Dissociable learning processes in comparative psychology

J. David Smith¹ · Barbara A. Church²

Published online: 10 August 2017
© Psychonomic Society, Inc. 2017

Abstract Comparative and cognitive psychologists interpret performance in different ways. Animal researchers invoke a dominant construct of associative learning. Human researchers acknowledge humans' capacity for explicit-declarative cognition. This article offers a way to bridge a divide that defeats productive cross-talk. We show that animals often challenge the associative-learning construct, and that it does not work to try to stretch the associative-learning construct to encompass these performances. This approach thins and impoverishes that important construct. We describe an alternative approach that restrains the construct of associative learning by giving it a clear operational definition. We apply this approach in several comparative domains to show that different task variants change—in concert—the level of awareness, the declarative nature of knowledge, the dimensional breadth of knowledge, and the brain systems that organize learning. These changes reveal dissociable learning processes that a unitary associative construct cannot explain but a neural-systems framework can explain. These changes define the limit of associative learning and the threshold of explicit cognition. The neural-systems framework can broaden empirical horizons in comparative psychology. It can offer animal models of explicit cognition to cognitive researchers and neuroscientists. It can offer simple behavioral paradigms for exploring explicit cognition to developmental researchers. It can

enliven the synergy between human and animal research, promising a productive future for both.

Keywords Associative learning · Explicit cognition · Learning systems · Comparative psychology

Introduction

For decades, comparative and cognitive psychologists, studying animals and humans, have lived divided by a less-than-open border. Comparative psychologists have taken associative learning to be the dominant learning mechanism in animal minds. Associative learning—encompassing classical conditioning and operant (instrumental) learning—grounded the theories of Pavlov (1927) and Thorndike (1911). It describes the mechanisms by which reinforcers bind stimuli to responses. It justifies the application of Morgan's (1906) Canon, ensuring low-level psychological interpretations of animals' performances. It is the primary interpretative framework for most comparative psychologists.

In contrast, important cognitive theories assume that there is more than one kind of learning and memory that humans can engage (i.e., multiple-systems views, Ashby & Maddox, 2011; Atkinson & Shiffrin, 1968; Baddeley & Hitch, 1974; Cohen & Squire, 1980; Knowlton & Squire, 1993; Schacter, 1990; Tulving, 1985; and multiple-process views, Craik & Lockhart, 1972; Erickson & Kruschke, 1998; Jacoby, 1991; Moscovitch, 1992; Nosofsky, Palmeri, & McKinley, 1994; Richardson-Klavehn & Bjork, 1988; Roediger & Blaxton, 1987; Yonelinas, 2002). Indeed, often cognitive researchers focus on humans' explicit-declarative cognition. They no longer try to make associative learning the dominant explanatory framework, given strong evidence that this explanation is insufficient (e.g., Neisser, 1967). Though important aspects of

✉ J. David Smith
jsmith395@gsu.edu

¹ Department of Psychology, Georgia State University, Atlanta, GA 30303-5010, USA

² Language Research Center, Georgia State University, Atlanta, GA 30303-5010, USA

human learning fall outside the realm of explicit-declarative cognition (for reviews see Aslin & Newport, 2012; Cleeremans, Destrebecqz, & Boyer, 1998; Watanabe & Sasaki, 2015; Wills, 2004), humans clearly can transcend associative learning and reactive behavior. Human cognition is often conscious and declarative. It uses explicit cognitive processes supported by the utilities of executive attention and working memory. It produces abstract rules and plans to achieve goals, not just conditioned responses to stimuli. Humans can be queried about these aspects of cognition, and they answer!

Comparative psychologists would never describe animals' performances as cognitive psychologists describe humans' performances. The reverse statement is equally true. This divide is likely not good, for there are human–animal continuities in cognition that theory must encompass on the way to also interpreting the contrasts. For example, animals have shown continuities with humans in these domains among others: category learning (Smith, Zakrzewski, Johnson, Valleau, & Church, 2016), magnitude-quantity estimation (Cantlon, Platt, & Brannon, 2009), conceptual processing (Harlow, 1949; Fagot, Wasserman, & Young, 2001), memory (e.g., Basile & Hampton, 2011), metacognition (Zakrzewski, Johnson, & Smith, 2017), and theory of mind (Call & Tomasello, 2008). Therefore, the sharp interpretative break across human and animal psychology stifles meaningful cross-talk and cross-pollination across the fields of animal and human research. It creates difficulties for developing meaningful animal models of human cognitive processes. It creates road blocks to understanding the neural/neurochemical underpinnings of these cognitive processes. It raises many questions about the points at which the animal and human explanatory frameworks may intersect. Do animals sometimes transcend associative learning to show forms of explicit cognition? How can this be demonstrated clearly? How much of human learning and behavior is mediated by and/or built upon associative learning? What is the nature of the threshold between associative processes and explicit cognition? Are they competing processes or different levels of a hierarchy? What are the minimal conditions that cause cognitive systems to operate not associatively, but explicitly? What are the simplest paradigms for exploring transitions from associative to explicit processing? When do children cross the threshold to explicit cognition, and by which first baby steps? What are the changes in the neural systems that support performances on opposite sides of the threshold? Do parallel neural changes apply to human and animal minds? Are there ways to perfectly unplug the associative-learning system, so that one knows the organism is operating on an explicit cognitive level? Could educational or clinical practitioners use approaches like this to ensure that children's or patients' learning occurs on an explicit, conceptual cognitive plane? This article's perspective bears on all these questions. Though it

has an especially clear message to deliver to comparative psychology, these questions concern cognitive, developmental, educational, clinical, and comparative psychologists, and, of course, neuroscientists.

Our article has this structure. The next section, *Stretched and Straining Concepts in Animal-metacognition Research*, raises concerns about the associative-learning construct. To do so, it draws on the animal-metacognition literature, wherein a 20-year debate has surrounded the appropriate extension of the associative-learning construct. The literature includes determined efforts to show that associative processes sufficiently capture the field's phenomena. Pursuing explanatory sufficiency, though, the associative-learning construct is sometimes seemingly stretched beyond principle. This stretching thins the construct and costs it theoretical clarity. This can leave the construct psychologically empty, untestable, and unfalsifiable, creating a break with associative-learning theory's distinguished and disciplined historical tradition.

Accordingly, the section *Associative Learning: An Illustrative Example* clarifies the appropriate limits on the associative-learning construct. It reasserts the first principles of associative learning. It helps the construct live within its means. It sets responsible limits on the ideas of “stimulus” and “reinforcement.” Thus, this section defines a boundary for associative learning and a threshold for higher-level cognition. It does so using an operant-learning task that instantiates crucial characteristics of the associative construct as understood for 100 years. If one draws the boundary line as this example proposes, associative learning becomes a powerful and transparent construct that explains many aspects of both animal and human behavior and will be sustainable into future decades.

This boundary is then an exciting thing, *because there might be something on the other side*. The section *Learning Processes in Discrimination Learning: A Theoretical Analysis* considers this possibility. Now we stress test our illustrative associative task. We show that it is obvious when the associative-learning construct stops stretching and snaps instead. We show that a seemingly small change in method dramatically changes the learning processes recruited by a task. To gloss over these differences, by trying to fit them (uncomfortably) into the construct of associative learning, undersells the importance of these qualitative learning transitions. This approach establishes the minimal conditions by which one can observe participants crossing a threshold away from associative learning. It will support strong tests of whether animals can sometimes cross that threshold, performing at a higher cognitive level.

But associative learning comprises operant learning and classical conditioning. Accordingly, *Learning Processes in Classical Conditioning: A Theoretical Analysis* extends our approach into the latter domain. This analysis is particularly illuminating because the relevant phenomena have been

understood for a century, but their interpretation has persistently undersold their theoretical importance. We show that operant and classical conditioning have basic similarities that are accommodated within the same limited, principled construct of associative learning. We show that in both domains the same minimal conditions disable associative processes, creating a processing vacuum that explicit cognition—in some organisms—may fill. We are excited about the empirical progress and theoretical development that may follow if researchers explore whether animals share aspects of humans' explicit cognitive system. In *Converging Techniques*, we note that the ingenuity of comparative researchers will likely lead to new empirical approaches that could be applied in other disciplines.

Finally, in section *A Learning-Systems Approach toward Animal Metacognition*, we return to apply our approach to the animal-metacognition literature. It translates well again, informing the dominant associative debate in that area.

Stretched and straining concepts in animal-metacognition research

Here, we illustrate the problem of stretching the associative-learning construct and the need for a disciplined framework. Our field has struggled toward this framework for decades. The animal-metacognition literature considers whether animals share humans' capacity for metacognition (cognitive self-awareness)—that is, whether they can reflect on the frailty or robustness of their perception, memory, or state of knowing. This area has been reviewed (e.g., Kornell, 2009; Smith, 2009; Smith, Beran, & Couchman, 2012; Smith, Couchman, & Beran, 2012) and research continues (e.g., Basile, Schroeder, Brown, Templer, & Hampton, 2015; Call, 2010; Foote & Crystal, 2007; Fujita, 2009; Kornell, Son, & Terrace, 2007; Paukner, Anderson, & Fujita, 2006; Roberts et al., 2009; Smith, Coutinho, Church, & Beran, 2013; Suda-King, 2008; Sutton & Shettleworth, 2008; Templer & Hampton, 2012—these references sample a much larger literature).

In many of these animal-metacognition studies, animals are given a mix of easy and difficult/uncertain trials. And, beyond the primary discrimination responses, they are given an extra “uncertainty” response that lets them decline to complete any trials they choose. They use this response selectively to fend off difficult trials. In an illustrative case, a dolphin (*Tursiops truncatus*) discriminated low tones (1200–2099 Hz) from high tones (always 2100 Hz). Figure 1 shows him at his true psychophysical threshold near 2085 Hz where the low and high curves cross. He responded uncertain to decline those difficult trials (Smith et al., 1995).

What was this animal thinking, or reacting to? Was he self-reflecting on mental states like doubt? This conclusion would take him beyond associative learning, across the threshold to

Fig. 1 A dolphin's auditory-discrimination performance in Smith et al. (1995). The horizontal axis indicates the frequency (Hz) of the trial. The High response was correct for tones at 2100 Hz. All lower-pitched tones deserved the Low response. The solid line represents the percentage of trials receiving the uncertainty response at each pitch level. The percentages of trials ending with the Low response (dashed line) or High response (dotted line) are also shown

explicit cognition. Had he grown averse to the concrete stimuli that instantiated hard trials, increasing errors, reducing rewards? This conclusion would hold the associative-learning construct safe, secure and sufficient. And so a 20-year associative-explicit debate ensued (e.g., Basile & Hampton, 2014; Carruthers, 2008; Hampton, 2009; Jozefowicz, Staddon, & Cerutti, 2009; Le Pelley, 2012, 2014; Smith, 2009; Smith, Beran, Couchman, & Coutinho, 2008; Smith, Beran, Couchman, Coutinho, & Boomer, 2009; Smith, Beran, et al. 2012; Smith, Couchman, et al. 2012; Smith, Couchman, & Beran, 2014; Staddon, Jozefowicz, & Cerutti, 2007).

But the associative construct faces challenges in this area, owing to careful experimentation by many scientists. The organizing challenge is that animals' performances keep forcing a drift away from the idea that associative stimuli trigger reactive responding and toward the idea that higher-level cognitive evaluations guide deliberate uncertainty responses.

