to the actual world than another. This is expressed by the following condition:

> P (Weak Pareto principle). If world *w* is more similar to the actual world than world *v* according to all aspects of similarity, then *w* is more similar overall to the actual world than *v*.

Further, a reasonable function should, when determining the comparative overall similarity of worlds *w* and *v* to the actual world, only take into account what the different aspectual

[3] Even more technically speaking, given *n* aspects of similarity, a profile is an *n*-tuple of aspectual similarity orderings.

similarity orderings say about *w* and *v*, and ignore where they place other worlds. That is, a reasonable function should satisfy:

> I (Independence of irrelevant alternatives). If two profiles do not differ with respect to worlds *w* and *v*, they determine the same overall similarity ordering with respect to *w* and *v*.

Another reasonable requirement is that the function be general enough to be able to handle all combinations of aspectual similarity orderings. That is, it should satisfy:

> U (Unrestricted domain). The domain of the function from aspects of similarities to overall similarity includes all possible profiles of orderings of worlds.

Finally, we should demand that there be no aspect of similarity that always overrules other aspects. In other words:

> D (Non-dictatorship). There is no aspect of similarity such that whenever a world *w* is more similar to the actual world according to this aspect than a world *v*, *w* is more similar overall to the actual world than *v*.

The requirement that D should hold can be illustrated by the parallel case of the comparative overall similarity between people. If D – more precisely, a condition analogous to D in terms of people instead of worlds – were violated, there would be one aspect of similarity such that resembling the target person more than another person does in this respect would entail resembling the target person more overall than this person. For instance, having a nose (say) that resembles Caesar's nose more closely than Pompey's nose does would by itself entail being more similar overall to Caesar than Pompey. This would seem absurd – Joe Bloggs might have little claim to similarity to Caesar apart from his Caesarian nose, leaving Pompey more similar to Caesar overall despite Bloggs's one aspectual advantage. More generally, the

more reasonable view is that even perfect similarity in one respect can always be outweighed by dissimilarity in other respects.[4] So D should hold.

Here comes the catch. The aspectual similarity orderings and the overall similarity ordering have the formal properties of preference orderings (viz. reflexivity, transitivity and completeness). Therefore, the determination of an overall similarity ordering by aspectual similarity orderings can be used as an interpretation of social choice theory. Standardly, social choice theory deals with individuals whose individual preferences with respect to certain alternatives need to be aggregated to a collective preference. In our case, the aspects of similarity play the role of the individual preferences, and overall similarity plays the role of the collective preference. Since the determination of an overall similarity ordering by aspectual similarity orderings is an interpretation of social choice theory, it is subject to the results of social choice theory. In particular, it is subject to Arrow's impossibility theorem. And according to Arrow's impossibility theorem, given that there are at least three worlds, there is no function that satisfies the four conditions P, I, U, and D simultaneously.[5] We have reached an impasse.

Before presenting our solution to this problem, further discussion of conditions U and D is in order. Is Lewis committed to a denial of condition D? Assuming determinism and restricting his attention to the "standard resolution of vagueness" (that is, to standard contexts), Lewis holds that overall similarity is determined by the following "system of weights or priorities" (1979: 472):

(1) It is of the first importance to avoid big, widespread, diverse violations of law.

(2) It is of the second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails.

(3) It is of the third importance to avoid even small, localized, simple violations of law.

[4] See also Morreau 2010: 473–80.

[5] Our presentation of Arrow's impossibility theorem follows Gaertner 2009: 19–21. For the original proof see Arrow 1951. Morreau 2010 contains a generalised result.

(4) It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly.

The aspects of similarity that are at work in (1)–(4) are (i) the absence of big, widespread, diverse violations of law (Lewis calls such violations of law "big miracles" (1986: 55–56)), (ii) the size of the spatio-temporal region in which there is perfect match of particular fact, (iii) the absence of small, localised, simple violations of law (also known as "small miracles" (*ibid.*)), and (iv) the degree of approximate similarity of particular fact.

