Unpacking Intuition

A Conjecture

Martin E.P. Seligman and Michael Kahana

University of Pennsylvania

ABSTRACT—*Can intuition be taught? The way in which faces are recognized, the structure of natural classes, and the architecture of intuition may all be instances of the same process. The conjecture that intuition is a species of recognition memory implies that human intuitive decision making can be enormously enhanced by virtual simulation.*

It has long been realized that many important decisions are not arrived at by linear reasoning, but by intuition (e.g., Djiksterhuis, Bos, Nordgren, & van Baaren, 2006; Gladwell, 2005; Hogarth, 2001; Lieberman, 2000). Intuitive decision making is a) rapid, b) not conscious, c) used for decisions involving multiple dimensions, d) based on vast stores of prior experiences, e) characteristic of experts, f) not easily or accurately articulated afterwards, and g) often made with high confidence (see Hogarth, 2001, for a review). As ubiquitous as intuitive decision making is, its cognitive architecture is essentially a mystery. We conjecture that the processes of facial recognition, of categorization, and of intuitive decision making are one and the same. If this is so, we believe that better intuition is eminently teachable by virtual simulation.

Wittgenstein (1953) wrestled with the issue of defining a concept in ordinary language, an issue we take to be identical with recognizing a member of a natural category. How does one recognize a chair or a game? Traditional epistemology, from Aristotle on, held that natural classes such as tables are similar to the class of mathematical objects (such as circles) in that they have necessary and sufficient conditions that define membership in the class. One recognizes a circle or a table by perceiving the necessary and sufficient conditions. Wittgenstein (1953) demolished this tired but venerable tradition once and for all:

> I am saying that these phenomena have no one thing in common which makes us use the same word for all,-but that they are related to one another in many different ways. And it is because of this

Address correspondence to Martin E.P. Seligman, Department of Psychology, University of Pennsylvania, 3701 Market Street, Philadelphia, PA 19104; e-mail: seligman@psych.upenn.edu.

relationship, or these relationships, that we call them all "language." I will try to explain this

66. Consider for example the proceedings that we call "games." I mean board-games, card-games, ball-games, Olympic games, and so on. What is common to them all? – Don't say: "There must be something common, or they would not be called 'games' "-but look and see whether there is anything common to all. – For if you look at them you will not see something that is common to all, but similarities, relationships, and a whole series of them at that. To repeat: don't think, but look! –

Look for example at board-games, with their multifarious relationships. Board games, what are some? Consider chess, of course, but think also of monopoly. Now pass to card-games; here you find many correspondences with the first group, but many common features drop out, and others appear.

When we pass next to ball-games, much that is common is retained, but much is lost.– Are they all 'amusing'? Compare chess with noughts and crosses. Or is there always winning and losing, or competition between players? Think of patience. In ball games there is winning and losing; but when a child throws his ball at the wall and catches it again, this feature has disappeared. Look at the parts played by skill and luck; and at the difference between skill in chess and skill in tennis.

Think now of games like ring-a-ring-a-roses; here is the element of amusement, but how many other characteristic features have disappeared!

If, unlike circles, there is no one property that all games have in common (no necessary condition) and no properties that distinguish a table from all other objects (no sufficient conditions), how then can one recognize that an activity is a game or an object a table? Wittgenstein's answer to this is "family resemblances."

One recognizes a table as a table in the same way that one can recognize a new face as a Churchill after having seen the faces of dozens of other members of the Churchill family. Wittgenstein's family resemblance is a characteristically seductive metaphor, but it is not an explanation; it explains one mystery merely by substituting another. The question simply transmutes into "By

what process does one recognize a new face as a member of a family?"

Significant progress has been made in the analysis of both natural and artificial category learning since Wittgenstein, and there have been several fruitful attempts to illuminate the process (e.g., Ashby & Maddox, 2005; Keil, 1989; Kruschke, 1992; Lamberts, 2000; Logan, 2002; Love, Medin, & Gureckis, 2004; Medin & Smith, 1984; Murphy & Medin, 1985; Nosofsky, 1992; Rehder, 2003; Rips, 1989). Modern accounts of human classification assign a major role to the concept of similarity (e.g., Verguts, Ameel, & Storms, 2004). These models conceive of the mental representation of an item (e.g., a particular face or a word) as consisting of a set of attribute values where the number of attributes may be very large. Such a conceptualization has a natural geometric interpretation, in which items are represented as points in a multidimensional space, with each dimension corresponding to an attribute. The representations are likely to be noisy (e.g., Ennis, 1988; Kahana & Sekuler, 2002), with the resulting representation being a multivariate probability distribution (or a "cloud") rather than a fixed vector (a point). Similarity is then a decreasing function of the distance between the representation of the items in the attribute space (e.g., Nosofsky, 1986; Shepard, 1987). Modern "machine-learning" techniques developed by computer scientists and statisticians have used very similar principles to solve large-scale classification problems (Hastie, Tibshirani, & Friedman, 2001).