First, note that the dolphin responded uncertain at his true threshold. The pitches to which he responded Uncertain and High were about 15 Hz—one-ninth of a diatonic halfstep—apart. It was B^{#####} vs. C. No creature (not even Yo Yo Ma) could hear out a concrete aversive stimulus at 2085 Hz. This is why behavioral analysts have concluded that the threshold state radically changes the rules of associative behavior and stimulus control (Boneau & Cole, 1967; Commons, Nevin, & Davison, 1991; Davison, McCarthy, & Jensen, 1985; Miller, Saunders, & Bourland, 1980; Terman & Terman, 1972). So is responding to a threshold state still associative responding, or have we already taken a step toward higher-level cognitive processing?

A related challenge is that a threshold stimulus—by definition—is terribly indeterminate. Because it equally suggests either of two responses (e.g., Low, High), it recommends neither. Shiffrin and Schneider (1977) thought that this indeterminacy (which they called inconsistent mapping) must cause automatic associative processes to fail, to be replaced by controlled / deliberate cognitive processes. But are there forms of

associative learning that are controlled and deliberate? Is this a contradiction in terms? A disciplined associative-learning construct would answer these questions. Providing that construct is our goal. Later sections will show how human cognitive neuroscience informs that construct. In turn, if animals show forms of explicit metacognition in these tasks, they could contribute valuable animal models to illuminate a fundamental human capacity and its neural/neuro-chemical underpinnings.

A third challenge to the associative-learning construct is that animals can report uncertainty about higher-level cognitive judgments. In Shields, Smith, & Washburn (1997), for example, macaques responded uncertain when they were unable to judge the status of an abstract Same–Different relation. Look at that sentence—there is no concrete stimulus that animals could react to associatively. So, does relational-judgment uncertainty transcend associative responding, crossing a threshold to explicit cognition? Or can the stimulus DIFFERENTNESS (e.g.,) trigger associative responding like the stimulus RED does (e.g., Debert, Matos, & McIlvane, 2007)? Our field has long needed a principled framework within which to explore these questions, so we introduce it here.

Alternatively, one might suggest that the stimulus DIFFICULTY triggers associative responding like the stimulus RED does. But now one sees that the thread to a disciplined construct of associative learning frays badly. For what sort of stimulus is DIFFICULTY, and how is it judged by the animal without metacognition?

The associative-learning construct is challenged further when the “stimulus” to be judged is only in the animal’s memory. In Hampton’s (2001) study, monkeys declined tests of their memory when they monitored that their memories were faint. There was no concrete stimulus that animals could react to associatively. Instead, the monkeys needed to reflect on whether—in their mind—they remembered the sample. Or they needed to initiate a deliberate search of possibly relevant locations in memory. Is a monitored, faint memory an associative stimulus? Is uncertainty reported following a memory search reactive responding? The need for a grounded, guiding theoretical framework emerges again.

Likewise, in research by Call and his colleagues (e.g., Call & Carpenter, 2001; Call, 2010), animals got to choose one of several hollow tubes that might hold food. On different trials, they had seen or not seen the food hidden. Animals spontaneously made information-seeking responses selectively on unseen trials by visually inspecting inside the tubes before choosing. Here, the animal’s associative-cue situation is identical in seen and unseen cases. Only by deliberately evaluating self-knowledge can the animal know when an information-seeking response is warranted. An associative-learning construct fits this deliberate self-evaluation poorly. An explicit-metacognition construct does so more naturally.

Thus, the associative-learning construct faces the problem of where the concepts of stimulus control and associative responding end. Many things can control humans’ and animals’ behavior. Stimuli like RED, threshold states like B#####, relations like Same, trial difficulty, faint short-term memories, episodic memories, pensive reflections on Roads Not Taken, New Year’s Resolutions, hopes for a secure retirement, beliefs in an afterlife, and on and on. Not all these are stimuli, not all exert stimulus control in the same way, not all are associative cues in the same way. So comparative psychology, as illustrated in our research area, has stood on an icy slope, wondering where the edge of associative learning is, and sliding farther down toward a less principled construct, having no map that gives the boundary. This map must be provided. We must draw disciplined lines that say: this is where stimulus control stops; this is where associative learning is no longer possible; this is where something else starts. In the next section, we will draw those lines, and then come to what the something else is.

This specification has implications for understanding human psychology, too. It would let us specify the associative in human performance, to acknowledge it, control it, transcend it. It would differentiate in a principled way the associative and the explicit in cognition. It would provide, as we will try to show, simple dissociative methodologies for separating these levels. It would let us map children’s journey across the threshold to explicit cognition. It would let us teach to different strengths in different special populations of children, and communicate on different levels to different clinical clients. It would ground these approaches in human neuroscience, while benefitting synergistically from a crucial set of animal models. Pursuing the disciplined construct of associative learning is constructive on many fronts, and it could benefit empirical progress and theoretical development in many areas of psychology and neuroscience.

Associative learning: an illustrative example

Figure 2a shows the structure of a basic discrimination-learning task. We will use this task as the raw materials with which to build a disciplined conception of associative learning. We will describe the task’s structure, its learning characteristics, its neuroscience basis. We will summarize why tasks like this instantiate a strong and sustainable associative-learning construct. (In a later section, we will provide a parallel description of classical conditioning, the other principal component of associative learning.)

In the task of Fig. 2a, the stimulus ellipses would be two perceptual categories, Category A and B. Within the ellipses, each symbol would correspond to a single perceptual stimulus that might be shown to participants, defined along two perceptual dimensions such as box size and box pixel density

Fig. 2 **a** The category structure for a hypothetical discrimination-learning task. The two stimulus ellipses comprise members of Category A (Red) and Category B (black). Each symbol in each category defines one two-dimensional category exemplar that might be presented to participants. **b** An illustrative task screen showing one such category

exemplar (top) and two alternative responses (A or B). In this task, the exemplars were rectangles that varied in size and internal pixel density (varying as shown in **a** along the horizontal and vertical axes, respectively)

(Fig. 2b). These would be the stimuli presented to participants. Exemplars from Category A and B would require the use of Category A and B responses. These would be the responses to be learned. Correct responses would be immediately rewarded, providing the reinforcement binding force by which stimulus–response pairs could strengthen their connections. Participants would learn these connections through repeated trials during contingency training. They would never see the whole stimulus array displayed as in Fig. 2, but just one stimulus at a time in a long series, each seen, responded to, and reinforced. This kind of task is known as an information integration (II) task in the literature on categorization and the neuroscience of category learning.

Figure 3a shows a human learning curve in this II task. It is increasing, curvilinear, negatively accelerating, with an asymptote. It has the familiar shape of instrumental and classical learning curves. Many participants employ the same basic associative processes in traditional associative tasks and in this category task. Animals' learning curves are like those of humans (Fig. 3b). Animals and humans seem to employ the same basic learning processes in this task. That is, humans often do not bring to this task their insight, their utilities for hypothesis testing and evaluation, their explicit strategizing, and so forth.¹ We have brought them down to a basic learning level that animals share. Confirming this, humans seem to lack conscious access to how they perform category tasks of this kind. They cannot declare the content of their category

knowledge to others. Their learning belongs to the sphere of implicit-procedural learning, akin to skill learning, habit learning, discrimination and instrumental learning. The learning mechanism that predominates in this task is basic enough that it does not reach up to contact working memory, consciousness, and declarative cognition.

Another strength of this task is that a lot is known about the brain systems that govern this kind of learning, even down to the synaptic level. This learning is probably actuated by processes linked to the basal ganglia. The basal ganglia have been proposed to underlie humans' skill, habit, and procedural learning (e.g., Mishkin, Malamut, & Bachevalier, 1984), and performance in instrumental-learning, perceptual-categorization, and some discrimination-learning tasks (Ashby & Ennis, 2006; Barnes, Kubota, Hu, Jin, & Graybiel, 2005; Divac, Rosvold, & Szwarcbart, 1967; Filoteo, Maddox, Salmon, & Song, 2005; Knowlton, Mangels, & Squire, 1996; Konorski, 1967; McDonald & White, 1993, 1994; Nomura et al., 2007; O'Doherty et al., 2004; Packard, Hirsh, & White, 1989; Pacard & McGaugh, 1992; Seger & Cincotta, 2005; Waldschmidt & Ashby, 2011; Yin, Ostlund, Knowlton, & Balleine, 2005). Categorization, discrimination, and other forms of associative learning are ancient, essential adaptations—it is natural that crucial learning systems might lie in phylogenetically older parts of the brain like the basal ganglia.

The basal ganglia are *necessary* for animals' reinforcement-based discrimination learning. In nonhuman primates, extrastriate visual cortex projects directly to the tail of the caudate nucleus—with massive convergence of visual cells onto caudate cells that project onward to premotor cortex (Alexander, DeLong, & Strick, 1986). The caudate is well placed to associate percepts through to actions, perhaps its primary role (Rolls, 1994; Wickens, 1993). Lesions of the tail of the caudate impair the learning of discriminations that require different responses to different stimuli (McDonald & White, 1993, 1994; Packard et al., 1989; Packard & McGaugh, 1992). These lesions may impair the formation of the stimulus–response associations that

¹ This description of human II learning and its neural basis should be qualified as follows. The description is apt when the task recruits in the learner the intended associative-learning system. However, humans sometimes insistently expect a task to have a clear cut classificatory rule. When that expectancy prevails, they can override the task's associative nature and impose upon it an adventitious rule structure that does not suit the task and that produces poor performance. When the II task is hijacked in this way, the description provided in this section will not be correct. Probably no task is ever perfectly process pure. Humans' occasional misbehavior in the II task is an interesting phenomenon of sharp relevance to the theoretical context of this article, because it shows their parallel access to associative-learning processes and explicit-rule processes.

Fig. 3 **a, b** Illustrative learning curves showing the performance of humans over 600 trials (**a**) and monkeys over 6000 trials (**b**) in a discrimination-learning task like that schematized in Fig. 2

mediate successful responding. For instance, Ashby, Alfonso-Reese, Turken, & Waldron's (1998) model assumes that caudate cells link visual-cortical cells to motor programs: therefore, lesioning this area would prevent the association of visual stimuli to motor responses.

The basal ganglia are largely *sufficient* for animals' discrimination learning. When paths out of visual cortex except to the caudate are lesioned (e.g., pathways to prefrontal cortex, hippocampus, and amygdala—Eacott & Gaffan, 1992; Gaffan & Eacott, 1995; Gaffan & Harrison, 1987) discrimination learning stays intact.

Discrimination-learning tasks like that shown in Fig. 2 rely on primary reinforcement, and the basal-ganglia system provides a mechanism by which reinforcement plays its role. Rewards cause dopamine release into the tail of the caudate (Hollerman & Schultz, 1998; Schultz, 1992; Wickens, 1993). The dopamine signal can strengthen recently active synapses that were plausibly participatory in reward (Arbuthnott, Ingham, & Wickens, 2000; Calabresi, Pisani, Centonze, & Bernardi, 1996). Consequently, the timing of the reward signal is critical to the strengthening of associations. The caudate's medium spiny cells can provide a brief synaptic memory, because the morphology of their dendritic spines allows a trace of recent activity to last briefly after response (Gamble & Koch, 1987; MacDermott, Mayer, Westbrook, Smith, & Barker, 1986). The spines can remain depolarized during this window, and the reinforcement signal can selectively strengthen appropriate synaptic connections to produce learning.

For many reasons, tasks like that shown in Fig. 2 exemplify the preferred construct of associative learning in comparative psychology today. These tasks display all aspects of that construct—for example, the concrete stimulus inputs, the simple behavioral outputs, and their joining (association) through learning. In caudate-mediated discrimination learning, whole stimulus representations (the caudate's direct inputs) are linked to adaptive responses (its indirect outputs). In this joining, reinforcement plays its ideal role as the binding agent, updating / strengthening stimulus–response associations.

