International Journal of Comparative Psychology, 2011, 24, 314-348.

Two Approaches to the Distinction between Cognition and 'Mere Association'

Cameron Buckner
Indiana University, Bloomington, U.S.A.

The standard methodology of comparative psychology has long relied upon a distinction between cognition and 'mere association'; cognitive explanations of nonhuman animals behaviors are only regarded as legitimate if associative explanations for these behaviors have been painstakingly ruled out. Over the last ten years, however, a crisis has broken out over the distinction, with researchers increasingly unsure how to apply it in practice. In particular, a recent generation of psychological models appear to satisfy existing criteria for both cognition and association. Salvaging the standard methodology of comparative psychology will thus require significant conceptual redeployment.

In this article, I trace the historical development of the distinction in comparative psychology, distinguishing two styles of approach. The first style tries to make out the distinction in terms of the properties of psychological models, for example by focusing on criteria like the presence of rules & propositions vs. links & nodes. The second style of approach attempts to operationalize the distinction by use of specific experimental tests for cognition performed on actual animals. I argue that neither style of criteria is self-sufficient, and both must cooperate in an iterative empirical investigation into the nature of animal minds if the distinction is to be reformed.

It is now commonplace in comparative psychology to find nonhuman animal behaviors explained in terms of "cognitive" capacities such as episodic memories, concepts, transitive orderings, and cognitive maps. When a comparative psychologist attempts to justify a cognitive explanation of animal behavior, however, it is widely-accepted that she must do more than merely show that the behavior is consistent with the exercise of a cognitive capacity. Rather, both proponents and skeptics of cognitive explanations agree that they can only be regarded as legitimate if "simpler" associative explanations of the behavior have been carefully considered.[1] As such, rigorous experimental controls ruling out "merely associative" explanations are routine and required component of any experiment purporting to show evidence of animal cognition. I will call this approach to comparative cognition research "Standard Practice."

Standard Practice has recently been challenged by some of its most prominent practitioners, who now argue that the distinction between cognition and mere association on which it relies is "oversimplified," "antiquated," and even

---

[1] Really, Standard Practice involves default concern for at least a "trichotomy," with even "simpler" processes like innate-releasing mechanisms placing even "lower" on the scale than mere association. However, for the sake of simplicity in this paper I will speak as others have of a dichotomy, as the distinction between such "innate" behaviors and learned ones merits its own extensive discussion.

Thanks go to Stanley Weiss and Aaron Blaisdell for organizing the focus session of the Winter Conference on Animal Learning and Behavior where an earlier version of this paper was first delivered. The paper also benefitted from feedback received with conference participants; special thanks are due to Andy Baker for extensive comments on an earlier version of this draft. This research was funded in part by the Center for the Integrative Study of Animal Behavior at Indiana University. Author is now a Visiting Assistant Professor, Department of Philosophy, University of Houston in Houston, TX, and Postdoctoral Researcher at Ruhr-Universität Bochum in Bochum, Germany. Correspondence should be addressed to Cameron Buckner, University of Houston, Department of Philosophy, 513 Agnes Arnold Hall, Houston, TX, 77204-3004, U.S.A. (cjbuckner@uh.edu).

"specious and unproductive" (Allen, 2006; Papineau & Heyes, 2006; Penn & Povinelli, 2007). Rather than presenting a systematic defense of Standard Practice against these critics — a task I have attempted elsewhere (Buckner, 2011) — I will here attempt to put this practice in historical context and provide a partial etiology of the current crisis. In particular, I suggest that amongst comparative psychologists there are actually two broad classes of approach to the distinction, and that improper coordination between these two perspectives has led to unnecessary disagreements and talking-past. On the one hand, there is a "model-based" approach, which tries to make out the distinction in terms of two different kinds of psychological model — for example, models constructed from rules and propositions vs. those built from links and nodes. On the other hand, there is a more behavior-based approach, which relies on tests of behavioral flexibility to operationalize the distinction in experimental practice — for example, the "short-cutting" test for cognitive mapping (O'Keefe & Nadel, 1978) or the 5-element series test for transitive inference (Bryant & Trabasso, 1971).

In this paper, I will discuss the role each of these perspectives has played in the history of the distinction, along the way drawing some general morals about the proper relationship between the two. While both perspectives must play an important role in the reform of Standard Practice, to do so each must be carefully distinguished and put in its proper place.

Standard Practice: From the Ancient to the Current Crisis

As the empirical and philosophical contributions to the recent volume *Rational Animals?* (Hurley & Nudds, 2006) demonstrate, a cloud of confusion now obscures the distinction between "cognition" and "mere association," with researchers from different backgrounds bringing to the table very different assumptions woven together from diverse historical threads. Over the centuries, associationism has been variously conflated with empiricism, behaviorism, and connectionism; and cognitivism has been conflated with rationalism, vitalism, introspectionism, the computational theory of mind, and, when applied to nonhuman animals, anthropomorphism. Any justification of Standard Practice will thus require some redeployment, as no account of the distinction could — or should — answer to all of these disparate uses.

The distinction we have today is the product of millennia of debates about the nature of animal psychological capacities and their similarity to those found in humans. These debates occurred in very different contexts, and as a result, our modern concept of cognition straddles many disparate elements: the ability to use logical rules, employ concepts or propositions, provide verbal justifications for actions, infer causes and effects, select the optimal course of action, and many others. Sorabji (1993) opines that this collection can seem "higgledy-piggledy" in the absence of the historical milieu from which it emerged.

Current debates regarding the nature of animal cognition often center on the question of whether animal behavior is best explained as a product of mere perceptual association or some form of "cognitive" reasoning. Sorabji (1993) argues persuasively that the basic shape of this dialectic coalesced long ago, when Aristotle denied that animals could possess reason (*logos*) and intellect (*nous*).

Aristotelian psychology focused on a hierarchy of three faculties — nutrition, perception (*aisthesis*), and intellect, with animals possessing only nutrition and perception. According to Sorabji, Aristotle's denial of intellect to animals[2] made it difficult for the Ancients to explain the intelligent flexibility apparent in animal behavior — known to the ancients primarily through a series of famous anecdotes, such as the tale of Chryssipus' dog sniffing two roads at a three-way fork and then straightaway taking the third. As a result, the content and inferential capabilities of perception had to be correspondingly expanded.[3] In Aristotle we find the roots of the modern conceptual challenges, for we see perhaps the first concerted attempt to make out a cluster of "middle ground" mental capacities which are powerful enough to explain the flexibility of animal behavior without thereby granting animals full rationality or reason.

Later commentators were greatly puzzled by the details of Aristotle's distinction between perception and intellect. On the one hand, Aristotle granted animals sophisticated perceptual content which could be propositional in nature (meaning, roughly, that their perceptual states can have the content that something is the case and that these contents could be assessed in terms of truth or falsity). In *Nicomachean Ethics*, for example, a lion is described as perceiving that the oxen is near and rejoicing that he will get a meal (Aristotle, trans. 1925, 3.10 1118a20-23); in *De Anima*, animals are credited as able to perceive that certain stimuli are pleasant or unpleasant, as well as to make judgments based on assessment of the "common" properties such as likeness and difference (Aristotle, trans. 1931, 3.2 426b12-427a14). On the other hand, he insisted that animals, lacking *logos*, could not possess beliefs or reasoning (Aristotle, trans. 1931, 3.3, 428a19-24, 3.10, 433a12). Yet since in Aristotelian thought memory is derived from perception and experience (*empeiria*) from memory (Aristotle, trans. 1928, 100a3-6), animals must also possess sophisticated capacities to perceive sensory impressions over time, compare those impressions to one another, remember and learn from those experiences, and act accordingly. Notably, these faculties may provide animals with all of the raw materials they need to form concepts (as specified in Aristotle's famous "battle metaphor," a *locus classicus* of empiricist concept acquisition).[4]

---

[2] For evidence of the denial, see Aristotle's *De Anima* (trans. 1931): 1.2, 404b4-6; 2.3, 414b18-19 with 23-3 and 415a7-8; 3.3, 428a19-24; 3.10, 433a12; 3.11, 434a5-11.

[3] For examples of the expanded content of perception, Aristotle allows sensory perception to report the "common sensibles" such as "movement, rest, number, figure, and magnitude" in *De Anima* 2.6; at 3.2 426b12-427a14, he suggests that assessments of sameness and difference are also assigned to pure perception (trans. 1931). For a more complete case, see Sorabji (1993, 17-20).

[4] "But when perception is present, in some animals the sense image (*aisthima*) comes to remain, and in some not. Where it does not at all or does not for certain things, there and for those things there is no awareness outside perception. But there is [such awareness] in those animals which can keep it in their minds after perception. And when many such [sense images] have come [to remain], there comes then to be a certain distinction, so that reason (*logos*) develops out of their remaining in the case of some animals, and in the case of others not. Thus out of perception arises memory, so we say, and out of memory often repeated of the same thing, experience (empeiria). For memories which are many in number are a single case of experience. And from experience, or from the whole universal come to rest in the mind, the one beside the many which is one and the same in all of them […] Thus dispositions are neither present as determinate entities nor developed out of cognitively superior dispositions, but out of perception. As when a rout occurs in a battle, if one man makes a stand, so does another and then another, until it has got back to the start of the rout." *An. Post.* 2.19 99b32-100a6. Quoted with editorial insertions from Sorabji (1993, p. 33).

Having apparently given animals so much, these commentators wondered, how could Aristotle consistently deny them full-fledged thought and reason?

The debate continued between those supporting and criticizing the Aristotelian position; Epicureans and Stoics tended to side with Aristotle, while Neoplatonists (most notably Porphyry) argued that at least some animals could possess genuine reason. These debates touched on a wide variety of criteria, and the conflation between "model-based" and "behavior-based" approaches to the distinction can, I suggest, already be found there. On the more model-based side, we find Aristotelian and Stoic suggestions that animals could not engage in reason due to an inability to represent the propositional content of their mental states linguistically (with the Stoics taking literally the Platonic suggestion that thought is "the soul conversing with itself"). This was significant because it left animals unable to place their reasoning in the form of a syllogism — at the time, the sole device available for formalizing inferences. On the more behavior-based side, proponents of animal reason retorted with anecdotes suggesting that animals can respond successfully to commands, themselves use verbal commands or signs, and exhibit evidence of deliberation and the ability to plan for future needs (based on criteria like hesitating and dithering when making choices and preparing environmental circumstances, such as lairs and refuges, for future needs).[5] The relationship between the two kinds of criteria was, then as now, hotly debated.

In the early modern period, the distinction reappeared in debates between rationalists like Descartes and Leibniz and empiricists like Locke and Hume. In the early modern discussion, however, there was less need for subtle criteria to distinguish reason from association for two reasons. First, subtle empirical criteria applicable to nonhuman animals were not needed because early modern empiricists and rationalists largely agreed that animals lacked any form of reason. Second, positions had hardened in the early modern debate; the issue became less a matter of how to distinguish two capacities which both humans (and perhaps animals) possessed, but rather a question of whether all thought was best conceived of as associative or propositional in nature.[6] This hardening illustrates a long-term pattern of oscillation between "monistic" and "dichotomous" approaches to the mind — with the more extreme swings often being driven by enthusiasm about developments in other areas, such as the syllogisms, Pavlovian nerve theory, or the advent of the digital computer.

Nevertheless, Hume's (1748/1993) associationism is worth mention because his famous "three principles of association" — similarity, contiguity, and cause and effect (the latter based on nothing more than "habit" together with constant conjunction and priority of cause to effect) — offered perhaps the first concerted attempt to codify the mechanisms of associative thought (Section III). Presaging behaviorism in both mechanism and method, Hume (1748/1993) steadfastly defended the position that animals possessed no inferential capacities beyond those provided by his three principles, as well as the optimistic appraisal

---

[5] See Porphyry's fascinating *On Abstinence from Killing Animals* (trans. 2000) for a summary of these ancient debates.