On the reading that is prima facie most natural, (1)–(4) state priorities. Thus, whenever a big miracle occurs in world *v* but not in world *w*, *w* will come out more similar overall to the actual world than *v* according to (1). If neither *w* nor *v* involves a big miracle while *w* has more match of particular fact with the actual world than *v*, then *w* will be more similar overall to the actual world than *v* according to (2); etc. On this reading, (1) implies that there is a dictatorial aspect of similarity, so that D is violated. Somewhat inelegantly, we may phrase the aspectual similarity ordering corresponding to this aspect as 'world *w* is more similar to the actual world than world *v* is with respect to the absence of big miracles in *w* but not in *v*'. Read as a system of priorities, (1)–(4) have been met with criticism.[6] It might be, however, that Lewis did not intend overall similarity to be governed by a system of priorities in the first place. Elsewhere he states merely that the weighting of similarities "*might* be nonarchimedean; that is, we *might* have a system of priorities rather than trade-offs" (1986: 54, emphasis added). And in *Counterfactuals*, he seems inclined towards trade-offs rather than priorities when he writes that trying too hard for similarity in one respect would result in "excessive differences in some other respect" (1973a: 9).

However, even if Lewis did intend (1)–(4) to state priorities and was correct, it does not follow that a general denial of condition D is justified. For (1)–(4) are restricted to the case of determinism and to standard contexts and thus are not guaranteed to afford a general

[6] See, for instance, Bennett 2003: chs. 12–14, Woodward 2003: sec. 3.6, Kment 2006, and Wasserman 2006. Most of the extant objections to Lewis focus on the claim that aspect (ii) overrules all other aspects given that the candidate worlds are on a par with respect to aspect (i), but prima facie condition (1) does not seem to be any more sacred than condition (2).

account of how overall similarity is determined. Notice also that (1)–(4) seem to stand their best chance of success for counterfactuals whose antecedents are about specific events (perhaps implicitly), such as the infamous 'If Nixon had pressed the button, there would have been a nuclear holocaust' (Fine 1975). It is far from clear that Lewis succeeds in generalising his treatment of these cases to counterfactuals whose antecedents are not about specific events, such as 'If kangaroos had no tails, they would topple over'.[7] For the reasons indicated above, we think that, if Lewis's restrictions are lifted and counterfactuals with antecedents that are not about specific events are allowed, it will be possible to outweigh an allegedly dictatorial aspect of similarity by other aspects. Thus, D holds at least in some cases, which is enough to generate the Arrovian impossibility (given that P, I, and U hold in these cases as well).

Let us turn to the unrestricted domain condition U. On the face of it, U may seem to be the most innocent of the four conditions P, I, U, and D.[8] It might, however, be questioned on the following grounds. In U, 'possible' means formally possible: the arguments our function can handle comprise all finite combinations of aspectual similarity orderings of worlds. Thus U can be regarded as a constraint on the choice of aspects of similarity: U requires them to be mutually independent. To illustrate this, consider the following case, where the requirement of independence is violated. Suppose that one relevant aspect of similarity is similarity with respect to *F*-facts and another similarity with respect to *G*-facts. The corresponding aspectual similarity orderings are 'world *w* is at least as similar to the actual world with respect to *F*-facts as world *v* is' (call this the *F-ordering*) and 'world *w* is at least as similar to the actual world with respect to *G*-facts as world *v* is' (call this the *G-ordering*). Suppose further that *G*-facts supervene on *F*-facts in the sense no two worlds differ with respect to *G*-facts without also differing with respect to *F*-facts. Finally, suppose that the same *F*-facts obtain in the two

[7] See Nute 1980: 104, Bennett 2003: 119–20, and Weatherson 2010: sec. 3.3. Counterfactuals with arbitrary antecedents are required for projects such as Williamson's (2007: ch. 5), which hold that counterfactuals subsume metaphysical modality. See also Kment 2006.

[8] Morreau does not even state condition U, even though he makes use of it in his proof (which is based on Geanakoplos 2005) by presupposing that it is always possible to choose an arbitrary similarity profile with certain desired features (see Morreau 2010: 489).

worlds *w* and *v*. Since the same *F*-facts obtain in *w* and *v*, *w* is as similar to the actual world as *v* is with respect to *F*-facts. By supervenience, the same *G*-facts obtain in *w* and *v*; so *w* also is as similar to the actual world as *v* is with respect to *G*-facts. In other words, both the *F*-ordering and the *G*-ordering have to treat *w* and *v* as on a par. This rules out a profile where the *F*-ordering ranks *w* and *v* as on a par while the *G*-ordering ranks *w* as more similar to the actual world than *v* (or vice versa). Such a profile, though formally possible, would be genuinely impossible, and it seems that we cannot demand that our function from profiles to overall similarity orderings be able to handle such profiles. So in the example condition U is violated.

