TWO ACCOUNTS OF SIMILARITY COMPARED

LIEVEN DECOCK and IGOR DOUVEN
VU University Amsterdam and University of Leuven

Abstract
Tversky's account of similarity could be said to reduce similarity to identity and thereby to reduce an allegedly philosophically problematic notion to an unproblematic one. In Gärdenfors's more recent account of similarity, similarity figures as a primitive, unreduced notion. We argue that this gives no reason for preferring Tversky's account to Gärdenfors's.

Both philosophical and psychological theorizing about similarity has long been dominated by what is sometimes called "the geometrical model of similarity" according to which similarity relations can be represented by means of a metric similarity space (see, e.g., Carnap [1928/1967], Coombs [1954], Shepard [1958], and Torgerson [1958], [1965]). A metric space is a pair (X, d) with X a set and d a metric (or distance function) on X, that is, $d: X \times X \rightarrow R$ with, for all a, b, and c in X:

Minimality:[1] $d(a, b) \geq 0$ and $d(a, b) = 0$ iff $a = b$
Symmetry: $d(a, b) = d(b, a)$
Triangle Inequality: $d(a, b) + d(b, c) \geq d(a, c)$

On this approach, a similarity relation is defined by reference to a fixed distance in the metric similarity space.[2] Specifically, objects a and b are said to be similar iff they are within some distance $\mathbf{t} \in R^{+}$:

$$\text{sim}(a, b) \Leftrightarrow d(a, b) \leq \mathbf{t}$$

---

[1] This condition is slightly stronger than the condition that Tversky [1977:328] calls "Minimality," which is this: $d(a, b) \geq d(a, a) = 0$. However, a function satisfying the latter condition together with Symmetry and Triangle Inequality is, contrary to what Tversky suggests, not guaranteed to be a metric, but only a pseudo-metric.
[2] Strictly speaking, the similarity function need not be the distance function; it may, for instance, also be an exponentially decaying function of the distance (see on this Shepard [1987], Nosofsky [1988], [1992], Hahn and Chater [1997]).

As plausible as the model might at first appear, in his influential "Features of similarity" ([1977]), the psychologist Amos Tversky shows that it fails dismally in accounting for certain empirical data. He therefore proposed to replace it by a set-theoretical approach to similarity that demonstrably does a much better job in explaining the relevant data. This approach has become hugely influential over the past two decades, and it is no exaggeration to say that at present it has the status of the received doctrine.

In the present paper, we contrast Tversky's popular account with a very recent proposal made by Peter Gärdenfors, which can be conceived as a refinement of the old geometrical model. While to date this refined model is not known to be empirically any less successful than Tversky's, the latter may seem to have the advantage that it reduces a notion that, following Nelson Goodman [1972], many philosophers have judged to be slippery – similarity – to one that is almost universally held to be "utterly simple and unproblematic" (Lewis [1986:192]), namely identity; by contrast, similarity figures as a primitive, unreduced – and for all we know irreducible – notion in Gärdenfors's account. We argue that, appearances to the contrary notwithstanding, this fails to give good reason for preferring Tversky's account to Gärdenfors's.

I

Tversky's main motive for dismissing the geometrical model of similarity is based on empirical evidence concerning the plausibility of the basic axioms of metric spaces when interpreted in terms of similarity. Minimality already poses a problem. For it entails reflexivity of similarity, that is, that S(*a*, *a*) is equal to S(*b*, *b*) for all objects *a* and *b*, whereas in psychological experiments the probability of judging two stimuli identical is not the same for all kinds of identical stimuli. Triangle Inequality is not corroborated by the data either; in some experiments in which objects were similar in entirely different respects, violation of Triangle Inequality occurred (Tversky and Gati [1982]). Nevertheless, in the first place, Tversky's critique concerns Symmetry. He demonstrated that people's similarity judgments are in many cases *not* symmetrical. For example, experiments reveal that North-Korea is typically judged to be more similar to China, than China to North-

Korea – which seems to provide a direct refutation of the geometrical model.

This led Tversky to develop a set-theoretical approach to similarity based on feature matching. In his ontology, objects in the domain $D = \{a, b, c,\ldots\}$ are not characterized by points in a geometrical space, but by a set of features; for example, an orange may be represented by a set of features $A = \{$round, orange, medium-sized, juicy, ripe,$\ldots\}$. Similarity is then defined in terms of certain set-theoretical relations obtaining (or otherwise) between sets of features representing different objects.