[6] Or, at least, nearly all thought — even Hume granted a limited role to reasoning in humans, to puzzle out pure relations between ideas (as in the fields of geometry, algebra, and arithmetic) or engage in practical means-end reasoning.

that these mechanisms were so powerful that appeal to more advanced forms of learning or reasoning was not required to explain the full intelligent flexibility of animal behavior (Section IX).[7]

Descartes (1637/1985) also denied any intellective faculties to animals, infamously regarding them as mindless automata (Part 5 57-59); this position was, in the history of Standard Practice, a double-edged sword. While Descartes' early accounts of the nervous system must be credited as foundational in the history of neuroscience, leading the way towards mechanistic explanations of animal behaviors, his identification of reason and mentality with an immaterial soul stunted materialist psychology for centuries. This identification drove many of the debates between vitalists/introspectionists and materialists in the 19th century, and the influence of this doctrine still surfaces in the odd place even today.[8]

Reacting to these introspectionists and vitalists, C. Lloyd Morgan sought to place comparative psychology on firmer empirical footing by basing hypotheses about animal abilities less on anecdote, introspection, and anthropomorphism and more on objective empirical observation. As a check against these influences, Morgan (1894) etched the modus operandi of Standard Practice into the foundations of comparative psychology with his famous "Canon": "In no case may we interpret an action as the outcome of the exercise of a higher psychical faculty if it can be interpreted as the outcome of the exercise of one which stands lower on the psychological scale" (pp. 53). Since its publication, nearly every word in the Canon has been the subject of disputed interpretation (Radick, 2000); the most pressing questions surround the default preference for explanations appealing to "lower" processes, with associative processes commonly taken to be lower on Morgan's scale than cognitive ones. The justification for this bias has often been interpreted as a matter of *parsimony* (sometimes taking the Canon to be a special case of Occam's Razor), implying that explanations appealing to association are simpler than those appealing to cognition. However, the relevant metric of simplicity is rarely specified,[9] and Sober (2005) has convincingly argued that the standard, parsimony-based justifications for the Canon do not withstand critical scrutiny.

I suggest that the Canon is just an extension of the standard scientific practice of ruling out alternative explanations of data, specialized for the context of psychology. A default concern for associative mechanisms is justified by the fact that they, unlike cognitive mechanisms, are supposed to be ubiquitous amongst animals. Basic forms of conditioning have been investigated in *Aplysia* and *Melanogaster*, and have even been demonstrated to occur in the spinal cords of rats after the connection to the brain has been severed (Allen, Grau, & Meagher, 2009).[10] The same is not true of cognition, which is presupposed to be a comparatively less common cause of behavior amongst animals which, when

---

[7] Notably, Hume also grants animals instinct; *cf.* footnote 1.

[8] E.g., Allen (2006) detects such a "Cartesian residue" in the apparent opposition between "associative mechanisms" and "intentional processes" set up by Clayton, Emery, and Dickinson (2006).

[9] Zentall (2001) does suggest that explanations positing unobservables are more complex than those that do not, though it is not clear how this form of simplicity should be weighed against others.

[10] One may retort that the effects observed in spinally-transected rats are not due to associative learning. Allen et al. (2009) argue extensively against this response.

present, depends upon specialized neural mechanisms. In short, the comparative ubiquity of associative mechanisms explains the ubiquity of the controls.[11]

We may still wonder, however, why we should not simply consider the presence of two competing explanations a draw. Sober (2005) suggests an interpretation of "higher" and "lower" that will work for present purposes: "One internal mechanism is higher than another if and only if the behavioral capacities entailed by the former properly include the behavioral capacities entailed by the latter" (pp. 236). If cognition is "higher" than mere association in this sense, then if the additional capacities of the cognitive mechanism are not observed when the organism is put in appropriate circumstances, then our failure to observe them is evidence against the cognitive explanation. The elaborate controls of comparative psychology can thus be understood as the attempt to ensure that the animal has been put in a situation where the "higher" (cognitive), but not "lower" (merely associative), capacities should manifest.

While Morgan's psychology was significantly more empirical than what came before, he reserved a place for introspection in theorizing about the mental capacities of nonhuman animals — recommending what he called the "double inductive" method, which combined empirical observation of animal behavior with human introspection when performing similar tasks. Radical behaviorists such as Watson (1928) and Skinner (1953) thus felt that Morgan's reforms had not gone nearly far enough. In his zeal to turn psychology into an objective, empirical science, Watson declared all mentalist concepts off-limits. Watson painted psychology as the study of behavior rather than the mind, rejecting mental states like beliefs and desires as subjectivist fictions. While the basic principles of Watson's radical behaviorism eventually found wide acceptance in the United States, his inheritors fiercely debated how to proceed. Some sought to reform mentalist constructs under a broadly behaviorist methodology; Hull (1943) is well-known for his work in quantifying various aspects of motivation, and "cognitive behaviorists" such as Tolman (1948) argued that some mental constructs such as goals and curiosity could be experimentally operationalized. Skinner, however, furthered Watson's anti-mentalism by explicitly banning reference to any intermediary processing states between stimulus and response. By the late 1950's, Skinner's camp had become the dominant — or at least most vocal — faction in American psychology.

After the broader "cognitive revolution" repudiated Skinner's (1953) strictures, Standard Practice gradually again became mainstream. As cognitivism about nonhuman animal behavior sought legitimacy, however, the behaviorist's mechanisms of instrumental and operant conditioning became the associative "null hypothesis" which must be ruled out for cognitive explanations to be deemed legitimate (Dennett, 1983; Wasserman & Zentall, 2006). In other words, the mechanisms studied by various strains of behaviorism were all conflated as "mere association" (which by now included at least all of the basic principles of classical Pavlovian and operant Skinnerian conditioning) in Standard Practice — despite the fact that many radical behaviorists had actively resisted the label of "associationism," which smacked to them of internal mental relations.

---

[11] A phyletic explanation of these distributions may redeem Morgan's contention that evolution legitimizes the Canon.

Completing the final step to the present crisis, I will eschew the standard bedtime story about the cognitive revolution which is, I suspect, familiar to most readers; but one brief qualification of this familiar fable is worth mentioning. It is commonly overlooked that the more advanced associationists were just as stymied by behaviorist epistemology as the cognitivists (Smith, 2000). While we emerged from the cognitive revolution with a fairly firm understanding of the most basic forms of conditioning, we had almost no precise understanding of the more advanced forms of learning involving internal interactions *between* stimulus elements. Many members of the latest generation of "associative" models do involve sophisticated between-stimulus interactions. The distinction's most prominent critics are explicitly driven by the worry that, using currently popular criteria, these models cannot be neatly classified as either cognitive or associative (e.g., Penn & Povinelli, 2007). The solution to the current crisis, I will suggest below, will require anchoring both our nascent model-based approaches to association and behavior-based approaches to cognition in something more stable.

The Modern Crisis

Many erstwhile proponents of Standard Practice have begun to call for the wholesale rejection of the distinction (Allen, 2006; Papineau & Heyes, 2006; Penn & Povinelli, 2007). The problem is that in the interim since the cognitive revolution, there has been a dramatic increase in the power and diversity of "associative" models of learning, and as a result, there seem to be few behaviors which cannot be modeled using some combination of associative tools. This poses an existential threat to the field of comparative cognition research, a threat which can be summarized by the following deductive argument:

1. Cognitive and associative explanations of behavior are exclusive alternatives.
2. It is highly likely that associative models will eventually explain all behavior.
3. Often implicit: If cognition does not explain behavior, it doesn't exist; there is no other work for it to do.
4. Therefore, it is highly likely that cognition does not exist.

Premise 1 is the contemporary interpretation of Morgan's Canon (Shettleworth, 2010). Premise 2 is taken from the "optimistic appraisal" inherited from Hume and seemingly borne out by the most recent generation of sophisticated "associative" models.[12] Premise 3 is a standard "realist" interpretation of the distinction (often left implicit), and premise 4 follows deductively from the others.

---

[12] The "optimistic" appraisal can be found in many places; for example, Mitchell, De Houwer, and Lovibond (2009) attest that "from experience we have learned that it is difficult to produce a pattern of data that cannot be explained by one or the other variant of these associative models" (pp. 194). Plausibility for this premise, however, can be derived from very general observations, such as that connectionist models trained by backpropagation are Turing-complete, or that the brain itself is a vastly more sophisticated kind of associative network.

Responses to the current crisis can be categorized in terms of how they respond to this argument.

Some associationists and neo-behaviorists may welcome the conclusion, arguing either that there is no role for cognitive theorizing in psychology generally or only in the study of non-human animals (Wynne, 2004). A more moderate version of this position rejects instead premise 3, holding that while cognition strictly speaking does not exist, cognitive theorizing may yet have some heuristic utility (Blaisdell, 2009; Zentall, 2001); I will not argue against these heuristic positions here, supposing merely that a realist account of the distinction is to be preferred if viable. Mitchell, De Houwer, & Lovibond (2009) reject premise 2, courting controversy by arguing that there is little evidence for the "pure link-forming" mechanism they use to characterize mere association and that even the simplest forms of conditioning are better interpreted as propositional in nature.

The most promising response to this troubling argument holds that its conclusion only follows if we commit a problematic equivocation. In particular, the tension is resolved by avoiding an equivocation on the word "associative" between premises 1 and 2. In other words, the sense in which the more sophisticated models are "associative" is not the sense of "associative" which Standard Practice takes to be mutually exclusive with cognition. The most popular form of this response suggests that cognitive capacities are "implemented by" or "emerge from" associative ones of a certain level of complexity. Crucially, these solutions are not available unless we avoid this equivocation, for they cannot be coherently described if the relevant senses of 'cognition' and 'association' are mutually exclusive.

To clarify this solution, we should here distinguish two senses of the words "cognitive" and "associative" (Buckner, 2011). On the one hand, there is the "exclusionary" use of the terms found in Standard Practice (and premise 1), which I will hereafter denote as "cognitive$_{ex}$" and "associative$_{ex.}$" These terms range over psychological processes, and diagnoses of cognition$_{ex}$ and mere association$_{ex}$ are mutually exclusive. A psychological process here is a series of events (which may or may not be causally responsible for some observed behavior) in an organism's mind, presumably a temporally-extended causal sequence taking place in the organism's nervous system. On the other hand, there is another use of these terms ranging instead over models of psychological processes, meaning roughly that those models are "constructed according to cognitive/associative principles"; this model-based sense of the distinction is the only sense in which premise 2 is plausible. In this latter sense, the distinction ranges not over psychological processes, but rather over our representations of those processes. This second use of these terms will hereafter be denoted by "cognitive$_{mod}$" and "associative$_{mod}$", and apply to models based on whether they are built out of things like rules and symbolic propositions on the one hand or nodes representing stimuli and links representing the associations between them on the other.[13] (Note that my goal in this paper is not to provide a new theory of cognition or association with specific

---

[13] Note that the "mere" modifier in this article's title is not accidental; in the vernacular of comparative psychology, it provides fairly reliable indication that the author intends to invoke the exclusionary interpretation of the distinction.

criteria for applying these terms, but rather to discuss general tensions found in the literature across a variety of diverse accounts.)

Separating these two distinctions helps us avoid fallacious inferences which can arise from confusing the natural phenomena studied by science with our representations of those phenomena. Just as a photograph of person has different properties than the subject of that photograph (e.g., being black-and-white, two-dimensional, or out-of-focus), psychological models are human-made artifacts that possess properties beyond those of the psychological processes they depict. For example, models can consist of diagrams, verbal descriptions, or a series of mathematical equations, whereas psychological processes are none of these things. Furthermore, models can omit aspects of the processes they are about for the sake of simplicity or in order to focus attention on particular aspects of those processes. While models can be evaluated in terms of the degree of correspondence between their properties and the properties of their target phenomena, abstractions, simplifications, and even assumptions which are outright false (such as frictionless surfaces in a physical model) may be beneficial depending on the purposes to which the model is put (Parker, 2011). Furthermore, the mapping from models to natural phenomena can be many-to-one, and different representations of the same process may be suitable for different purposes.