How likely is it that the relevant aspects of similarity will thus violate condition U? In the analogous case of comparative overall similarity between people, the aspects of similarity that might come to mind – say similarities with respect to different physical features – do not seem to exhibit any supervenience-induced dependence. The best candidate for such dependence in the case of comparative overall similarity between worlds seems to be similarity with respect to laws of nature and similarity with respect to particular facts on the assumption that laws supervene on particular facts.[9] However, even if laws do thus supervene and both similarity with respect to laws and similarity with respect to particular facts are relevant for the determination of comparative overall similarity in some cases, the problem raised by Arrow's impossibility theorem will persist. For it suffices that there is some collection of aspects of similarity that are independent as required by U for Arrow's impossibility theorem to apply (given that P, I, and D hold for them as well). A single counterfactual in an appropriate context will guarantee the existence of such a collection. We shall provide an example in the next section.

Another counterexample to condition U might seem to come from the position of the actual world in the aspectual similarity orderings. One might hold that, according to any aspect of similarity, the actual world is at least as similar to itself as every other world, or that,

[9] Note that the similarity with respect to laws invoked here is distinct from aspects (i) and (iii) mentioned above, since the latter concern violations of the *actual* laws of nature in different worlds, while here similarity with respect to laws of nature compares the laws of nature that hold in different worlds.

according to any aspect of similarity, the actual world is more similar to itself than every other world.[10] The former condition rules out profiles where some aspectual similarity ordering ranks some non-actual world more similar to the actual world than the actual world itself; the latter condition additionally rules out profiles where some aspectual similarity ordering ranks some non-actual world equally similar to the actual world as the actual world itself. However, if one of these conditions holds, this does not solve the threat posed by Arrow's impossibility theorem but merely shifts the problem. Even if the position of the actual world is fixed in all aspectual similarity orderings, we still need to determine the comparative overall similarity of all the non-actual worlds from their comparative aspectual similarities. The Arrovian impossibility re-appears here, for by the independence of irrelevant alternatives I, how the non-actual worlds are ordered vis-à-vis one another is not affected by how they are ordered vis-à-vis the actual world.

A third difficulty with condition U arises when we take into account more comprehensive relations of comparative similarity. So far we have only considered comparative similarity to the actual world. Thus, the relations of comparative similarity discussed in section 2 have been two-place relations, viz. 'world *w* is at least as similar overall to the actual world as world *v* is' and, for a fixed aspect of similarity *i*, the relations 'world *w* is at least as similar to the actual world according to aspect of similarity *i* as world *v* is'. We could equally well have used the asymmetric relations 'world *w* more similar overall to the actual world than world *v* is' and 'world *w* is more similar to the actual world according to aspect of similarity *i* than world *v* is', since they can be defined in terms of the former relations (and vice versa). Now, the relation Lewis invokes is in fact the three-place relation 'world *w* is more similar overall to world *u* than world *v* is' (see 1973a: 48–50); this allows him to evaluate counterfactuals not just at the actual world, but also at non-actual worlds. In the spirit of the minimal model of how comparative overall similarity is determined, this three-place relation of comparative overall similarity should be taken to be a function of a number of three-place relations of the form 'world *w* is more similar to world *u* according to

[10] These conditions correspond, respectively, to the conditions of "Weak centering" and "Centering" for overall similarity (Lewis 1973a: 120–21).

aspect *i* than world *v* is'; call these relations *three-place aspectual similarity orderings*. On the face of it, invoking three-place aspectual similarity orderings does not seem to affect the problem from Arrow's impossibility theorem. For we can conceive of a three-place aspectual similarity ordering as a family of standard two-place aspectual similarity orderings with one such two-place ordering for each fixed *u*. Each member of this family, it seems, can still be required to satisfy the four conditions P, I, U, and D.

However, one might hold that

(a) *w* is more similar to *u* according to aspect *i* than *v* is

and

(b) *v* is more similar to *w* according to aspect *i* than *u* is

imply

(c) *w* is more similar to *v* according to aspect *i* than *u* is.

