COOPERATION CAME FIRST: EVOLUTION AND HUMAN COGNITION

STEVEN C. HAYES AND BRANDON T. SANFORD

UNIVERSITY OF NEVADA, RENO

Contextual behavioral perspectives on learning and behavior reside under the umbrella of evolution science. In this paper we briefly review current developments in evolution science that bear on learning and behavior, concluding that behavior is now moving to the center of evolution studies. Learning is one of the main ladders of evolution by establishing functional benchmarks within which genetic adaptations can be advantaged. We apply that approach to the beginning feature of human cognition according to Relational Frame Theory: derived symmetry in coordination framing. When combined with the idea that cooperation came before major advances in human cognition or culture, existing abilities in social referencing, joint attention, perspective-taking skills, and relational learning ensure that the behavioral subcomponents of symmetrical equivalence relations would be reinforced. When coordination framing emerged and came under arbitrary contextual control as an operant class, a template was established for the development of multiple relational frames and the emergence and evolutionary impact of human cognition as we know it. Implications of these ideas for translational research are briefly discussed.

Key words: relational frame theory, language, evolution, symmetry, cooperation, eusociality

When viewed as a contextual behavioral science, behavior analysis is part of evolution science. In major areas, its arguments have been made on that basis, and developments in evolutionary thinking are taken seriously. In the area of human language and cognition that has been true as well (e.g., Catania, 2001; Skinner, 1981). This paper examines some of the recent developments in evolution science, and attempts to apply them to the central idea in Relational Frame Theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001), a contextual behavioral theory of language and cognition. It is our argument that symbolic behavior is a recent evolutionary adaptation that coopted operant learning and social behavioral processes that were already established, but that required a highly cooperative species for this adaptation to be selected. Stated simply, cooperation came first, and the core behavioral process underlying human language and cognition from an RFT perspective was selected as an extension of cooperation.

Contingency Learning and Evolution

Operant and classical conditioning are likely about 520–545 million years old. The main reason to suggest that age is that sponges, fish, comb jellies, placozoa, or other organisms that evolved before the Cambrian period do not show evidence of contingency learning, while animals that evolved after that time and have been properly tested, do (Ginsburg & Jablonka, 2010). All neural animals show long term sensitization and habituation (Razran, 1971) but the adaptations to the environment these more primitive learning processes provide are too limited for major changes in behavior to occur within the lifetime of individuals. Before the arrival of complex forms of learning, the behavior of organisms dominantly had to be modified based on differential survival and reproduction across individual lifetimes. With the arrival of contingency learning and the possibility of selection by consequences in an ontogenetic sense, however, the relation between organisms and their environment fundamentally changed (Dennett, 1995; Schneider, 2012). Organisms could seek out or avoid situations that provided specific positive or negative consequences of importance, such as approaching new food sources, or adopting new behavioral means of avoiding predation.

This effect of learning on behavior places it at the center of evolution itself (Plotkin, 1988; Schneider, 2012). This is not a new idea but we are beginning to understand how profoundly it is true. Operant conditioning meant that organisms could select and construct their own environments. Schneider provides a useful example of the importance of niche selection in the evolution of the flamingoes’ beak (2012,
A similar but more complex process is shown in niche construction, in which learned behavior creates long-lasting changes in environmental conditions, which in turn alter the selective effects of the environment (Odling-Smee, Laland, & Feldman, 2003). Many examples of niche construction are dependent on learning, particularly in humans (Flynn, Laland, Kendal, & Kendal, 2013). For example, the development of lactose tolerance in human adults has co-evolved with the learned ability to domesticate milk-producing animals (Mace, Jordan, & Holden, 2003). This is hardly an isolated example. As Laland, Odling-Smee, and Myles (2010) note:

Over the past 50,000 years, humans have spread from Africa around the globe, experienced an ice age, begun to exploit agriculture, witnessed rapid increases in densities, domesticated hundreds of species of plants and animals and, by keeping animals, experienced a new proximity to animal pathogens. Each of these events represents a major transformation in human selection pressures, and all (except the ice age) have been self-imposed (p. 140).

Behavior and learning are also central to strategies of predation and the avoidance of predation. A recent experimental evolutionary study showed that changes in predators’ behavior can establish dramatically different selection pressures that favor members of the population with more behavioral flexibility in responding to the new strategies employed by the predators, which in turn can drive rapid morphological change (Losos, Schoener, Langerhans, & Spiller, 2006). These researchers concluded: “...behavioral flexibility indeed may often be the key in driving extremely rapid reversals in evolution” (p. 1111).

Learning and the environment are known to have a direct and long-lasting impact on the epigenetic regulation of genes themselves. It is now unquestioned that epigenetics is an inheritance system in its own right (Jablonka & Lamb, 2005). Epigenetic processes such as DNA methylation and histone acetylation appear to be key features of the biological processes that regulate the long-term effects of learning (Miller, Campbell, & Sweatt, 2008). Some of these effects are multigenerational; epialleles can be stable across many generations even in the absence of extensive DNA variation (Johannes et al., 2009). Mice exposed to chronic and unpredictable aversive events, for example, not only show characteristic methylation in the promoter of several genes in the germline, their offspring continue to show comparable patterns of methylation and resulting changes in gene expression (Franklin et al., 2010). Similarly dramatic examples exist in the processes that have been shown to underlie the impact of the environment on an ability to learn. When young knock-out mice without a gene known to support the ability to learn were exposed to an enriched environment containing novel objects, elevated social interactions, and voluntary exercise, not only did they show epigenetic changes leading to an enhanced ability to learn despite the genetic defect, so too did their offspring (Arai, Li, Hartley, & Feig, 2009).

Due in part to the known relation of epigenetics to health and disease (Feinberg, 2007), the increase in attention to epigenetics in biological science is startling: a search of the Web of Science shows that the number of articles on the topic has increased from a handful in the early and mid-1990’s to more than 1,400 in 2012; citations to this work were virtually absent until the late 1990’s—but there were more than 30,000 citations to it in 2012 alone. From a behavioral point of view what is most remarkable is that biologists themselves recognize that understanding and controlling these epigenetic effects will require knowledge about the role of environment and behavior: “Epigenetics represents the long sought after molecular interface mediating gene–environmental interactions during critical periods throughout the lifecycle” (Mehler, 2008, p. 305). The epigenetic changes...
produced by learning appear to be part and parcel of larger patterns of evolutionary developmental processes (West-Eberhard, 2003) that link phenotypic plasticity to genetic assimilation (Pigliucci & Murren, 2003). Even before the mechanistic details are determined, the larger picture is rapidly becoming clear. Learning is a key part of the scaffolding of evolution in complex organisms. In essence, this recognition puts contextual behavioral approaches into the center of evolution science itself.