These general morals carry important implications for the present context. For example, if, as many have suggested, the same psychological process can be aptly described at one “level of analysis” by an associative$_{\text{mod}}$ model and at another “level of analysis” by a cognitive$_{\text{mod}}$ model, then the verdicts of associative$_{\text{mod}}$ and cognitive$_{\text{mod}}$ are not necessarily mutually exlusive. If, as some have suggested (e.g., Fodor & Pylyshyn, 1988), a cognitive$_{\text{ex}}$ process is one which admits of precise description with a cognitive$_{\text{mod}}$ model, then this entails that associative$_{\text{mod}}$ and cognitive$_{\text{ex}}$ are also not mutually exclusive. This possibility succinctly describes a solution to the “modern crisis”: Many of the most recent associative$_{\text{mod}}$ models actually depict processes which are cognitive$_{\text{ex}}$, rather than associative$_{\text{ex}}$, in nature.

This solution has, in broad strokes, long been familiar in certain areas of cognitive science — in particular, in the debate about cognitive architecture between connectionists and classical cognitive scientists which raged from the ‘80s to mid-‘90s (e.g., Fodor & Pylyshyn, 1988; Smolensky, 1988). However, the pressing pragmatic question still concerns how and where to draw the line between the associative$_{\text{mod}}$ models depicting merely associative$_{\text{ex}}$ processes and those depicting cognitive$_{\text{ex}}$ ones. For this kind of strategy to be of any practical help to comparative psychologists, we must characterize cognition$_{\text{ex}}$ and mere association$_{\text{ex}}$ with enough precision such that the two can be experimentally distinguished. This is a difficult challenge, for there are many model-based and behavior-based candidates to choose from (many of which are mutually-inconsistent), and no clear winner is favored by the current state of psychological data.

The key way that model-based and behavior-based approaches must work together, I suggest, is in helping us choose which of the candidate criteria will be the most conducive to future productive science. I here only suggest that this role is pragmatic or methodological; I remain agnostic on the troubled metaphysical question of the intrinsic nature of psychological states, e.g., as to whether their "essences" are behavioral, functional, or neural. In the following two sections, I will review elements of the contemporary debate which focus on more model-based and behavior-based approaches to the distinction, concluding with some suggestions as to how the two might be usefully integrated in future research.

General Comments about the Two Approaches

The difference between the two approaches that I will discuss can be expressed simply enough: "Model-based" criteria attempt to constrain the class of models that could be used to aptly describe cognitive$_{ex}$ (and/or merely associative$_{ex}$) processes, and "behavior-based" criteria attempt to specify the behavioral capacities required of processes which are cognitive$_{ex}$ (and/or merely associative$_{ex}$). However, a few general caveats must be added to this simple picture before reviewing specific model- and behavior-based approaches to the distinction.

First, unlike the distinction between cognition$_{ex}$ and mere association$_{ex}$ which they are designed to illuminate, the two approaches are not intended to be mutually exclusive. Quantifiable psychological models will entail that particular behaviors will be produced in response to particular stimuli; and particular behavioral tests for cognition$_{ex}$ may only be satisfiable by systems describable using a limited range of models. The degree of "distance" between a model-based and a behavior-based approach to the distinction depends in part on the degree of abstraction present in the model or behavioral criteria ("abstraction" here signifying "lack of detail"). There may be a tight relationship between the specification of the behavioral capacities and the model; one might argue that to specify behavioral criteria with enough precision just is to specify a model. Some models, such as the purely input-output models of the behaviorist, provide only very weak modeling constraints, placing few restrictions on what happens "inside" the system; yet these models also provide specific behavioral constraints as to how the system will respond in a wide variety of situations. On my scheme, these input-output models would thus fall in-between pure model-based and pure behavior-based accounts. These examples illustrate that the distinction I wish to draw between the two styles of approach is a continuum or smear, with the most abstract specifications of families of models at one pole and most abstract behavioral criteria on the other.

Second, though the distinction between cognition and mere association in comparative psychology is often tied to "dual systems" theories from human psychology, this identification should be regarded with suspicion. A variety of such theories draw the boundaries of the two systems differently; here, I shall only focus on a few of the most prominent and oft-cited. An obvious problem is that many dual-system theorists suppose their "cognitive" system (often referred to as "System II") to be uniquely human, and in such cases it is almost guaranteed that the distinction drawn there and that of Standard Practice do not coincide. Rather, despite the fact that these systems are often labeled as "associative" and "cognitive", both systems satisfy standard criteria for cognition$_{ex}$ in Standard Practice (Mitchell et al., 2009; Smolensky, 1988).

Finally, while I have defined ‘association$_{ex}$’ and ‘cognition$_{ex}$’ as properly ranging over *processes* and not *models*, I will at times speak of *model*-based approaches to the distinction between these two ($_{ex}$) terms. This way of speaking should be understood as shorthand for “a way of making out the distinction between cognitive$_{ex}$ and associative$_{ex}$ *processes* based on whether those processes are aptly describable using model(s) picked out using certain model-based criteria.” This must be distinguished from criteria which are merely designed to make out a distinction between associative$_{mod}$ and cognitive$_{mod}$ models, lest we fall back into the problematic equivocation pointed out above.

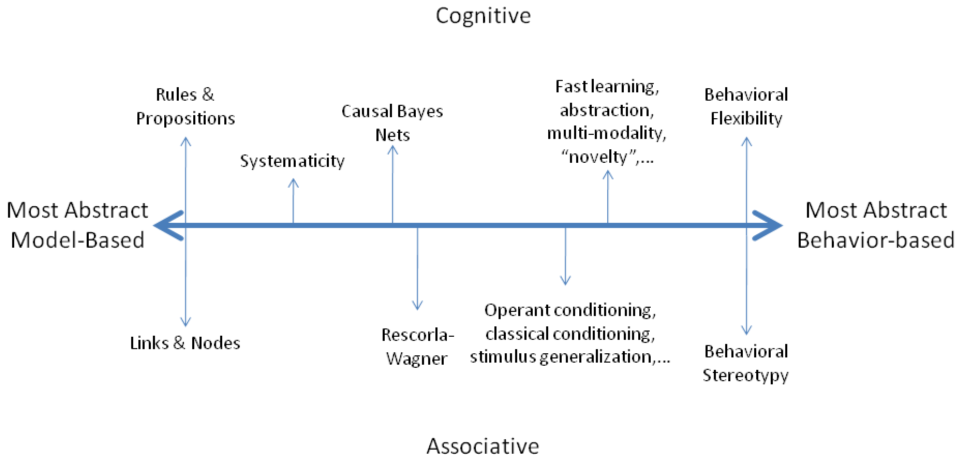

Figure 1. Example of some standard model- and behavior-based criteria arranged on a continuum from most abstract model-based to most abstract behavior-based. Criteria typically considered cognitive are above the line, those typically considered associative are below. More specific models and behavioral criteria (which imply both behavioral and modeling constraints) are located in the middle. Arrangement of criteria on the continuum is meant to be monotonic, but distances from each other and from continuum line are not significant.

Model-Based Approaches

Examples

In almost all standard accounts of the distinction between cognition and mere association, one can find appeals to some model-based criteria. Where comparative psychologists offer accounts of the distinction, they are often parasitic upon “dual systems” accounts in human psychology just mentioned. In these accounts, model-based criteria proliferate, as the accounts emerged from the modeling debate which raged in the ‘80s and ‘90s between the classical computationalists -- typified by Chomsky (1965), Newell and Simon (1976), and Fodor and Pylyshyn (1988) and connectionists -- led by McClelland and Rumelhart (1986), Smolensky (1988), and Elman (1996). Like the empiricists and rationalists in the 17th and 18th centuries, positions in these debates tended towards monistic extremes and thus away from subtle criteria which could help disentangle the “middle ground” capacities so troubling to Standard Practice. Connectionists often claimed (and some still do) that all thought was better understood on the model of connectionist networks, and classical computationalists responded that any mental activity deserving the label “cognition” must instead be modeled on the rule-based perspective of Newell and Simon’s Physical Symbol System hypothesis (1976).

As these two factions fought more or less to a stalemate and productive research continued to emerge from both camps, interest in the early ‘90s began to grow in ecumenical compromises. As a result, “dual system” views on cognition began to proliferate, a perspective which perhaps remains the dominant view in human psychology today. The early roots of the two-systems views can perhaps be found in Smolensky’s (1988) “connectionist proposal” that the brain contains both a conscious rule-interpreter which is properly described at one “level of analysis” as a physical symbol system, and a subsymbolic connectionist system which admits of no accurate classical description. Notably, Smolensky took both systems to be implemented on fundamentally connectionist architecture, so the two systems could not be distinguished using this associative$_{\text{mod}}$ model-based criterion. Rather, his “reflective” system differed from his “intuitive” system in that only the former admitted of precise description at the “symbolic” level.

In later accounts of the two systems views, more abstract model-based criteria proliferate. Sloman’s (1996) influential account of the two systems view consists almost entirely of the abstract model-based variety; the associative and rule-based systems are distinguished along the dimensions of “principles of operation,” “sources of knowledge,” “nature of representations,” “relations,” “nature of processing,” and “illustrative cognitive functions.” The associative system is characterized as operating on similarity and contiguity, as manipulating representations consisting of concrete images, stereotypes, and feature sets, and as relating representations with associations and “soft constraints.” Some more behavior-based features are included in the list as “illustrative cognitive functions;” such functions for System 1 including a capacity to engage in fantasy, creativity, imagination, and associative memory tasks, for System II deliberation, analysis, ascription of function, and strategic memory. Notably, Gigerenzer and Regier (1996) worry that these many criteria cross-cut one another and thus are unsuitable for drawing a coherent dichotomy.

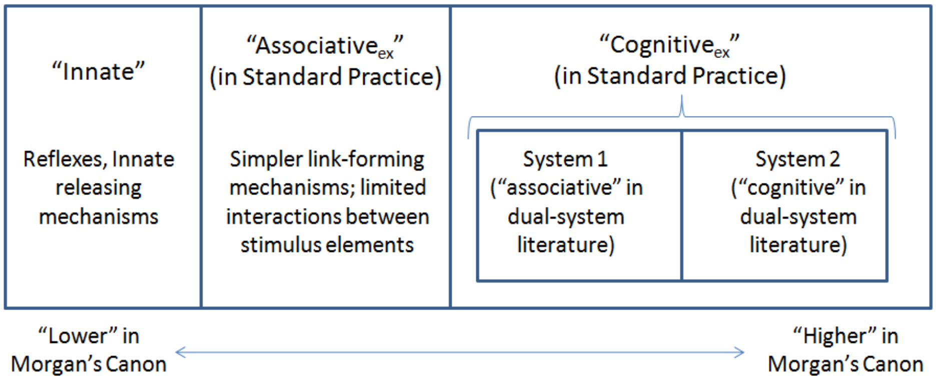

Figure 2. Proposed mapping of terminologies from literatures on Standard Practice, dual-systems theory, and Morgan’s Canon.