If this implication holds, it will no longer be the case that each two-place aspectual similarity ordering satisfies condition U. For (a) and (b), by implying (c), will then rule out that the two-place ordering for the comparative aspectual similarity to *v* ranks, say, *u* as more similar than *w*. It is not entirely clear whether the implication from (a) and (b) to (c) holds.[11] If it turned out that it did, we would have to use, instead of Arrow's impossibility theorem, a generalised impossibility theorem that allows two-place aspectual similarity orderings to constrain one another in accordance with the implication from (a) and (b) to (c). As far as we know, whether such a generalised impossibility theorem exists has not yet been studied in social choice theory. However, since Arrow's impossibility theorem has proved rather robust in the face of weakenings of condition U (see Campbell and Kelly 2002: 64), it seems likely that the required generalised impossibility holds.

To sum up the results reached so far, it seems plausible that a function that aggregates aspectual similarity orderings of worlds to an overall similarity ordering should satisfy the

[11] See Lewis 1973a: 51–52 and Williamson 1988: 459–61 for discussion.

weak Pareto principle P, the independence of irrelevant alternatives I, the condition of unrestricted domain, and the non-dictatorship condition D. However, by Arrow's impossibility theorem, there can be no such function. What are the options available in response to this problem? One might want to uphold the minimal model of how comparative overall similarity is determined. In this case, one will have to give up at least one of the four conditions P, I, U, and D after all. Perhaps D is the condition most likely to be given up, since one might hope to make a case that Lewis's conditions (1)–(4) should be read as priorities and can be extended beyond their limited scope of application. If one drops just one of the four conditions, Arrow's impossibility theorem still has an interesting application, since it will imply that giving up the condition in question is non-discretionary: if the other three conditions are satisfied, the condition in question has to be false (given there are at least three worlds). Since we take all four conditions to have a good standing, we shall not pursue this strategy, however. In social choice theory Sen (1970) proposed to circumvent Arrovian impossibility by enriching the informational basis: individuals are equipped with real-valued utility functions instead of mere preference orderings. It is then possible to aggregate these utility functions. We shall show that an analogous solution is available in our case if comparative similarity is represented by ranking functions (see Spohn 1988). As we shall argue in the next section, ranking functions allow us to avoid Arrovian impossibility in much the same way as real-valued utility functions do.

3 Ranking Functions

The general strategy we would like to advocate in response to the Arrovian problem for the aggregation of overall similarity is to represent similarity in a numerical way, particularly to start from numerical representations of aspectual similarity, thus providing more information than the information that worlds are ordered thus-and-so with respect to their aspectual similarity to the actual world. Our vehicles of choice for this numerical representation are ranking functions.

Ranking functions are functions r from the set of possible worlds W into the class of ordinal numbers such that at least one world w is assigned value 0. The value of a ranking function for a given world is called the world's *rank*. In this paper we consider only ranking functions with natural numbers as ranks. While this is not their original interpretation, we can think of ranking functions as encoding worlds' similarity, aspectual as well as overall, to the actual world. In particular, world w is at least as similar overall to the actual world as world v is just in case the rank encoding w's overall similarity to the actual world, $r(w)$, is not greater than the rank encoding v's overall similarity to the actual world, $r(v)$. Aspectual similarity can in turn be represented by ranking functions r_i so that world w is at least as similar to the actual world according to aspect i as world v is just in case $r_i(w)$ is not greater than $r_i(v)$.[12] Ranking function r can be determined by ranking functions r_i as the weighted sum of the latter. Thus, we can define an aggregation function f such that

$$r(w) = f(r_1(w), ..., r_n(w)) = m_1r_1(w) + ... + m_nr_n(w) - \text{MIN},$$

where the m_i are natural numbers and MIN is a 'normalisation' parameter, the minimum of the numbers $m_1r_1(w) + ... + m_nr_n(w)$ for the worlds w, that guarantees that the resulting ranking function r is well-defined and maps at least one world to 0.

How does switching from a purely ordinal framework to ranking functions solve the problem from Arrow's impossibility theorem? Since the four conditions P, I, U, and D are formulated in terms of purely ordinal representations of aspectual and overall comparative similarity, it does not make sense to say that our aggregation function f satisfies these conditions. However, there is an f that does satisfy the following analogous conditions:[13]

---

[12] Note that ranking functions, in contrast to real-valued similarity measures, do not impose a constraint on the cardinality of the set of worlds which differ from each other in their degree of similarity to the actual world (see Lewis 1973a: 50–52, Williamson 1988: 458–59).