There are other trends in evolution science converging on the centrality of learning and behavior but niche selection, niche construction, response flexibility in predation and predation avoidance, epigenetic inheritance, and genetic assimilation are excellent examples of the kinds of areas in evolution science that have emerged to change the intellectual landscape. Historically, evolutionists have sometimes tended to view learning and behavior primarily as the output of genetic evolution. To the contrary, we now know that learning and behavior are drivers of genetic evolution itself. Indeed, major evolutionary biologists now argue that learning was itself responsible for the “Cambrian explosion”—the increase in speciation and phenotypic diversity that occurred 520–540 million years ago (Ginsburg & Jablonka, 2010).

This is a profound change from a time when evolutionary reasoning was widely used to diminish the importance of general learning processes. Seligman (1970), in his highly influential article on the “generality of the laws of learning,” used evolutionary ideas to claim that phenomena such as taste aversion (e.g., Garcia, Ervin, & Koelling, 1966) show that “we have reason to suspect that the laws of learning discovered using lever pressing and salivation may not hold” (p. 417). The version of general process learning theory that Seligman challenged was one that never had many thoughtful modern advocates: that all events are equally associable regardless of the evolutionary history of the stimuli or organisms involved. Nevertheless, criticisms such as these had an impact: they pushed those interested in general learning processes unwillingly away from the evolution science table.

Advocates of Evolutionary Psychology (Tooby & Cosmides, 1992) took this idea a step further by opposing general process ideas from either behavioral or cognitive perspectives, arguing that “massive modularity” was the only possible way to understand the development of complex behavior. In their hands, human cognitive abilities were based on “a set of information-processing machines that were designed by natural selection to solve adaptive problems faced by our hunter-gatherer ancestors” (Cosmides & Tooby, 1997, p. 1).

Today many evolutionists view “massive modularity” as empirically unsupported (e.g., Bolhuis, Brown, Richardson, & Laland, 2011) and evolutionary biologists themselves openly advocate the integration of general process learning principles and evolutionary thinking (e.g., Bolhuis et al., 2011; Jablonka & Lamb, 2005). This change is not intellectually surprising to behavior analysts and contextual behavioral scientists, although perhaps the speed of it is a bit disorienting. Behavior analysis has always thought of itself as linked to the evolution sciences writ large (Catania, 2001; Donahoe, 2012), based on the scientific integration via selection by consequences across levels of analysis. Skinner (1981) argued that thinking of behavior in evolutionary terms was a key step in rising to the challenge of the human condition:

Selection by consequences is a causal mode found only in living things or in machines made by living things. It was first recognized in natural selection, but it also accounts for the shaping and maintenance of the behavior of the individual and evolution of cultures. In all three of these fields, it replaces explanations based on the causal modes of classical mechanics. The replacement is strongly resisted. Natural selection has now made its case, but similar delays in recognizing the role of selection in the other fields could deprive us of valuable help in solving the problems that confront us. (p. 501)

Contextual behavioral approaches bring a great deal to the table that can build on this perspective, particularly now that the political tide is turning. The precision and scope of behavior-analytic methods and concepts is an advantage, as is the large, systematic body of work. Nevertheless, changes are also needed to take advantage of this opportunity and there are lessons that needed to be learned.

The first needed change is the active embrace of a form of multidimensional evolutionary thinking that includes the possibility of modifications of general learning processes themselves. In psychology those who have most openly
declared an interest in evolution science have tended to over-emphasize the specialized properties of closed behavioral systems and the evolutionary history that might have established those properties (e.g., Barkow, Cosmides, & Tooby, 1992; Pinker, 1997; 2003), while under-emphasizing or even dismissing their possible relationships to open learning processes (Pinker & Bloom, 1990). Evolution was for many years dramatically gene-centric, with the exception of cultural evolution in the form of memes. Ontogenetic evolution was virtually ignored, not just in learning but in other developmental processes (West-Eberhard, 2003).

Historically, at times behavior analysis seems to have made the opposite error, at least in emphasis. A good example is provided by the transcript of the recorded interview between B. F. Skinner and E. O. Wilson, in which almost every specialized, evolutionarily established behavior put forward by Wilson was promptly interpreted by Skinner in operant terms (Naour, 2009). In examining the history of the odd breach between behavior analysis and evolutionary perspectives in psychology, a good starting point is “everyone was wrong and progress requires movement on all sides” (Wilson, Hayes, Biglan, & Embry, in press, ms page 23).

In defending the importance of general learning processes, behavioral psychologists have at times failed to acknowledge that these processes themselves can be modified by evolution in specific contexts. This goes beyond the more passive but important point we will return to later, that what is specialized is often built on structures or processes that initially had other functions; rather, general learning processes are especially suited to evolutionary modification. This is why: Learning provides overall functionality to complex sequences of behavior in a way that would be difficult to assemble piecemeal via genetically driven behavioral adaptations. When learned patterns can be repeated across lifetimes they become an “adaptability driver” for genetic assimilation by providing an end-point against which genetically and epigenetically-driven adaptations can be compared via selection (Bateson, 2005; 2013). This kind of benchmarking appears to be one of the functional processes that explain the so-called “Baldwin effect” (Dennett, 2003). Once a functionally important learned behavioral pattern exists, if more efficient or reliable genetic or epigenetically driven short cuts emerge within that pattern, conditions are established that are analogous to those that have long been known in the experimental laboratory to lead to genetic assimilation (e.g., Waddington, 1953).