More than in any other operational definition of associative learning that we have encountered, the present definition takes the idea of stimulus–response bonds literally.

These tasks include sustained contingency training that is the hallmark of associative-learning tasks. They produce learning curves—typical of associative performances in classical and instrumental domains—that reflect gradual learning and gradual association strengthening toward asymptote (Fig. 3). Fitting Morgan's Canon and the reductionistic imperative in comparative psychology, these tasks do not benefit from hypothesis testing or explicit rules. Learning can and does stay implicit and procedural, seemingly inaccessible to consciousness and awareness and removed from the reportable sphere of declarative cognition.

Our definition has value added because it incorporates a detailed neuroscience of associative learning, including the neural basis of stimulus registration, motor outputs, associative connections, and the reinforcement signal. Through incorporating these neuroscience elements, one modernizes the construct of associative learning and builds a bridge between neural and behavioral levels of analysis. By incorporating the dopamine reinforcement system as an important component of associative learning, our definition provides one of the clearest operational definitions of reinforcement ever incorporated into a construct of associative learning. By specifying the mechanism through which associations are strengthened (i.e., synaptic improvement), the definition gains additional operational precision.

The components of the neural learning system described in this section co-occur stably, making this a coherent, integrated system of learning to which comparative psychologists should attend sharply. Indeed, this is probably an evolutionarily old system of learning by which vertebrates accomplished learning and behavioral regulation. We will see next that these components also co-vanish in concert given different conditions. Even in their co-vanishing, they confirm themselves as constituting a coherent and integrated system of learning.

This coherent brain system for learning probably underlies humans' and animals' performance in many instrumental, categorization, and discrimination tasks, along with tasks of procedural learning (skills, habits, etc.). It is a dominant neural system for learning in humans and animals. It must be acknowledged as such and given a prominent construct label. We believe the natural construct label is: associative learning. By this we mean that this brain system ideally represents the associative-learning construct. We do not mean that it exhausts that construct (see section *Learning Processes in Classical Conditioning: A Theoretical Analysis*). However, even with only this system assimilated to the construct of associative learning, that construct is already broad, central to comparative psychology, with great explanatory power. We do not believe this assimilation diminishes that construct at all.

Thus, we propose that the previous paragraphs serve as an important element of a modern-day definition of associative learning. They crystallize many crucial aspects of that construct. They ground these aspects in neuroscience.

An equally important aspect of this definition is that it imposes theoretical restraint and discipline on the construct of associative learning, which is sometimes subject to inflationary pressures in its application. It establishes boundaries on what the components of associative learning reasonably can be. It grants the potential that research may explore beyond those boundaries, to consider what alternative learning processes animals may also use in some circumstances. We believe this bounded definition brings sharp clarity to a dominant theoretical construct, strengthening it for the long haul for all the substantive explanatory work it does. But, by also delimiting that construct, that definition opens up the field of comparative psychology to important new theoretical directions.

Learning processes in discrimination learning: a theoretical analysis

Now we illustrate the theoretical gain from a clearly bounded definition of associative learning. We do so by considering a variation on the task in the preceding section. We assume now that there is a temporal delay between the stimulus–response pairing and the feedback/reinforcement. This is a subtle change that comparative psychologists could consider just a methodological tweak. What should be the effect of reinforcement delay?

The previous section showed that associative processing in discrimination-learning tasks depends on a time-locked cascade: stimulus–response–reward. The timing of the reward signal is crucial. Reinforcement can only improve synaptic connections—that is, strengthen associations—in the brief window during which synapses preserve the relevant

stimulus–response activation pattern. One sees that this reinforcement-learning system could be disabled, unplugged, and associative learning severely impaired, by disrupting the time-locked cascade. What would we see then?

If associative learning were eliminated by delay, and no other learning process steps in, then the acquisition of the discrimination should fail. But, if some other learning system steps up instead, there might still be some successful acquisition. Ideally, there might be observable indications of the disabling of associative processes and the substitution of other processes.

Exploring the disabling of associative learning within discrimination tasks, Maddox, Ashby, & Bohil (2003) and Maddox & Ing (2005) gave participants the task discussed in the previous section under conditions of immediate and delayed reinforcement. Both studies found learning impaired by delay.

In a striking convergence, Yagishita et al. (2014) visualized, at the synaptic level, the operation of the dopamine reinforcement signal, including during reinforcement delay. They used optogenetic methods to stimulate sensori-motor inputs and dopaminergic inputs separately, gaining precise control over the temporal asynchrony between stimulus presentation and delivery of the reward signal. Dopamine failed to promote strengthened synapses if delayed beyond 2.0 s. Remarkably, these authors imaged dendritic spine improvement but only saw it given immediate reinforcement. That is, they showed that under reinforcement delay the reinforcement signal was no longer able to play its critical function of strengthening the active synapses that had likely produced the reward. The likely reason for this is that the system had returned to baseline during the delay and so the dopamine signal had no differential pattern of synaptic activation on which to operate.

The waning effectiveness of dopamine with delay that Yagishita et al. (2014) found mirrors the waning category knowledge of humans learning with reinforcement delays in Maddox et al. (2003) and Maddox and Ing (2005). This convergence across behavioral and neuroscience levels confirms that reinforcement delay is not just a subtle variation in method. To the contrary, it may disable one of the brain's primary reinforcement-learning systems.

What do participants facing this disabling do? Is there another learning process to be swapped in when this type of associative learning is shut down? Exploring this substitution, Maddox et al. (2003) and Maddox and Ing (2005) fit formal models to determine the placement of participants' decision boundaries as they perform these tasks. The models indicated that participants under reinforcement delay turned to explicit unidimensional rule strategies. In effect, they imposed vertical or horizontal decision boundaries onto the II stimulus space shown in Fig. 2a. One can see that such a category rule would support only poor accuracy levels. Maddox et al. (2003) found that the proportion of rule strategies increased by 200% under

delayed reinforcement. Maddox and Ing (2005) found that the proportion of rule strategies under delay increased by 250%.

This learning transition occurs though the rule strategy is distinctly nonoptimal. Participants do not adopt rule strategies to achieve any benefit. They adopt them because associative learning has been disabled, and now it is a question of any port in a storm. That we can reason so clearly about this transition stems from defining associative learning so transparently.

Yet this transition seems natural from a learning-systems perspective based in neuroscience. Humans swap in the learning process they can still use, which is to entertain hypotheses about what might be going on in the task, and then evaluate these rules against the trials as they unfold. These hypotheses/rules are naturally of low dimensionality—in essence, they are often unidimensional rules. The use of this rule-based process is still possible under reinforcement delay, because the working rule (e.g., the big ones are Category B) can be maintained in working memory through the delay until delayed reinforcement arrives. Associative learning is dependent on immediate reinforcement within strict time limits. Explicit rule learning is not.

Systems neuroscience would explain the transition under reinforcement delay in this way. Participants have switched over now to use their explicit-declarative cognitive system. This system includes the prefrontal cortex, the anterior cingulate gyrus, and the hippocampus. Explicit rules might reverberate in the form of working-memory loops between prefrontal cortex and thalamus (Alexander et al., 1986). The anterior cingulate might choose provisional hypotheses to seed into working memory. There could be a capability to switch between dimensional hypotheses. Research shows that patients with frontal dysfunction are impaired in tasks that have verbalizable rule solutions (Brown & Marsden, 1988; Cools, van den Bercken, Horstink, van Spaendonck, & Berger, 1984; Kolb & Whishaw, 1990; Robinson, Heaton, Lehman, & Stilson, 1980). Converging fMRI research shows that participants performing this kind of category task activated the right dorsal-lateral prefrontal cortex and the anterior cingulate (Rao et al., 1997). Other studies have viewed these areas as components of working memory and executive attention (Fuster, 1989; Goldman-Rakic, 1987; Posner & Petersen, 1990), both of which could support rule formation in explicit categorization. Other imaging results have suggested the anterior cingulate as a rule-generation site (Elliott & Dolan, 1998).

On looking closely at participants' learning under reinforcement delay, we would also find that their rules are attentionally narrow (unidimensional), resident in working memory, conscious, verbalizable, and declarable to others. Thanks to extensive work by Ashby, Maddox, and their colleagues, there are now other dissociative symptoms that differentiate associative learning under immediate reinforcement from this form of explicit-declarative learning (e.g., Maddox & Ashby, 2004). Reinforcement delay is only one possible approach.

It is the bounded definition of associative learning that lets us predict its collapse under delay, interpret that failure when it occurs, and understand by contrast the explicit-declarative learning processes that step in instead.

That definition can bring comparative psychology many benefits. To qualitatively shut down one of the brain's primary reinforcement systems—to disable a dominant component of associative learning—is an extraordinary empirical tool. To our knowledge, comparative psychologists have never reckoned with this possibility. We can use this technique to ask whether some animals, like humans, can transition to different learning processes. We can ask which animals. We can contrast the alternative learning processes that humans and nonhumans recruit. We can ask about the affordances of language in facilitating these alternative learning processes. These studies can provide a close look by neuroscientists at an elemental form of explicit cognition and they can provide animal models of it.

But here is the point that we think comparative psychologists will find most useful once it is incorporated into their theoretical perspective. Delayed reinforcement fundamentally changed learning. Category knowledge narrowed to a single dimension, so that performance became guided by a dimensional category rule that partitioned the categories using a vertical or horizontal decision boundary. Category knowledge entered awareness. It became verbal. It became declarative. Its brain locus was transformed. The intervention was simple. The effects on learning dramatic. The two kinds of learning are opposites along every information-processing continuum. It would not be responsible science to include performance under immediate and delayed reinforcement as instances of the same learning process or as belonging under any unitary construct label, no matter what that learning process was called.

In particular, if one tried to call both kinds of learning associative, that construct label would lose all meaning. Then we would not be studying learning, or cognition, or anything else. For to study those capacities means to characterize things, to differentiate things, to draw meaningful distinctions. But here, we would be claiming that even the most striking and qualitative differentiation is still no differentiation to us. No matter how different the experience, the behavior, or its neural instantiation, we would be expressing our willingness just to blur everything together and ignore differences to preserve our assumption that all learning is the same. It is important that we all reflect on this theoretical point.

This point has implications for human cognitive researchers, too. Comparative psychology has had a strong unitarian impulse in its theoretical determination to explain animals' performances parsimoniously by relying on the construct of associative learning. But cognitive psychologists have often had an analogous unitarian impulse, as when theorists doggedly pursued unitary-code theory in the imagery

literature, or theorists strongly defended unitary exemplar theory in the categorization literature, or theorists sharply doubted the idea of multiple, dissociable systems or processes in the memory literature. The hope for parsimony runs deep in cognitive science. We hope this article is generally useful as a case study in drawing the boundaries that may delineate the processing systems of mind. Indeed, this article—bringing to bear the converging cognitive and neuroscientific dimensions of mental representation, awareness level, brain locus, and reinforcement mechanism—contains one of the most transparent exercises in making this delineation, an exercise that is also instructive because it reaches a very clear result.

Learning processes in classical conditioning: a theoretical analysis

However, the associative-learning construct extends beyond discrimination tasks as considered in the preceding two sections. The primary reinforcement systems in the brain extend beyond the dopaminergic functions described there, and brain regions other than the basal ganglia are essential in various classical-conditioning tasks (see below). Thus, the burden falls on us to extend our theoretical analysis, to ask whether, in classical conditioning, too, changing the timing and sequencing of stimuli and reinforcers produces profound changes in the character of learning.