While, again, it is unlikely that this distinction corresponds to the one found in Standard Practice, these debates show that there are, broadly, two kinds of model-based approach one could adopt. The first kind of approach focuses on one or two specific models as exemplars of cognition or mere association. This kind of approach might draw the line between cognition and mere association by taking, for example, a classical model of associative learning as such as Estes (1950) model of elemental conditioning or the Rescorla-Wagner model (1972) of cue competition as the paradigm associative model which is supposed to define the limits of associative$_{ex}$ learning. The utility of this kind of approach is limited, however, in that there are a number of different associative models and it is not clear from behavioral evidence alone which (if any) is the right one. Models of associative learning often have non-overlapping strengths and weaknesses; models can differ in their perceived "simplicity," the number of free parameters which must be fit, the breadth or type of phenomena they explain, and their degree of biological plausibility. It is clear from the current literature that nearly all associative theorists agree that it would be premature, given the current state of evidence, to nominate one of these models as "the model" of associative$_{ex}$ learning. What is much less clear, unfortunately, is what future evidence should be gathered that would help us make such a selection.

To avoid these complications, the second style of model-based approach instead defines classes of models by focusing on more abstract properties of those models, such as whether they are constructed from things like nodes representing stimuli and links representing the associative strength between them, or rules and propositions. Discussion of individual models can still be relevant to this approach, but only insofar as they illuminate principles uniting a whole class of models. For example, Wallace and Fountain (2002) consider their Sequential Pairwise Associative Memory (SPAM) model as a paradigm stand-in for mere association$_{ex}$ (i.e., they presume the behavioral capacities of SPAM and mere association$_{ex}$ to be equivalent), but argue that SPAM is equivalent to a broad class of models that share common features (see also Fountain and Doyle, this issue). Popular contemporary candidates for such abstract criteria for cognitive$_{mod}$ models include a tendency to use rules and propositions (Mitchell et al., 2009; Smolensky, 1988) or deploy representations vehicled in a "language of thought" (Fodor & Pylyshyn, 1988); for associative$_{mod}$ models, criteria include being based on "contiguities between cues" (Penn & Povinelli, 2007) and/or similarity (Gigerenzer & Regier, 1996), or resemblance to a simple link-forming mechanism (Mitchell et al., 2009). (As I emphasized in Section 3, we should here exercise extreme caution in reasoning from cognitive$_{mod}$/associative$_{mod}$ to cognitive$_{ex}$/associative$_{ex}$.)

Problems with Purely Model-Based Approaches

Model-based approaches are saddled by a number of challenges which make it very unlikely that a purely model-based approach could provide a solution to the current crisis. Many of the problems issue from a kind of trade-off faced by model-based approaches: they must be both general enough to capture a wide enough class of phenomena to leave cognition$_{ex}$ and mere association$_{ex}$ with enough empirical significance to merit their default roles in Standard Practice, and yet specific enough to issue precise, quantifiable predictions which could be verified through a wide variety of experiments. Let us call these desiderata "Generality" and "Empirical Reference" ("generality" here signifying "applies to more phenomena"). Let us thus summarize these three desiderata as follows:

Desideratum #1: Generality. The criteria nominated to characterize association$_{ex}$ (and cognition$_{ex}$) should cover the full range of associative$_{ex}$ (and cognitive$_{ex}$) processes.

Desideratum #2: Empirical Reference. The criteria nominated to characterize association$_{ex}$ (and cognition$_{ex}$) should issue quantifiable predictions which could allow a hypothesis that a behavior is the result of "association$_{ex}$" (and "cognition$_{ex}$") to be empirically confirmed or falsified.

To the two desiderata in this trade-off, we should also add the constraint that all cognitive$_{ex}$ phenomena should have much in common with one another, and all associative$_{ex}$ phenomena should have much in common with one another. This desideratum is justified by the appraisal that if cognitive$_{ex}$ or associative$_{ex}$ phenomena were instead radically dissimilar (and merely shared a label for, say, historical reasons), then it would not be useful to group them into unitary classes in Standard Practice. It could be, for example, that the processes enabling the different "cognitive" capacities such as episodic memory, cognitive mapping, transitive inference, conceptual abilities, metacognition, mindreading, and so on share nothing in common beyond the fact that behaviorists have difficulty explaining them. In this case, it would then be more useful to concede to the distinction's critics that it cannot answer to the needs of comparative psychology and thus should be rejected.

Desideratum #3: Inductive Unity. The criteria nominated to characterize association$_{ex}$ (and cognition$_{ex}$) should allow a significant number of inductive generalizations to be true of all associative$_{ex}$ processes (and a significant number to be true of all cognitive$_{ex}$ processes).[14]

Problem 1: Picking the Winning Horse

Model-based approaches which focus on particular models as exemplars of cognition$_{ex}$ or mere association$_{ex}$ tend to do a decent job on satisfying the Empirical Reference desideratum but have a comparatively tenuous grip on Generality. As mentioned above, model-based approaches which focus on particular models as exemplars of cognition$_{ex}$ or association$_{ex}$ overreach. We simply do not know enough at present about the distribution of behavioral properties amongst animals to say whether specific models carve nature at its joints. Many models seem clearly too specific because they make predictions only about a limited range of phenomena in particular experimental settings. For example, the input and output nodes of De Lillo, Floreano, and Antinucci's (2001)

---

[14] Note that there is work to do to ensure that these properties are "projectible" and not disjunctive, but I will not draw these specifications here.

connectionist model of transitive inference are specialized for a dyadic choice situation (the only output nodes being “left response” and “right response”), leaving it unclear what the model would say, if anything, about more complex decision environments. Other models with a more general reach, such as Rescorla-Wagner, might be *too* general, crossing the line from mere association$_{ex}$ to claim some territory better described as cognitive$_{ex}$ — a charge pressed against this model by Mitchell et al. (2009), for example.

Problem 2: Runaway Abstraction

Model-based approaches which try to delineate general classes of models might easily include all the properties we attribute to cognition$_{ex}$ and mere association$_{ex}$, but only at the cost of being too abstract to satisfy the Empirical Reference criterion. Very abstract model-based criteria such as “links & nodes” and “rules & representations” fit into this category. As is often observed, many abstract model-based criteria possess too much representational power to effectively draw a distinction. Penn and Povinelli (2007) press this charge against Blaisdell et al.’s (2006) Causal Bayes Net approach to causal cognition, calling the framework a “*lingua franca* that can give posthoc explanations for nearly any nonpathological causal inference” (pp. 105). Smolensky (1988) emphasizes that both subsymbolic computational processes and theorem provers can, in principle, be implemented in connectionist models. At the very least, such abstract criteria must be constrained in further ways if they are to be taken as model-based criteria for cognition$_{ex}$. Obvious additions would include implementation constraints, such as that we should not consider “associative$_{ex}$” anything which a connectionist network could in principle be trained to do if given an infinite number of nodes, an optimal training set, and an infinite amount of time to master it. Rather, only those tasks which biologically-relevant architectures could perform when trained in biologically-relevant learning conditions should be considered relevant. A similar modification to the Causal Bayes Net approach might precisely specify how causal models themselves are acquired (a process which may or may not itself be Bayesian), in order to constrain the class of causal models which could be induced from a given training regime.

Problem 3: Contingent Notational Properties

The greatest danger posed by model-based criteria is that they encourage us to focus on contingent notational properties of models. As I use the phrase, a “contingent notational property” is a property of a model which is irrelevant to the predictions that model makes about the phenomena it purports to be about. In other words, the same model could be expressed without any reference to those properties; those properties could be changed or excised, and the model would make the same behavioral predictions. Without further constraints or elaboration, “rules & propositions” vs. “links & nodes” are, I suggest, just such contingent criteria. The Rescorla-Wagner model is often taken as an exemplary associative model; but it has been expressed using rules and symbols in a PROLOG program (Shanks & Pearson, 1987). In fact, the representational power of both rules &

propositions and of links & nodes ensure that, without additional constraints, almost any model could be expressed using either. When appealing to model-based criteria in an evaluation of the distinction in Standard Practice, we must carefully base our classification not on contingent properties of models but rather on deeper properties of the processes those models describe.

Behavior-Based Approaches

Due to the difficulties in outlining purely model-based approaches to the distinction, one might suppose our attention to be better-directed towards the phenomena of association$_{ex}$ and cognition$_{ex}$ themselves, rather than our (preliminary, partial, diverse) models of them. This approach would suppose that we could identify some rough behavioral criteria which could be used to experimentally distinguish cognition$_{ex}$ and mere association$_{ex}$, and then gradually home in on the specifics and sharpen the boundaries as our level of knowledge improved. While modeling could be seen to provide important steps on the road to the ultimate account of the distinction, all modeling would be regarded with some suspicion as tentative, and researchers would place more trust in experimental assessments of the properties of cognition$_{ex}$ and mere association$_{ex}$ themselves. The mantra of this approach to the distinction might be that "the animal itself is its own best model."

Admittedly, this "pure" way of presenting the approach is an oversimplification; behavioral criteria are often initially proposed to distinguish between two kinds of model, though they can quickly thereafter take on a life of their own. Whatever their origin, behavior-based approaches to the distinction rely on a series of operational criteria which can be assessed in experimental settings to determine whether some behavior is driven by cognition$_{ex}$ or mere association$_{ex}$. Behavior-based approaches to mere association$_{ex}$ often stray little from the behaviorist's conditioning methodologies, so I will comment little on them here. By contrast, the last four decades have seen a great surge of interest in behavioral tests for cognition$_{ex}$, especially those which, by requiring particular forms of behavioral flexibility, rule out associative$_{ex}$ explanations of the data.

These criteria can be formulated simultaneously at two levels. In a pessimistic appraisal of the utility of grand dichotomies, Newell (1973) described the nature of psychological research as simultaneously investigating the level of "phenomena" and the level of "conception." The "phenomenal" level pertains directly to observed effects that can be reliably produced in experimental settings; Newell produced a list of 59 such effects, including things like categorical conceptualization, subitizing, chunking, hierarchical effects in long-term memory, serial position effects in recall, backward associations, reversal learning, and so on. At the same time, however, psychologists attempt to tie their "low-level" experimental or modeling research into conceptualizations "at the high level of grand theory" — which Newell supposed to unfold "mostly…by the construction of [binary] oppositions" (of which cognition vs. association was one of the most prominent).

While the list of phenomena and our models of them might have evolved during the intervening decades, this basic two-tiered approach is still alive and well in comparative cognition research. The important point for present purposes is that we can direct our attention towards formulating very specific behavioral criteria for individual cognitive$_{\text{ex}}$ or associative$_{\text{ex}}$ capacities, or towards very abstract criteria at the level of the grand dichotomy. As with the model-based criteria, we find a trade-off between the desiderata of Generality and Empirical Reference. In order to also satisfy Inductive Unity, the criteria at the two levels should not be unrelated; for example, the criteria for assessing whether a behavior was driven by a particular cognitive$_{\text{ex}}$ capacity should be seen as specific expressions of the characteristics attributed to cognition$_{\text{ex}}$ more generally.

Newell's Levels	Capacities	Operational Tests
"Conceptual" Level	Cognition vs. Association	Behavioral Flexibility vs. Behavioral Stereotypy
"Phenomenal" Level	• Cognitive mapping • Transitive Inference • Conceptual abilities • Episodic memory • Mindreading • Metacognition • Self-recognition • … • Blocking •Value transfer • Overshadowing • …	• Latent learning, novel routes… • 5-element series • Abstraction, "sharp" generalization gradients • Dissociation of current & future motivations • Deception detection, false belief task • "Bail-out" options in uncertainty monitoring • Dot-touching paradigm • … • Blocking •Value transfer • Overshadowing • …

Figure 3. A conceptual map of current research into Newell's "levels," with examples of some common operational criteria.

Most of the abstract behavioral criteria for cognition$_{\text{ex}}$ can be seen to follow from the intuition that cognition$_{\text{ex}}$ permits a kind of behavioral flexibility which is unavailable to mere association$_{\text{ex}}$. This general intuition is expressed by Smolensky's (1988) criteria for cognition (which includes both his "intuitive" and "reflective" systems):

> A cognitive system can, under a wide variety of environmental conditions, maintain a large number of goal conditions. The greater the repertoire of goals and variety of tolerable conditions, the greater the cognitive capacity of the system. (p. 15)

Mitchell et al. (2009) rely on indicators of "effort" and "attention" as behavioral criteria for "propositional" processes. Penn and Povinelli (2007) regard the ability "to understand the 'web of possibilities' that connects causes and effects" as a sufficient criterion for cognition, as this is critical to their argument that the extended comparator hypothesis is both cognitive and associative.