It is important to hold on to evolutionary insights—that what is new often coopts what is old; that what is special is often a modification of what is not. Behavior analysis has an enormous knowledge base to apply because basic behavior analytic research over the last 75 years, particularly with nonhuman animals, has been so successful in elucidating the properties of those 500+ million-year-old processes of operant and classical conditioning. A bottom-up principle-based account of learning that has focused on isolating and manipulating simple responses, such as lever presses and key pecks, has provided a framework within which evolutionary modification can be seen. Evolutionists often misunderstand this research strategy, because from the outside it appears to instantiate a kind of evolutionarily naïve “blank slate” assumption in which the properties of stimuli or responses, or the differences between species do not matter (e.g., Pinker, 2003). To the contrary, looking for generality across the tips of evolutionary branches using fairly simple responses was a research strategy, not a naïve or careless assumption about the lack of differences between species, stimuli, or forms of responding (Hayes & Hayes, 1992). Skinner explicitly agreed that “the behavior of any species could not be adequately understood, predicted, or controlled without knowledge of its instinctive patterns, evolutionary history, and ecological niche” (1979, p. 209). But that does not mean that general learning processes do not exist—and if they do, that the more specifically focused research strategies of ethologists were unlikely to discover them.

Once it is realized that in complex areas it is far easier to construct adaptive specialization by building on the functionality of general processes, a focus on general learning processes hardly shows the evolutionary naïveté critics claim. To the contrary, such a focus has placed behavior analysts’ hands on a primary ladder of evolution, provided only that they are willing to acknowledge that once that ladder is in place, in key contexts evolution can replace the rungs in ways that no long rely on ontogenetic adaptations.

Over the last decade there has been a notably increased willingness to think of evolution in a multidimensional way that goes beyond genes,
or genes and memes, to include epigenetics, ontogenetic change in behavior, symbolic evolution, and the impact of behavior on the fixed environment (Jablonka & Lamb, 2005). Maintaining a focus on all of these dimensions helps situate learning into the unified fabric of evolutionary development, and makes it natural to include possible modifications of general learning processes as a result. What is required is a more systemic approach, but such a systemic focus is deeply resonant with the assumptions of contextual behavioral perspectives because it means maintaining one’s grip on the whole organism acting in and with a situational and historical context:

While the behavior of whole animals can be informed by knowledge of the underlying mechanisms, the process of reassembly can only be conducted at the level of the whole organism. Except in a small number of cases, the notion of genes, however they are defined, coding for an organism’s characteristics has been discredited. Referring to genes as being adapted to the environment no longer makes any sense. Adaptation is at the level of the phenotype (Bateson, 2013, p. 8).

A telling example of how learning processes can be thought of in a multidimensional way is provided by taste aversion, the very area which initially seemed so far outside the known parameters that it raised questions about the relevance of general process learning research. As it was studied, data on taste aversion rapidly emerged that both extended the phenomenon (e.g., Coombes, Revusky, & Lett, 1980; Kalat & Rozin, 1973) and better linked it to other concepts in conditioning such as overshadowing, latent inhibition, blocking, potentiation, or interference (e.g., Bouton, 1993). Today, taste aversion is so comfortably considered to be a special form of classical conditioning that it is routinely used as a preparation for studying learning and memory and its biological substrates (Welzl, D’Adamo, & Lipp, 2001). Furthermore, by understanding how the parameters of general learning processes were modified in several areas based on the evolutionary relevance of these changes, researchers have been able to apply this knowledge to human problems in areas such as the development of food preference or the avoidance of conditioned nausea in cancer treatment (Birch, 1999; Stockhorst, Wiener, Klosterhalfen, Klosterhalfen, Aul, & Steingruber, 1998). In essence, what was specific is now seen to have evolved as a modification of what was general. It is a theme of this Perspectives on Behavior piece that the challenge of human language and higher cognition can be addressed in much the same way.

A second change that appears to be needed for behaviorists to take advantage of the current zeitgeist in evolution science is to take multilevel selection more seriously. Changes that mark major evolutionary transitions, such as the emergence of eukaryotic cells, multicellular organisms, or eusocial species, often appear to involve a shift in the unit of selection from individuals to groups. This can happen when competition between groups dominates over within-group competition, and when means are found to dampen individual selfishness. The interests of the individual never disappear—in the case of the “group” represented by a eukaryotic cell, mitochondria will occasionally produce male cytoplasmic sterility; in the case of the “group” represented by multicellular organisms, individual cells will occasionally become cancerous—but like a teeter-totter in an evolutionary time frame the group as a unit of selection can hit a tipping point causing a fundamental and permanent shift (Wilson & Wilson, 2007; 2008). As this old idea has been revived and extended empirically and analytically, multilevel thinking has also become much more dominant in areas such as the evolution of altruism (Nowak, Tarnita, & Wilson, 2010). Multilevel selection is also key to contextual behavioral perspectives on derived relational responding as the core of human language and cognition, a topic to which we now turn.

The Challenge of Human Language and Cognition

The single biggest challenge to behavior analysis being seen as a member of the evolution sciences writ large is the same challenge Skinner (1938, p. 442) anticipated over 75 years ago: human language. The present issue of the Journal of the Experimental Analysis of Behavior touches on the kinds of empirical findings that are relevant to this topic, but a more profound change is needed beyond additional research on stimulus–stimulus relations, derived relational responding, or any other single research area. Behavior analysts need a comprehensive and evolutionarily sensible approach to human
language. The contextual behavioral science wing of behavioral thinking has been pursuing RFT as that vehicle, and the present article will elaborate on a key aspect of the RFT analysis, but regardless of one’s approach to the topic an evolutionarily sensible analysis in this area is central (Catania, 2001).

It was not just specialized learning phenomena that helped push behavioral psychology away from the evolution science table. Perhaps even more central was the conclusion that general learning processes were irrelevant in accounting for human language and cognition. Here Seligman’s key article (1970) is again illustrative. After reviewing studies that seemed to show that human language required little specific training, he concluded, “instrumental and classical conditioning are not adequate for an analysis of language” (p. 414). Behavior analysts have fought against that conclusion, but Skinner’s solution (1957) never captured the imagination or research attention of mainstream scientists interested in the evolution of human language and cognition. More than 50 years later there are few reasons to believe it ever will.