The classical-conditioning situation has several components: an initially neutral concrete stimulus like a tone (the conditioned stimulus, CS), a prepotent, biologically relevant reinforcer like a corneal air puff or lingual meat powder (the unconditioned stimulus, UCS). These prepotent reinforcers spontaneously—with no training—elicit behavioral reactions (blinking, salivating, unconditioned responses—UCR). Then, through repeated CS-UCS (UCR) pairings, the neutral stimulus is granted the eliciting power to produce a response (the conditioned response—CR) that is like the UCR.

The neural mechanisms of classical conditioning are diverse and different from those in operant learning. For learning conditioned behaviors, the cerebellum and related structures are often crucial (e.g., Thompson, 1990). For learning conditioned emotions, the amygdala and related structures are often crucial (e.g., Kim & Jung, 2006). Either way, it is not certain that our theoretical analysis would extend to this different form of associative learning.

Yet classical conditioning has many similarities to operant learning, which is why these two types of learning have jointly defined associative learning for 100 years. The classical task also features neutral, concrete stimuli that gain the power to elicit behaviors. This power is conferred by the immediate, binding presence of a reinforcer. In the classical task, as in the operant task, learning is supported by a careful sequencing of stimuli and reinforcers. In both tasks, both predictive

contingency and temporal contiguity are generally crucial. Operant and classical tasks very often depend on trial repetition that fosters gradual learning, not sudden insights or rule discoveries. Classical learning curves are curvilinear and negatively accelerated (Rescorla & Wagner, 1972), as already shown for an operant task (Fig. 3). In both cases, this shape may reflect the principle that reinforcement's incremental effect on a trial is given by the prediction error in the system that the reinforcer produces (e.g., Aguado, 2003).

Those shared properties compose a strong family resemblance. Accordingly, given the strong learning transition we demonstrated in the preceding sections, we sought converging evidence for a similar learning transition in the other domain of associative learning—classical conditioning. We asked whether this form of learning, too, might have the same boundary conditions, the same breaking point, past which a threshold was crossed that required the intervention of different learning systems and different neural structures. We describe this transition now.

Reinforcement delay provides again one means to test for a transition. For example, one can introduce time intervals between the termination of the CS (e.g., the tone in an eye-blink experiment) and the onset of the reinforcing UCS (e.g., the air puff that brings the blink). One procedure that contains this temporal gap—this reinforcement delay—is called trace conditioning (only the trace of the CS remains when the reinforcing UCS arrives). These delays impair conditioning. This has been known since Pavlov (1927). Even small time intervals impair learning.

Reinforcement delay also alters the neural systems that are recruited in learning (Han et al., 2003; Kryukov, 2012; Raybuck & Lattal, 2014), consistent with findings summarized in *Learning Processes in Discrimination Learning: A Theoretical Analysis*. In particular, trace conditioning, unlike its non-delay counterpart, uniquely requires the hippocampus and the prefrontal cortex (Kim, Clark, & Thompson, 1995; Kronforst-Collins & Disterhoft, 1998; Moyer, Deyo, & Disterhoft, 1990; Powell, Skaggs, Churchwell, & McLaughlin, 2001; Solomon, Vander Schaaf, Thompson, & Weisz, 1986; Weible, McEchron, & Disterhoft, 2000; Weiss, Bouwmeester, Power, & Disterhoft, 1999).

Learning in trace conditioning is also allied to the processes of declarative memory. In patients with amnesia who present with hippocampal damage, trace conditioning is severely disrupted at a 1.0 s trace interval (McGlinchey-Berroth, Carrillo, Gabrieli, Brawn, & Disterhoft, 1997). The hippocampus is an important structure in the formation of declarative memories. Learning in trace conditioning also shows a consolidation pattern like that seen in declarative memory (Kim et al., 1995; Squire, Clark, & Knowlton, 2001).

At least in humans, learning in trace conditioning is bound up with the participant's explicit awareness of the content of the learning, another reflection of explicit-declarative

cognitive processes. In a trace-conditioning procedure, this knowledge would be of the precedence of the CS in the trial, its signaling role, and its predicting contingency. Intriguing studies have documented the awareness relationship as described now.

In a study of humans' differential trace conditioning that included assessments of awareness, only participants who became aware of the task's contingencies successfully conditioned. Moreover, four amnesic participants with hippocampal damage failed to become aware of the CS-UCS contingency and failed to condition. In contrast, in closely related non-delay procedures, nonaware participants and amnesics both successfully conditioned (Clark & Squire, 1998, 1999; Manns, Clark, & Squire, 2002). (Note to readers: in differential trace conditioning, there are two auditory signals. One CS signal—for example, the tone—signals the future UCS as just described. Another signal—for example, a burst of noise—signals the absence of the UCS on that trial).

With older participants, who do not gain awareness of the CS-UCS contingency so easily, another study manipulated awareness deliberately by explaining the stimulus contingencies before the conditioning trials. Now awareness scores improved post-experiment, supporting more successful trace conditioning (Clark & Squire, 1999).

Other participants were given a distractor task during the trace-conditioning procedure. These participants failed to condition, and failed to bring the contingencies into conscious awareness. By contrast, participants given this same distraction task during a comparable non-delay procedure conditioned as well as participants who were not distracted (Clark & Squire, 1999).

Similar findings exist for the simplest delay procedure—single-cue trace conditioning in which there is only an affirmative CS signaling the UCS. Participants in middle age participated in a trace-conditioning procedure while either watching a silent movie or performing an attentionally demanding concurrent digit-monitoring task. Movie watchers gained more awareness of the conditioning contingency and conditioned more strongly. Moreover, awareness of the contingency early in the experiment predicted the strength of ultimate conditioning (Manns, Clark, & Squire, 2000; Woodruff-Pak, 1999). These awareness findings were not found for closely related non-delay procedures (i.e., a delay-conditioning procedure in which the CS and UCS overlap in time—e.g., Frcka, Beyts, Levey, & Martin, 1983; Grant, 1973; Hilgard & Humphreys, 1938; Manns, Clark, & Squire, 2001; Papka, Ivry, & Woodruff-Pak, 1997; Weiskrantz & Warrington, 1979).

Clark, Manns, & Squire (2002) concluded that trace conditioning—compared to non-delay procedures—is uniquely dependent on higher-level cognitive processes and uniquely related to awareness and declarative knowledge of the contingencies. This conclusion is similar to that in the section on

Learning Processes in Discrimination Learning: A Theoretical Analysis about explicit-declarative discrimination learning. Reinforcement delay takes cognitive processing toward the conscious, explicit, declarative pole of cognitive functioning. The change is qualitative and transforming. Reinforcement immediacy takes cognitive processing toward the unconscious, implicit, procedural, associative pole of behavioral functioning. The label of associative learning fits this latter case beautifully, as it has done for 100 years in comparative psychology. Crucially, both aspects of the construct of associative learning—instrumental learning and classical conditioning—reveal the same boundary conditions and the same qualitative transition to forms of learning that are more explicit, declarative, and, possibly, conscious.

There is extensive evidence that these conclusions apply also to nonhuman species (Kim et al., 1995; Kronforst-Collins & Disterhoft, 1998; Moyer et al., 1990; Powell et al., 2001; Solomon et al., 1986; Weible et al., 2000; Weiss et al., 1999). That is, the neural systems involved in different conditioning procedures are probably largely shared across vertebrate species. This suggests the possibility that nonhumans may pass through awareness transitions of their own. This recommends further research to understand when humans gain awareness during conditioning and when not. By replicating these conditions with animals, we might manipulate their states of awareness, too. In this way, even some conditioning results could become interpretable according to our knowledge about higher-level cognitive functions related to awareness. This is similar to the hope we expressed in *Learning Processes in Discrimination Learning: A Theoretical Analysis*, and it is an exciting possibility. It is a distinctive theoretical benefit of our perspective that, on giving associative learning its proper bounded definition, suddenly we become positioned to see explicit, and even possibly aware, modes of cognition in animals.

In the end, this analysis shows that even the tight-knit family of classical-conditioning procedures must be fractioned into those that instantiate different processes of learning. This fractioning is essential if we are to group together experimental procedures that are similar psychologically, while distancing these from others that are very different psychologically. This fractionation shows that there is no unitary view of classical conditioning that can turn out to be useful and comprehensive. There cannot be, because the delay and non-delay procedures require such different psychological and neuropsychological descriptions. The preceding section reached the identical conclusion about delayed and nondelayed discrimination tasks. But the conditioning literature—because it has for so long grounded the hope for a unitary associative-learning construct—provides an especially sobering reminder that the construct of associative learning must be allowed to break constructively along its natural, psychological, fracture planes.

Converging techniques

To support further work in this area, in this section we broaden our methodological perspective. This section introduces other methods by which one might shut down associative learning and require humans and animals to transition toward explicit-declarative cognitive processes. Given our colleagues' ingenuity in asking animals difficult questions, a comprehensive study of these issues could be forthcoming using additional converging measures. We will illustrate two additional methods now.

Smith, Boomer, et al. (2014) showed that a reinforcement regimen called deferred reinforcement could also qualitatively disable associative learning in tasks like those described in our discussion of discrimination learning. Human adults completed trial blocks without reinforcement. At the end of a block, they received the reinforcements from all correct trials clustered and then the timeouts from all errors clustered. Reinforcement was delayed temporally *and* scrambled out of trial-by-trial order. Associative learning was doubly disrupted. Humans could not know which stimulus–response pairs had been completed correctly or which stimulus–response bonds to strengthen. Reinforcement was also delayed beyond the useable temporal window that we have discussed.

Figure 4a shows the result from participants under immediate reinforcement. Formal modeling showed that they partitioned the two sets of exemplars according to the appropriate diagonal decisional boundary (see stimulus distributions in Fig. 2). Each line in the figure shows one participant's apparent decision boundary, modeled based on the stimulus–response associations they revealed to us through their performance.

To be clear, participants did not learn the diagonal boundary. That is not how learning occurs or performance unfolds in

this task. They learned stimulus–response associations for many of the stimuli in the two category distributions, and the result of these learned associations was that their behavior could be modeled using that boundary. In a discussion of associative learning, this distinction between boundary learning and associative learning that can be depicted as a boundary is very important to convey.

Figure 4b shows that no one under deferred reinforcement learned specific stimulus–response associations that reflected the task's true (diagonal) reinforcement contingency. Under deferred reinforcement, there was no back door, no workaround to achieve associative learning. Associative learning, the stimulus-to-response mapping process, was completely unavailable.

Yet Fig. 4b also shows that performance under these circumstances was not haphazard. Humans adopted a new cognitive strategy when associative learning collapsed. They substituted their own rule, reflecting an attempt to self-construct the task explicitly when they could not response-map its reinforcement signals. They solved the task the only way they could—through rule generation. The evidence is strong that humans make this turn to explicit-declarative cognition when associative learning in this task is undermined. Though Smith, Boomer, et al. (2014) made this demonstration with deferred reinforcement, we saw above that humans make this same turn facing simple reinforcement delays (Maddox et al., 2003).