These discussions of behavioral flexibility are often too abstract to be useful, however; such criteria must be cashed out in specific experimental tests to be empirically meaningful.[15] More interesting and illuminating criteria arise from the behavioral tests for specific cognitive capacities that have been worked out by comparative cognition researchers. A case study of these criteria can reveal the specific forms of behavioral flexibility attributed to cognition$_{ex}$ by comparative psychologists.

Examples of Behavior-Based Approaches

Case Studies: Cognitive Maps. Cognitive (or mental) mapping is a paradigm cognitive$_{ex}$ phenomena. Like real maps, they are characterized as *representing* the environment which they model. Furthermore, work on cognitive maps was historically influential in shifting attention in animal psychology from the "behavioral stereotypy" emphasized by radical behaviorists to the "behavioral flexibility" that has been the focus of comparative cognition research (Olton, 1979). While Chomsky's research in human linguistics is often cited as the most important work driving the cognitive turn in human psychology, Tolman's (1948) work on cognitive maps was perhaps more influential amongst researchers on nonhuman animals.

In his classic work *Cognitive maps in rats and men* (1948), Tolman insisted that the "stimulus-response" school of learning could not account for the navigation behaviors of his rats. Instead, he suggested the existence of more sophisticated learning mechanisms:

> The stimuli, which are allowed in, are not connected by just simple one-to-one switches to the outgoing responses. Rather, the incoming impulses are usually worked over and elaborated in the central control room into a tentative, cognitive-like map of the environment. (p. 193)

The primary distinguishing factor of this latter form of learning is its "selectivity" and the way in which stimuli are affected by each other and by what is already known (especially about other cues). On Tolman's more sophisticated form of learning, stimuli do not simply act individually and independently on the rat's nervous system. Instead, only contingencies between stimuli deemed *relevant* and *informative* are attended to and integrated into a global, map-like representation of the environment — the selection and integration processes being based in large part on what is already known.

Tolman cited five categories of behavioral criteria which he used to operationalize his "cognitive control room" (leading to his classification as a "cognitive behaviorist"): "latent learning," "vicarious trial and error," "searching for the stimulus," "hypothesis," and "spatial orientation," each of which investigated the rat's abilities to make use of information in ways which were more

---

[15] To their credit, some proponents of abstract criteria have often attempted to discharge this burden. Mitchell et al. (2009) dismiss purported evidence for "link-forming mechanism" explanations of learning data by using diminishment under cognitive load as an operational criterion for propositional learning; Penn and Povinelli (2007) suggest that the cognitive understanding of causes and effects can be operationalized in terms of retrospective reevaluation effects, supporting their contention that the Extended Comparator Hypothesis satisfies cognitive criteria.

flexible than would be predicted by the S-R behaviorism of the day. In his famous experiments on latent learning, for example, Tolman demonstrated that rats could acquire navigational knowledge about a maze through idle exploration, in the absence of explicit reinforcement.

Further experiments on cognitive mapping since Tolman have continued to emphasize the importance of being able to flexibly and adaptively use information gained on one task in different contexts or to achieve different goals. For example, Walker and Olton (1979) conducted a series of experiments showing that rats could spontaneously deploy spatial information gained from exploring one part of a maze to navigate to the goal from a novel starting position at which they had never before been placed. Menzel (1973, 1978) showed that chimpanzees could also use spatial information flexibly in a variety of tasks in a way which could not be explained by simple conditioning. For example, in one case experimenters carried chimpanzees around a room in crisscrossing and convoluted paths while hiding food in 18 locations; they then released the chimps to see if they would replicate the tortured paths or recover the food in a more efficient manner. The chimps tended to utilize novel, efficient routes around the perimeter of the enclosures to recover the food, rarely crossing their paths.

The concept of a cognitive map was perhaps developed into its most sophisticated form by O'Keefe and Nadel in their influential interdisciplinary work, *The Hippocampus as a Cognitive Map* (1979). There, the cognitive map concept is described in detail and contrasted with more basic systems, such as "instinctual," single-cue, and route-based strategies for spatial navigation (cognitive maps have also since been distinguished from path integration and dead reckoning (see Mackintosh, 2002) though it is not clear that all authors would endorse a neat division). Again, O'Keefe and Nadel (1979) claim that the distinguishing feature of cognitive maps is the behavioral flexibility they support:

> The first striking feature of a map is its flexibility. Whereas a route specifies a starting point, a goal, and a particular direction of movement from the former to the latter, a map specifies none of these, either in its construction or its usage. It can be used with equal facility to get from any particular place to any other. Additional flexibility derives from the freedom from specific objects and behaviours. If one path is blocked another can be easily found and followed. (p. 87)

O'Keefe and Nadel unpack the flexibility of cognitive maps into five characteristics (all but #4 of which are straightforwardly behavior-based — and operationalizations of #4 have been proposed) which they hold can be used to distinguish cognitive maps from other navigation strategies:

1. Learning is driven by "curiosity" and "novelty-seeking," rather than by mere exposure to stimulus co-occurrences or repetition. Redundant or uninteresting cues may be ignored entirely, whereas novel or surprising cues are sought out.

2. The representation is suitable to reach a range of goals and from a range of starting points, many of which may have never been explored. When using a cognitive map, the creature can get from many points on the map to many other points, whereas routes lead to only one goal.

3. Maps are robust in the face of environmental change and navigational error. They do not depend on any particular subset of landmarks or turns. Routes and instinctual strategies are brittle in the face of change, depending upon a small subset of cues or a particular series of turns or distances, which, if missed through mistake, environmental change, or lapse in attention, render the goal inaccessible until the creature can re-locate a specific position on that route.

4. Maps are extremely efficient and non-redundant. Many routes may terminate in the same sequence of turns, for example, but these sequences will be redundantly represented separately in each route. Maps, on the other hand, represent an enormous number of possible routes with a single, integrated representation.

5. Maps are at least potentially multi-modal; they do not depend on any one sensory modality. Involving a true representation of space, navigation is possible by sight, touch, and (for many species) smell, sound, and taste.

Case Studies: Episodic Memory and Future Planning. Episodic memory, another paradigm cognitive$_{ex}$ capacity, is typically characterized by the ability to flexibly deploy a mental record of a specific experienced event in terms of its “what,” “when,” and “where.” Furthermore, episodic memory is almost universally distinguished from other forms of associative$_{ex}$ learning such as procedural memory — and thus serves as an illustrative case study for present purposes.

Tulving’s (1983, 1985) influential account of episodic memory has been the launching point for many studies on the faculty. According to Tulving, episodic memory was built on earlier memory systems, the procedural (stimulus-response) and the semantic (which involves the construction of “response-neutral,” “third-person” mental models of the world). Episodic memory is to be distinguished from these two earlier systems by virtue of its ability to enable what Tulving (1985) called “mental time travel”:

> Episodic memory affords the additional capability of acquisition and retention of knowledge about personally experienced events and their temporal relations in subjective time and the ability to mentally "travel back" in time. (p. 387)

In short, an organism with episodic memory is able to “re-experience” remembered events and use information acquired from that re-experiencing to make decisions. The ability to engage in mental time travel is purportedly enabled by a difference in the kind of representations involved in episodic memory, which records particular experienced events in their autobiographical context.

According to these theorists, procedural memory allows only rather inflexible behavior; responding according to a relatively rigid format to stimuli determined at the time of learning is obligatory (Hirsh, 1974; Mishkin & Petri, 1984). On the other hand, knowledge acquired in both semantic and episodic memory can be expressed flexibly, in a variety of distinct behavioral and perceptual circumstances. In influential studies on episodic memory, Clayton and Dickinson (2009) have proposed to investigate these forms of flexibility in caching bird species like scrub jays. Clayton and Dickinson capitalize on the idea that one of the proposed functions of an ability to relive past experiences is to utilize that information in anticipating and planning for similar events occurring in the future. As they put it,

> The function of episodic memory lies primarily in its constructive rather than reconstructive ability: its purpose is to mentally simulate multiple future scenarios *by flexibly recombining details from past events without having to physically engage in the actual behavior*. Inevitably there is a trade-off between flexibility and stability, and therefore the cost is that the episodic memory system is much more vulnerable than the semantic knowledge system to memory errors such as misattribution and false recognition. (p. 62, emphasis added).

As such, they have investigated the ability of caching birds to recall details of past caching activities. In these tests, experiments are designed to determine whether birds are capable of recording this information rapidly and deploying it flexibly to plan for the future.

For example, one of the key experiments that Clayton and Dickinson (2009) performed was designed to demonstrate their ability to plan for the future by differentially caching a food item that would be preferred in anticipated future conditions, but not preferred at the time of caching. In one experiment (Raby, Alexis, Dickinson, & Clayton, 2007), scrub jays were confined at breakfast time in one of two enclosures, and given the opportunity to learn that they would receive either no food or a particular type of food in one enclosure, and another type of food in the other. One evening, the birds were then given the opportunity to cache a particular type of food; if they were operating with “episodic-like” memory, we would predict that they would preferentially cache the type of food in the enclosure where they had not previously received it, remembering their previous deprivation and anticipating their future breakfast needs. This prediction was confirmed; across several different variations of the experiment, the birds preferentially cached foods in the enclosures where they had not previously received it before.

Why should we suppose that the explanation of this behavior merits appeal to episodic memory? Again, the contrast class for Clayton and Dickinson (2009) were various forms of “instinctive” behaviors and associative$_{ex}$ learning. First, the experimenters noted that the abilities must not be found to depend invariantly upon a handful of cues and be extremely limited in their means of expression; as they point out, relatively “hard-wired” behaviors like migration, nest-building, and hibernation, despite being apparently “future-oriented,” would not count as cognitive$_{ex}$ unless we can “rule out simpler accounts in terms of behavior triggered by seasonal cues or previous reinforcement of the anticipatory act” (p. 65). Previous experiments had found significant flexibility in scrub jay caching abilities, such as a sensitivity to the type of item cached and an ability to learn and utilize information about the rates at which these cached items degraded (reviewed in Clayton & Dickinson, 2009), ruling out the simplest “instinctual” explanations. It remained to be proven, however, that the bird’s apparent abilities to plan for the future were not based in past conditioning or in current motivational states. Their experiments thus included elaborate controls to rule out such “simpler” explanations, which included that:

1. The birds were confined the same number of times in each compartment in an attempt to rule out an associative preference for one over the other.

2. The birds were fed to satiety prior to the caching event, to rule out an explanation in terms of their current motivational state.

3. The birds had never before been given the opportunity to cache in either enclosure, and thus caching in those locations was unexpected and novel.

These controls have convinced some who were skeptical of earlier studies — for example, Shettleworth (2007) has suggested two requirements for evidence of future planning in nonhuman animals, that “the behavior involved should be a novel action or combination of actions” and “it should be appropriate to a motivational state other than the one the animal is in at that moment.” On surveying the controls in the Raby et al. (2007) experiment, Shettleworth (2007) conceded that it comprised “the first observations [on nonhuman animals] that unambiguously fulfill both requirements’’ (p. 825).[16]

This is not to say that all skeptics have been convinced. Suddendorf and Corballis (2007) have challenged the conclusions of Raby et al. (2007) on the basis of — again — a lack of the appropriate kind of flexibility for cognitive$_{ex}$ ascriptions. One complaint stems from the fact that a “genuine” manifestation of episodic memory and future planning should be able to exhibit itself domain-generally, and there is not currently any evidence that jays can exhibit episodic memory in any domain other than caching. While the “flexibility criteria” of Clayton and her colleagues (for a review, see Clayton, Russel, & Dickinson 2009) have focused on the rapidity of learning from caching episodes and the ability to generalize the learned information to novel circumstances, Suddendorf and Corballis instead emphasize a need to manifest memory and future-oriented abilities across different. Suddendorf and Corballis (2007) suggest that absent further information, a simple associative mechanism may explain the data and that the “ability in linking their caching and retrieval they have demonstrated may be more akin to Garcia-type learning than to human mental time travel” (p. 2). Rapid one-trial learning recording what-when-where data, they suggest, may not be

---

[16] Shettleworth has since expressed some additional skepticism; see Shettleworth, 2010.

sufficiently flexible to count as even episodic-*like* memory if the context of application is so limited.