It has to be admitted that there is a kind of disconnect between evolutionists interested in human cognition and evolutionary analyses drawn from comparative approaches more generally. Nonhuman animals do extraordinary things that appear similar to human language, and that likely overlap with some of its properties. A border collie has learned to respond to 1,000 names and can respond to them in a variety of sequences (Pilley & Reid, 2011); Irene Pepperberg’s African-Grey parrot shows numerical abilities similar to chimpanzees or very young children (Pepperberg & Gordon, 2005); the careful work of Duane and Sue Savage-Rumbaugh or of Allen and Trixie Gardner has documented a number of abilities in apes such as delay of gratification (Beran, Savage-Rumbaugh, Pate, & Rumbaugh, 1999) or the use of categories when describing events or answering questions (Gardner, VanCauterfort, & Gardner, 1992). These are remarkable and interesting behavioral performances, and modern behavior analysts have at times been willing to describe them as demonstrating language or even referential language (Schneider, 2012, p. 145).

Conversely, after over 25 years of study, behavior analysts have been unable to find systematic and replicable evidence even of symmetry in nonhuman animals assessed the way symmetry is assessed in stimulus equivalence or other derived relational research (Lionello-DeNolf, 2009), an ability that is readily shown in human infants (Luciano, Gómez-Becerra, & Rodríguez-Valverde, 2007; Peláez, Gewirtz, Sanchez, & Mahabir, 2000). Behavior analysts have found in many studies that stimulus equivalence or other derived relational responses appear to relate to human linguistic and cognitive performance in both typically developing and disabled populations, both infants and adults (e.g., Devany, Hayes, & Nelson, 1986; McLay, Sutherland, Church, & Tyler-Merrick, 2013; O’Hora, Pelaez, Barnes-Holmes, Rae, Robinson, & Chaudhary, 2008; Peláez et al., 2000). Furthermore, training in derived relational responding has been consistently shown to increase verbal and cognitive performance in normal and disabled populations, both infants and adults (e.g., Cassidy, Roche, & Hayes, 2011; Luciano et al., 2007; Persicke, Tarbox, Ranick, & St. Clair, 2012; Weil, Hayes, & Cappuro, 2011) Thus, behavior analysis as a field seems torn about how to proceed conceptually and pragmatically: Nonhuman animals are regularly said to be showing language skills, but the skills known to be key to language in human subjects, nonhuman animals seemingly do not display.

The purpose of the present paper is more humble than that of reaching a resolution to this conundrum. For over 20 years (e.g., Hayes & Hayes, 1992), those behavior analysts interested in RFT have argued that the core of human language and cognition is learned operant patterns of derived relational responding under arbitrary contextual control. Whatever else one may say about it, there is no doubt that RFT is a vigorous area of applied and basic contextual behavioral research (Dymond, May, Munnelly, & Hoon, 2010). It has been applied to a wide range of language phenomena in both basic and applied areas (Hayes et al., 2001; Roche & Dymond, 2013), so far without findings that contradict its central claims. In the remainder of this paper we will attempt to link derived relational responding to the evolutionary ideas we have been developing, and briefly contrast it with the evolutionary perspective within Skinner’s (1957) approach.

**Evolution and Derived Relational Responding**

Evolutionary steps represent changes to existing phenotypes—not *de novo* creation. For example, the formation of the Giant Panda’s
Ailuropoda melanoleuca) pseudo-thumb, which has been noted to be an especially refined and unique manipulation system among mammals (Endo et al., 1999), has been used to illustrate the process by which environmental changes drive changes in morphology which were previously used for an entirely different function (Gould, 1980). The Giant Panda’s false thumb is actually an extension of the radial sesamoid, a small bone in the wrist, and grants the animal prehensile abilities which would not otherwise be possible given the fixed position of the true thumb. This adaptation is thought to be an exaggeration of the already enlarged radial sesamoid bone present in ursid species and evolved in response to a more herbivorous diet which selected for increases in dexterity (Davis, 1964; cf., Salesa, Anton, Peigne, & Morales, 2006). The existing morphology of ursids (i.e., the large radial sesamoid bone), in the context of an increasingly herbivorous diet and a fixed thumb, led to a situation in which increasingly exaggerated successions of the wrist bone could be selected.

Metaphorically, we can carry this example to the phenotypic trait called human language. Behavior analysts have long struggled to come to an approach to human language that works, but they have almost always approached the topic within a broadly evolutionary framework, in which the task is to identify what is characteristic of the domain and to identify the conditions that could have given rise to those properties. Metaphorically they have sought out the “enlarged radial sesamoids”—that is, the behavioral, morphological, or cultural processes—that could have supported the evolution of human language.

If operant conditioning is 500+ million years old, human language is by comparison just an eye-blink. Modern human beings are almost certainly under 200 thousand years old (McDougall, Brown, & Fleagle, 2005), and analyses of reliable changes in language structure suggest that human language is perhaps only about 100,000 years old (Nichols, 1992). Even if other hominids had forms of human language, operant conditioning is incomparably more ancient than human language.

Skinner (1938) wondered early on if operant conditioning could account for language. At its most basic, his eventual account of the domain (1957) answered “yes” by combining three forms of what Skinner termed “levels of variation and selection” (1981, p. 502): contingencies of survival, reinforcement, and cultural evolution. Verbal behavior was behavior shaped by consequences mediated by others who were specifically trained to do so. The behavior of the listener qua listener was not verbal: it was established culturally and was regulated by verbal stimuli, viewed simply as the products of the speaker’s verbal behavior (Skinner, 1957, p. 34). We will not here trot out our behavioral concerns over the specifics of this approach (those arguments have been presented in detail elsewhere, such as in Hayes et al., 2001; or Hayes & Hayes, 1992). What is more important in the current context is to note that Skinner’s analysis was an evolutionary account, and it was argued entirely as an evolutionist would argue it. He appealed to the interaction between genetically established processes, ontogenetic evolution, and cultural evolution:

The human species presumably became much more social when its vocal musculature came under operant control. Cries of alarm, mating calls, aggressive threats, and other kinds of vocal behavior can be modified through operant conditioning, but apparently only with respect to the occasions upon which they occur or their rate of occurrence. The ability of the human species to acquire new forms through selection by consequences presumably resulted from the evolution of a special innervation of the vocal musculature, together with a supply of vocal behavior not strongly under the control of stimuli or releasers—the babbling of children from which verbal operants are selected. No new susceptibility to reinforcement was needed because the consequences of verbal behavior are distinguished only by the fact that they are mediated by other people. The development of environmental control over the vocal musculature greatly extended the help one person receives from others. By behaving verbally people cooperate more successfully in common ventures.