To be more exact, the similarity relation, on Tversky's proposal, must fulfil several conditions. The most important of these are the matching condition and the monotonicity condition. The first says that

$$S(a, b) = F(A \cap B, A{-}B, B{-}A),$$

for some real-valued ternary function F. In other words, the similarity of two objects a and b is strictly a function of the set of their shared features ($A \cap B$) and the two sets of their distinctive features ($A{-}B$, the set of features that belong to a but not to b, and $B{-}A$, the set of features that belong to b but not to a). According to the monotonicity condition,

S(a, b) $\geq$ S(a, c) whenever (i) $A \cap C \subseteq A \cap B$; (ii) $A{-}B \subseteq A{-}C$; and (iii) $B{-}A \subseteq C{-}A$, with the inequality being strict iff at least one of the three inclusion relations is proper.

Less formally, an object b is more similar to an object a than another object c is if either the features a and b share are among the ones shared by a and c, or the features *not* shared by a and b are among the ones *not* shared by a and c, or both. Thus, similarity increases with addition of common features and/or deletion of distinctive features. In addition to these, there are also three more technical conditions: independence, solvability, and invariance; we refer the reader to Tversky [1977:351] for the definitions.

Tversky then proves a representation theorem to the effect that, if S(. , .) fulfils the foregoing conditions, there is guaranteed to exist a similarity scale s and a nonnegative scale f such that for all a, b, c, d in the domain:

(i) $s(a, b) \geq s(c, d)$ iff $S(a, b) \geq S(c, d)$;
(ii) $s(a, b) = \theta f(A \cap B) - \alpha f(A-B) - \beta f(B-A)$, for some $\theta, \alpha, \beta \geq 0$;
(iii) f and s are interval scales.

This representation of the similarity relation has been named "the contrast model": similarity is expressed as a linear combination, or a contrast, of the measures of the common and the distinctive features.[3]

Based on this rough exposition, it should already be clear that the contrast model easily explains the experimentally established asymmetry in people's similarity judgments. For nothing in the model requires that $\alpha = \beta$ (the task presented to the participants in an experiment may be what Tversky calls "directional"), so that if the complement sets of features do not have the same measure (i.e., if $f(A-B) \neq f(B-A)$), we have the asymmetrical $S(a, b) \neq S(b, a)$. This simple fact has done much to propagate the popularity of Tversky's model.

II

It is also worth stressing that the representation theorem does not characterize a unique similarity scale s, but rather a family of such scales. For instance, if $\alpha = \beta = 0$, then we have a similarity relation that only depends on the shared features of two objects. If $\theta = 0$, we have a similarity relation that only depends on the distinctive features between the two objects. In addition to this, a function f must be fixed. This function is most naturally construed as a salience function. Certain sets of common or distinctive features contribute more to the similarity scale than others, and are thus more salient in the comparison of the objects.

This is a further virtue of Tversky's approach, for it helps in dealing with what had generally been recognized as another major obstacle – next to the asymmetry problem – for the geometrical model, to wit the context-dependence of similarity judgments. One of the first to emphasize this context-dependence, often approvingly cited in the literature (see, e.g., Medin et al. [1993:254] and Gärdenfors [2000:109]), was Goodman:

---

[3] Besides the contrast model, there are other models based on feature matching. An interesting alternative is the ratio model, according to which $S(a, b) = f(A \cap B)/[f(A \cap B) + \alpha f(A-B) + \beta f(B-A)]$; see Tversky [1977:333].