Whatever the fate of the Raby et al. (2007) conclusions, from this brief review of the dispute on episodic memory and future planning, it is supposed that behaviors driven by episodic memory:

1. Should deviate from what would be predicted from simple conditioning and other invariant S-R mechanisms, and should deviate in ways which are flexible and adaptive. (How intelligent? Adaptive in what circumstances?)

2. Should manifest themselves across a diverse range of tasks and domains (How many?)

3. Should manifest in novel perceptual circumstances on tasks never before attempted. (How novel?)

4. Should cease to manifest when no longer appropriate/needed (always, or only when there is some cost — and if so, how quickly and how costly?).

Each criterion here is faced with a problem which limits its objective applicability: they require an evaluative judgment for which precise criteria are unspecified, indicated in the parentheticals above. Different researchers often disagree on the evaluative yardstick to be applied; this difficulty will be commented on again in at the end of this section.

Case Studies: Transitive Inference. The final cognitive$_{ex}$ capacity which I will discuss here is transitive inference. The ability to pick out the easiest prey, most nourishing food, or dominant troop mate is of obvious adaptive significance. Two kinds of solution to this problem are possible. First, animals can learn rankings by rote, encountering each possible set of options and learning their relative rank(s) by trial and error. Secondly, and more efficiently, animals can draw inferences about novel sets of options (that is, between sets never before encountered) by flexibly integrating information learned in previous encounters. In other words, animals can use encounters with two or more distinct choice sets which involve common elements to rank items on a common monotonic dimension (tastiness, defensive potential, dominance, etc.). For example, if the *R* relation is transitive and I know that *aRb* and *bRc*, I can conclude that *aRc* without needing to evaluate an *ac* pair. Recognizing this would allow an organism to display correct behavior on the perceptually novel pair the first time it is encountered, without needing to go through the potentially costly evaluation process. This latter solution to the ranking problem has been considered another paradigm cognitive$_{ex}$ capacity.

The "gold standard" behavioral test for transitive inference has, for many years, been the 5-element series (Bryant & Trabasso, 1971; McGonigle & Chalmers, 1977). In the 5-element series, subjects are trained on four successive discrimination dyads — A+B-, B+C-, C+D-, and D+E-. Subjects are then tested on the novel BD dyad, and above-chance responding to "B" is taken as evidence that animals are capable of transitive inference. The key to this test is that B and D have been rewarded an equal number of times, and thus should have a similar elemental reinforcement value. A correct choice on a novel AE dyad, for example, would provide no evidence of transitive inference, for A was rewarded 100% of the time and E 0%. On the other hand, the organism responding on the basis of mere elemental conditioning should show no preference for B over D (this is why all authors agree that we must, at minimum, go to a five-element series to find genuine evidence of transitive inference).

At the time of writing, a number of species have "passed" the 5-element test, including several species of monkey, rats, and pigeons. This is not to say, however, that skeptics have been satisfied that all these animals are implementing the "cognitive" solution to the problem. von Fersen, Wynne, Delius, and Staddon (1991) suggested an "associative" mechanism of "value transfer" to account for animal performance on the 5-element task. On the value transfer account, stimuli can get a reward "boost" from co-occurring with more frequently rewarded stimuli. So while A has been rewarded 100% of the time and B 50% of the time, when A and B co-occur, B gets a boost in its own reinforcement value for having co-occurred with a more highly-rewarded A. Value transfer is here construed as a merely associative$_{ex}$ mechanism, given that it is based only on the spatiotemporal contiguity of cues, and thus to provide a "simpler," deflationary explanation for observed animal data on the 5-element task.

Models built on the value transfer principle (De Lillo et al., 2001; Frank, Rudy, & O'Reilly, 2003) have since purportedly done a better job of accounting for animal data than some of the more traditional, rule-based approaches to transitive behaviors (e.g., Dusek & Eichenbaum, 1997). Van Elzakker, O'Reilly, and Rudy (2003), for example, collected additional data on a six-element task (A+B-, B+C-, C+D-, D+E-, E+F-), arguing that the traditional, rule-based conclusion should predict no difference in responding between BD and BE pairs (since each correct choice receives equal reward), whereas they observed much stronger responding to B when paired with E than when paired with D. However, more sophisticated versions of the cognitive$_{mod}$ models of transitive inference have predicted a Symbolic Distance Effect (McGonigle & Chalmers, 2006), where animals find it "easier" to make discriminations amongst items spaced far apart on the monotonic scale than those ranked nearby (and thus can make them more quickly). Since B and E would be farther apart on the integrated preference ranking, the thought goes, we would predict quicker and more successful discriminations on BE than on BD. Van Elzakker et al. (2003) retort, however, that even if we allow the cognitivist the normalized symbolic distance scales that allow them to make the right predictions about BD and BE on the 6-element task, the account seems unable to explain why the observed BE responding in their 6-element task was essentially identical to the BD responding in the 5-element task. We must accept that a 0.1 value difference between the normalized distances between B and D on the 5- and 6-element scales (0.5 vs. 0.4) translates into a 30% difference in response time.

Should we thus conclude the discussion with a clear associationist win? Nowhere are the conceptual tensions surrounding the distinction between cognition$_{ex}$ and mere association$_{ex}$ more apparent than in the debate surrounding transitive inference, for we have come farther with the experiments, models, and underlying neuroscience here than with any other capacity. While Frank, Rudy, and O'Reilly (2003) present a sophisticated neural (and apparently associative$_{mod}$) model of transitive inference in a companion paper to Van Elzakker et al. (2003), how are we certain that they have offered a deflationary associative$_{ex}$ alternative for the cognitive$_{ex}$ accounts of transitive inference, as opposed to a neural implementation of them? While Van Elzakker et al.'s value transfer account makes slightly different predictions about the shape of learning curves than the Symbolic Distance account of McGonigle and Chalmers (1992), they concede that BD responding would improve on the 6-element task with additional training. Their model also posits a variety of mechanisms for flexible representation which go well beyond what one might expect of simpler associative$_{ex}$ approaches to learning — for example, they interpret the hippocampal system as "construct[ing] conjunctive representation[s] of the choice stimuli that resolves the ambiguity of their associative values." Furthermore, their model takes the hippocampal system to record and constantly update excitatory strengths for each stimulus by tracking the reward and co-occurrence history in these conjunctive representations, effectively creating a ranking which will roughly reflect the position of those stimuli as ordered along a monotonic dimension. Their account thus requires a variety of processes which would seem cognitive$_{ex}$ according to many criteria. The question remains as to whether their model counts as *implementing* an integrative, cognitive$_{ex}$ solution to transitive inference, or rather specifies a sophisticated associative$_{ex}$ alternative to it.

Problems with Purely Behavior-Based Criteria

Problem 1: It Matters How You Do It

The first observation, mirroring those found in the case of purely model-based criteria, is that purely behavior-based criteria cannot be taken as self-standing indicators of cognition$_{ex}$. To illustrate this point, consider the debate over John Searle's (1980) (in)famous "Chinese Room" thought experiment. Searle invites us to imagine a man who speaks no Chinese in a room with an input slot, and output slot, drawers full of Chinese symbols, and a rulebook. The rulebook contains a series of instructions about which Chinese symbols to put in the output slot in response to certain symbols being presented at the input slot. We are also invited to imagine that this rulebook is so comprehensive that native Chinese speakers cannot distinguish the responses to queries released in the output slot by the man in the room from those of a native Chinese speaker.

The intuition Searle (1980) expects us to have is that the man in the room does not understand Chinese, because the procedure he uses to generate his responses in fact involve no real intelligence. While Searle intended to generalize this intuition to support a pessimistic appraisal of the prospects for true "artificial intelligence," others have derived from it a broader and more moderate lesson. Block (1981), for example, has argued that the moral of the story is not that no artificial system could be intelligent, but rather that it matters for the purported intelligence of the system how it implements the behavioral capacities it appears to display. If that system displayed a degree of internal functional sophistication that went beyond the series of simple input-output rules suggested by Searle's rulebook, then it becomes increasingly more plausible to say that, in fact, the whole system of the room — that is, the man together with the rulebook — does understand Chinese.

The same moral carries over to behavioral criteria for cognition$_{ex}$ in the present context, as behavior-based methods can be seen as input-output tests. The problem arises from the Inductive Unity desiderata, for simple behavioral tests could be passed by different processes which internally had little in common with one another. Really, the idea of a merely associative$_{ex}$ "stimulus-response automaton" that produced behavior truly indistinguishable, in all circumstances, from that of a cognitive$_{ex}$ organism is not a live empirical possibility worth worrying about. However, if we are to engage in an iterative investigation where our behavioral criteria begin as fairly thinly-specified, this kind of concern can remind us to regard these initial criteria with humility. It may be possible for a "stimulus-response automata" to produce behavior which passes some of the behavioral tests nominated above if we are not careful to implicitly constrain these tests by appeal to certain forms of internal processing. In short, one must go further than specify *what* an organism must do to count as cognitive$_{ex}$; we must also link these tests to some constraints on *how* the organism can do it.

Problem 2: The Motley Crew

A further problem for the behavior-based criteria for cognition$_{ex}$ and mere association$_{ex}$ is that these criteria can appear quite diverse. While many philosophers and psychologists have tended to focus on one or two criteria for cognition$_{ex}$ such as systematicity or effortfulness in isolation, there seems little a priori reason why these properties should truck together. Why would systematic thought necessarily be effortful? The situation seems even more puzzling when we make closer contact with the experimental work surrounding the distinction in Standard Practice, as we can come up with scores of additional experimental tests from a systematic review of the literature. Why, for example, should "fast learning" and "ability to detect abstractions" both be part of cognition$_{ex}$? Why should an ability to detect abstractions permit context-sensitivity? Why should any of these properties imply the use of multiple modalities? The questions become even more baffling if we move to specific behavioral criteria at Newell's (1973) "phenomenal" level discussed in the case studies of the previous section; why should an ability to pass the 5-element series imply an ability to find novel routes in mazes or dissociate current from future motivations when planning for breakfast?[17]

Again, it matters for the Inductive Unity desiderata whether they do. While it may be possible to answer this question through brute force, by systematically

---

[17] It is a live possibility — perhaps even a likely one — that these properties only co-occur due to historical reasons. For example, they may all be due to the fact that evolutionary homologies cause certain functionally-distinct brain structures to regularly develop together across a wide swathe of the animal kingdom. That the properties attributed to cognition$_{ex}$ may only co-occur for historical reasons, however, is no bar to its legitimacy as a respectable scientific posit — for example, see Millikan (1999) and Boyd (1999).

investigating whether these properties co-occur across the full range of behaviors and species of relevance to comparative psychology, we should surely avoid such an inefficient method if possible. It is not clear that there are more efficient methods available at the purely behavioral level, however. Rather, obtaining answers to these questions more efficiently may require appeal to processing stories inspired by reflection on the internal workings of particular models, which can make predictions about the behavioral capacities which would be exhibited under a wide variety of circumstances.