(Skinner, 1981, p. 502)

Lest the reader fail to appreciate what was being argued, Skinner later in the same paper summed up his entire approach to complex human behavior, including verbal behavior, in terms no evolutionist could misunderstand:

In summary, then, human behavior is the joint product of (i) the contingencies of
survival responsible for the natural selection of the species and (ii) the contingencies of reinforcement responsible for the repertoires acquired by its members, including (iii) the special contingencies maintained by an evolved social environment. (Ultimately, of course, it is all a matter of natural selection, since operant conditioning is an evolved process, of which cultural practices are special applications.)


For Skinner, what was new about verbal behavior was not the modification of conditioning processes in any way, but their extension to vocal musculature, and their occurrence in the context of the cultural evolution of a verbal community. Cooperation was a result of these processes—not the other way around. The focus on operant control over the vocal musculature has continued among behavior analysts (e.g., Schneider, 2012) but it is not entirely clear why since deaf children can readily learn human language through signs, and operant control over mere clicks provides the ability to communicate through Morse code. It is also worth noting that Skinner was never very clear about how the cultural evolution of a verbal community occurred. This is important because behavioral “accounts of language origins must show not only how ontogenic selection could have been selected phylogenetically but also how ontogenic and cultural selection could have evolved concurrently” (Catania, 2001, p. 55).

It is here we can return to the issue of multilevel selection. Evolutionary biologists argue that there are at least three distinctive (but with the possible exception of cognition, not unique) characteristics of human evolution: cognition, culture and cooperation—the so called “three C’s” (Wilson, 2007). It is not uncommon to argue, as Skinner does here, that cooperation is a human extension of cognition and culture. But the re-emergence of the plausibility of group selection in evolution science raises another possibility: cooperation came first and thus “the three C’s of human evolution are all manifestations of one C—cooperation” (Wilson, 2007, p. 154). That possibility has important implications for derived relational responding, as we have discussed previously (Hayes & Long, 2013) but will expand on here.

Multilevel selection theory (O’Gorman, Wilson, & Sheldon, 2008) suggests that selection can act at different levels of organization (such as genes, cells, individuals, groups) and that the balance tips toward higher levels of organization only when competition occurs at that level and selfishness at lower levels of organization is restricted. Thus, from the point of view of multilevel selection, cooperation, prosociality and altruism can be accounted for by the degree to which these behaviors promote success of competitive groups in contexts that restrict within-group competition. Experimental research has confirmed that differential social outcomes occur under these selection conditions (e.g., Muir, 2009).

A useful example of extreme sociality that appears to be driven by multilevel selection is provided by eusocial species such as ants or bees. Eusocial species have cooperative forms of brood care in which overlapping generations divide care for offspring among reproductive and nonreproductive individuals (Crespi & Yanega, 1995; Nowak et al., 2010; Sherman, Lacey, Reeve, & Keller, 1995). While the level of relatedness (i.e., as proposed by kin selection) is relevant to the development of cooperative groups, they may also form among unrelated members of a species (Clutton-Brock, 2009; Nowak et al., 2010).

Eusociality is incredibly rare with only 15 of the 2,600 currently living taxonomic families containing eusocial species (Gadagkar, 2001; Michener, 1974; Wilson, 1971). When it occurs it is tremendously adaptive, however. As an illustrative example, “... although [eusocial] ants and termites together compose only 2% of the ≈900,000 insect species known globally, they make up more than half the insect biomass” (Wilson & Hölldobler, 2005, p. 13370).

Human beings are arguably a eusocial species (Foster & Rattrieks, 2005; Nowak et al., 2010). Whether or not that is ultimately agreed to be the case, it is clear that human beings are by far the most cooperative primates (Jaeggi, Burkart & Van Schaik, 2010). Great apes are much more individualistic than human beings in areas such as foraging and sharing (Tomasello, Melis, Tennie, Wyman, & Herrmann, 2012). Humans evolved in the context of small groups, bands, and tribes (Chapais, 2013) and human society is fundamentally based on cooperation (Nowak, 2006). In line with some defining features of eusociality, in tribal human groups, postmenopausal women share in childcare after losing reproductive abilities, benefiting the child-rearing
of the group as a whole and allowing adult females more time to rear younger offspring (Hawkes, O’Connell, Blurton Jones, Alvarez, & Charnov, 1998; Shanley & Kirkwood, 2001). The high level of alloparenting, or cooperative maternal care, seen in humans similarly requires a sacrifice in individual reproductive opportunity while benefitting the group (Hrdy, 2009; Burkart, Hrdy, & van Schaik, 2009). Evolutionary biologists believe these conditions had a profound effect: “… as egalitarianism became sufficiently established, genetic evolution started to reshape our minds and bodies to function as team players rather than competing against members of our own group” (Wilson, 2007, p. 165).

If we assume that cooperation came first, we can explore the degree to which cooperation and the other behavior it impacts set an evolutionary context in which processes of operant learning itself could have been coopted to produce human language, defined as it has been by RFT. The evidence is unquestionable that verbal humans do in fact derive relations even among arbitrary events, mutually and in combinations, with the functions of these events changing as a result (Hayes et al., 2001). The core phenomenon was shown with stimulus equivalence (Sidman, 1971), but it has expanded over the years to include many other types of derived relations, such as difference, opposition, comparison, and deictic relations such as I-You, Here-There, or Now-Then that are central to perspective taking (see Roche & Dymond, 2013, for a recent review).

As a practical example of arbitrarily applicable relational responses, consider comparative relations. Very young children who know that coins can be used to buy candy will prefer a quarter over a nickel over a dime based on the nonarbitrary relation of relative size. When comparative relations become arbitrarily applicable, that same child will prefer a quarter over a dime over a nickel. Both in typically functioning and disabled children, the performances are no longer dependent on nonarbitrary stimulus relations once the core relational skill is trained (Berens & Hayes, 2007; Gorham, Barnes-Holmes, Barnes-Holmes, & Berens, 2009); even adults with deficits in relational reasoning benefit from training programs that focus on the flexibility and fluency of arbitrary comparative relations (Vitale, Campbell, Barnes-Holmes, & Barnes-Holmes, 2012).