[13] Note that the existence of an aggregation function f that satisfies these conditions is not threatened by another impossibility result due to Seidenfeld *et al.* (1989). The latter prove a theorem to the effect that two Bayesian agents, represented by pairs of different probabilities and utilities (P_1, U_1) and (P_2, U_2) respectively, cannot find a Bayesian compromise (P, U) respecting the strict preferences of both (P_1, U_1) and (P_2, U_2) that

PR If $r_i(w) < r_i(v)$ for all i, $1 \leq i \leq n$, then $r(w) < r(v)$.

IR For all w in W, $r(w)$ is determined solely by $r_1(w), \ldots, r_n(w)$.

UR f is defined for all possible n-tuples of ranking functions.

DR There is no r_i, $1 \leq i \leq n$, such that $r(w) < r(v)$ if $r_i(w) < r_i(v)$.

One of us (Huber manuscript) has recently proposed a semantics for counterfactuals in terms of ranking functions. Aggregating comparative overall similarity of worlds by way of ranking functions does not commit one to endorse such a semantics, however. For ranking function r induces an overall similarity ordering. This was already implicit in the informal characterization above: world w is at least as similar overall to the actual world as world v is just in case $r(w)$ is not greater than $r(v)$. We can also formulate four conditions analogous to P, I, U, and D in terms of this induced overall similarity ordering (where 'wOv' abbreviates 'w is at least as similar to the actual world as v is'):

PR* If $r_i(w) < r_i(v)$ for all i, $1 \leq i \leq n$, then wOv.

IR* For all w, $v \in W$, whether or not wOv is determined solely by $r_1(w)$, $r_1(v)$, …, $r_n(w)$, $r_n(v)$.

UR* f is defined for all possible n-tuples of ranking functions.

DR* There is no r_i, $1 \leq i \leq n$, such that wOv if $r_i(w) < r_i(v)$.

It is easy to see that PR*, IR*, UR*, and DR* are true if PR, IR, UR, and DR are. Thus, one might well adhere to Lewis's truth-conditions for counterfactuals in terms of a purely ordinal notion of comparative overall similarity if one is willing to replace the minimal model of

---

does not coincide either with (P_1, U_1) or with (P_2, U_2). This new impossibility theorem cannot be avoided in the ways Arrow's impossibility result can be avoided. In particular, the new impossibility result cannot be avoided by "add[ing] structure to the representation of individual preferences" (*ibid*.: 227), which is the strategy we have pursued above. The reason our approach is not threatened by this new impossibility result is that our aggregation function f does not result in a Bayesian compromise (P, U) with probabilities P and utilities U. Instead, it results in a ranking function, which by its nature is not subject to Bayesian constraints such as expected utility maximisation.

aggregating comparative overall similarity from purely ordinal comparative aspectual similarity with the aggregation involving ranking functions.

Let us illustrate this somewhat dry treatment of similarity aggregation by means of ranking functions by an example. (Note that, in this example, the individual aspects of similarity are independent in the sense of condition U.) The department was filling a research position on the sole basis of the candidates' publications. Three candidates, Alice, Brenda, and Charlotte, were on the shortlist. In the first scenario, Alice had one book with a decent publisher and forty articles in top journals; Brenda had no book, but forty articles in top journals; and Charlotte had one book with a decent publisher, but no articles. Alice was hired. In the second scenario, everything is as in the first scenario, except that Alice and Brenda had only two articles each. In both scenarios, with respect to books, Charlotte's publication list was more similar to Alice's than Brenda's. In both scenarios, with respect to articles, Brenda's publication list was more similar to Alice's than Charlotte's.

	Books	Articles
Alice	1	40
Brenda	0	40
Charlotte	1	0

First Scenario

	Books	Articles
Alice	1	2
Brenda	0	2
Charlotte	1	0

Second Scenario

In the first scenario, the following is true. If Alice had declined while the department had still made an offer, Brenda would have got it. Furthermore, if both Alice and Brenda had declined while the department had still made an offer, Charlotte would have got it. In the

second scenario, by contrast, the following is true. If Alice had declined while the department had still made an offer, Charlotte would have got it. Furthermore, if both Alice and Charlotte had declined while the department had still made an offer, Brenda would have got it.

