ELSEVIER

Neuropsychologia 44 (2006) 1519–1523

NEUROPSYCHOLOGIA

www.elsevier.com/locate/neuropsychologia

You can play 20 questions with nature and win: Categorical versus coordinate spatial relations as a case study

Stephen M. Kosslyn*

832 William James Hall, 33 Kirkland Street, Harvard University, Cambridge, MA 02138, United States

Available online 19 April 2006

Abstract

Alan Newell famously asserted that "You can't play 20 questions with nature and win" (Newell, A. (1973). You can't play 20 questions with nature and win. In W.G. Chase (Ed.), *Visual information processing*. New York: Academic Press.), and specifically focused on the futility of studying binary distinctions. However, the distinction between categorical and coordinate spatial relations representations has turned out to be fruitful. In this brief article, the categorical/coordinate distinction is treated as a case study, as a way to address a more general point, namely how to play 20 questions with nature and win. The key to studying binary distinctions may lie in the ways this one differs from previous ones. First, from the outset this distinction was cast within the context of a theory of a more general processing system; second, it was formulated from the perspective of multiple levels of analysis within a processing system, and thereby bridges characteristics of information processing with characteristics of the brain.
© 2006 Elsevier Ltd. All rights reserved.

1. You can play 20 questions with nature and win

Alan Newell famously asserted that "You can't play 20 questions with nature and win" (Newell, 1973). In the game of 20 questions, one player thinks of an object or situation, and the others attempt to guess it by asking a series of binary questions: is it living? is it an animal? is it domesticated? Each question reduces the search space, and eventually a questioner can pounce on just the right answer. In his classic paper, Newell argued that this game is a bad model for how science should be conducted. Specifically, he decried the tendency of psychologists to formulate and test binary distinctions—such as those between episodic versus semantic memory, serial versus parallel search, and gradual versus all-or-none learning. He argued that more often than not such distinctions are illusory, and after an enormous amount of research ultimately all we know is that nature resists clear-cut binary divisions.

At the time that Newell wrote his article, he could make a strong argument for his case; the outcomes of the great debates that centered around such binary distinctions typically were inconclusive. Thus, in his view, this strategy was not advancing psychological science, knowledge was not accumulating. Newell urged us to abandon efforts to study binary distinctions, and instead suggested another approach. We should consider how to fit the available data together into a single coherent story. And in his view, the best way to do this is to attempt to build computer simulation models that mimic human performance.

I wonder how Newell would have regarded the growing research literature on the distinction between categorical versus coordinate spatial relations representations. On the face of things, we seem to have a plethora of evidence against his general point. The distinction between the two types of spatial relations representations is a binary distinction if there ever was one, but it is leading to solid evidence that two distinct types of representations exist. But even so, I do not believe that Newell was entirely off-base. This particular binary distinction, and the research that supports it, differs in crucial ways from most previous efforts to formulate and investigate binary distinctions in psychology.

In this brief article, I wish to take a step back, and consider why the distinction between categorical and coordinate spatial relations representations has turned out to be fruitful. I treat the categorical/coordinate distinction as a case study, as a way to address a more general point, namely how to play 20 questions with nature and win. I will argue that this binary distinction differs from previous ones in two fundamental ways. First, from the outset this distinction was cast within the context of a theory of a

* Fax: +1 617 496 3122.
E-mail address: smkosslyn@wjh.harvard.edu.

0028-3932/$ – see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2006.01.022

more general processing system; second, it was formulated from the perspective of multiple levels of analysis within a processing system, and thereby bridges characteristics of information processing with characteristics of the brain.

2. Drawing distinctions within processing systems

There was a fundamental problem with most (if not all) of the binary distinctions that Newell railed against (summarized in a table on page 288 of his article). Namely, the distinctions were formulated independently of concerns about how the putative representations or processes would operate within the context of a more general processing system. In contrast, from the start the distinction between categorical versus coordinate spatial relations representations was cast within the context of a more general processing system.

In what follows I will begin with two abstract principles that guided the theorizing effort, and then briefly outline the theory itself.

2.1. Two guiding principles

The approach to theorizing rested in part on the following two principles.

2.1.1. Divide-and-conquer

The first principle is *divide-and-conquer*. According to this principle, complex tasks never are accomplished by a single process, all in one swoop. Rather, most tasks are treated as if they are combinations of simpler sub-tasks, each of which is grappled with by a separate aspect of the overall processing system.