The simplest of all the response forms in arbitrarily applicable derived relational responding is stimulus equivalence. This is so in one sense because it is the only derived relational performance in which the trained and derived relations are the same regardless of the size of the resulting relational network. It also emerges first developmentally (Lipkens, Hayes, & Hayes, 1993).

From an RFT point of view the key response feature leading to stimulus equivalence is relational framing: deriving a mutually entailed relation under arbitrary contextual control, that combines into networks, and changes the response functions of related events (Hayes et al., 2001). Before we explore the evolutionary sense of that claim, however, we need to bring in a small set of related behavioral skills, some of which are supported by cooperation and eusociality. Although even human infants generally outperform nonhuman primates on each of the tasks we will describe in the sections that follow (e.g., Caron, Kiel, Dayton, and Butler, 2002; Mumme, Fernald, & Herrera, 1996; Talkington, Taglialatela, & Lewis, 2013), these abilities are shared with other nonhuman animals.

Use of vocalizations to regulate the behavior of others. Primates as well as many other nonhuman animals use vocalization as a method of social regulation (Masataka, 2003; Snowdon, 1997). For example, attacked chimpanzees emit specific vocalizations that function to solicit support from bystanders (de Waal & van Hooff, 1981). These vocalizations are context-specific and can vary systematically based on the severity of the aggression and audience (Slocombe & Zuberbühler, 2007). Nonhuman primates also use meaningful combinations of vocalizations for these social purposes (Arnold & Zuberbühler, 2006; 2008; Schel, Tranquilli, & Zuberbühler, 2009). Nevertheless, the nonhuman primate vocalizations are a fixed set, linked to emotional situations, and are difficult or impossible to modify with training (Tomasello, 2008; 2009). It is important not to anthropomorphize the meaning of these vocalizations, given the tight stimulus control under which they are emitted. For example, macaques emit an alarm call when threatened, but a macaque mother seeing a predator threaten her offspring will not emit the call unless or until she herself is at risk (Cheney & Seyfarth, 1990).

Social referencing. Social referencing — seeking of information from another individual, so as to
use that information to respond to an object or event is well documented in human infants (e.g., Adamson, 1996) but it is also evident in chimpanzees (Russell, Bard, & Adamson, 1997) and dogs (with the latter most likely being due to the coevolution of canines with humans—Merola, Prato-Previde, & Marshall-Pescini, 2012). For example, when young chimpanzees are presented with novel objects they look to caregivers and their response to the object is guided by the caregiver’s emotional expression (Russell et al., 1997).

**Joint attention and nonverbal forms of perspective taking.** Human infants are adept in responding to the gaze, reaching, or pointing of others as stimuli for directing shared attention or inferring intentionality (Tomasello, Carpenter, Call, Behne, & Moll, 2005; Woodward, 2005). Chimpanzees are less skilled but have several aspects of “theory of mind” skills (Call & Tomasello, 2008) such as being able to follow the gaze of others (Okamoto-Barth, Call, & Tomasello, 2007) or to return their attention to the looker if a salient stimulus is not then discriminated (Povinelli & Eddy 1996; Tomasello, Hare, & Fogleman, 2001). Great apes discriminate intentionality from the behaviors of others under some conditions, but do so at a lower level than young humans. For example, if a restrained human competitor unsuccessfully reaches for one of several buckets that visibly contains a banana, great apes will run to that bucket and retrieve the banana (Hare & Tomasello, 2004). Conversely, if the banana is hidden and a disinterested human third party reaches toward the correct bucket, examinations of the bucket fall to random levels (Tomasello, Call, & Gluckman, 1997).

**Nonarbitrary relational learning.** All complex animals show forms of relational learning. Monkeys trained to select the taller of two stimuli over a series of training trials will avoid a previously “correct” stimulus and select an even taller novel stimulus, showing that responding was controlled by relational rather than absolute properties of the stimuli (Harmon, Strong, & Pasnak, 1982). A wide variety of perceptual relations of this kind can be learned by nonhumans, especially primates (see Penn, Holyoak, & Povinelli, 2008 for a review and methodological criticism). What nonhumans apparently do not do, however, is to abstract from these relations in a higher-order way that is structurally systematic and inferentially productive (Penn et al., 2008). RFT is focused on precisely that difference.

**Relational Framing and the Extension of Human Cooperation**

Although cognition, culture, and cooperation are all characteristic of humans, the phylogeny of relational framing seems more readily understandable if it is assumed that cooperation came first (Hayes & Long, 2013; Wilson, 2007). The core process of relational framing is deriving a symmetrical or mutually entailed relation. Equivalence between a name and an object provides the simplest example and one that is critical to symbolic communication. A member of a nonhuman species learning to select an object when hearing or seeing a characteristic sign, will not thereby know to produce that sign when seeing the object; even “language trained” chimpanzees fail this task (Dugdale & Lowe, 2000), but human infants do not (Lipkens et al., 1993; Luciano et al., 2007; Peláez et al., 2000). Our task is to explain this difference as a possible evolutionary sequence.

Suppose a speaker sees an apple and says “apple.” The ability to emit characteristic vocalizations under the control of characteristic stimuli is not at issue; that is close to behaviors that exist in many nonhuman species. Suppose later that same speaker sees an apple that is out of reach and reaches for it or points toward it while saying “apple.” In a highly cooperative and eusocial primate, listeners within arm’s reach would likely then reach for that apple and give it to the speaker.

This is not far from things human infants do before they acquire significant linguistic ability. If an adult who has been playing with a baby points to a toy on the floor while cleaning up the toys, the baby is likely to pick it up and put it in the clean-up basket; if a third party enters the room and points to the same toy, the baby will look, and may pick it up, but will not put it away (Liebal, Behne, Carpenter, & Tomasello, 2009). Stated loosely, the baby understands that one of the adults, but not necessarily the other, “wants the toy to be put away.”