Presently we are testing another reinforcement regimen we hypothesize may produce similar dissociative phenomena. This method borrows from the *n*-back working-memory task in which subjects respond relative to the stimulus they saw *n* trials ago. We apply the *n*-back logic to the task's reinforcement. Following Trial 2, participants receive feedback on the stimulus–response pair that was Trial 1. Following Trial 3,

Fig. 4 a,b The decision bounds that provided the best fits to the responses of humans performing the discrimination-learning task illustrated in Fig. 2. **a** Under conditions of immediate, trial-by-trial reinforcement, many participants responded optimally, so that the best fit was achieved by performance that reflected a decisional boundary along the major diagonal of the stimulus space. **b** Under conditions of deferred reinforcement, as explained in the text, no participants responded

optimally. Instead, participants applied explicit but inappropriate unidimensional rules, strategies that were best fit by vertical and horizontal decision boundaries. From *Deferred Feedback Sharply Dissociates Implicit and Explicit Category Learning*, by J. D. Smith, A. C. Zakrzewski, J. Roeder, B. A. Church, & F. G. Ashby, *Psychological Science*, 25, 453. Copyright 2013 by the article's authors. Reprinted with permission

they receive feedback on the stimulus–response pair that was Trial 2. And so on. Now, if the participant learns associatively, he or she will connect just-received reinforcement to the just-responded trial—leading to a very bad learning outcome because those things are non-contingent. Instead, participants have to break that associative barrier, reach back to remember the previous stimulus and response, to consider the just-received reinforcement in that light. For some species, one-back reinforcement may be more natural than asking animals to persist through whole blocks of trials before receiving any reinforcement. However, both tasks are cognitively effortful—you can feel this as you perform them. This is exactly because they disable associative learning, rule out the low-level shaping of behavioral responses, and require cognitive processing instead at a more explicit-declarative level.

We believe that our innovative colleagues in comparative psychology will be able to derive other paradigms of this kind, and that these paradigms can produce a period of rapid empirical and theoretical development. The technique of shutting down associative learning qualitatively, leaving space for animals' explicit cognitive processes to reveal themselves instead, raises important theoretical questions in comparative psychology and makes these approachable.

A learning-systems approach toward animal metacognition

Emerging from the main sections of this article, colleagues in the animal-metacognition area may wish us to comment on the implication of this neural-systems approach for that area. We will do so. In brief, we believe the hypothesis is warranted that many first-order perceptual-discrimination responses in the metacognition area (like the dolphin's Low and High responses—Fig. 1) are instances of reinforcement-based associative learning. Those responses are true to the associative-learning construct as it is framed in this article. In contrast, we believe that uncertainty responses and other metacognitive responses may be the output of a qualitatively different cognitive process, one allied to the explicit processes we have been discussing. We consider two recent results that illustrate the merit of these proposals.

First, Smith et al. (2013) gave macaques a concurrent memory load while they performed a Sparse-Dense discrimination task with an uncertainty response also available to them (Fig. 5).

Smith et al. hypothesized that the task's primary perceptual responses (Sparse, Dense) would be based in associative learning as traditionally understood, making few working-memory demands and being barely affected by the load. In contrast, they hypothesized that uncertainty responses would be more an expression of explicit-declarative cognition, and thus uncertainty responses might be dependent on working

Fig. 5 Illustrating a trial in a Sparse-Uncertain-Dense task. In Smith et al. (2013), macaques saw a pixel box filled to one of 60 levels of proportional pixel density. Each level had 1.8% more pixels than the last. They touched the *S* or *D* icons with the joystick-controlled cursor to report that the box on a trial was Sparse (Stimulus Levels 1–30) or Dense (Stimulus Levels 31–60). The most difficult trials clustered around the discrimination breakpoint (Stimulus Levels 30–31). They touched the ? icon to decline the trial, thus fending off any difficult trials they did not wish to complete

memory and strongly affected by the load. In fact, uncertainty responses were disrupted by the concurrent task far more than Sparse or Dense responses (cf. Figs. 6b and a). It is an important observation that the metacognitive responses of macaques may be working-memory intensive. This observation dovetails with research on humans' metacognitive states, showing that for them, also, memory loads can strongly affect metacognitive judgments, especially decreasing tip-of-the-tongue experiences (Schwartz, 2008) and reducing the use of unpracticed uncertainty responding (Coutinho et al., 2015).

In an additional dissociation, Smith et al. (2013) found that Middle responses in a Sparse-Middle-Dense discrimination, like Sparse and Dense perceptual responses, were minimally affected by a working-memory load (cf. Figs. 6d and c). This is also true of human performance under load (Coutinho et al., 2015). In this case, the Middle response mapped to a discrete set of concrete stimulus levels (interestingly, just the same set of stimulus levels to which the uncertainty response mapped). Middle responses to those stimuli were immediately rewarded, creating the ideal conditions for the functioning of the associative-learning system. This result strengthens the proposal that the primary perceptual responses in metacognitive tasks—that is, Sparse, Dense, and Middle responses—are the products of something like associative-learning processes. But the uncertainty response is not.

Second, Paul, Valentin, Smith, Barbey, & Ashby (2015) placed humans in a Sparse-Uncertain-Dense task similar to that just described. Stimuli were presented across a Sparse to Dense continuum, with half the trials defined as Sparse, half as Dense, with a discrimination breakpoint at the continuum's center, and with the difficult/uncertain Sparse and Dense trials naturally clustering around that breakpoint. Then, using rapid event-related fMRI, they showed that the neural activity

Fig. 6 **a, b** Percentage of uncertainty responses (solid line), sparse responses (dashed line), and dense responses (dotted line) made by macaques Murph and Lou in their baseline performance and in their first phase of concurrent-load testing. **c, d** Percentage of middle responses (solid line), sparse responses (dashed line), and dense responses (dotted line) made by macaques Hank and Gale in their

baseline performance and in their first phase of concurrent-load testing. From *Executive-Attentional Uncertainty Responses by Rhesus Macaques (Macaca mulatta)*, by J. D. Smith, M. V. C. Coutinho, B. A. Church, & M. J. Beran, *Journal of Experimental Psychology: General*, 142, 472. Copyright 2013 by the American Psychological Association. Reprinted with permission

patterns recruited during humans' uncertainty responses are distinctively different from those recruited during humans' primary perceptual responses (Sparse, Dense).

In particular, Fig. 7 (top) shows the simple activation pattern when participants made correct Sparse or Dense responses in lieu of uncertainty responses. There was selective activation in the occipital lobe bilaterally, and in the caudate and nucleus accumbens. This is consistent with the view that Sparse and Dense responses are first-order perceptual events in which concrete visual stimuli elicit well-associated behavioral responses.

In contrast, Fig. 7 (bottom) shows the complex activation pattern when participants made uncertainty responses in lieu of Sparse or Dense responses. Uncertainty responding activated a distributed network including prefrontal cortex, anterior and posterior cingulate cortex, anterior insula, and posterior parietal areas. Thus, uncertainty monitoring benefits from a large-scale cognitive control network including recently evolved brain regions such as the anterior dorsolateral and medial prefrontal cortex. Were uncertainty responding just

an instance of associative learning, it would not depend on such an elaborate neural network or produce activation consistent with that network. But if we adopt a neural learning-systems framework instead, we would naturally conclude that the uncertainty response's processing network is allied to the explicit-declarative pole of cognitive processing that we have been describing.

In Paul et al. (2015), task uncertainty changed the cognitive system's mode of processing—in that case toward the explicit-declarative level. Daw and his colleagues have described similar control-switching processes based on uncertainty (Daw, Niv, & Dayan, 2005; Glascher, Daw, Dayan, & O'Doherty, 2010). They distinguish model-free and model-based learning systems that have similarities to the functions of associative learning and explicit cognition described here, including their neural locus (dorsolateral striatum and prefrontal cortex, respectively). Daw and his colleagues focus much needed attention on the questions of which learning system governs ongoing performance, and when, and why. That is, how in a Sparse-Uncertain-Dense task (Smith et al.,

Fig. 7 *Top* Whole-brain results from the contrast Correct Sparse-Dense Response > Uncertainty Response projected on an inflated lateral and medial cortical surface. Table 2 in Paul et al. (2015) listed the coordinates and number of voxels for every significant cluster. Images were cluster thresholded (correcting for multiple comparisons) at $z > 3.54$, $P < .01$. *Bottom* Whole-brain results from the contrast Uncertainty Response > Correct Sparse-Dense Response projected on an inflated lateral and medial cortical surface. Table 3 in Paul et al. (2015) listed the coordinates and number of voxels for every significant cluster. Images were cluster thresholded (correcting for multiple comparisons) at $z > 3.54$, $P < .01$. From *Neural Networks of the Psychophysical Uncertainty Response*, by E. J. Paul, V. Valentin, J. D. Smith, A. K. Barbey, & F. G. Ashby, *Cortex*, 71, 316 (top), 317 (bottom). Copyright 2015 Elsevier Ltd. Reprinted with permission

2013) is the perceptual information deemed inadequate for a primary response so that an uncertainty response ensues? When do humans facing delayed reinforcement in an II task (see above), and the failure of associative learning, override those learning efforts so that the cognitive system adopts instead an adventitious category rule? Daw and his colleagues discuss these interactions and arbitrations and simulate them elegantly. As suggested in Paul et al. (2015), they note the possible involvement of the anterior cingulate cortex in managing these monitoring and arbitration functions (Botvinick, Cohen, & Carter, 2004; Holroyd & Coles, 2002).

Our overall conclusion about this research area fits with everything discussed in this article. Considering the animal-metacognition results, there is no unitary associative-learning construct that can explain all the performance dissociations and the psychological uniqueness of the uncertainty response or other seemingly metacognitive behaviors (e.g., information-seeking responses). We will need to divide to

conquer this literature, too, creating a dissociative framework that allows for the involvement of multiple brain/cognitive systems in organizing animals' metacognitive response patterns. This dissociative framework can help resolve the associative debate in animal-metacognition research, while opening possibilities for research to fully specify the character of animals' explicit uncertainty-monitoring and other metacognitive capabilities.

Summary and conclusion

Summary

We described challenges to the associative-learning construct in comparative psychology. It is a theoretical weakness to address these challenges by stretching and thinning the meaning of associative learning. These defensive associative interpretations are psychologically empty and they dull the sharpness of associative-learning theory. Addressing this concern, we showed that the associative-learning construct is still profoundly important for capturing dominant forms of learning by animals and humans—if only it is granted a constrained operational definition. Our article honors that construct by trying to strengthen and sustain it. It does not diminish it in any way.

Illustrating an appropriate boundary for associative learning, we showed that different discrimination-learning procedures—even procedures that seem closely related—in reality dissociate in their reliance on qualitatively different learning processes and neural systems. Different task variants change—in concert—representational content, the dimensional breadth of knowledge, the level of awareness, the declarative nature of knowledge, the brain systems that organize learning, and the involvement of phylogenetically older vs. newer brain structures. We showed that, even though different evolutionarily older brain systems (i.e., cerebellum, brain stem, and amygdala) may be involved, classical-conditioning procedures that are closely related to each other dissociate in the same way.

These theoretical analyses express a fundamental point. One cannot responsibly assimilate two performances that differ diametrically along every axis of cognitive functioning and simply call them both associative learning, or for that matter call them both anything else. Science demands that we differentiate contrastive cognitive performances and codify the contrasts. This has been productive in the study of human cognition (e.g. Ashby and Maddox, 2011; Atkinson & Shiffrin, 1968; Baddeley & Hitch, 1974; Craik & Lockhart, 1972; Cohen & Squire, 1980; Jacoby, 1991; Moscovitch, 1992; Richardson-Klavehn & Bjork, 1988; Roediger & Blaxton, 1987; Schacter, 1990; Tulving, 1985), and we believe it will be equally productive in comparative psychology.

Implications for animal psychology

The approach developed here has many implications for comparative psychology. It constrains the construct of associative learning, preventing the bracket creep in that construct that has not been constructive, and helping define a disciplined and sustainable associative-learning construct. It establishes the boundary, the limit, the fracture plane, at which reinforcement systems fail and other modes of information processing must take their place. It defines the threshold for the transition to modes of information processing that are more like explicit cognition. It emphasizes sensitivity to method: small methodological changes in the sequencing and timing of reinforcement have dramatic information-processing consequences, altering qualitatively the character of learning. Both operant learning and classical conditioning are included in the present framework, sharing similar boundary conditions and transition zones.