[C]omparative judgments of similarity often require not merely selection of relevant properties but a weighting of their relative importance, and variation in both relevance and importance can be rapid and enormous. Consider baggage at an airport checking station. The spectator may notice shape, size, color, material, and even make of luggage; the pilot is more concerned with weight, and the passenger with destination and ownership. Which pieces are more alike than others depends not only upon what properties they share, but upon who makes the comparison, and when. … Circumstances alter similarities. (Goodman [1972:445])

For Goodman, the extreme context-dependence of the similarity relation was a reason to dismiss similarity as "a pretender, an impostor, a quack" ([1972:437]). Name calling aside, it is hard to see how the geometrical model could deal with this context-dependence, lacking, as it does, any parameter that might allow for contextual variation. By contrast, Tversky's model can be regarded as being an appropriate reply to Goodman's worries. Goodman argued that every use of the notion of similarity is in need of a frame of reference. Tversky's account can easily accommodate this kind of relativity precisely because, as intimated, the contrast model provides a *class* of similarity scales, and it is easily imaginable how, for instance, the salience of certain sets of features (the function f) may change from one context to another. Moreover, similarity, on the current proposal, is relative to a particular selection of features. The set of features that represents an object will in general be only a subset of the set of all the properties the object has. For the representation of the objects in the domain, only a limited set of features is considered relevant, and the selection of these features is relative to certain interests and purposes. For instance, in Goodman's example, the selected features of baggage are different for the bystander, the pilot, and the passenger. Needless to point out, these interests and purposes may vary per context. Further, the selection of the domain may be context-sensitive, which may have implications for the similarity judgments, as Tversky's extension effect illustrates (Tversky [1977:343]; see also Medin et al. [1993:361]). If all the objects under consideration share a feature, it becomes neutral in the comparison. However, if one adds an object not sharing it in a comparison task, the feature at once becomes salient. The activation of this feature may shift the similarity judgments between the former objects that do share the feature. In this way, adding a non-European country in a comparison task involving only European countries may have an impact on previous similarity judgments.

In fact, similarity judgments may be even more sensitive to context than Tversky anticipated. In his view, it is possible, at least in experimental set-ups, to prevent context shifts from occurring, and therefore also to select a priori a similarity scale appropriate for the context of the experiment relative to which the subjects' similarity judgments can be assessed. But, backed by experimental findings, Medin et al. [1993] doubt that this is correct. As they say:

> It is natural to assume that, to constrain similarity comparison appropriately, the representation of each of the constituent terms must be rigid [i.e., context-insensitive]. In contrast, our observations suggest that the effective representations of the constituents are determined in the context of the comparison, not prior to it. It is as if the two terms were dancers: Each dancer may have a repertoire of stylistic preferences, but the actual performance depends on an interaction between the two. For asymmetrical comparisons, the "base dancer" takes the lead and the "target dancer" follows. The result is appropriately constrained even though the constituents are quite flexible. (Medin et al. [1993:275])

In other words, the "respect" in which two objects are similar is, or at least may be, selected in the process of comparing the objects, and not (necessarily) a priori. Accordingly, Medin et al. put forward a double explanation of the asymmetry of certain similarity judgments. As in Tversky's account, the similarity scale s_C can be asymmetrical in context C. Furthermore, the order of presentation of the objects may determine which object "takes the lead." Reversing this order may result in the selection of a different context. Hence, in a particular situation, the similarity judgment $S(a, b)$ may be based on $s_C(a, b)$, while the similarity judgment $S(b, a)$ is based on $s_{C'}(b, a)$. This difference in context can either be a difference in the domain, a difference in selected features, a difference in salience of the selected features, a difference in the contrast parameters θ, α, β, or a combination of the foregoing.

Clearly, it is not difficult to adapt Tversky's contrast model so that it also accommodates this further kind of context-sensitivity: as the above already suggests, replacing the context-insensitive similarity scale $s(. , .)$ by a set of scales $\{s_C(. , .)\}$, containing one scale for each context C, should already do the job. Equally clearly, Medin et al.'s observation might seem to be the nail in the coffin of the geometrical model. Still, more context-dependence can only be more bad news for a model that, as we saw, appears unable to represent any context-dependence at all.

III

And yet some authors have recently tried to resuscitate the geometrical approach. Specifically, they have tried to modify the geometrical model in a way which enables it to account for both the asymmetry effect reported in Tversky's work and that pointed to by Medin et al. The result, we contend, is a serious rival of the contrast model.