In the case of visual perception, the brain has clearly divided processing of object properties, such as shape and color, from processing of spatial properties, such as location. Gross and Mishkin (1977) argued long ago that this division of labor afforded a neat solution to two problems. On the one hand, we need to be able to recognize objects when their images fall in different parts of the visual field (or on the retina). This is accomplished by ignoring location in the system that processes object properties—the so-called "ventral system", which runs from the occipital lobe to inferior temporal cortex. The ventral system plays a central role in recognizing objects. On the other hand, we need to know where an object is located in order to reach for it or navigate with respect to it (approaching or avoiding it, as appropriate). Such processing is accomplished by using the very information that is discarded by the ventral system. Location is registered by a system that processes spatial properties—the so-called "dorsal system", which runs from the occipital lobe to posterior parietal cortex.

Thus, the two problems (recognizing objects in different locations and being able to specify location) have contradictory requirements—and it is rather elegant that the brain deals with each in a separate system.

2.1.2. Weak modularity

The second principle is *weak modularity*. The brain has numerous specialized systems, which exist either courtesy of evolution, learning, or a combination of the two. But these systems are not "modules" of the sort proposed by Fodor (1983). Fodor's modules are independent, in the sense that the workings of one cannot affect the inner workings of another. However, given the nature of the neuroanatomy of the brain, we are better off conceptualizing processing in terms of neural networks—which may share some cortex and some types of processing. Moreover, we should expect "leakage" between these systems.

2.2. Aspects of a theory of high-level vision

These two principles (as well as others reviewed by Kosslyn & Koenig, 1992), led me to formulate a sketch of a general theory of high-level vision. This theory was rooted in basic findings about neuroanatomy and neurophysiology (Kosslyn, 1994). For present purposes, only four aspects of the theory are relevant:

1. *Visual buffer*: First, visual input during perception is organized in a series of brain areas in the occipital lobe, which I have grouped into a single function structure called the visual buffer. These areas are topographically organized, such that the pattern of activation over the surface of the cortex (roughly) preserves the pattern of activation on the retina. Most of the connections among neurons in these areas are short and inhibitory, which facilitates demarcating the edges of objects (which is important for figure-ground segregation). The output from the visual buffer is a representation of edges and regions of an object (or objects, if a scene is being seen).
2. *Object properties processing system*: Output from the visual buffer flows into the ventral system, where it is compared to stored visual memories. If a match is found, the object (or part of an object) is recognized.
3. *Spatial properties processing system*: Output from the visual buffer also flows into the dorsal system, where location and other spatial properties are computed.
4. *Long-term associative memory*: The outputs from the object properties processing and spatial properties processing systems converge on long-term associative memories. Such memories specify the spatial relations among objects or parts of objects. For example, a "landmark" consists of an object in a specific location relative to other objects. To identify a landmark, information from both processing streams must be used.

The theory also specifies subsystems used in to carry out top-down search, should first-pass bottom-up processing not be sufficient. Although such mechanisms are crucial when objects are viewed in impoverished conditions (e.g. partially occluded, in poor lighting), they are not central to the present discussion and so I will not address them further.

2.3. A problem in vision and a possible solution

The distinction between the ventral and dorsal systems makes sense from the perspective of the two principles briefly outlined earlier, divide-and-conquer and weak modularity. We saw earlier that the existence of the ventral and dorsal systems allows the visual system to solve certain fundamental problems, namely recognizing objects in different locations (by discarding location information) while at the same time promoting navigation and reaching (by specifying location information). Let us now extend the logic of this approach to another problem that confronts the visual system. Many objects can be contorted in different ways, and thus project numerous shapes. For example, consider a human form crouching, standing on tip-toes with arms raised high, or balancing on one leg. How can the visual system identify objects when they can project an almost infinite number of images?

One possibility is suggested by essential features of the theory just reviewed. First, note that no new parts are added to the image when the object is contorted in its many and varied ways, although some parts may be occluded (entirely or partially). Thus, if a sufficient number of individual parts can be recognized, this is a strong indication that a specific object is present. Second, and this is the crucial part, the spatial relations between parts remain constant *if they are described in a relatively abstract way*. Consider the spatial relation between the forearm and upper arm when a person assumes different postures. If that relation is described as "connected by a hinge", the same spatial relation applies when the arm is bent a little bit, straight, or bent more than 90°—or when it is at any other angle. Thus, if the forearm and upper arm are described as "connected by a hinge" (with the parts being recognized in the ventral system and the spatial relation produced in the dorsal system), outputs from the ventral and dorsal systems can produce the same descriptions for different contortions of the form—and those descriptions in turn can match the same representation in long-term associative memory.