The overall similarity between worlds governing the counterfactuals in these scenarios coincides with the overall similarity in publications. In the first scenario, worlds where Brenda got an offer are more similar to the actual world (where Alice got an offer) than worlds where Charlotte got an offer. In the second scenario, worlds where Charlotte got an offer are more similar to the actual world (where again Alice got an offer) than worlds where Brenda got an offer.

Let similarity in published books and similarity in published articles be represented, respectively, by ranking functions r_b and r_a. In the first scenario, let ranking function r_b assign the following ranks: worlds where Alice or Charlotte got the offer are assigned rank 0, and worlds where Brenda got an offer are assigned rank 1. Further, let ranking function r_a assign the following ranks: worlds where Alice or Brenda got an offer are assigned rank 0, and worlds where Charlotte got an offer are assigned rank 40. In the second scenario, let r_b be as before; and let r_a assign rank 0 to worlds where Alice or Brenda got an offer, and rank 2 to worlds where Charlotte got an offer. Then any weighted sum of r_b and r_a where one book with a decent publisher counts more than two articles in top journals, but less than forty, results in an overall ranking function r that validates the counterfactuals of both scenarios. Formally, this weighted sum is described by an aggregation function f such that for all worlds w in W, $r(w) = f(r_a(w), r_b(w)) = m_a r_a(w) + m_b r_b(w)$ with $3 \leq m_b \leq 39$ if, for simplicity, we let m_a be 1.

What about a case where the department applied different criteria in assessing the candidates? Suppose that the qualifications of Alice, Brenda, and Charlotte are as in the second scenario. Suppose further that there is a fourth candidate, Deborah, who does not have any publications, so that worlds where Deborah gets the offer are assigned rank 2 by r_a and rank 1 by r_b. Suppose, lastly, that the department's sole criterion is that candidates have at least one publication. Then it seems reasonable that worlds where Alice, Brenda, and Charlotte, respectively, get the offer should each be assigned overall rank 0, while worlds where Deborah gets the offer should be assigned an overall rank greater than 0. If the overall

ranking function r is determined as described at the end of the previous paragraph, it will be impossible to produce these new overall ranks from the original aspectual ranking functions r_a and r_b.[14] For only if weighting factors m_a and m_b are both set to 0 will worlds where Alice, Brenda, and Charlotte, respectively, get the offer receive the desired overall rank 0; but then worlds where Deborah gets the offer will also – wrongly – be assigned overall rank 0. This case is no counterexample to the aggregation of overall similarity that we suggested or to the representation of similarity by ranking functions, however. Rather, it illustrates that different aspects of similarity can be in play in different cases. In the original scenarios, the details of the candidates' publication records were relevant for the similarity aspects, yielding the ranking functions r_a and r_b. In the new case, the details are no longer relevant. This yields a new ranking function for aspectual similarity which assigns rank 0 to worlds where Alice, Brenda, and Charlotte, respectively, get the offer, and a non-zero rank to worlds where Deborah gets the offer. In the absence of further aspects of similarity, this ranking function coincides with the ranking function for overall similarity. It is not our aim to present an account of how the relevant similarity aspects are determined by the context here, or of the epistemology of this determination relation. But assuming that the relevant aspects are somehow determined by the context, there is no obstacle to their being represented, and aggregated, by ranking functions.

In this section, we have argued that the problem for a purely ordinal framework of similarity posed by Arrow's impossibility theorem can be overcome by representing similarity in a numerical way, particularly by employing ranking functions. Let us now turn to an argument against numerical representations of comparative similarity in general.

4 Against Incommensurability

Morreau (2010) endorses not just the applicability of Arrow's impossibility theorem to the determination of overall similarity; he makes the stronger claim that overall similarity cannot be determined even when metric concepts (such as real-valued utility functions or ranking

[14] Thanks to an anonymous referee for drawing out attention to this issue.

functions) are applied because the different aspects of similarity are incommensurable. In this section we shall argue that, contra Morreau, there is no such incommensurability. Where aspects of similarity seem to be incommensurable, this is due to their aggregation being vague.

It is prima facie plausible that we make judgments about the overall similarity between things (between different people, different cars, etc.) very often, and that these judgements are somehow based on judgements of aspects of similarity. Thus, the burden of proof should be on someone who advocates incommensurability rather than on someone who denies it. Morreau (2010: 480–83) takes up this burden and provides arguments for his incommensurability thesis. We shall attempt to show, however, that his arguments fall short of establishing what they purport to establish.