Classification learning can be modeled as learning about which items cluster together, where the clustering can be defined based on their similarity relations or based on a set of rules that carve the space up into regions. Although conceptually distinguishable, these two ideas (rule-based and similarity-based clustering) often give rise to similar quantitative predictions about classification performance (Kahana & Bennett, 1994; Maddox & Ashby, 1993). Some theorists argue that the above conceptualizations are not sufficiently rich to capture the subtleties of natural categories (Rehder & Murphy, 2003). They argue that there may be a causal structure underlying the individual features that cannot be accounted for by simple similarity relations.

To make the ideas concerning exemplar models more concrete, consider the universe of objects that all people agree are tables. Using prior knowledge of the world, one will note a great many features of tables that are potentially relevant (but neither necessary nor sufficient singly or jointly) to being a table (e.g., flatness of the surface, number of legs, capacity for supporting other objects, function, compatibility with chairs). Each of these features can be assigned a binary (present vs. absent) or continuous value. The lack of a necessary condition means that different instances of tables will have different values along several of the dimensions (e.g., some tables, like dining room tables are flat, whereas others, like pool tables, have pockets). This means that the process of categorization is stochastic in nature. Upon observing a new object, one can decide whether it

is a table by comparing its features with the features of stored tables in memory. If the sum of its similarity to all of the tables in memory is higher than the sum of its similarity to other objects (e.g., chairs, animals) then one will predict that it, too, is a table. If the summed similarities of the table exemplars exceed the summed similarity for the nontable exemplars, the decision process will be fast and certain. When the summed similarities do not strongly favor one category, the decision will be slow and uncertain. Under this condition, nonintuitive reasoning processes, such as logical analysis, may come into play. Thus, the recognition of an object as an instance of a category reflects a computational process involving a comparison of the features of a given item with the features of all items stored in memory.

It is an important feature of this theory that such a process for categories (or concepts) also explains family resemblances in just the same way: Consider the faces of several dozen Churchills and a large number of people who are clearly not Churchills. Lacroix, Murre, Postma, and van den Herik (2006) have shown how models based on the dimensional analysis of human faces can account for the detailed data on face recognition. These models assume that each face is characterized by a relatively small number of dimensions, each of which has an assigned value (e.g., Winston Churchill has large, prominent jowls and a smallish pug nose).

Now consider two instances of what is generally considered intuitive decision making: a lieutenant recognizing a likely ambush or a surgeon coping with an unexpected ruptured artery. In both of these cases, professional legend has it that there are some expert "eagles" who intuitively just know what to do and many more "turkeys," who try to go by the book and flounder lethally. Each of these cases, we believe, is amenable to the computational modeling of recognition analysis above.

The lieutenant who intuitively and correctly decides that this patch of forest is a likely ambush site is an easy case. Here, in theory, is what this eagle has stored in her memory. She has a list of the dimensions detailing what constitutes an ambush site versus a nonambush site. She has values along each of these dimensions for each of the ambush and nonambush sites that she has experienced or learned about. She has a mental model that assigns weights to each of these basic dimensions or features (and to higher order features, such as the interaction between two dimensions). On the basis of past experiences with similar sets of features, she determines whether the present features more closely resemble those associated with ambushes or nonambushes.

The same logic applies to "moral intuition" (Haidt, 2001). How do we know that a given action we contemplate is "right" or "wrong?" The action ("do it" or "refrain" for simplicity) can be dimensionalized (e.g., the act is forbidden by the Ten Commandments, the nation will be better off, selfish gain, enhancement of reputation, likelihood of being caught, and victimlessness). For each dimension, the individual has a value in memory for similarity to past "right" and past "wrong" exem-

plars. The individual has a mental model in which each dimension and each higher order interaction has a weight. If the summed similarity of the action to past "right" actions exceeds the summed similarity to past "wrong" actions, the course is recognized as right (right vs. wrong, rather than right vs. not wrong, again for simplicity). To the degree that the summed similarities for right exceed the summed similarity for wrong, the moral decision will be fast and confident. When the summed similarities are close, the decision will be uncertain and "higher" processes, such as moral reasoning, will come into play.

The eagle surgeon goes through a similar computation, but there is an important additional factor: In the case of the ambush and in the case of the moral decision, the issue was to recognize an ambush or to recognize a "right" course. For the surgeon there is an explicit action term. We conjecture that this is also a recognition problem. Here, in theory, is what the eagle surgeon has stored in his memory. He has a list of the dimensions relevant to two (the number could be greater, but we use only two for simplicity) actions he might take given the artery rupture. He has a value for each of the dimensions and each of their interactions for Action A and for Action B. He has a model that assigns weights to each of these values, and the decision criterion for these weights is the possibility of saving a life. He now plugs the present values into the predictive model and determines that Action A is more likely to save a life than will Action B, so he prefers Action A and does it.

Our conjecture strongly implies that intuition is teachable, perhaps massively teachable. There are two old-fashioned ways of teaching intuition: through brute force experience and through verbal explanation. Simple repeated experience with forced choice seems to build intuition, and the well-known chicken-sexing endeavor is an example of such brute force. Professional Japanese chicken-sexers can tell male from female chicks at a glance, and they cannot articulate how they do it. With many forced choice trials with feedback, however, ordinary people can be trained to very high accuracy, and they too are unable to report how they do it (Myers, 2002).