Problem 3: How much? Vague and Relational Criteria

Another persistent threat to the independence of behavior-based criteria is that they are often specified in vague or relational terms. “Fast learning” is often attributed to cognition$_{ex}$ in tests of learning set or reversal learning, for example; but how fast must the learning be to count as cognitive$_{ex}$? These vague, relational criteria were found to be present in each of the criteria which could be derived from the debates over the Raby et al. (2007) studies, for example. If we had particular models which we could appeal to in order to provide anchor points for the comparatives, we might be able to answer this challenge. We could specify, for example, how many trials an organism will require to master a learning set, or the range of different contexts in which an organism should be able to satisfy a goal, for the processes behind those behaviors to count as cognitive$_{ex}$.

Admittedly, simple answers which would apply invariantly across all species or classes is unlikely to be in the cards; sloths might simply take longer to cognize$_{ex}$ than hummingbirds. What is needed is not only the ability to reference a particular model of cognition$_{ex}$ to garner an estimate of, for example, the number of trials required to master a learning set, but rather also some indication of how to fit the model to the particular abilities possessed by disparate species. As I will suggest in the conclusion, having some idea as to how to tie the cognitive$_{ex}$ abilities to underlying neuroanatomy and then comparing the similarities and differences of the neuroanatomical structures possessed by the different species might provide some guidance in this respect — showing both what different organisms have in common and making quantitative predictions as to how they might differ.

Problem 4: Mixed, “Unmoored” Criteria — Where Things Go Really Awry

Here is a pattern that has been repeated several times in the debate surrounding the distinction which led directly to the current crisis: A comparative cognition researcher begins investigating some behavioral capacity at the level of Newell’s (1973) “phenomena.” At least two possible explanations for a phenomena are proposed, one cognitive and the other associative. A behavior-based criterion is proposed which would discriminate between those explanations, and a successful experimental paradigm designed around that behavioral criteria leads to a cottage industry in determining which animals can “pass the test.” Later, however, that behavior-based test — initially justified in virtue of its ability to distinguish between two specific model-based explanations — becomes "unmoored" from the explanations from which it originally derived its justification. For example, one or both of the original models might have fallen out of favor. In other cases, a new associative($_{\text{mod}}$) model is proposed which is a slight modification or tweak of the prior associative($_{\text{mod}}$) model for which the behavioral criterion was devised, but which can now "pass" the test for cognition$_{\text{ex}}$ that the prior associative$_{\text{mod}}$ model could not (and, often, do little extra besides). It is then proclaimed that the original capacity was non-cognitive$_{\text{ex}}$, or, worse, that the distinction itself is untenable and should be rejected.

Examples of this pattern are not difficult to find: traditional associative models were augmented with "value transfer" (von Fersen et al., 1991) to pass the 5-element transitive inference test; the comparator hypothesis was "extended" with higher-order comparisons to allow it to show retrospective revaluation effects (Denniston, Svastano, Blaisdell, & Miller, 2001); and multi-stimuli "elements" were "added-" or "replaced-" to elemental models of learning to allow them to solve certain configural and non-linear classification tasks (Wagner, 2008). While it is always good to try to devise models that explain *more* data, all other things being equal, turning around and then using such developments to criticize the distinction itself, based on the fact that the augmented model can now pass a behavioral test nominated only for its (now obsolete) ability to distinguish between two specific explanations, is surely a mistake.

Certainly, it is possible that an increasing number of tweaks defeating traditional tests for cognition$_{\text{ex}}$, if derived from some unified, independent justification, could indicate that a "dichotomous" approach to the psychology is destined for the dustbin. The fact that there seems to be little structurally in common across the tweaks mentioned in the previous paragraph, however, does not offer the kind of "concerted accumulation of evidence" that would support this narrative. Rather, it looks like a number of models can be extended with a number of (possibly artificial) tweaks which help those models pass specific behavioral tests. While again one should not fault psychologists for attempting to produce models which can accommodate more data, I think we should be wary when "unmoored" behavioral tests are deployed in criticisms of the distinction without further argument justifying those behavioral tests themselves. For many of these tests, I suggest that no further argument can be given. Divorced from specific cognitive$_{\text{ex}}$ and associative$_{\text{ex}}$ models between which it can distinguish, there is little reason to suppose that "passing" the five-element series test establishes anything of significance.

No defensible justification is typically given, in these dialectics, to support the conclusion that the tweaked associative$_{\text{mod}}$ model which passes the (now unmoored) behavioral test for cognition$_{\text{ex}}$ in fact still describes capacities which are associative$_{\text{ex}}$. A significant possibility is that the tweaked model in fact describes a process which is cognitive$_{\text{ex}}$, or — much worse — is a gerrymandered representation which does not really correspond to any actual process. If this is right, then such models should not be offered as evidence for the triumphant conclusion that all thought is really associative$_{\text{ex}}$ in nature.

What is needed are criteria to distinguish between "virtuous" and "vicious" tweaks to models which would allow us to place the "genuine advances" to one side and the "ad hoc tweaks" to the other. Standard scientific principles already supply us with much to draw upon here — we should investigate whether the modified models match other features of the "performance" data such as response times, make novel predictions about behavioral capacities which are unrelated to the behavioral tests they are explicitly designed to defeat, involve fewer free parameters, are more biologically plausible, match patterns of deficits found in real organisms or lesion models, and so on. How well the recent generation of "tweaked" models will fare on these broader criteria for model selection remains to be seen. At present, I simply mean to temper the enthusiasm encouraged by the mere existence of an associative$_{mod}$ model which can defeat a behavioral test for cognition$_{ex}$.

Conclusion: How Comparative Psychology Might get by with a Little Help from its Friends

From behaviorism, we inherited a fairly firm understanding of the most basic forms of association$_{ex}$. From classical computationalism (e.g., Newell & Simon 1976), we also possess a fairly firm understanding of the most advanced forms of cognition$_{ex}$. The problem — the basic shape of which has changed little from the time of Aristotle — is that we still don't know how to divide up the great middle ground between them, which surely constitutes the majority of the mental capacities of humans and animals, and neither model- nor behavior-based criteria are alone up to the task. Yet we have already conceded that both behavior- and model-based criteria must be continuously revised as our continual accumulation of experimental evidence allows us to improve our tenuous grasp on the distinction, resulting in better models and behavioral predictions. How can we engage in this iterative inquiry, constantly revising both model- and behavior-based criteria, without, as Newell (1973) feared, having the distinction constantly shift under our feet, preventing the accumulation of conceptual progress?

The rough solution is to carve up the middle ground such that both association$_{ex}$ and cognition$_{ex}$ simultaneously fare as well as possible on the aforementioned desiderata of Generalization, Empirical Reference, and Inductive Unity. The challenge, of course, is that we do not yet know the best way to do this given the current state of empirical knowledge available to us in psychology. What is needed is thus a way to structure an iterative inquiry which makes optimal use of what little we do know to gradually sharpen the distinction while minimizing the chance that Newell's (1973) fears will be realized.

My suggestion is that the best way to do this is to be guided by other sciences which have often been supposed "irrelevant" to the ontology of psychology. Much ink has been spilt about the purported "autonomy" of psychology (Fodor, 1974, 1997; Jones, 2004). It is important to note, however, that these autonomy arguments, even if successful, establish only the freedom of psychology from metaphysical reduction to lower-level sciences, and even then only on relatively strong conceptions of reduction. All that is needed in the present context are methods to estimate the suitability along the aforementioned desiderata of various ways of grouping together behavioral properties and psychological models under the labels "cognitive$_{ex}$" and "associative$_{ex}$." We already know a great deal about comparative neuroanatomy which generates predictions about the presence or absence of behavioral capacities. The same goes, to perhaps a lesser degree, with evolutionary biology, under the assumption that species with shared phyletic backgrounds and ecological pressures will tend to have similar psychological capacities and behavioral abilities. Attempting to assess the legitimacy of the distinction in Standard Practice without making use of this clearly-relevant data seems to me like trying to run a marathon with one leg hobbled.

Being guided by neuroscience and evolutionary biology is not a panacea, of course. In many ways, the interdisciplinary move simply introduces additional moving parts into the debate, and issues of Generality, Empirical Reference, and Inductive Unity arise again in other sciences. Neuroanatomical regions can themselves be described at varying levels of abstraction, and it is not always clear which level is appropriate when linking a behavioral capacity to its neural substrates (Craver, 2009). Though I have elsewhere defended the importance of medial temporal lobe structures as the locus of representational flexibility characteristic of cognition$_{ex}$ (Buckner, 2011), there have been significant difficulties in cleanly delineating this region from nearby brain structures, and there continue to be heated debates as to the region's function — for example whether it is best construed as primarily spatial (O'Keefe & Nadel, 1978), as supporting all forms of relational memory (Dusek & Eichenbaum, 1997), or as an all-purpose declarative memory system (Squire, 1992).

Progress in neuroscience, however, is arguably more stable than progress in psychology. While our divisions at the level of gross neuroanatomy have changed dramatically over the last hundred years, some accumulation of information has clearly been achieved. The exaggerated rhetorical differences between competing research groups working on medial temporal lobe function obscures the vast amount of ground now shared between them. There is now little doubt that the medial temporal lobes play a critical role in a variety of cognitive$_{ex}$ capacities such as cognitive mapping, episodic memory, and transitive inference. Similarly, we have made some significant progress in recent years in establishing homologies in brain structures between disparate classes, such as the verdict of homology between the mammalian and avian hippocampus (Columbo & Broadbent, 2000; Reiner et al., 2004). These hard-won verdicts summarize vast amounts of anatomical and behavioral data, have been endorsed by a consensus of the neuroscientific community, and carry significant implications about the cognitive$_{ex}$ capacities that disparate species are likely to possess. These neuroscientific findings can thus ground an iterative inquiry with a reasonable hope of avoiding stagnation in Newell's (1973) quagmire.

The best way to make gradual progress on our quarry is to begin with model-based criteria which are focused on large-scale differences in gross neuroanatomy — at a balanced level of abstraction which is specific enough to illustrate compelling differences in the functional capacities of different areas of the brain while still abstracting away from details likely to be unique to particular species and models which will be revised as our future neuroscience evolves. We should then develop behavior-based criteria which distinguish between these models, and then use the evidence we accumulate by testing animals on these criteria to elaborate those models. I have suggested (Buckner, 2011) that the distinction in the cortico-hippocampal model of Gluck and Myers (2001) provides an example of how such a balance might be struck, with learning processes mediated by the medial temporal lobes (and functional homologues in other classes) uniting the cognitive$_{ex}$ side of the distinction and less flexible learning systems such as the cortices and spinal cord uniting the associative$_{ex}$ side. Other neurally-based schemes should of course be explored — though I know of no story with as wide and relevant a scope as that based on the medial temporal lobes.

Like any model, Gluck and Myers' (1994) involves a number of idealizations which may be discarded as our knowledge develops. Nevertheless, their model is based in gross differences in underlying neuroanatomy — in particular on the well-established flow of highly-processed sensory information through the entorhinal cortex, to the CA3 region, and back out to the same stimulus input stream — which are common across many species and likely to be conserved in future discoveries. However, their model is still specific enough to issue a variety of predictions which appear to be borne out by currently-available experimental data for diverse species and learning paradigms. Thus, a distinction supported by this or similar neuroanatomically-guided models would seem to fare well on our three desiderata.

That all being said, monistic or pluralistic perspectives on animal psychology may still turn out to be correct — and Standard Practice might turn out to be misguided. The comments in this concluding section were meant as a defense of how we should conduct our investigations to obtain a fair assessment of the hypothesis that the dichotomous perspective presumed by Standard Practice is cogent. If we had the "optimally predictive global model of the mind" which outperformed all other models on all tasks, we could simply read off the legitimacy of the distinction from that model by seeing whether it posits one, two, or many basic kinds of psychological process. In the absence of this optimal model, however, we should continue to simultaneously explore all plausible perspectives — and to the winner go the spoils. What we should not continue to do, I have argued above, is haphazardly mix model- and behavior-based approaches in our assessments of the cogency of Standard Practice.