One of Morreau's arguments draws on an example by Keynes (1921: 36) about the similarity between books with respect to the colour and the kind of leather of their binding. According to Morreau (2010: 480–81), a book bound in blue morocco "bears some overall likeness" to one bound in red morocco. "You can decrease this likeness by changing the material of its binding from morocco to calf, while keeping the color the same. But you cannot regain the original overall likeness to the book in red morocco by subsequently changing the color of the calf binding from blue to red. More similarity in respect of color will not make up for less similarity in respect of the kind of leather." While this argument is plausible, it does not establish that similarity in respect of the colour of a binding and similarity in respect of the kind of leather of a binding are incommensurable. It shows, instead, that there is some covariation of overall similarity with similarity in respect of the kind of leather since the overall likeness is decreased by changing the latter. It also shows that, if we take into account only colour and the kind of leather, similarity in respect of the kind of leather is dictatorial since the original likeness cannot be regained by any change in colour (even though in the example *some* overall likeness is regained by the change in colour). Neither result amounts to showing that the two aspects of similarity that are in play are incommensurable.

Another argument Morreau puts forward uses the case of overall similarity between people and the aspects of similarity with respect to weight on the one hand, and of similarity with respect to temperature on the other. His argument is as follows. Suppose that a certain person resembles you more closely, overall, than another person. Let the former person become slightly less like you with respect to weight, say by gaining a pound or so. Morreau (2010: 481) poses "these questions: How much warmer or cooler should he become to restore the original overall comparison? How much more similar in respect of his height? What about his income or his wisdom or hairstyle? That there might be factual answers to these questions is hard to believe."

We agree that Morreau's questions are hard to answer. We hold, however, that the explanation for this is not any incommensurability between different aspects of similarity, but rather ordinary vagueness. How much warmer or cooler the person in question has to become (or how he would have to change with respect to height, income, or wisdom) seems just as unclear as how many hairs are required to turn a bald person into one that is not bald. Explaining vagueness-induced unclarity is a difficult task for philosophy, but there is no special difficulty involved in Morreau's example.[15]

Morreau anticipates this objection and presents data that are supposed to rule out ordinary vagueness as the explanation of his case. He holds that, were vagueness to blame, we should be able to make out at least rough tradeoffs between, say, similarity with respect to weight and similarity with respect to temperature. That is, there should be trade-offs that we could express by using suitably vague language involving expressions such as 'only a little' or 'quite a lot'. But no such rough trade-offs are available, Morreau holds: "When someone has become less like you in respect of his weight […], we cannot say that he will need to become 'only a little' or 'quite a lot' or 'vastly' more similar in respect of his temperature in order to regain his earlier overall likeness to you. […] [A]s far as we can tell, there are no rates of exchange here" (2010: 482). We disagree. Suppose that, originally, Jones is heavier and much hotter than you, and that he gains one kilogram of weight. Then it seems perfectly natural to say that he would need to become 'only a little' cooler to regain his earlier degree of overall

[15] See Williamson 1994 and Keefe 2000 for recent discussions of the problem of vagueness.

similarity to you, while we would say that he would have to become 'vastly' cooler, had he gained a hundred kilos. Notice also that the unclarity involved in questions about trade-offs is not constant. The answer to the question 'Would a 1 °C decrease of temperature make up for a 100 kg increase in weight?' is more clearly negative than the answer to the question 'Would a 1 °C decrease of temperature make up for a 1 kg increase in weight?'. This, too, suggests that vagueness rather than incommensurability is at work, since typical cases of vagueness involve inconstant degrees of unclarity (compare 'Is someone with 10 hairs bald?' versus 'Is someone with 10,000 hairs bald?'). By contrast, if two dimensions are genuinely incommensurable, their incommensurability and the corresponding unclarity in our judgements should not depend on what pair of values along these dimensions we are considering.

Thus, we hold that, when comparing the likeness of books or the similarity between people, one is facing questions of the same kind as in the example from the previous section. What number of articles in top journals are enough to make up for one book with a decent publisher? Two are not enough; forty are; but what about ten? This is difficult to answer, but the difficulty is due to vagueness, not incommensurability.