Thus, if combined with perspective taking and similar skills reviewed above, cooperation makes it likely that if a human listener hears a speaker say something characteristic while reaching unsuccessfully for an object, the speaker’s action is likely to be reinforced by the provision of the object. Indeed, human infants will readily
reinforce requests for sharing even at the cost of personal reinforcers (Schmidt & Sommerville, 2011). Thus the see object \(\rightarrow\) emit symbol relation of the speaker will be reinforced. Instances of cooperation on the part of the listener are in turn likely to be reinforced. Infants are more likely to give rewards to a puppet that was observed being helpful toward others while taking away rewards from a puppet that had previously acted harmfully toward others (Hamlin, Wynn, Bloom & Mahajan, 2011). Thus the hear symbol \(\rightarrow\) provide object relation of the listener will be reinforced. The entire exchange will build cooperation, perspective taking, and joint attention as patterns that are maintained within the group because it is a functionally useful communication exchange.

If we unpack this highly likely sequence it means that in the context of high levels of cooperation, and adequate skills in joint attention, social referencing, and perspective taking, any characteristic vocalization in the presence of a desired object would likely lead to reinforced instances of symmetry or mutual entailment. In the absence of high levels of cooperation, however, this sequence would not occur. Thus, it is critical to our analysis that we assume that cooperation comes first and be widespread within the group. The initial rise of cooperation, however, can be explained by the impact of multilevel selection on human evolution (Nowak et al., 2010): human cognition is not required. Thus, the sequence of events in this analysis does not appeal to any process that is implausible from an evolutionary perspective.

In the absence of perspective taking, joint attention, social referencing, and cooperation this interlocking contingency alone apparently does not lead to a relational operant, since each side of this process has been extensively trained for many years in so called “language trained” chimpanzees through the artificial intervention of humans without it leading to derived symmetry (Dugdale & Lowe, 2000). But in the context of sufficient perspective taking and related skills, exposure to the operant contingencies and these subcomponents of the speaker (object \(\rightarrow\) name) and listener (name \(\rightarrow\) object) roles could become integrated into a single relational unit in which each side of the exchange immediately implies the other for both participants. Trained A \(\rightarrow\) B relations imply B \(\rightarrow\) A relations at the neurobiological level in humans even before equivalence testing of the derived relation (Wang & Dymond, 2013; Yorio, Tabullo, Wainselboim, Barttfeld, & Segura, 2008). The operant contingency is mutual, and the skills noted earlier (especially perspective taking, but also social referencing, joint attention, nonarbitrary relational abstraction) may help make the relation itself mutual in an integrated psychological sense. Some degree of genetic assimilation was likely needed to increase the speed or efficiency of the derivation as well, but because cooperation and perspective-taking skills appear to have existed in hominids (Dunbar, 2003; Hare, 2011) these learning contingencies could have been in place for a very long time. Such a scenario meets the conditions under which learning provides a kind of benchmark within which evolutionary adaptations can occur. Bateson (2005) describes it this way:

The existence of a phenotype, acquired by learning, sets an end-point against which phenotypes that develop in other ways must be compared. The chances that all the necessary mutations and genetic recombinations would arise at the same time are very small indeed. In the natural world, if a spontaneously expressed phenotype is not as good as the learned one in the sense that it is not acquired more quickly or at less cost, then nothing will happen. If it is better, evolutionary change is possible. The question is whether the spontaneously expressed phenotype could evolve without the comparison. If learning involves several subprocesses, as well as many opportunities for chaining (the discriminative stimulus for one action becomes the secondary reinforcer that can strengthen another) then the chances against a spontaneously expressed equivalent appearing in one step are very small. However, with the learned phenotype as the standard, every small step that cuts out some of the plasticity with a simultaneous increase in efficiency is an improvement. (p. 36)

It is important to realize that genetic assimilation is not all-or-none. Coordinated communication of the sort described would be functional even before an abstracted relational operant existed. Early in the development of these patterns a speaker could ask a listener across a canyon or around a corner to be told what is there or to bring back an object, for example. Thus, this beginning core of human cognition would increasingly extend human cooperation itself.
These conditions meet the criteria for an “adaptability driver”: learning that provides functionality as a context for evolutionary adaptations. Just as in taste aversion, modifications of responsivity to preexisting units of learning would happen incrementally. Reinforcement of the subelements of a higher order relational operant (cf., Catania, 1996) are all likely in a eusocial (or at least highly cooperative) species with reasonable levels of joint attention, social referencing, and perspective taking. Evolutionary adaptations could then strengthen the genetic and epigenetic basis of the integration of these sub-elements into a relational response class under arbitrary contextual control. For example, as relevant communication practices within the group and perspective-taking skills strengthen, and individuals engage in many reinforced instances of both speaker and listener relating (object → name and name → object, respectively) seeing the object and knowing its name would psychologically imply hearing the name and envisioning the object, and vice versa. When one relational response entails the other by derivation as a learned operant class, arbitrary paralinguistic cues or verbal cues (such as “is” in the sentence “this is an apple”) that reliably predict the mutual relation could readily come to control the relational response, since there were no formal properties of the related events that were determining the interlocking contingencies between speakers and listeners in the first place.

This is in essence a description of the abstraction and contextual regulation of the key feature of RFT’s “frame of coordination,” regulated by cues established by social convention rather than by the form of the related events. Coordination framing alone would result in speaking and listening roles merged into a single communication system. Nonhuman speakers and listeners communicate in interlocking systems but “listeners acquire information from signalers who do not, in the human sense, intend to provide it” (Seyfarth & Cheney, 2003, p. 168; see Tomasello, 2008) in the sense that the vocalizations are under tight stimulus control rather than being reinforced by the listener. In humans, conversely, speaking with meaning and listening with understanding stimulates the same regions of the brain (Menenti, Gierhan, Segaert, & Hagoort, 2011; Segaert, Menenti, Weber, Petersson, & Hagoort, 2012), but because perspective-taking, joint attention, social referencing and similar processes are also being assimilated, it simultaneously explains why differences in human cognition are not restricted to semantic, lexical, or syntactic domains (Penn et al., 2008).

