BBS COMMENTARY	
ABSTRACT	65 words
MAIN TEXT	813 words
REFERENCES	241 words
ENTIRE TEXT	1119 words

Theories or fragments?

Nick Chater[a1] and Mike Oaksford[a2]

[a1] Behavioural Science Group, Warwick Business School, University of Warwick
Coventry, CV4 7AL, UK
+44 (0)24 7652 4506
Nick.Chater@wbs.ac.uk
http://www.wbs.ac.uk/about/person/nick-chater/

[a2] Room 531, Department of Psychological Sciences,
Birkbeck, University of London, Malet Street, London WC1E 7HX
+44 (0)20 7079 0879
m.oaksford@bbk.ac.uk
http://www.bbk.ac.uk/psychology/our-staff/mike-oaksford

ABSTRACT:

Lake et al argue persuasively that modelling human-like intelligence requires flexible, compositional representations in order to embody world knowledge. But human knowledge is too sparse and self-contradictory to be embedded in “intuitive theories.” We argue instead that knowledge is grounded in exemplar-based learning and generalization, combined with high flexible generalization, a viewpoint compatible both with non-parametric Bayesian modelling and sub-symbolic methods such as neural networks.

MAIN TEXT:

Lake et al make a powerful case for the modelling human-like intelligence depends on highly flexible, compositional representations, to embody world knowledge. But will such knowledge really be embedded in "intuitive theories" of physics or psychology? This commentary argues that there is a paradox at the heart of the "intuitive theory" view point---that has be-devilled analytic philosophy and symbolic artificial intelligence: human knowledge is both (i) extremely sparse and (ii) self-contradictory (e.g., Oaksford & Chater 1991).

The sparseness of intuitive knowledge is exemplified in Rozenbilt and Keil's (2002) discussion of the "illusion of explanatory depth." We have the feeling that we understand how a crossbow works, how a fridge stays cold, or how electricity flows around the house. Yet, when pressed, few of us can provide much more than sketchy and incoherent fragments of explanation. Thus, our causal models of the physical world appear shallow. The sparseness of intuitive psychology seems at least as striking: indeed, our explanations of our own and other's behavior often appear to be highly ad hoc (Nisbett & Ross 1980).

Moreover, our physical and psychological intuitions are also self-contradictory. The foundations of physics and rational choice theory has consistently shown how remarkably few axioms (e.g., the laws of thermodynamics; the axioms of decision theory) completely fix a considerable body of theory. Yet our intuitions about heat and work, or probability and utility, are vastly richer and more amorphous—and cannot be captured in any consistent system (e.g., some of our intuitions may imply our axioms; but others will contradict them). Indeed, contradictions can also be evident even in apparent innocuous mathematical or logical assumptions (as illustrated by Russell's paradox, which unexpectedly exposed a contradiction in Frege's attempted logical foundation for mathematics, Irvine & Deutsch 2016).

The sparse and contradictory nature of our intuitions explains why explicit theorizing requires continually ironing out contradictions, making vague concepts precise, and radically distorting or replacing existing concepts. And the lesson of two and half millennia of philosophy is arguable that clarifying even the most basic concepts, such as 'object' or 'the good' can be entirely intractable, a lesson re-learned in symbolic AI. In any case, the raw materials for this endeavor—our disparate intuitions—may not properly be viewed as organized as theories at all.

If this is so, how do we interact so successfully in the physical and social worlds? We have *experience*, whether direct or by observation or instruction, of crossbows, fridges and electricity, to be able to interact with them in familiar ways. Indeed, our ability to make sense of new physical situations often appears to involve creative extrapolation from familiar examples: e.g., assuming that heavy objects will fall faster than light objects, even in a vacuum, or where air resistance can be neglected. Similarly, we have a vast repertoire of experience of human interaction, from which we can generalize to new interactions. Generalization from such experiences, to deal with new cases, can be extremely flexible and abstract (Hofstadter 2001). For example, the perceptual system uses astonishing ingenuity to construct complex percepts (e.g., human faces) from highly impoverished signals (e.g., Hoffman 2000; Rock 1983) or interpret art (Gombrich 1960).

We suspect that the growth and operation of cognition is more closely analogous to case law than it is to scientific theory. Each new case is decided by reference to the facts of that present case, and to ingenious and open-ended links to precedents from past cases; and the history of cases creates an intellectual tradition which is only locally coherent, often ill-defined, but surprisingly effective in dealing with a complex and ever-changing world. In short, knowledge has the form of a loosely inter-linked history of reusable fragments, each building on the last, rather than being organized into anything resembling a scientific theory.

Recent work on construction-based approaches to language exemplify this viewpoint in the context of linguistics (e.g., Goldberg 1995). Rather than seeing language as generated by a theory (a formally specified grammar) and the acquisition of language as the fine-tuning of that theory, such approaches see language as a tradition, where each new language processing episode, like a new legal case, is dealt with by reference to past instances (Christiansen & Chater 2016). In both law and language (see Blackburn 1984), there will be a tendency to impose local coherence across similar instances, but there will typically be no globally coherent theory from which all cases can be generated.

Case-, instance- or exemplar-based theorizing has been widespread in the cognitive sciences (e.g., Kolodner 1993; Logan 1988; Medin & Shaffer 1978). Exploring how creative extensions of past experience can be used to deal with new experience (presumably by processes of analogy and metaphor rather than deductive theorizing from basic principles) provides an exciting challenge for artificial intelligence, whether from a non-parametric Bayesian standpoint or a neural network perspective, and is likely to require drawing on the strengths of both.

REFERENCES:

Blackburn, S. (1984). *Spreading the word: Groundings in the philosophy of language*. Oxford, UK: Oxford University Press.

Christiansen, M. & Chater, N. (2016). *Creating language*. Cambridge, MA: MIT Press.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago, IL: University of Chicago Press.

Gombrich, E. (1960). *Art and Illusion*. New York: Pantheon Books.

Hoffman, D. D. (2000). *Visual intelligence: How we create what we see*. WW Norton & Company.

Hofstadter, D. R. (2001). Epilogue: Analogy as the core of cognition. In D. Gentner, K. J. Holyoak, & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 499–538). Cambridge, MA: MIT Press.

Irvine, A. D. & Deutsch, H. (2016). Russell's Paradox. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), URL = <https://plato.stanford.edu/archives/win2016/entries/russell-paradox/>.

Kolodner, J. (1993). *Case-based reasoning*. San Mateo, CA: Morgan Kaufmann.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*(*4*), 492-527.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(*3*), 207-238.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Oaksford, M., & Chater, N. (1991). Against logicist cognitive science. *Mind and Language*, *6*(*1*), 1-38.

Rock, I. (1983). *The logic of perception*. Cambridge, MA: MIT Press.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, *26*(*5*), 521-562.