Cognitive Science 27 (2003) 327–329

Letter to the Editor

COGNITIVE SCIENCE

http://www.elsevier.com/locate/cogsci

"Effective systematicity" in, "effective systematicity" out: a reply to Edelman and Intrator (2003)

Edelman and Intrator (2003) claim that their "what + where" representation is effectively systematic, which is meant to imply that the CoF model overcomes the inability of view-based models to account for the human capacity to represent and process spatial relations—a limitation pointed out in Hummel (2000). However, the claim that the CoF is "systematic" is meaningless. "Systematicity is defined as the correlation between the top row of the representation on the left and the bottom row on the right and vice versa" (Fig. 10 caption). That is, according to Edelman and Intrator, a representation is systematic if a given shape produces correlated patterns of activation on different parts of the representation. One problem with this definition is that the question of how the visual system knows such correlations is nothing less than the question of how vision works. Edelman and Intrator's solution to this problem, based on computing the marginals and tracing the fragments back to the what + where units, is inadequate both because, as the authors note, it cannot work with distributed representations (e.g., as generated by novel inputs) and because, by relying on what + where units to represent locations, the model neither represents relations explicitly nor even represents locations independently of what resides there (the same location is represented by different what + where units depending on the value of "what"). Moreover, any pixel-based image (including a retinal image, or the image the model takes as input) satisfies the authors' definition of "effectively systematic", so the CoF itself contributes nothing to the model's effective systematicity. And although the authors have assumed away the problem of systematicity, the question of how we perceive spatial relations remains a real and important one.

The authors' claim that CoF ". . . is effectively systematic. . . [because] it has the ability. . . to recognize objects that are related through a rearrangement of middle-scale parts" (principle C5; p. 8). The ability to recognize such objects bears no relation to the limitations of view-based models detailed in Hummel (2000). My argument was *not* (and *is* not) that a model should be able to recognize such objects (it is trivially true that a view-based model can recognize such objects), or that a model that can ". . . make sense of object $A =$ (circle *above* square) should also be able to make sense of $B =$ (square *above* circle)" (p. 3).

My argument was (and remains) that view-based models are fundamentally inconsistent with the human ability to perceive spatial relations—an ability that manifests itself, among countless other ways, in our ability to understand what a small circle atop a large square has in common with (and how it differs from) a small square atop a large circle (the example

doi:10.1016/S0364-0213(03)00006-5

in Hummel, 2000). Shown a pair of such "objects", people effortlessly appreciate that, in terms of shape, the small circle in the first object corresponds to the large circle in the second; but in terms of relative size and location, the small circle in the first object corresponds to the small square in the second. Nothing in the authors' arguments or simulations suggests that their model is even remotely capable of computing these alternative correspondences. (It is telling that in their version of this example, the circle in one object is of the same size as the circle in the other [as are the squares]—a departure from my example necessitated by the fact that the CoF model cannot respond equivalently to circles [or squares] of different sizes.) More generally, nothing in their arguments or simulations suggests that their model is capable of accounting for *any* of the hallmarks of relational processing (see Hummel, 2000).

The authors' claims about the limitations of the structural description approach (SD) are as inaccurate as their claims about the strengths of their own. The central tenet of SD is that *the visual system represents spatial relations explicitly* (at least in some circumstances; see Hummel, 2001). It is not an "... ontological stance of reifying discrete parts..." (p. 4), nor does it seek to recognize "... a random rearrangement of [animal parts]..." as an animal (p. 4), nor is it committed to "... generic... parts... along with 'crisp' categorical spatial relations..." (p. 5). These claims are attacks on a distorted caricature of the theory. They are wrong in both principle and practice. Even the original Hummel and Biederman (1992) model is based on distributed (i.e., "neural-like") representations of object parts; more recent versions of the theory are still more flexible (Hummel, 2001; Hummel & Stankiewicz, 1998). For example, Hummel and Stankiewicz (1998; which the authors do not cite) describe a structural description based on non-categorical relations among parts with non-rigid properties. SD as a whole is committed only to the claim that spatial relations are an explicit component of the representation of object shape. The entities that enter into those relations can be anything that the visual system represents explicitly, from points, to simple features, to "middle-scale parts", up to complete objects and collections of objects (a point made by Hummel (2000) and others, and ignored by Edelman and Intrator).

Over a decade ago, Biederman and I detailed the limitations of conjunctive (e.g., what + where) coding—and the resulting need for dynamic binding—as a basis for representing relational structures (Hummel & Biederman, 1992). Nothing has changed since then. This is bad news for models that eschew dynamic binding. The good news is that dynamic binding is easy to accomplish in artificial neural networks, and plausible in natural ones. Moreover, the need for dynamic binding provides a natural account of numerous phenomena that, at first blush, do not even seem related to relational processing (including the role of attention and working memory, effects of various kinds of brain damage on perceptual and cognitive tasks, relations between perception and "higher" cognition, among others; see Hummel & Holyoak, 1997, in press). Edelman and Intrator dismiss dynamic binding as an "engineering convenience" that no self-respecting neuroscientist would touch with a 10 ft pole. Apparently, the brain has more respect for engineering than neuroscience.

References

Edelman, S., & Intrator, N. 2003. Towards structural systematicity in distributed, statically bound visual representations. *Cognitive Science*.

Hummel, J. E. (2000). Where view-based theories break down: The role of structure in shape perception and object recognition. In E. Dietrich & A. Markman (Eds.), *Cognitive dynamics: Conceptual change in humans and machines* (pp. 157–185). Mahwah, NJ: Erlbaum.

Hummel, J. E. (2001). Complementary solutions to the binding problem in vision: Implications for shape perception and object recognition. *Visual Cognition*, *8*, 489–517.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, *99*, 480–517.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*, 427–466.

Hummel, J. E., & Holyoak, K. J. (in press). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*.

Hummel, J. E., & Stankiewicz, B. J. (1998). Two roles for attention in shape perception: A structural description model of visual scrutiny. *Visual Cognition*, *5,* 49–79.

John E. Hummel
Department of Psychology
UCLA, 405 Hilgard Ave.
Los Angeles, CA 90095-1563, USA
Tel.: +1-310-206-9262; fax: +1-310-206-5895
E-mail address: jhummel@psych.ucla.edu