References

Allen, C. (2006). Transitive inference in animals: Reasoning or conditioned associations? In S. Hurley & M. Nudds (Eds.), *Rational animals?* (pp.175-185). Oxford, UK: Oxford University Press.

Allen, C., Grau, J., & Meagher, M. (2009). The lower bounds of cognition: What do spinal cords reveal? In J. Bickle (Ed.), *The Oxford handbook of philosophy of neuroscience* (pp. 129-142). Oxford, UK: Oxford University Press.

Aristotle. (1925). Nicomachean ethics. Trans. W. D. Ross. In R. McKeon (Ed.), *Introduction to Aristotle* (pp.151-249). New York: Random House.

Aristotle. (1928). *Posterior analytics.* Trans. G. R. G. Mure. In R. McKeon (Ed.), *Introduction to Aristotle* (pp.151-249). New York: Random House.

Aristotle. (1931). *De anima.* Trans. J. A. Smith. In R. McKeon (Ed.), *Introduction to Aristotle* (pp.151-249). New York: Random House.

Blaisdell, A. (2009). The role of associative processes in spatial, temporal, and causal cognition. In S. Watanabe, L. Huber, A. P. Blaisdell, & A. Young (Eds.), *Rational animal, irrational human*. Tokyo, Japan: Keio University Press.

Blaisdell, A., Sawa, K., Leising, K., Waldmann, M. (2006). Causal reasoning in rats. *Science, 311*, 1020-22.

Block, N. (1981). Psychologism and behaviorism. *Philosophical Review, 90*, 5 - 43.

Boyd, R. (1999). Kinds, complexity and multiple realization. *Philosophical Studies, 95*, 67-98.

Bryant, P., & Trabasso, T. (1971). Transitive inferences and memory in young children. *Nature, 232*, 456-458.

Buckner, C. (2011). *Learning from mistakes: Error-correction and the nature of cognition.* (Doctoral dissertation). ProQuest/UMI. (Publication No. 3456443)

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Clayton, N., Emery, N., & Dickinson, A. (2006). The rationality of animal memory: The cognition of caching. In S. Hurley & M. Nudds (Eds.), *Rational animals?* (pp. 197–216). Oxford, UK: Oxford University Press.

Clayton, N., & Dickinson, R. (2009). Are animals stuck in time or are they chronesthetic creatures? *Topics in Cognitive Sciences, 1*, 59-71.

Columbo, M., & Broadbent, N. (2000). Is the avian hippocampus a functional homologue of the mammalian hippocampus? *Neuroscience and Biobehavioral Reviews, 24*, 465-484.

Craver, C. (2009). Mechanisms and natural kinds. *Philosophical Psychology, 22*, 575-594.

De Lillo, C., Floreano, D., & Antinucci, F. (2001). Transitive choices by a simple, fully connected, backpropagation neural network: Implications for the comparative study of transitive inference. *Animal Cognition, 4*, 61-68.

Dennett, D. (1983). Intentional systems and cognitive ethology — the 'Panglossian paradigm' defended. *Behavioral and Brain Sciences, 6*, 343-390.

Denniston, J., Svastano, H., Blaisdell, A., & Miller, R. (2001). The extended comparator hypothesis: Learning by contiguity, responding by relative strength. In R. Mowrer & S. Klein (Eds.), *Handbook of contemporary learning theories* (pp. 65-117). Mahway, NJ: Earlbaum.

Dusek, J., & Eichenbaum, H. (1997). The hippocampus and memory for orderly stimulus relations. *Proceedings of the National Academy of the Sciences, 94*, 7109-7114.

Elman, J. (1996). *Rethinking innateness: A connectionist perspective on development.* Cambridge, MA: MIT Press.

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review, 57*, 94-104.

Fodor, J. (1974). Special sciences, or the disunity of science as a working hypothesis. *Synthese, 28*, 97-115.

Fodor, J. (1997). Special sciences: Still autonomous after all these years. *Philosophical Perspectives, 11*, 149-163.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28*, 3-71.

Fountain, S., & Doyle, K. (this issue). Association and abstraction in sequential learning: 'What is learned?' revisited. *International Journal of Comparative Psychology.*

Frank, M., Rudy, J., & O'Reilly, R. (2003). Transitivity, flexibility, conjunctive representations and the hippocampus: II. A computational analysis. *Hippocampus, 13*, 341-354.

Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution*. New York: Basic Books.

Gigerenzer, G., & Regier, T. (1996). How do we tell an association from a rule? Comment on Sloman (1996). *Psychological Bulletin, 119*, 23-26.

Gluck, M., & Myers, C. (1994). A neurocomputational theory of hippocampal function in stimulus representation and learning. In S. Zournetzer, J. Davis, T. Mckenna, & C.

Lau (Eds.), *An introduction to neural and electronic networks*, 2^{nd} ed. (pp. 77-90). San Diego, CA: Academic Press.

Gluck, M., & Myers, C. (2001). *Gateway to memory: An introduction to neural network modeling of the hippocampus and learning*. Cambridge, MA: MIT Press.

Hirsh, R. (1974). The hippocampus and contextual retrieval of information from memory: A theory. *Behavioural Biology, 12*, 421-444.

Hull, C. (1943). *Principles of behavior*. New York: Appleton-Century.

Hume, D. (1748/1993). *An enquiry concerning human understanding*. Trans. E. Steinberg. Indianapolis, IN: Hackett Publishing.

Hurley, S. & Nudds, M. (2006). *Rational animals?* Oxford, UK: Oxford University Press.

Jones, T. (2004). Special sciences: Still a flawed argument after all these years. *Cognitive Science, 28*, 409-432.

Mackintosh, N. (2002). Do not ask whether they have a cognitive map, but how they find their way about. *Psicológica, 23*, 165-185.

McGonigle, B., & Chalmers, M. (1977). Are monkeys logical? *Nature, 267,* 694-696.

McGonigle, B., & Chalmers, M. (1992). Monkeys are rational. *Quarterly Journal of Experimental Psychology, 45*, 189-228.

McGonigle, B., & Chalmers, M. (2006). Ordering and executive functioning as a window on the evolution and development of cognitive systems. *International Journal of Comparative Psychology*, *19*, 241-267.

Menzel, E. W. (1973). Chimpanzee spatial memory organization. *Science, 822,* 943-945.

Menzel, E. W. (1978). Cognitive mapping in chimpanzees. In S. H. Hulse, H. F. Fowler, & W. K. Honig (Eds.), *Cognitive processes in animal behavior* (pp. 375-422). Hillsdale, NJ: Erlbaum.

Millikan, R. (1999). Historical kinds and the special sciences. *Philosophical Studies*, *95,* 45-65.

Mishkin, M., & Petri, H. L. (1984). Memories and habits: Some implications for the analyses of learning and retention. In L. R. Squire & N. Butters, (Eds.), *Neuropsychology of memory* (pp. 287-296). New York: The Guilford Press.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*, 183-246.

Morgan, C. L. (1894). *An introduction to comparative psychology*. London: Walter Scott.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing* (pp. 283-308). New York: Academic Press.

Newell, A & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the Association for Computing Machinery, 19(3),* 113–126.

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as cognitive map*. Oxford, UK: Clarendon Press.

Olton, D. (1979). Mazes, maps, and memory. *American Psychologist*, *34*, 583-596.

Papineau, D., & Heyes, C. (2006). Rational or associative? Imitation in Japanese quail. In S. Hurley & M. Nudds, (Eds.), *Rational animals?* (pp. 198-216). Oxford, UK: Oxford University Press.

Parker, W. (in press).Scientific models and adequacy for purpose. *The Modern Schoolman, 87*.

Penn, D., & Povinelli, D. (2007). Causal cognition in human and nonhuman animals: A comparative, critical review. *Annual Review of Psychology, 58*, 97-118.

Porphyry. (2000). *On abstinence from killing animals*. Trans. Gillian Clark. London: Duckworth Press.

Raby, C., Alexis, D., Dickinson, A., & Clayton, N. (2007). Empirical evaluation of mental time travel. *Behavioral Brain Sciences*. *30*, 330-331.

Radick, G. (2000). Morgan's canon, Garner's phonograph, and the evolutionary origins of language and reason. *The British Journal for the History of Science, 33,* 3-23.

Reiner, A., Perkel, D., Bruce, L., Butler, A., Csillag, A., Kuenzel, W. . . .Jarvis, E. D. (2004). Revised nomenclature for avian telencephalon and some related brainstem nuclei. *The Journal of Comparative Neurology, 473*, 377-414.

Rescorla, R. & Wagner, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. Black & W. Prokasy (Eds.), *Classical Conditioning II* (pp. 64-99). New York: Appleton-Century-Crofts.

Rumelhart, D., McClelland, J., & the PDP research group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume I*. Cambridge, MA: MIT Press.

Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*, 417-457.

Shanks, D., & Pearson, S. (1987). A production system model of causality judgment. *Proceedings of the Ninth Annual Conference of the Cognitive Science Society* (pp. 210-220). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Shettleworth, S. (2007). *Cognition, evolution, and behavior*, 1st ed. Oxford, UK: Oxford University Press.

Shettleworth, S. (2010). *Cognition, evolution, and behavior*, 2nd ed. Oxford, UK: Oxford University Press.

Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.

Smith, L. (2000). Avoiding association when it's behaviorism you really hate. In R. Golinkoff & K. Hirsh-Pasek (Eds.), *Breaking the word learning barrier* (pp. 169-174). Oxford, UK: Oxford University Press.

Sloman, P. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3-22.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, *11*, 1-23.

Sober, E. (2005). Comparative psychology meets evolutionary biology – Morgan's Canon and cladistic parsimony. In L. Daston & G. Mitman (Eds.), *Thinking with animals – New perspective on anthropomorphism* (pp. 85-99). New York: Columbia University Press.

Sorabji, R. (1993). *Animal minds and human morals*. New York: Cornell Press.

Squire, L. (1992). Memory and the hippocampus: A synthesis of findings with rats, monkeys, and humans. *Psychological Review, 99*, 195-231.

Suddendorf, T., & Corballis, M. C. (2007). New evidence of animal foresight. *Animal Behaviour*, *75*, e1-e3.

Tolman, E. (1948). Cognitive maps in rats and men. *The Psychological Review*, *55*, 189-208.

Tulving, E. (1983). *Elements of episodic memory*. Oxford, UK: Clarendon Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, *26*, 1-12.

Van Elzakker, M., O'Reilly, R., & Rudy, J. (2003). Transitivity, flexibility, conjunctive representations, and the hippocampus: I. An empirical analysis. *Hippocampus, 13(3)*, 334-340.

von Fersen, L., Wynne, C., Delius, J., & Staddon, J. (1991). Transitive inference formation in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *17*, 334-341.

Wagner, A. (2008). Evolution of an elemental theory of Pavlovian conditioning. *Learning and Behavior*, *36*, 253-265.

Wallace, D., & Fountain, S. (2002). What is learned in sequential learning? An associative model of reward magnitude serial-pattern learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *28*, 43-63.

Walker, J. & Olton, D. (1979). The role of response and reward in spatial memory. *Learning and Motivation, 10*, 73-84.

Wasserman, E., & Zentall, T. (2006). Comparative cognition: A natural science approach to the study of animal intelligence. In E. Wasserman & T. Zentall (Eds.), *Comparative cognition* (pp. 619-636). Oxford, UK: Oxford University Press.

Watson, J. B. (1928). *The ways of behaviorism*. New York: Harper.

Wynne, C. (2004). The perils of anthropomorphism. *Nature, 428*, 606.

Zentall, T. (2001). The case for a cognitive approach to animal learning and behavior. *Behavioural Processes, 54*, 65-78.