I called this sort of abstract relation a *categorical spatial relation*, because it specifies categories (Kosslyn, 1987). A category is an equivalence class; for instance, if you hold one hand next to the other, the first will remain left or right of the second no matter how high, low, or far away it is from the other hand. Once assigned to the category, the spatial relations are treated as equivalent, with any differences (e.g. between a bent versus outstretched arm) ignored.

However, the dorsal system cannot compute only categorical spatial relations representations. Such representations are useless for another key role of the dorsal system, namely reaching and navigation. Knowing that a table is "in front of" you (a categorical spatial relation) will not help you walk around it, or pull your chair up to it. In these cases you need precise metric information, and you need such information relative to your body, a part of your body, or relative to another object that serves as an "origin". I have called this sort of representation a *coordinate spatial relation*, given that it requires a metric coordinate system.

A central part of the theory is the idea that categorical spatial relations representations typically can be captured by a word or two, and the left cerebral hemisphere is better than the right at such processing. In contrast, coordinate spatial relations representations are essential for navigation, and the right cerebral hemisphere is better than the left at such processing. Thus, one of the first predictions tested hinged on assumptions of how the two sorts of processing may be implemented in the brain. And in fact, many studies have shown that the left hemisphere is relatively better at encoding categorical spatial relations, and the right is relatively better at encoding coordinate spatial relations (Laeng, Chabris, & Kosslyn, 2003).

In short, here is an example of a situation where 20 questions seems to be working. At the first cut, we divided the entire system into two coarsely defined subsystems, distinguishing between the object-properties-processing ventral system and the spatial-properties-processing dorsal system. At the second cut, we focused on the dorsal system, and now divided it into two more finely characterized subsystems, which compute categorical versus coordinate spatial relations representations.

So far, so good. This approach does seem to allow us to play 20 questions with nature and make progress. But why? In the following section, I consider fundamental assumptions that underlie this approach.

3. Bridging levels of analysis

So far, Newell might not take issue with the thrust of my observations; after all, he strongly argued that researchers should formulate theories of general processing systems. But now we definitely part ways. A second major factor contributed to the failure of the earlier distinctions, and to the success (so far!) of the categorical/coordinate one. This factor rests on the type of theory—namely one that spans multiple levels of analysis. Newell had no particular use for the brain in this theorizing, but as was just illustrated, I have found neuroscientific data to play crucial roles—both in formulating and in testing theories of cognitive processing.

3.1. Levels of analysis

The approach I advocate is based largely on that of Marr (1982), but adapted in various ways to be more appropriate for cognitive processing rather than vision per se. Following Marr, a fundamental characteristic of a theory of a processing system is that it begins with an analysis of the task to be accomplished. In the case of the categorical/coordinate distinction, the hypothesis that separate processing systems are used in the two cases rests on an analysis of the purposes of each sort of representation. Marr (1982) stressed the importance of developing a *theory of the computation*. In Marr's conceptualization, the theory of the computation is a theory of *what* a process computes. Such a theory is an argument for the existence of a specific process. However, and this is a key point, the argument for the existence of a specific process was cast within the context of a system as a whole. Marr stressed that separate processes work together to accomplish any complex task.

The theory of the computation can be conceptualized as specifying a black box, which takes a specific input and produces a specific output; this output in turn is used as input to yet other processes. But to understand the nature of the transformation from input to output, we need to consider what goes on inside this box. The theory must also specify the nature of the internal representations and the processes used to interpret and transform those representations.

Marr treated this aspect of the enterprise as another level of analysis, which required a different sort of theory than that cast at the most abstract level. According to Marr, whereas a theory of the computation describes *what* is computed, *a theory of the algorithm* specifies *how* it is computed. An algorithm consists of a step-by-step procedure that guarantees that a certain output will be produced on the basis of a certain input.

