Representational Distortion as a Theory of Similarity

Nick Chater and **Ulrike Hahn**
Dept. of Experimental Psychology
University of Oxford, England
{nick,ulrike}@psy.ox.ac.uk

A central drawback of extant psychological theories of similarity is that they are typically defined over very limited classes of representations such as points in a multidimensional space or feature-sets. Our new account, *representational distortion*, can deal with arbitrary representations. Here, similarity between two representations is defined by the amount of distortion required to transform one representation into another.

This theory is based on the *information distance* (Li & Vitanyi, 1993) between representations. Intuitively, the information distance between two representations measures how many instructions must be followed to transform one into the other. The information distance, e.g. between the sequences 1 2 3 4 5 and 2 3 4 5 6 is small, because the simple instruction "add 1 to each digit" transforms one into the other. The information distance between 8 5 7 4 9 and 8 7 9 0 5 is larger, because there is no simple instruction here. This approach is general, because the length of the shortest set of instructions required to transform one represention to another is well-defined for all representations, whether these are feature vectors, a structural description of perceptual input, a parsed sentence, a schema of general knowledge, a pictorial represention or a sequence of motor commands.

Representational distortion can be quantified by applying ideas from a branch of computer science and mathematics known as *Kolmogorov complexity* (Li & Vitanyi, 1993).[1] The fundamental idea of Kolmogorov complexity theory is that the complexity of any mathematical object, x, can be measured by the length of the shortest computer program that is able to generate that object. This length is the Kolmogorov complexity, $K(x)$ of x, a notion which has fruitfully been applied in statistics, inductive inference and machine learning. To quantify representational distortion we use the conditional Kolmogorov complexity, $K(x|y)$, of one object, x, given another object, y. This is the length of the shortest program which produces x as output from y as input—the shorter the program, the simpler the relation between the two objects, and therefore the less representational distortion is required.

This gives a simple account of similarity, with a number of interesting properties (Hahn and Chater, 1996). **Generality**: it applies to representations of all kinds— spatial, feature-based or, crucially, structured representations. **Flexibility**: Similarity is defined over representations of objects, and the goals and knowledge of the subject may affect the representations which are formed, hence, allowing for the great flexibility of human similarity judgements. **Self-similarity**: it is maximal, as no program at all is required to transform an object into itself. **Asymmetry**: Representational distortion allows for asymmetry in similarity judgements: $K(x|y)$ is not in general equal to $K(y|x)$. This asymmetry is particularly apparent when the representations being transformed differ substantially in complexity, as is observed experimentally (Tversky, 1977). Symmetrical judgements are captured by the average of the distances in either direction: $D(x, y) = 1/2(K(x|y) + K(y|x))$. **Background knowledge**: this may radically affect the program length required to transform two objects. 1 5 3 7 2 3 9 0 6 and 3 0 7 4 4 7 8 1, e.g., are very simply related (and, hence, their similarity high) –*if* it is recognized that, as base 10 numbers, the second is double the first. **Shepard's Universal Law of Generalization**: the account provides a derivation of the Universal Law of Generalization (Chater and Hahn, 1996).

[1] This was previously applied in a psychological context as a framework for perceptual organization (Chater, in press).

References

Chater, N. (in press) Reconciling simplicity and likelihood principles in perceptual organization. Psychological Review.

Chater, N. and Hahn, U. (1996). Representational Distortion and the Universal Law of Generalization. *Unpublished manuscript*.

Hahn, U. and Chater, N. (1996). Concepts and Similarity. *In:* Lamberts, K. & Shanks, D. R. (Eds.), *Knowledge, Concepts, and Categories*. London: UCL Press.

Li, M. & Vitanyi, P. (1993). An introduction to Kolmogorov complexity and its applications. New York: Springer-Verlag.

Shepard, R. N. (1987) Toward a universal law of generalization for psychological science. Science, 237, 1317-1323.

Tversky, A. (1977). Features of Similarity. *Psychological Review*, 84, 327—352.