The present perspective bridges a theory gap between animal and human psychology. Only in the latter case has a learning-systems framework been deeply engaged and elaborated. It bridges a methodological gap, because the occasional studies relevant to this framework have been technically complex, staged affairs on topics such as latent learning and reinforcer devaluation (Otto, Gershman, Markman, & Daw, 2013). In contrast, the paradigms described here are very friendly to animals' participation. For example, 1-back feedback as described in *Converging Techniques* is no technical burden to an operant researcher, and it might be no burden to a primate participant, either.

Above all, the techniques that may qualitatively unplug associative learning, and qualitatively require a transition to explicit forms of cognition, are potentially very powerful tools. Comparative psychologists have essentially never considered that associative learning might have an off switch, freeing researchers to explore other modes of information processing that animals may possess. By closing this switch, we might persuade animals to transcend their associative training histories and raise their cognitive game to the explicit level. In that way, we might discover the true top of animals' capacity for abstraction, conceptualization, and symbolic functioning, topics that have been central to comparative psychology for decades. Moreover, explicit cognition is attended by conscious awareness—in humans! It is a fascinating question whether this is a singular confluence in human minds, or whether this is an intrinsic emergent property of explicit cognitive systems generally. The perspective offered here could put comparative psychology in a strong position to systematically explore animal awareness.

Using these techniques, researchers could draw the species map of the vertebrate lines that do and do not have forms of explicit cognition. They could resolve long-standing questions about the cognitive capacities of pigeons vs. primates, and

monkeys vs. apes. Through this research, they would also map the earliest evolutionary roots of humans' explicit cognitive capacity.

Thus, for many reasons, we believe the approach outlined in this article is potentially transformative in comparative psychology.

Implications for human psychology

The present perspective has implications for human psychology as well. It potentially offers a useful theoretical and methodological perspective to a variety of psychological and neuroscience disciplines.

First, if comparative psychologists have not reckoned with explicit cognition, cognitive psychologists of late have seldom reckoned with associative learning. A thousand undergraduate paradigms have featured trial-by-trial reinforcement, with inevitable learning-systems consequences. But theory rarely accommodates these consequences, or models these associative influences, and method rarely controls or factors away those influences.

Second, in fact, there do exist many concurrent-task approaches that block the explicit, the declarative, the executive in cognitive processing. However, scant attention has been given to the idea of blocking the associative influences on learning and performance. This is the possibility that the present perspective offers, and we believe it could have many empirical and theoretical uses within experimental psychology and cognitive neuroscience.

Third, our perspective could give human psychology a powerful set of animal models concerning the most basic forms of explicit declarative cognition in categorization, discrimination, rule learning, decision making, and so forth. We can study the neuroscience underpinnings of explicit cognition and search for neurochemical facilitators and enhancers.

Fourth, cognitive theory is almost too casual in its understanding of explicit-declarative cognition. For example, there has been a perennial conflation between explicit cognition and human propositional thought and language. On reflection, though, one sees that this conflation is neither necessary nor established truth. There could be language-less thought propositions. These could be declared behaviorally, not linguistically. By studying animals in the same explicit-cognitive contexts as humans, one could establish the affordances of language, and explore the possibility of wordless, language-less, explicit cognition. This work would go to the fundamental nature of explicit cognition, a question that should span vertebrate phylogeny.

Fifth, developmental psychologists could adapt the present perspective to examine the earliest steps that young children take as they cross the threshold to explicit cognition. The paradigms described here are possibly the simplest with which

to explore explicit cognition directly and purely behaviorally. But for the crucial change in reinforcement's sequencing, they are closely related to operant designs, and they are very friendly to participation by young children. Developmental researchers have already adapted the behavioral paradigms of animal-metacognition research to test young children (e.g., Balcomb & Gerken, 2008).

Sixth, we dare one comment about educational and clinical practice. There are diverse training programs that aim toward behavior modification by targeting deliberately the associative level of learning. These have their role in helping some children and patients manage difficult and destructive behavioral habits. But this led us to wonder about the converse learning strategy. For example, imagine a child learning new math skills, but now only rewarded through a 1-back reinforcement regimen that lifts the task off the associative plane. It seems this approach might suppress some rote, reactive, automatic response habits that children might (and do!) learn in acquiring strong math algorithms. It seems this approach might efficiently teach children at a higher, explicit, conceptual level. A similar approach might even be used to help clinical patients learn new interpretations of and responses to their feelings of anxiety and/or depression.

Conclusion

In the end, our article supports an interdisciplinary hope. It did not have to be, perhaps it should not have been, that animal and human psychology diverged in some behaviorist wood, creating lasting divisions. There is a rich and pressing need for cross talk, for animal models, for research synergies, for correlated neuroscience across species, though these interdisciplinary interactions have been remarkably sparse. Our hope is that the theoretical framework offered here may be useful to both research areas, especially as each discipline—exploring the threshold of explicit cognition—reaches out toward the other.

Acknowledgements The preparation of this article was supported by Grants HD-060563 and HD-061455 from NICHD.

References

- Aguado, L. (2003). Neuroscience of Pavlovian conditioning: A brief review. *The Spanish Journal of Psychology*, 6, 155–167. doi:10.1017/S1138741600005308
- Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, 9, 357–381. doi:10.1146/annurev.ne.09.030186.002041
- Arbuthnott, G. W., Ingham, C. A., & Wickens, J. R. (2000). Dopamine and synaptic plasticity in the neostriatum. *Journal of Anatomy*, 196, 587–596. doi:10.1046/j.1469-7580.2000.19640587.x
- Ashby, F. G., & Ennis, J. M. (2006). The role of the basal ganglia in category learning. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 46, pp. 1–36). San Diego: Academic.
- Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences*, 1224, 147–161. doi:10.1111/j.1749-6632.2010.05874.x
- Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, 105, 442–481. doi:10.1037/0033-295X.105.3.442
- Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science*, 21, 170–176. doi:10.1177/0963721412436806
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89–195). New York: Academic.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, 8, 47–89. doi:10.1016/S0079-7421(08)60452-1
- Balcomb, F. K., & Gerken, L. (2008). Three-year-old children can access their own memory to guide responses on a visual matching task. *Developmental Science*, 11, 750–760. doi:10.1111/j.1467-7687.2008.00725.x
- Barnes, T. D., Kubota, Y., Hu, D., Jin, D. Z., & Graybiel, A. M. (2005). Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature*, 437, 1158–1161. doi:10.1038/nature04053
- Basile, B. M., & Hampton, R. R. (2011). Monkeys recall and reproduce simple shapes from memory. *Current Biology*, 21, 774–778. doi:10.1016/j.cub.2011.03.044
- Basile, B. M., & Hampton, R. R. (2014). Metacognition as discrimination: Commentary on Smith et al. (2014). *Journal of Comparative Psychology*, 128, 135–137. doi:10.1037/a0034412
- Basile, B. M., Schroeder, G. R., Brown, E. K., Templar, V. L., & Hampton, R. R. (2015). Evaluation of seven hypotheses for metamemory performance in rhesus monkeys. *Journal of Experimental Psychology: General*, 144, 85–102. doi:10.1037/xge0000031
- Boneau, C. A., & Cole, J. L. (1967). Decision theory, the pigeon, and the psychophysical function. *Psychological Review*, 74, 123–135. doi:10.1037/h0024287
- Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, 8, 539–546. doi:10.1016/j.tics.2004.10.003
- Brown, R. G., & Marsden, C. D. (1988). Internal versus external cues and the control of attention in Parkinson's disease. *Brain*, 111, 323–345. doi:10.1093/brain/111.2.323
- Calabresi, P., Pisani, A., Centonze, D., & Bernardi, G. (1996). Role of Ca²⁺ in striatal LTD and LTP. *Seminars in Neurosciences*, 8, 321–328. doi:10.1006/smns.1996.0039
- Call, J. (2010). Do apes know that they could be wrong?. *Animal Cognition*, 13(5), 689–700.
- Call, J., & Carpenter, M. (2001). Do apes and children know what they have seen? *Animal Cognition*, 4, 207–220. doi:10.1007/s100710100078
- Call, J., & Tomasello, M. (2008). Do chimpanzees have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, 12, 187–192. doi:10.1016/j.tics.2008.02.010
- Cantlon, J. F., Platt, M. L., & Brannon, E. M. (2009). Beyond the number domain. *Trends in Cognitive Sciences*, 13, 83–91. doi:10.1016/j.tics.2008.11.007
- Carruthers, P. (2008). Meta-cognition in animals: A skeptical look. *Mind and Language*, 23(1), 58–89. doi:10.1111/j.1468-0017.2007.00329.x