5 Conclusion

We have argued that a semantics for counterfactuals that requires aggregating comparative overall similarity in a purely ordinal fashion faces a formal limitation due to Arrow's impossibility theorem from social choice theory. We have then shown how to escape this formal impossibility by representing comparative similarity in terms of ranking functions. Further, we have argued that the additional limitation claimed by Morreau (2010), viz. incommensurability, does not exist. The phenomena he presents as evidence for incommensurability are best explained by ordinary vagueness.[16]

[16] We are grateful to Franz Dietrich, Luke Glynn, Brian Leahy, Robert Michels, Michael Morreau, Samir Okasha, Martin Rechenauer, Tobias Rosefeldt, Wolfgang Spohn, Timothy Williamson, and two anonymous referees for helpful comments and suggestions. Franz Huber's research was supported by the German Research Foundation through its Emmy Noether program.

References

Arrow, K. J. 1951. *Social Choice and Individual Values*. New York: John Wiley.

Bennett, J. 2003. *A Philosophical Guide to Conditionals*. Oxford: Clarendon Press.

Campbell, D. E., and Kelly, J. S. 2002. Impossibility theorems in the Arrovian framework. In *Handbook of Social Choice and Welfare, Volume 1*, ed. K. J. Arrow, A. K. Sen, and K. Suzumura, Amsterdam: Elsevier.

Edgington, D. 1995. On conditionals. *Mind* 104: 235–329.

Edgington, D. 2004. Counterfactuals and the benefit of hindsight. In *Cause and Chance*, ed. P. Dowe and P. Noordhof. London: Routledge.

Fine, K. 1975. Critical notice of D. Lewis, *Counterfactuals*. *Mind* 84: 451–58.

Gaertner, W. 2009. *A Primer in Social Choice Theory*. Oxford: Oxford University Press.

Geanakoplos, J. 2005. Three brief proofs of Arrow's impossibility theorem. *Economic Theory* 26: 211–15.

Huber, F. (manuscript). New foundations for counterfactuals.

Keefe, R. 2000. *Theories of Vagueness*. Cambridge: Cambridge University Press.

Keynes, J. M. 1921. *A Treatise on Probability*, London: Macmillan and Co.

Kment, B. 2006. Counterfactuals and explanation. *Mind* 115: 261–309.

Lewis, D. 1973a. *Counterfactuals*. Oxford: Blackwell.

Lewis, D. 1973b. Causation. *Journal of Philosophy* 70: 556–67.

Lewis, D. 1973c. Counterfactuals and comparative similarity. *Journal of Philosophical Logic* 2: 418–46.

Lewis, D. 1979. Counterfactual dependence and time's arrow. *Noûs* 13: 455–76.

Lewis, D. 1986, Postscripts to 'Counterfactual dependence and time's arrow'. In his *Philosophical Papers, Volume II*. New York: Oxford University Press.

Morreau, M. 2010. It just does not add up: trouble with overall similarities. *Journal of Philosophy* 107: 469–90.

Nute, D. 1980. *Topics in Conditional Logic*. Dordrecht: Reidel.

Okasha, S. 2011. Theory choice and social choice: Kuhn versus Arrow. *Mind* 120: 83–115.

Seidenfeld, T., Kadane, J. B., and Schervish, M. J. 1989. On the shared preferences of two Bayesian decision makers. *Journal of Philosophy* 86: 225–244.

Sen, A. K. 1970. *Collective Choice and Social Welfare*. San Francisco: Holden-Day.

Spohn, W. 1988. Ordinal conditional functions: a dynamic theory of epistemic states. In *Causation in Decision, Belief Change, and Statistics II*, ed. W. L. Harper and B. Skyrms. Dordrecht: Kluwer.

Wasserman, R. 2006. The future similarity objection revisited. *Synthese* 150: 57–67.

Weatherson, B. 2010. David Lewis. In *The Stanford Encyclopedia of Philosophy* (Summer 2010 Edition), ed. E. N. Zalta. URL = <http://plato.stanford.edu/archives/sum2010/entries/david-lewis/>.

Williamson, T. 1994. *Vagueness*. London: Routledge.

Williamson, T. 1988. First-order logics for comparative similarity. *Notre Dame Journal of Formal Logic* 29: 457–81.

Williamson, T. 2007. *The Philosophy of Philosophy*. Oxford: Blackwell.

Woodward, J. 2003. *Making Things Happen*. New York: Oxford University Press.