Similarity *can* be thought of as having a geometrical structure, the said authors urged. But according to the original geometrical model, and in line with what some philosophers have claimed,[4] a single general metric similarity space underlies all similarity judgements. And that is a mistake, as is already amply testified by the manifest context-dependence of similarity judgments. To rectify the mistake, in recent psychological literature, most notably in Gärdenfors's *Conceptual Spaces* ([2000]), a *contextual* geometrical notion of similarity has been developed. The point of departure of these endeavors has been the plausible claim that general similarity judgments are not possible, and that, first, objects can only be similar in certain respects, and second, different respects can be salient in different contexts. They then argue that the relevant notion of respect can be cashed out by reference to conceptual spaces. For example, if one compares objects with respect to color, one invokes a color space. Another, perhaps even simpler, example of a conceptual space is a three-dimensional Euclidean space with a Euclidian metric serving to represent proximity in visual space.[5] Still further examples are the one-dimensional temporal space, the two-dimensional auditory space, olfactory and tactile spaces, and even shape spaces (see Gärdenfors [2000:94-98]). In short, one crucial idea is that, instead of a general similarity space, we have at our avail a multitude of conceptual spaces (or similarity spaces). The second crucial idea is that in each context typically only a subset of all these spaces is activated, to wit, those corresponding to the respects that are salient in the given context.

---

[4] In Carnap [1967], a complete reconstruction of our knowledge of the world is based on a single relation *Rs* (recollection of similarity). This relation is described as "the given." By means of *Rs*, one can easily define similarity itself. It is clear that a geometrical conception of similarity looms in the background. Though there is room for scholarly debate on the precise role of these geometrical notions in Carnap's *Aufbau*, the constitution system seems to imply the existence of a general geometrical notion of similarity.

[5] One often finds the term "phenomenal visual space" for this space (e.g., in Shepard [2001]). For a discussion of the metaphysical status of this space, see Decock [2006].

As a matter of fact, this is only one way in which the refined geometrical model is able to accommodate context-dependence. In addition to selecting, per context, different conceptual spaces, one can also rescale, per context, the distances within any single conceptual space, and thus obtain a related but different similarity relation. As a result, similarity judgments can also be relative to a particular distance function on the selected conceptual spaces.

Evidently, the foregoing furnishes adequate answers to the objections that Goodman and Medin et al. had levelled against the uncontextualized geometrical model. Just as importantly, it also answers Tversky's asymmetry objection, for nothing precludes that a comparison of China with North-Korea selects a different set of conceptual spaces than a comparison of North-Korea with China, or alternatively – or in addition – that the two tasks lead to different selections of distance functions on the same (or partly the same) conceptual spaces.

IV

It has now come to appear that the notion of similarity can be analyzed in at least two – very different – ways. Further, from a metaphysical viewpoint the two accounts are very different. Gärdenfors's account explains similarity by means of primitive conceptual or similarity spaces. For all he has shown, or we are aware of, contextualized geometrical similarity cannot be reduced to other notions. In Tversky's contrast model, on the other hand, similarity is a function of the sets of shared and distinct features. This effectively amounts to a reduction of the notion of similarity to basic set-theoretical concepts, the concept of context (or that of salience), and the concepts of sharedness and distinctness, where these last concepts can be further reduced to that of identity (that objects *a* and *b* share a feature means that there is a feature of *a* that is identical to a feature of *b,* and the set of distinct features consists of those that one of the objects has without the other having the identical feature). In brief, Tversky's account seems to have the distinctive virtue of reducing an allegedly problematic notion to an uncontested one (identity) and one that at least nowadays many analytic philosophers tend to regard as relatively unproblematic (the notion of context).

Therefore, even if the two accounts of similarity considered above are both empirically adequate – which they are, for all that is presently

known[6] – then, purely on the basis of theoretical considerations, Tversky's account might still be said to be superior. However, this would be rash, for at least two reasons.

Firstly, even supposing the foregoing to be an advantage of Tversky's account, it may well be offset by advantages that Gärdenfors's account has in comparison with its rival. Without going into details, we point in this connection to the fact that conceptual spaces can at the same time partake in an explanation of categorization and concept formation. This is important, as in many cases there is an interdependence of categorization and similarity judgments (see, e.g., Gärdenfors [2000], Hahn and Ramscar [2001], and references therein). In Tversky [1977], this point is obfuscated by the fact that in all the experiments reported in that paper it was obvious which were the features relevant to the given comparison task, where moreover these features were always artificially precise (see for instance the highly stylized faces in Figure 2 on p. 331 of Tversky's paper). In real-life comparison tasks, by contrast, the relevant features are almost invariably of a more complex nature, and the selection of those features may not be straightforward at all.