Once mutuality was well established, the other defining features of relational framing—combinatorial entailment, and transformation of stimulus functions (Hayes et al., 2001) would readily occur. In equivalence class formation combinatorial entailment is in a sense the repeated application of a mutually entailed relational response, and not many reinforced instances seem needed in humans for it to emerge once mutuality is robust (Lipkens et al., 1993). Thus, relational networks would emerge as a matter of operant learning and cultural evolution based on that learning, once derived symmetry was integrated into a response class. In the same way, a transformation of stimulus functions (i.e., altering perceptual, operant, or classically conditioned functions of events based on their derived relation with other events) would be extended and contextually controlled without additional evolutionary adaptations being required. Mutual entailment already includes such functional transformations. For example, once part of a derived stimulus relation, hearing “apple” has some of the visual functions of apples: the person looking for the object already “sees” it before it is seen. Bringing these under specific forms of contextual control should occur merely as a matter of operant abstraction.

A group with even a few speakers and listeners competent in coordination framing would be advantaged in their ability to compete with other groups due to the verbal extension of cooperation provided by these abilities. Once coordination framing was common within the group, equivalence relations would provide a template for other types of relational framing (comparison, opposition, hierarchical, deictic, and so on) as the initially merely cooperative community became a verbal community. In this analysis derived mutuality initially is based on the ability to see the speaker’s role from the point of view of the listener and vice versa as part of a cooperative act, but the core relational ability that goes across all relational frames is seeing each relata from the point of view of the other. That template is the sense in which big entails small, or here entails there, and so on. Thus, the perspective-taking skills inside
cooperative coordination framing provides a road map for relational framing in general.

If that is correct, the ability of groups to compete through better communication would over time focus more on the extension of relational learning to the temporal, conditional, and comparative framing central to problem solving, as a result of the competitive advantages between groups linked to the fluency and diversity of relational learning and cultural adaptations to support it. Human language enormously expands possible variation, and especially when environmental contexts change rapidly, cognitive sources of variation could be crucial. If a nonhuman primate learns eight separate sign → object relations, only these eight are available. If verbal humans learn these eight trained relations, several thousand derived relations emerge; every sign and relationship between and among the signs can be related one to the other in all directions, and to every object and relationship between and among the objects in all directions (cf., Deacon, 1997). These relational responses in turn would support skills in problem solving, group planning, prediction, and so on that would greatly advantage groups with more language skills. From the enormous increase in variation that language afforded cooperative groups, successful cultural practices would be selected and expanded (Barnes-Holmes et al., 2001).

This preliminary analysis shares some features with that of Catania (2001) who used the idea of shared echoics as setting the occasion for detecting mutual correspondence between sounds and events. In this view, echoics led to the detection of relations. While he did not call them frames of coordination, or extend their implications as relation responses, the analysis overlaps to a degree with the present account:

The significant consequences of verbal behavior range from such direct outcomes as getting something one has asked for to such indirect ones as hearing a remark relevant to something one had just said. If among these consequences we include the correspondences between sounds one has heard and sounds one has produced oneself, it is not too great a leap for us to extend such correspondences from phonemic to semantic and syntactic properties of verbal behavior. It is presumably important to discover that the relations among words and between words and things in one’s own behavior correspond to those relations in the behavior of others.

(p. 62)

Translational Research

Behavioral psychology has had a great deal of success in translating basic contingency learning principles into applied domains. Translating basic findings into application from research in stimulus equivalence (e.g., Fields et al., 2009) or derived relational responding is just beginning, but the outlines of success are unmistakable. As examples, RFT has been used successfully to assess and begin to remediate perspective-taking deficits in autistic children (Rehfeldt, Dillen, Ziomek, & Kowalchuk, 2007); to successfully predict dropout from drug treatment centers (Carpenter, Martínez, Vadhan, Barnes-Holmes, & Nunes, 2012); to predict social anhedonia (Vilardaga, Estévez, Levin, & Hayes, 2012); to increase intellectual performance (Cassidy et al., 2011); to guide the development of psychotherapy methods (Hayes, Strosahl, & Wilson, 2011); and to refine clinical components (Luciano et al., 2011).

There are three main reasons to suppose that translational research based on a contextual behavioral approach to human language and cognition is likely to be important. First, human language and cognition is an inheritance system in its own right (Jablonka & Lamb, 2005). Human language can produce enormous changes in behavioral variation, selection (e.g., via values and goals), and retention (via books, tapes, digital media and the like). Thus the rate of cultural development is likely to be far greater if this relatively new evolutionary stream is scientifically studied and better managed (Wilson et al., in press). Second, human language and cognition operates on other dimensions of evolution, impacting genes, epigenes, and behavior regulated by contingency learning, both within the lifetime of individuals and culturally. Relational operants can alter both classical conditioning processes (e.g., Dougher, Hamilton, Fink, & Harrington, 2007) and operant conditioning processes (e.g., Whelan & Barnes-Holmes, 2004). These interactions can create analytic problems but they can also create behavioral change: they need to be studied. Finally, unlike traditional cognitive perspectives, as a contextual behavioral approach to human
language and cognition, RFT retains a firm grasp on the history and circumstances that regulate the occurrence and form of human cognition and its impact on other events. This is a major advantage for translational work because applied workers (such as clinical psychologists) cannot change the behavior (including the verbal behavior) of others directly—all behavior change occurs by altering the context of action (Hayes & Brownstein, 1986).

**Conclusion**

There is nothing in the present analysis that is hostile to the animal learning tradition within traditional behavior analysis; indeed, the present analysis is dependent on findings from it. But evolution science itself suggests that human symbolic behavior has unique properties (Penn et al., 2008). No nonhuman animal has yet shown the defining features of relational framing, and the centrality of relational framing to complex human behavior is very evident in the growing research base of contextual behavioral approaches both applied and basic (Roche & Dymond, 2013). Those data cannot be ignored, despite their implications: human language and cognition is based on and intertwined with operant learning but it is not the same as operant learning. The ~100,000-year-old process differs from the 500 million-year-old process. The continuity assumption suggests that new contains old, but evolution itself makes no sense if we suppose that old contains new. The present analysis suggests that the core of human cognition was only a very small step forward, based on the extension of functional interactions within human cooperative groups. Nevertheless, we as a species took that step, and that has made all the difference.

**References**


Received: November 19, 2013
Final Acceptance: November 20, 2013
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.