- Clark, R. E., & Squire, L. R. (1998). Classical conditioning and brain systems: The role of awareness. *Science*, 280, 77–81. doi:10.1126/science.280.5360.77
- Clark, R. E., & Squire, L. R. (1999). Human eyeblink classical conditioning: Effects of manipulating awareness of the stimulus contingencies. *Psychological Science*, 10, 14–18. doi:10.1111/1467-9280.00099
- Clark, R. E., Manns, J. R., & Squire, L. R. (2002). Classical conditioning, awareness, and brain systems. *Trends in Cognitive Sciences*, 6, 525–531. doi:10.1016/S1364-6613(02)02041-7
- Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences*, 2, 406–416. doi:10.1016/S1364-6613(98)01232-7
- Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science*, 210, 207–209. doi:10.1126/science.7414331
- Commons, M. L., Nevin, J. A., & Davidson, M. C. (1991). *Signal detection: Mechanisms, models, and applications (quantitative analyses of behavior series)*. Richmond: Ergodebooks.
- Cools, A. R., van den Bercken, J. H., Horstink, M. W., van Spaendonck, K. P., & Berger, H. J. (1984). Cognitive and motor shifting aptitude disorder in Parkinson's disease. *Journal of Neurological and Neurosurgical Psychology*, 47, 443–453. <http://jnnp.bmj.com/content/jnnp/47/5/443.full.pdf>
- Coutinho, M. V. C., Redford, J. S., Church, B. A., Zakrzewski, A. C., Couchman, J. J., & Smith, J. D. (2015). The interplay between uncertainty monitoring and working memory: Can metacognition become automatic? *Memory and Cognition*, 43, 990–1006. doi:10.3758/s13421-015-0527-1
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671–684. <http://homepage.westmont.edu/bsmith/documents/CraikandLockhart1972.pdf>
- Davison, M., McCarthy, D., & Jensen, C. (1985). Component probability and component reinforcer rate as biasers of free-operant detection. *Journal of Experimental Analysis and Behavior*, 44, 103–120. doi:10.1901/jeab.1985.44-103
- Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8, 1704–1711. doi:10.1038/nn1560
- Debert, P., Matos, M. A., & McIlvane, W. (2007). Conditional relations with compound abstract stimuli using a go/no-go procedure. *Journal of Experimental Analysis of Behavior*, 87, 89–96. doi:10.1901/jeab.2007.46-05
- Divac, I., Rosvold, H. E., & Szwarcbart, M. K. (1967). Behavioral effects of selective ablation of the caudate nucleus. *Journal of Comparative and Physiological Psychology*, 63, 184–190. doi:10.1037/h0024348
- Eacott, M. J., & Gaffan, D. (1992). Inferotemporal-frontal disconnection: The uncinate fascicle and visual associative learning in monkeys. *European Journal of Neuroscience*, 4, 1320–1332. doi:10.1111/j.1460-9568.1992.tb00157.x
- Elliott, R., & Dolan, R. J. (1998). Activation of different anterior cingulate foci in association with hypothesis testing and response selection. *NeuroImage*, 8, 17–29. doi:10.1006/nimg.1998.0344
- Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, 127, 107–140.
- Fagot, J., Wasserman, E. A., & Young, M. (2001). Discriminating the relation between relations: The role of entropy in abstract conceptualization by baboons (*Papio papio*) and humans (*Homo sapiens*). *Journal of Experimental Psychology: Animal Behavior Processes*, 27, 316–328. doi:10.1037/0097-7403.27.4.316
- Filoteo, J. V., Maddox, W. T., Salmon, D. P., & Song, D. D. (2005). Information-integration category learning in patients with striatal dysfunction. *Neuropsychology*, 19, 212–222. doi:10.1037/0894-4105.19.2.212
- Foote, A., & Crystal, J. (2007). Metacognition in the rat. *Current Biology*, 17, 551–555. doi:10.1016/j.cub.2007.01.061
- Frcka, G., Beyts, J., Levey, A. B., & Martin, I. (1983). The role of awareness in human conditioning. *Pavlovian Journal of Biological Science*, 18, 69–76. doi:10.1007/BF03001857
- Fujita, K. (2009). Metamemory in tufted capuchin monkeys (*Cebus apella*). *Animal Cognition*, 12, 575–585. doi:10.1007/s10071-009-0217-0
- Fuster, J. M. (1989). *The prefrontal cortex* (2nd ed.). Philadelphia: Lippincott-Raven.
- Gaffan, D., & Eacott, M. J. (1995). Visual learning for an auditory secondary reinforcer by macaques is intact after uncinate fascicle section: Indirect evidence for the involvement of the corpus striatum. *European Journal of Neuroscience*, 7, 1866–1871. doi:10.1111/j.1460-9568.1995.tb00707.x
- Gaffan, D., & Harrison, S. (1987). Amygdectomy and disconnection in visual learning for auditory secondary reinforcement by monkeys. *Journal of Neuroscience*, 7, 2285–2292. www.jneurosci.org/content/7/8/2285
- Gamble, E., & Koch, C. (1987). The dynamics of free calcium in dendritic spines in response to repetitive synaptic input. *Science*, 236, 1311–1315.
- Glascher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, 66, 585–595. doi:10.1016/j.neuron.2010.04.016
- Goldman-Rakic, P. S. (1987). Circuitry of the prefrontal cortex and the regulation of behavior by representational knowledge. In F. Plum & V. Mountcastle (Eds.), *Handbook of physiology* (pp. 373–417). Bethesda: American Physiological Society.
- Grant, D. A. (1973). Cognitive factors in eyelid conditioning. *Psychophysiology*, 10, 75–81. doi:10.1111/j.1469-8986.1973.tb01086.x
- Hampton, R. R. (2001). Rhesus monkeys know when they remember. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 5359–5362. doi:10.1073/pnas.071600998
- Hampton, R. R. (2009). Multiple demonstrations of metacognition in nonhumans: Converging evidence or multiple mechanisms? *Comparative Cognition and Behavior Reviews*, 4, 17–28. doi:10.3819/ccbr.2009.40002
- Han, C. J., O'Tuathaigh, C. M., van Trigt, L., Quinn, J. J., Fanselow, M. S., Mongeau, R., & Anderson, D. J. (2003). Trace but not delay fear conditioning requires attention and the anterior cingulate cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 13087–13092. doi:10.1073/pnas.2132313100
- Harlow, H. F. (1949). The formation of learning sets. *Psychological Review*, 56, 51–65. doi:10.1037/h0062474
- Hilgard, R. E., & Humphreys, L. G. (1938). Delay conditioning in humans can occur without individuals becoming aware of the relationship between the CS and the US. *Journal of General Psychology*, 19, 111–125.
- Hollerman, J. R., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience*, 1, 304–309. doi:10.1038/1124
- Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109, 679–709. doi:10.1037/0033-295X.109.4.679
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30, 513–541. doi:10.1016/0749-596X(91)90025-F
- Jozefowicz, J., Staddon, J. E. R., & Cerutti, D. T. (2009). Metacognition in animals: How do we know that they know? *Comparative Cognition and Behavior Reviews*, 4, 29–39. doi:10.3819/ccbr.2009.40003

- Kim, J. J., & Jung, M. W. (2006). Neural circuits and mechanisms involved in Pavlovian fear conditioning: A critical review. *Neuroscience & Biobehavioral Reviews*, 30, 188–202. doi:10.1016/j.neubiorev.2005.06.005
- Kim, J. J., Clark, R. E., & Thompson, R. F. (1995). Hippocampectomy impairs the memory of recently, but not remotely, acquired trace eyeblink conditioned responses. *Behavioral Neuroscience*, 109, 195–203. doi:10.1037/0735-7044.109.2.195
- Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel memory systems for item memory and category-level knowledge. *Science*, 262, 1747–1749. doi:10.1126/science.8259522
- Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, 273, 1399–1402. doi:10.1126/science.273.5280.1399
- Kolb, B., & Whishaw, I. Q. (1990). *Fundamentals of human neuropsychology* (3rd ed.). New York: Freeman.
- Konorski, J. (1967). *Integrative activity of the brain*. Chicago: University of Chicago Press.
- Kornell, N. (2009). Metacognition in humans and animals. *Current Directions in Cognitive Science*, 18, 11–15. doi:10.1111/j.1467-8721.2009.01597.x
- Kornell, N., Son, L., & Terrace, H. (2007). Transfer of metacognitive skills and hint seeking in monkeys. *Psychological Science*, 18, 64–71. doi:10.1111/j.1467-9280.2007.01850.x
- Kronforst-Collins, M. A., & Disterhoft, J. F. (1998). Lesions of the caudal area of the rabbit medial prefrontal cortex impair trace eyeblink conditioning. *Neurobiology of Learning and Memory*, 2, 147–162. doi:10.1006/nlme.1997.3818
- Kryukov, V. I. (2012). Towards a unified model of pavlovian conditioning: Short review. *Cognitive Neuroscience*, 6, 377–398. doi:10.1007/s11571-012-9195-z
- Le Pelley, M. E. (2012). Metacognitive monkeys or associative animals? Simple reinforcement learning explains uncertainty in nonhuman animals. *Journal of Experimental Psychology: Learning, Memory, and Cognition (Washington, DC)*, 38, 686–708. doi:10.1037/a0026478
- Le Pelley, M. E. (2014). Primate polemic: Commentary on Smith, Couchman and Beran (2014). *Journal of Comparative Psychology*, 128, 132–134. doi:10.1037/a0034227
- MacDermott, A. B., Mayer, M. L., Westbrook, G. L., Smith, S. J., & Barker, J. L. (1986). NMDA-receptor activation increases cytoplasmic calcium concentration in cultured spinal cord neurones. *Nature*, 321, 519–522. doi:10.1038/321519a0
- Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioral Processes*, 66, 309–332. doi:10.1016/j.beproc.2004.03.011
- Maddox, W. T., & Ing, A. D. (2005). Delayed feedback disrupts the procedural-learning system but not the hypothesis testing system in perceptual category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition (Washington, DC)*, 31, 100–107. doi:10.1037/0278-7393.31.1.100
- Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition (Washington, DC)*, 29, 650–662. doi:10.1037/0278-7393.29.4.650
- Manns, J. R., Clark, R. E., & Squire, L. R. (2000). Parallel acquisition of awareness and trace eyeblink classical conditioning. *Learning & Memory*, 7, 267–272. doi:10.1101/lm.33400
- Manns, J. R., Clark, R. E., & Squire, L. R. (2001). Single-cue delay eyeblink conditioning is unrelated to awareness. *Cognitive, Affective, & Behavioral Neuroscience*, 1, 192–198. doi:10.3758/CABN.1.2.192
- Manns, J. R., Clark, R. E., & Squire, L. R. (2002). Standard delay eyeblink classical conditioning is independent of awareness. *Journal of Experimental Psychology: Animal Behavior Processes*, 28, 32–37. doi:10.1037//0097-7403.28.1.32
- McDonald, R. J., & White, N. M. (1993). A triple dissociation of memory systems: Hippocampus, amygdala, and dorsal striatum. *Behavioral Neuroscience*, 107, 3–22. doi:10.1037/0735-7044.107.1.3
- McDonald, R. J., & White, N. M. (1994). Parallel information processing in the water maze: Evidence for independent memory systems involving dorsal striatum and hippocampus. *Behavioral and Neural Biology*, 61, 260–270. doi:10.1016/S0163-1047(05)80009-3
- McGlinchey-Berroth, R., Carrillo, M. C., Gabrieli, J. D., Brawn, C. M., & Disterhoft, J. F. (1997). Impaired trace eyeblink conditioning in bilateral, medial-temporal lobe amnesia. *Behavioral Neuroscience*, 111, 873–882. doi:10.1037/0735-7044.111.5.873
- Miller, J. T., Saunders, S. S., & Bourland, G. (1980). The role of stimulus disparity in concurrently available reinforcement schedules. *Animal Learning & Behavior*, 8, 635–641. doi:10.3758/BF03197780
- Mishkin, M., Malamut, B., & Bachevalier, J. (1984). Memories and habits: Two neural systems. In G. Lynch, G. J. L. McGaugh, & N. M. Weinberger (Eds.), *Neurobiology of human learning and memory* (pp. 65–88). New York: Guilford.
- Morgan, C. L. (1906). *An introduction to comparative psychology*. London: Scott.
- Moscovitch, M. (1992). Memory and working-with-memory: A component process model based on modules and central systems. *Journal of Cognitive Neuroscience*, 4, 257–267. doi:10.1162/jocn.1992.4.3.257
- Moyer, J. R., Deyo, R. A., & Disterhoft, J. S. (1990). Hippocampectomy disrupts trace eye-blink conditioning in rabbits. *Behavioral Neuroscience*, 104, 243–252. doi:10.1037/0735-7044.104.2.243
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century Crofts.
- Nomura, E. M., Maddox, W. T., Filoteo, J. V., Ing, A. D., Gitelman, D. R., Parrish, T. B., Mesulam, M. M., & Reber, P. J. (2007). Neural correlates of rule-based and information-integration visual category learning. *Cerebral Cortex*, 17, 37–43. doi:10.1093/cercor/bhj122
- Nosofsky, R. M., Palmeri, T. J., & McKimley, S. K. (1994). Rule-plus exception model of classification learning. *Psychological Review*, 101, 53–79. doi:10.3758/BF03208813
- O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, 304, 452–454. doi:10.1126/science.1094285
- Otto, A. R., Gershman, S. J., Markman, A. B., & Daw, N. D. (2013). The curse of planning: Dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychological Science*, 24, 751–761. doi:10.1177/0956797612463080
- Packard, M. G., & McGaugh, J. L. (1992). Double dissociation of fornix and caudate nucleus lesions on acquisition of two water maze tasks: Further evidence for multiple memory systems. *Behavioral Neuroscience*, 106, 439–446. doi:10.1037/0735-7044.106.3.439
- Packard, M. G., Hirsh, R., & White, N. M. (1989). Differential effects of fornix and caudate nucleus lesions on two radial maze tasks: Evidence for multiple memory systems. *Journal of Neuroscience*, 9, 1465–1472. <http://www.jneurosci.org/content/jneuro/9/5/1465>
- Papka, M., Ivry, R. B., & Woodruff-Pak, D. S. (1997). Eyeblink classical conditioning and awareness revisited. *Psychological Science*, 8, 404–408. <http://www.jstor.org/stable/40063221>
- Paukner, A., Anderson, J. R., & Fujita, K. (2006). Redundant food searches by capuchin monkeys (*Cebus apella*): A failure of metacognition? *Animal Cognition*, 9, 110–117. doi:10.1007/s10071-005-0007-2
- Paul, E. J., Valentin, V., Smith, J. D., Barbey, A. K., & Ashby, F. G. (2015). Neural networks of the psychophysical uncertainty response. *Cortex*, 71, 306–322. doi:10.1016/j.cortex.2015.07.028

- Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. London: Oxford University Press, Humphrey Milford.
- Posner, M. I., & Petersen, S. E. (1990). Attention systems in the human brain. *Annual Review of Neuroscience*, 13, 25–42.
- Powell, D. A., Skaggs, H., Churchwell, J., & McLaughlin, J. (2001). Posttraining lesions of the medial prefrontal cortex impair performance of Pavlovian eyeblink conditioning but have no effect on concomitant heart rate changes in rabbits (*Oryctolagus cuniculus*). *Behavioral Neuroscience*, 115, 1029–1038. doi:10.1037/0735-7044.115.5.1029
- Rao, S. M., Bobholz, J. A., Hammeke, T. A., Rosen, A. C., Woodley, S. J., Cunningham, J. M., Cox, R. W., Stein, E. A., & Binder, J. R. (1997). Functional MRI evidence for subcortical participation in conceptual reasoning skills. *Neuroreport*, 27, 1987–1993. doi: 10.1097/00001756-199705260-00038
- Raybuck, J. D., & Lattal, K. M. (2014). Bridging the interval: Theory and neurobiology of trace conditioning. *Behavioral Processes*, 101, 103–111. doi:10.1016/j.beproc.2013.08.016
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II* (pp. 64–99). New York: Appleton-Century Crofts.
- Richardson-Klavehn, A., & Bjork, R. A. (1988). Measures of memory. *Annual Review of Psychology*, 36, 475–543. doi:10.1146/annurev.ps.39.020188.002355
- Roberts, W. A., Feeney, M. C., McMillan, N., MacPherson, K., Musolino, E., & Petter, M. (2009). Do pigeons (*Columba livia*) study for a test? *Journal of Experimental Psychology: Animal Behavior Processes*, 35, 129–142. doi:10.1037/a0013722
- Robinson, A. L., Heaton, R. K., Lehman, R. A. W., & Stilson, D. W. (1980). The utility of the Wisconsin Card Sorting Test in detecting and localizing frontal lobe lesions. *Journal of Consulting and Clinical Psychology*, 48, 605–614.
- Roediger, H. L., III, & Blaxton, T. A. (1987). Effects of varying modality, surface features, and retention interval on priming in word fragment completion. *Memory & Cognition*, 15, 379–388. doi:10.3758/BF03197728
- Rolls, E. T. (1994). Neurophysiology and cognitive functions of the striatum. *Revue Neurologique*, 150, 648–660.
- Schacter, D. L. (1990). Perceptual representation systems and implicit memory: Toward a resolution of the multiple memory systems debate. In A. Diamond (Ed.), *Annals of the New York academy of sciences: Development and neural bases of higher cortical functions* (Vol. 608, pp. 543–571). New York: New York Academy of Sciences.
- Schultz, W. (1992). Activity of dopamine neurons in the behaving primate. *Seminars in Neurosciences*, 4, 129–138. doi:10.1016/1044-5765(92)90011-P
- Schwartz, B. L. (2008). Working memory load differentially affects tip-of-the-tongue states and feeling-of-knowing judgments. *Memory and Cognition*, 36, 9–19. doi:10.3758/MC.36.1.9
- Seger, C. A., & Cincotta, C. M. (2005). The roles of the caudate nucleus in human classification learning. *Journal of Neuroscience*, 25, 2941–2951. doi:10.1523/JNEUROSCI.3401-04.2005
- Shields, W. E., Smith, J. D., & Washburn, D. A. (1997). Uncertain responses by humans and rhesus monkeys (*Macaca mulatta*) in a psychophysical same-different task. *Journal of Experimental Psychology: General*, 126, 147–164. doi:10.1016/S0010-0277(96)00726-3
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, 84, 127–190. doi:10.1037/0033-295X.84.2.127
- Smith, J. D. (2009). The study of animal metacognition. *Trends in Cognitive Sciences*, 13, 389–396. doi:10.1016/j.tics.2009.06.009
- Smith, J. D., Schull, J., Strote, J., McGee, K., Egnor, R., & Erb, L. (1995). The uncertain response in the bottlenosed dolphin (*Tursiops truncatus*). *Journal of Experimental Psychology: General*, 124, 391–408. doi:10.1037/0096-3445.124.4.391
- Smith, J. D., Beran, M. J., Couchman, J. C., & Coutinho, M. V. C. (2008). The comparative study of metacognition: Sharper paradigms, safer inferences. *Psychonomic Bulletin Review*, 15, 679–691. doi:10.3758/PBR.15.4.679
- Smith, J. D., Beran, M. J., Couchman, J. J., Coutinho, M. V. C., & Boomer, J. (2009). Animal metacognition: Problems and prospects. *Comparative Cognition and Behavior Reviews*, 4, 40–53. doi:10.3819/ccbr.2009.40004
- Smith, J. D., Beran, M. J., & Couchman, J. J. (2012). Animal metacognition. In T. Zentall & E. Wasserman (Eds.), *Comparative cognition: Experimental explorations of animal intelligence* (pp. 282–304). Oxford: Oxford University Press.
- Smith, J. D., Couchman, J. J., & Beran, M. J. (2012). The highs and lows of theoretical interpretation in animal metacognition research. *Philosophical Transactions of the Royal Society B*, 367, 1297–1309. doi:10.1098/rstb.2011.0366
- Smith, J. D., Coutinho, M. V. C., Church, B. A., & Beran, M. J. (2013). Executive-attentional uncertainty responses by rhesus macaques (*Macaca mulatta*). *Journal of Experimental Psychology: General*, 142, 458–475. doi:10.1037/a0029601
- Smith, J. D., Boomer, J., Zakrzewski, A. C., Roeder, J., Church, B. A., & Ashby, F. G. (2014). Deferred feedback sharply dissociates implicit and explicit category learning. *Psychological Science*, 25, 447–457. doi:10.1177/0956797613509112
- Smith, J. D., Couchman, J. J., & Beran, M. J. (2014). Animal metacognition: A tale of two comparative psychologies. *Journal of Comparative Psychology*, 128, 115–131. doi:10.1037/a0033105
- Smith, J. D., Zakrzewski, A. C., Johnson, J. M., Valteau, J. C., & Church, B. A. (2016). Categorization: The view from animal cognition. *Behavioral Sciences*, 6, 12. doi:10.3390/bs6020012
- Solomon, P. R., Vander Schaaf, E. R., Thompson, R. F., & Weisz, D. J. (1986). Hippocampus and trace conditioning of the rabbit's classically conditioned nictitating membrane response. *Behavioral Neuroscience*, 100, 729–744. doi:10.1037/0735-7044.100.5.729
- Squire, L. R., Clark, R. E., & Knowlton, B. J. (2001). Retrograde amnesia. *Hippocampus*, 11, 50–55. doi:10.1002/1098-1063(2001)11:1<50::AID-HIPO1019>3.0.CO;2-G
- Staddon, J. E. R., Jozefowicz, J., & Cerutti, D. (2007). Metacognition: A problem not a process. *PsyCrit*, 1–5. http://www.psyrcrit.com/Targets/2007-Foote/2007-03-15_Staddon
- Suda-King, C. (2008). Do orangutans (*Pongo pygmaeus*) know when they do not remember? *Animal Cognition*, 7, 239–246. doi:10.1007/s10071-007-0082-7
- Sutton, J. E., & Shettleworth, S. J. (2008). Memory without awareness: Pigeons do not show metamemory in delayed matching-to-sample. *Journal of Experimental Psychology: Animal Behavior Processes*, 34, 266–282. doi:10.1037/0097-7403.34.2.266
- Templer, V. L., & Hampton, R. R. (2012). Rhesus monkeys (*Macaca mulatta*) show robust evidence for memory awareness across multiple generalization tests. *Animal Cognition*, 15, 409–419. doi:10.1007/s10071-011-0468-4
- Terman, M., & Terman, J. (1972). Concurrent variation of response bias and sensitivity in an operant-psychophysical test. *Perception & Psychophysics*, 11, 428–432. doi:10.3758/BF03206285
- Thompson, R. F. (1990). Neural mechanisms of classical conditioning in mammals. *Philosophical Transactions: Biological Sciences*, 329, 161–170. <http://www.jstor.org/stable/76874>
- Thorndike, E. L. (1911). *Animal intelligence*. New York: Macmillan.
- Tulving, E. (1985). How many memory systems are there? *American Psychologist*, 40, 385–398. doi:10.1037/0003-066X.40.4.385

- Waldschmidt, J. G., & Ashby, F. G. (2011). Cortical and striatal contributions to automaticity in information-integration categorization. *NeuroImage*, 56, 1791–1802. doi:[10.1016/j.neuroimage.2011.02.011](https://doi.org/10.1016/j.neuroimage.2011.02.011)
- Watanabe, T., & Sasaki, Y. (2015). Perceptual learning: Toward a comprehensive theory. *Annual Review of Psychology*, 66, 197–221. doi:[10.1146/annurev-psych-010814-015214](https://doi.org/10.1146/annurev-psych-010814-015214)
- Weible, A. P., McEchron, M. D., & Disterhoft, J. F. (2000). Cortical involvement in acquisition and extinction of trace eyeblink conditioning. *Behavioral Neuroscience*, 114, 1058–1067. doi:[10.1037/0735-7044.114.6.1058](https://doi.org/10.1037/0735-7044.114.6.1058)
- Weiskrantz, L., & Warrington, E. K. (1979). Conditioning in amnesic patients. *Neuropsychologia*, 17, 187–194. doi:[10.1016/0028-3932\(79\)90009-5](https://doi.org/10.1016/0028-3932(79)90009-5)
- Weiss, C., Bouwmeester, H., Power, J. M., & Disterhoft, J. E. (1999). Hippocampal lesions prevent trace eyeblink conditioning in the freely moving rat. *Behavioural Brain Research*, 99, 123–132. doi:[10.1016/S0166-4328\(98\)00096-5](https://doi.org/10.1016/S0166-4328(98)00096-5)
- Wickens, J. (1993). *A theory of the striatum*. New York: Pergamon.
- Wills, A. J. (2004). *New directions in human associative learning*. London: Routledge. doi:[10.4324/9781410612113](https://doi.org/10.4324/9781410612113)
- Woodruff-Pak, D. S. (1999). New directions for a classical paradigm: Human eyeblink conditioning. *Psychological Science*, 10, 1–3. doi:[10.1111/1467-9280.00096](https://doi.org/10.1111/1467-9280.00096)
- Yagishita, S., Hayashi-Takagi, A., Ellis-Davies, G. C., Urakubo, H., Ishii, S., & Kasai, H. (2014). A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science*, 345, 1616–1620. doi:[10.1126/science.1255514](https://doi.org/10.1126/science.1255514)
- Yin, H. H., Ostlund, S. B., Knowlton, B. J., & Balleine, B. W. (2005). The role of the dorsomedial striatum in instrumental conditioning. *European Journal of Neuroscience*, 22, 513–523. doi:[10.1111/j.1460-9568.2005.04218.x](https://doi.org/10.1111/j.1460-9568.2005.04218.x)
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46, 441–517. doi:[10.1006/jmla.2002.2864](https://doi.org/10.1006/jmla.2002.2864)
- Zakrzewski, A. C., Johnson, J. M., & Smith, J. D. (2017). The comparative psychology of metacognition. In J. Call, G. M. Burghardt, I. M. Pepperberg, C. T. Snowdon, & T. Zentall (Eds.), *APA handbook of comparative psychology: Perception, learning, and cognition* (pp. 703–721). Washington, DC: American Psychological Association.