Secondly, philosophical criticisms of the notion of similarity seem, at least in part, to have exploited the deficiencies of the traditional geometrical model, all of which seem to have been remedied in Gärdenfors's contextualized geometrical model. More generally, it seems to us that Gärdenfors has succeeded in establishing similarity as a scientifically kosher notion and thereby also – a lesson from naturalism that, we suppose, is broadly accepted nowadays – a philosophically respectable one.[7] At a minimum, Gärdenfors's important work should make the critics of similar-

---

[6] Our discussion is confined to accounts that aim to model similarity relations between single objects. As accounts of similarity relations between what in the literature are sometimes called "multipart scenes" (which typically involve multiple objects or figures), both Tversky's and Gärdenfors's account may fare less well; see Goldstone and Son [2005:22ff] and references given there.

[7] Note that this is not necessarily to deny Goodman's [1977] claim that similarity cannot do a lot of philosophically useful work. But, first, even if this claim is correct, that does not automatically make similarity philosophically suspect. For instance, if deflationists are right, then truth does not do a lot of philosophical work. Yet they do not claim that truth is philosophically suspect. Second, the remarks to be made in the final section of this paper, and more clearly the arguments given in Douven and Decock [2009], suggest that, pace Goodman, similarity does do important work for philosophers, if for no other reason then because it helps to dissolve the so-called paradoxes of identity.

ity want to reconsider their earlier harsh verdicts (recall Goodman's earlier-cited words: "Similarity … is a pretender, an impostor, a quack"). In light of this, it is much less clear that it is advantageous for Tversky's account that it reduces similarity to identity (and other ostensibly unproblematic notions).

One might respond here that even if similarity has been shown to be a philosophically respectable notion, we still prefer a metaphysics on which one of the notions of similarity and identity can be reduced to the other to a metaphysics that makes both come out as being fundamental. A proper assessment of this objection would require us to go much more deeply into the general topic of reductionism than we can do here. But even granting – what we doubt – that it is generally a good strategy to reduce as many notions as possible to as few as possible, we would still be inclined to disagree that the fact that Tversky has managed to reduce similarity to identity gives grounds for favoring his account over Gärdenfors's. The reason is that even if the latter does not reduce similarity to identity, it may enable us to go in the other direction and reduce identity to similarity. We briefly explain this in the final section.

V

The paradoxes of identity – puzzle cases involving change over time and questions about constitution – exhibit that people's identity judgments have properties that cannot directly derive from the properties of the identity predicate, as standardly conceived. For instance, they show that identity judgments are context-sensitive, that they may fail to be transitive, that they can be vague or indeterminate, and that intuitions as to the correctness of a given identity judgment may vary greatly among people.

To give a well-worn example, few think that replacing one plank of a ship by a new plank yields a different ship. However, when more, possibly all, planks get replaced, people become more inclined to think that the resulting ship *is* different from the one that had all the original planks in place. More generally, we typically judge that small changes to an object preserve its identity, but a series of small changes may add up to a big change, and big changes are often not judged to preserve identity. This suggests that people's identity judgments may fail to respect the putative transitivity of the identity predicate.

Moreover, none of those who think that the final ship is no longer identical to the original ship will be able to specify an exact number n such that, for them, after the nth plank had been replaced, the new ship came into existence. Rather, they will say that, for some part of the transformation process, the ship was *more or less* identical to the original ship. It would thus seem that identity judgments can be vague to some extent.

In connection with this same example, Plutarch already remarked that identity judgments display a considerable interpersonal variability: he noted that, in discussing the example, philosophers reached different verdicts concerning the identity (or otherwise) of the ship with the original planks and the ship with the new planks. In fact, anyone who has ever taught an introductory metaphysics course will have first-hand knowledge of the point him- or herself.

The context-sensitivity of identity judgments is also illustrated nicely by some of the paradoxes of identity. One may think here of the various paradoxes of personal identity: in some contexts it may be perfectly all right to say that Harry at sixteen is not identical to Harry at sixty, whereas in other contexts that may seem wrong (see Douven and Decock [2009, Sect.I]). Or think of the famous puzzle concerning the statue and the lump of bronze that makes it up: in a context in which we are attending to the different modal and historical properties the statue and the lump of bronze may have, we are inclined to judge these to be different objects, but when the said properties are ignored, as often they are, then we may be as much inclined to judge them to be identical.