Finally, algorithms are implemented in hardware (on a computer) or "wetware" (in a brain). The level of the *implementation* specifies how an algorithm is physically realized. For example, the fact that the posterior parietal lobes and dorsolateral prefrontal lobes are richly interconnected with large fiber tracks implies that the two areas work together (in at least some circumstances). This observation seems particularly relevant to the encoding and use of spatial relations representations (e.g. Baciu et al., 1999; Kosslyn et al., 1998).

3.2. Interdependence among levels

Marr sometimes wrote as if a theory at one level of analysis could be formulated with only weak links to theories at the other levels. However, computations rely on algorithms, and those algorithms have to operate in a brain that does some things well and other things not so well (e.g. as fast as it is, the speed of brain processes pales when compared to computer operations). In addition, as evolution progressed, older parts of the brain often were relatively preserved—new areas were added, but the old ones rarely were redesigned from scratch. If you want to rebuild a boat at sea, there's only so much you can do about changing the hull—you have got to keep afloat all the while you modify the vessel. Thus, the newer portions had to work with the older ones, which may not have been optimal for the final product (cf. Allman, 1999).

In short, characteristics at each of the levels of analysis affect theorizing at the other levels—and hence a powerful approach to theorizing about cognition requires that all three levels of analysis be considered at the same time. It is important not to lose sight of the fact that these are levels of analysis of the same thing—the brain. Even the most abstract level is still a description of brain function. Hence, it should not be surprising that the different levels of analysis must be viewed as interdependent.

Nevertheless, Marr's approach underscores the importance of theorizing about what processing is for, of specifying the relation between abstract characterizations of processes and the representations and processes that carry them out, and of characterizing the way representations and processes are implemented in the brain—and, more generally, the importance of discovering the relations among events at different levels of analysis.

And here is where I think the approach to studying categorical versus coordinate spatial relations most strongly parts ways from the approaches adopted for studying the dichotomies in Newell's "do-not-do" list. The current approach depends on theorizing at all three levels of analysis, and on coordinating those aspects of the theory. At the level of what is computed, the analysis of the purposes of the two kinds of representations was key; the goal of producing the same representation for objects in different contortions versus providing input to guide motor systems was at the heart of the distinction between the two kinds of representation. At the level of the algorithm, conceptualizing processing within the context of the larger system played a central role; the fact that object properties and spatial properties are processed separately provided a key constraint on the theory of what is computed and how such computation proceeds. And, finally, the idea that the two cerebral hemispheres would differ for the two kinds of processing not only helps to specify the nature of the representations and processes, but also offers one way to test the hypothesis.

4. Leveraging multi-level theories

Why is it important that scientists be able to play 20 questions with nature and win? One reason is simple: cognitive processing is extraordinarily complex, and we must find ways to gain traction in studying it. Think of a really large apple, which is too big to get your mouth around; to eat it, you need to cut it up into small, bite-size morsels. If we cannot ask relatively simple questions, such as those posed by binary oppositions, we may never be able to get a grip on the overwhelming complexity of the human mind.

In this context, I argue that multi-level theories, which bridge from information processing to the brain, should play a special role in playing the science game of 20 questions. Such theories have three properties that I find especially inviting.

First, they lead researchers to collect different sorts of data. For example, the categorical/coordinate distinction has been studied using connectionist models (e.g. Jacobs & Kosslyn, 1994; Rueckl, Cave & Kosslyn, 1989), neuroimaging (e.g. Baciu et al., 1999; Kosslyn et al., 1998), studies of brain-damaged patients (Laeng, 1994), studies of patients who have had one or the other cerebral hemisphere temporarily anaesthetized (Slotnick, Moo, Tesoro, & Hart, 2001), and divided-visual-field behavioral methods (e.g. Hellige & Michimata, 1989; Kosslyn et al., 1989 for review, see Laeng et al., 2003). Although any one method has problems, different methods have different problems—and thus converging evidence from different methods has particular power. To the extent that very different sorts of data converge to document a distinction, that distinction garners strong support.

Second, when theorizing on the basis of such varied types of data, there are more constraints on the theory. Moreover, multi-level theories must respect qualitatively different sorts of constraints simultaneously. Perhaps paradoxically, the more constraints that are available the easier it is to theorize, even though it is more difficult to fit all the constraints together within a common framework. The added difficulty is a virtue, which leads the theorist to reject what could have been possible contenders. Moreover, it is easier to formulate studies to distinguish among relatively few theories than a very large number.