Our conjecture, which argues that the mind can weight main effects and all relevant higher order interactions, implies that intuition can also be taught the old-fashioned didactic way: Consider explaining verbally a triple interaction for recognizing an ambush to platoon leaders. The main effect is danger: a thick line of trees at the top of a rise. In the double interaction, this tree line is not dangerous if it was inspected in the last 24 hr. In the triple interaction, however the area is dangerous once again if there have been enemy soldiers spotted near the hill since the last inspection 24 hr ago.

Our conjecture does more than account for brute force experiential and verbal teaching of intuition. It argues that intuition may be teachable virtually and on a massive scale. Most battle commanders and surgeons must go through quite a bit of bloody experience to develop a mental model that is robust enough to accurately predict an ambush or to determine which of two actions to take in a surgical emergency. Unfortunately, many patients and soldiers will have to die for a commander or a surgeon to have sufficient relevant life experience. There is, however, in our theory, a way around this: virtual simulation of ambush and no-ambush situations in a war or of the results of Action A versus Action B in surgery. A sufficient number of simulations with enough variations to allow a buildup of the mental model will result in a commander or a surgeon who has "seen it before" virtually and will take the life-saving action at zero prior cost in blood when confronted with the situation in real life.

Just as it is a waste of training to simulate obvious decisions, it is crucial to closely model and overtrain "close calls," the scalpel-edge cases that yield the slowest response times and are most prone to error. The computational modeling derives a decision contour, along which close calls occur. Using virtual simulation, one can systematically morph material along the decision contour and thereby overrepresent cases near the boundary (e.g., Lacroix et al., 2006).

Finally we note that a simulator generating many trials of virtual experience is also a selection device. One can select for asymptotic performance or for speed of acquisition in order to pick the commanders, moral agents, and surgeons needed for especially difficult cases. These will be the future eagles.

REFERENCES

Ashby, F., & Maddox, W. (2005). Human category learning. *Annual Review of Psychology, 56*, 149–178.

Dijksterhuis, A., Bos, M.W., Nordgren, L.F., & van Baaren, R.B. (2006). On making the right choice: The deliberation-without-attention effect. *Science, 311*, 1005–1007.

Ennis, D.M. (1988). Confusable and discriminable stimuli: Comment on Nosofsky (1986) and Shepard (1986). *Journal of Experimental Psychology: General, 117*, 408–411.

Gladwell, M. (2005). *Blink*. Boston: Little, Brown.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814–834.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer-Verlag.

Hogarth, R. (2001). *Educating intuition*. Chicago: University of Chicago Press.

Kahana, M.J., & Bennett, P.J. (1994). Classification and perceived similarity of compound gratings that differ in relative spatial phase. *Perception and Psychophysics, 55*, 642–656.

Kahana, M., & Sekuler, R. (2002). Recognizing spatial patterns: A noisy exemplar approach. *Vision Research, 42*, 2177–2192.

Keil, F. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Kruschke, J. (1992). An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22–44.

Lacroix, J., Murre, J.M.J., Postma, E.O., & van den Herik, H.J. (2006). Modeling recognition memory using the similarity structure of natural input. *Cognitive Science*, *30*, 121–145.

Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review*, *107*, 227–260.

Lieberman, M. (2000). Intuition: A social cognitive neuroscience approach. *Psychological Bulletin*, *126*, 109–137.

Logan, G.D. (2002). An instance theory of attention and memory. *Psychological Review*, *109*, 376–400.

Love, B., Medin, D., & Gureckis, T. (2004). Sustain: A network model of category learning. *Psychological Review*, *111*, 309–332.

Maddox, W.T., & Ashby, F.G. (1993). Comparing decision bound and exemplar models of categorization. *Perception and Psychophysics*, *53*, 49–70.

Medin, D., & Smith, E. (1984). Concepts and concept formation. *Annual Reviews in Psychology*, *35*, 113–138.

Murphy, G., & Medin, D. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*, 289–316.

Myers, D. (2002). *Intuition: Its powers and perils*. New Haven, CT: Yale University Press.

Nosofsky, R. (1986). Attention, similarity and the identification–classification relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Nosofsky, R. (1992). Similarity scaling and cognitive process models. *Annual Reviews in Psychology*, *43*, 25–53.

Rehder, B. (2003). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1141–1159.

Rehder, B., & Murphy, G. (2003). A knowledge-resonance (KRES) model of category learning. *Psychonomic Bulletin & Review*, *10*, 759–784.

Rips, L.J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21–59). New York: Cambridge University Press.

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

Verguts, T., Ameel, E., & Storms, G. (2004). Measures of similarity in models of categorization. *Memory & Cognition*, *32*, 379–389.

Wittgenstein, L. (1953). *Philosophical investigations*. Oxford, United Kingdom: Macmillan.