Patently, these features of our identity judgments cannot be directly explained, and even seem to put some pressure on, what one might call "the logical notion of identity". According to this, after all, the identity relation is context-invariant and transitive, and it holds categorically: its holding cannot be a matter of degree. The standard approach to explaining the said features has been to attribute misconceptions on our part of other metaphysical notions, like that of an object or that of a property. In Douven and Decock [2009], we have proposed a different approach. The basic idea is that the notion of identity at play in the identity judgments that appear to conflict in the paradoxes of identity is not the logical notion of identity but rather what we called "the folk notion of identity", according to which "identity" is to be interpreted as meaning "highly similar in all relevant respects," where both "highly" and "respect" are to thought of as context-dependent terms. We left open the possibility that this is the correct interpretation of the identity predicate *everywhere*, at least outside of mathe-

matics. That this is so indeed, is still a conjecture, the case for which is yet to be made in full. But if the conjecture is correct – as we hope to show in future work – then Gärdenfors's account could be said to do as well as Tversky's also from a reductionist perspective.

We conclude that, at a minimum, the current orthodoxy concerning similarity has recently come to have a serious contender, and that reductionist considerations may not offer any help in deciding between the two.

REFERENCES

Carnap, R. [1928/1967]. *The Logical Structure of the World*, Berkeley: University of California Press.
Coombs, C.H. [1954]. "Method for the Study of Interstimulus Similarity", *Psychometrika*, 19, pp.183–194.
Decock, L. [2006]. "A Physicalist Interpretation of 'Phenomenal' Spaces", *Phenomenology and the Cognitive Sciences*, 5, pp.197–225.
Douven, I. & Decock, L. [2009]. "Identity and Similarity", *Philosophical Studies*, in press.
Gärdenfors, P. [2000]. *Conceptual Spaces*, Cambridge MA: Bradford.
Goldstone, R.L. & Son J.Y. [2005]. "Similarity", in: K.J. Holyoak & R.G. Morrison (eds.), *The Cambridge Handbook of Thinking and Reasoning*, Cambridge: Cambridge University Press, pp.13–36.
Goodman, N. [1972]. "Seven Strictures on Similarity", in: *Problem and Projects*, Indianapolis/New York: Bobbs-Merrill, pp.437–446.
Hahn, U. & Chater, N. [1997]. "Concepts and Similarity", in: K. Lambert & D. Shanks (eds.), *Knowledge, Concepts, and Categories*, East Sussex: Psychology Press, pp.43–92.
Hahn, U. & Ramscar, M. [2001]. *Similarity and Categorization*, Oxford: Oxford University Press.
Lewis, D. [1986]. *On the Plurality of Worlds*, Oxford: Blackwell.
Medin, D.L., Goldstone, R.L., & Gentner, D. [1993]. "Respects for Similarity", *Psychological Review*, 100, pp.254–278.
Nosofsky, R.M. [1988]. "Similarity, Frequency, and Category Representations", *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, pp.54–65.
Nosofsky, R.M. [1992]. "Similarity Scaling and Cognitive Process Models", *Annual Review of Psychology*, 43, pp.25–53.
Shepard, R.N. [1958]. "Stimulus and Response Generalization: Tests of a Model Relating Generalization to Distance in Psychological Space", *Journal of Experimental Psychology*, 6, pp.509–523.
Shepard, R.N. [1987]. "Toward a Universal Law of Generalization for Psychological Science", *Science*, 237, pp.1317–1323.

Shepard, R.N. [2001]. "Perceptual-cognitive Universals as Reflections of the World", *Behavioral and Brain Sciences*, 24, pp.581–601.

Torgerson, W.S. [1958]. *Theory and Methods of Scaling*, New York: Wiley.

Torgerson, W.S. [1965]. "Multidimensional Scaling of Similarity", *Psychometrika*, 30, pp.379–393.

Tversky, A. [1977]. "Features of Similarity", *Psychological Review*, 84, pp.327–354.

Tversky, A. & Gati, I. [1982]. "Similarity, Separability, and the Triangle Inequality", *Psychological Review*, 89, pp.123–154.