However, perhaps the most important aspect of this approach is that it invites accumulation of research findings and theoretical distinctions. I expect future research on this topic to become more deeply integrated into general theories of object recognition and reasoning (e.g. Laeng, Shah & Kosslyn, 1999), and to make deeper contact with theories of language. In fact, I would not be surprised if the distinction between categorical and coordinate spatial relations provides insight into how linguistic categories bridge to perceptual representations.

Newell was troubled not simply by the failure of most binary distinctions to lead to fruitful research, but also by the lack of accumulation of such results. He had the sense that research was not accumulating to paint a coherent overall picture, but instead isolated fragments of knowledge were being collected. The multi-level approach avoids this problem in part because it is rooted in the brain. To the extent that we can begin to illuminate what the brain is doing during a specific type of task, we are likely to be able to integrate those results with what it is doing in other types of tasks. The brain is, after all, a single organ.

Acknowledgements

Preparation of this article was supported by grants from the National Space Biomedical Research Institute [NCC 9-58-193], National Science Foundation [BCS-0322611, REC-0411725], National Institutes of Health [2 R01 MH060734-05A1], and Defense Advanced Research Projects Agency [FA8750-05-2-0270].

References

Allman, J. M. (1999). *Evolving brains*. New York: W.H. Freeman.

Baciu, M., Koenig, O., Vernier, M. P., Bedoin, N., Rubin, C., & Segebarth, C. (1999). Categorical and coordinate spatial relations: fMRI evidence for hemispheric specialization. *Neuroreport*, *10*, 1273–1378.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Gross, C. G., & Mishkin, M. (1977). The neural basis of stimulus equivalence across retinal translation. In S. Harnad, R. Doty, J. Jaynes, L. Goldstein, & G. Krauthamer (Eds.), *Lateralization in the nervous system* (pp. 109–122). New York: Academic Press.

Hellige, J. B., & Michimata, C. (1989). Categorization versus distance: Hemispheric differences for processing spatial information. *Memory and Cognition*, *17*, 770–776.

Jacobs, R. A., & Kosslyn, S. M. (1994). Encoding shape and spatial relations: The role of receptive field size in coordinating complementary representations. *Cognitive Science*, *18*, 361–386.

Kosslyn, S. M. (1987). Seeing and imagining in the cerebral hemispheres: A computational approach. *Psychological Review*, *94*, 148–175.

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate.* Cambridge, MA: MIT Press.

Kosslyn, S. M., & Koenig, O. (1992). *Wet mind: The new cognitive neuroscience*. New York: Free Press.

Kosslyn, S. M., Koenig, O., Barrett, A., Cave, C. B., Tang, J., & Gabrieli, J. D. E. (1989). Evidence for two types of spatial representations: Hemispheric specialization for categorical and coordinate relations. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 723–735.

Kosslyn, S. M., Thompson, W. L., Gitelman, D. R., & Alpert, N. M. (1998). Neural systems that encode categorical vs. coordinate spatial relations: PET investigations. *Psychobiology*, *26*, 333–347.

Laeng, B. (1994). Lateralization of categorical and coordinate spatial functions: A study of unilateral stroke patients. *Journal of Cognitive Neuroscience*, *6*, 189–203.

Laeng, B., Chabris, C. F., & Kosslyn, S. M. (2003). Asymmetries in encoding spatial relations. In R. Davidson & K. Hugdahl (Eds.), *The asymmetrical brain*. Cambridge: The MIT Press.

Laeng, B., Shah, J., & Kosslyn, S. M. (1999). Identifying objects in conventional and contorted poses: contributions of hemisphere-specific mechanisms. *Cognition*, *70*, 53–85.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: W.H. Freeman.

Newell, A. (1973). You can't play 20 questions with nature and win. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press.

Rueckl, J. G., Cave, K. R., & Kosslyn, S. M. (1989). Why are "what" and "where" processed by separate cortical visual systems? A computational investigation. *Journal of Cognitive Neuroscience*, *1*, 171–186.

Slotnick, S. D., Moo, L. R., Tesoro, M. A., & Hart, J. (2001). Hemispheric asymmetry in categorical versus coordinate visuospatial processing revealed by temporary cortical deactivation. *Journal of Cognitive Neuroscience*, *13*, 1088–1096.