A semiotic perspective based on the philosophy of Charles Sanders Peirce is offered to open up new directions to the current debate over basic emotions. While explaining in a systematic way contested questions such as causal chain, association, and dissociation among the components of emotion, this semiotic analysis suggests that preoccupation with these building blocks type of questions masks and distracts attention from the more global problems that plague affective science—the essentialism that drives the debate, and the tendency to explain behavior in terms of isolated organisms.

John Searle is of the opinion that most research on consciousness is taking the “building blocks” approach with corresponding neglect of the “unified fields” question (Faw, 2005). In the field of emotion research, the same may be said of the multitude of papers generated by the basic emotions debate, including the otherwise rather comprehensive critique by Zachar (2006). My comments on Zachar’s contribution to the basic emotions debate have therefore a twofold purpose and intent: (a) fill a vacuum in his argument, and (b) continue and expand along the lines of pragmatism that Zachar used productively in his theoretical analysis. More specifically, I approach the larger picture of the debate from the perspective of semiotics as formulated by Charles Sanders Peirce, the father of pragmatism. One of the larger issues I focus on concerns the possibility of a paradigm shift in emotion research, a possibility first opened by James Russell’s (2003) critique of the basic emotions paradigm.

A well known story about another Russell, Bertrand (Bakker, 2008), has it that once when he tried to show the infinite regress involved in traditional cosmology, he posed the question that if we are to believe that the world is carried on the backs of giant turtles, then what carries the turtles, he asked. An old lady in the audience said to him, “Very smart, young man, but this won’t do. It’s turtles all the way down.” As
James Russell (2003) points out that emotion is not an "entity" but an inference, i.e., a sign, this paper argues, with the old lady, that emotions are turtles (read "signs") all the way down. Put another way, the quest for the basic building blocks of emotions is "misplaced concreteness"—to borrow a felicitous phrase from Whitehead. The central argument of this paper is that the basic emotions debate is symptomatic of a field that is hampered by a conventional but inadequate understanding of the nature of mental representations, and that a more sophisticated understanding of signs, one that is informed by the semiotics of Charles Sanders Peirce, can help to clarify things and open up new directions for the debate.

Representation According to Charles Peirce

What is a sign? A sign, according to Charles Peirce, is anything that represents another thing to a mind. This perspective differs from the conventional dyadic signifier-signified formulation of signs by the inclusion of a third element, the interpretation of signs by a mind. This third element is referred to by Peirce as the *interpretant*. The interpretant entails two cognitive functions: First, the reference making process of the mind that makes interpretation of signs possible. For instance, interpretation of the sign as an emotion entails the inference of internal states—intentions, experience—and agents, persons capable of internal states. Second, translation of the sign into equivalent forms giving rise to another sign, which is to be interpreted by another interpretant, and the process continues *ad infinitum*. Because of the potentially unbroken cycle of signs and their interpretations, Peirce claims that the mind is a sign generator.

The far reaching implications of this matrix of sign relations for emotion theories in general, and the basic emotions debate in particular, constitute the central thread of argument in this paper.

*From Representation In the Mind to Representation To the Mind*

The Peircean formulation can be meaningfully compared with the conventional notion of the sign in terms of the distinction, made by Karmiloff-Smith (1995), between representation *in* the mind and representation *to* the mind. Representation *in* the mind is characteristic of modules and mechanisms that are context-free, i.e., interpretant independent. Under the aegis of this paradigm, affective states under the sway of automaticity, akin to trees that fall in the woods unheard by anyone, have been meticulously catalogued by countless studies in affective science.

In the Peircean framework, by contrast, emotion functions as representation *to* the mind, a *signal for* further mental operation, a sign that
is meant to be read. Consider a concrete example: The baby is making angry grimaces.

... the parent responds to the baby’s angry looks with a soft soothing look of ‘what’s the matter’ and, with hands out, an offer to pick him up ... The baby responds with a softening of his grimace and anger and a look of expectation. The parent then responds with another gesture ... and the baby now begins to break into a smile ... A second later, the parent is holding the baby ... and the baby relaxes. The tension in his body dissipates and he has a look of calm. (Greenspan & Shanker, 2004, pp. 31-32)

In this scenario, the child’s expressions function as signals to the mind—that of the parent’s. As Trevarthen (1998) points out, adults play the role of “intelligent interpreters of child’s naïve reaching out and vocalizing ‘as if’ these acts were linguistic, transforming ‘actions’ into ‘gestures,’ and then into ‘symbols’” (p. 27). Once initiated by the care-taker into the art of mind reading, the child will eventually be able to do the same with his or her own mind, such that representation in the mind—“hearing a sound and reacting immediately with actions, such as attacking, fleeing, or avoiding”—can be replaced with representation to the mind—“registering the sound, holding it in mind, and using it as a signal or basis for additional mental operations” (Greenspan & Shanker, 2004, p. 264). One of these additional mental operations that we perform on our own emotion experiences is savoring (Sundararajan, 2008a; Frijda & Sundararajan, 2007), in which experiences become object of aesthetic contemplation instead of triggers for direct action.

As Bråten (1998) and others have observed, inter- and intra-subjectivity conform to the same operational format—the mind that reads one’s sign needs not be other’s; it can be one’s own. Both intra- and intersubjectivity locate the emergence of emotions in the intermental arena—the mind minding the mind (Bogdan, 2000). Frith (1995) claims that the content of consciousness is shareable knowledge, and that mechanisms—such as our respiratory system—that are not meant to be shared knowledge will remain nonconscious. By the same token, I would argue that the reason why emotions can become conscious is because they are signs meant to be read/shared. The major contribution of the representation to the mind paradigm lies therefore in the suggestion that emotion is not subjective so much as an inter- and intra-subjective phenomenon; and correspondingly, that emotion is not mental so much as an intermental phenomenon.

The difference between representation in and to the mind does not fall along the divide between animals and humans. Rather, the differ-
ence lies in the researcher's approach, or "decision" as Zachar (2006) puts it, either to consider a dog's growling as the mechanistic operation of a hard-wired module, or as a signal to cue the mental operation of another mind (the master's or another dog's). These two approaches roughly correspond to current debates in the origin of language, between Chomskyan linguistics—which shares with basic emotions theories the assumption of a hard-wired module of operation, capable of the production of a universal grammar, and so on, on the one hand; and theories of gesture, on the other. Pollick and de Waal (2007), for instance, claim that ape gestures imply a capacity to define/interpret signals. Consistent with but broader in scope than Pollick and de Waal's (2007) account is the Peircean framework, according to which, the signal value of the dog's growling, for instance, would depend upon the inference making capacity of the interpretant—it might mean anger to the master, but perhaps dominance to another dog.

Implications of this paradigm shift from mental (representation in the mind) to intermental (representation to the mind) are spelt out in the following pages.

**Mediated Relation between Terms**

One radical implication of the sign as representation to the mind is that there is no one to one translation between terms of a sign. The relationship between the two terms—facial expression and anger—is not direct, but rather mediated by the third term, the mind, or interpretant. Why such a round about way of stating the obvious? The emphasis on the interpretation and translation of signs brings to light the interactive dimension of meaning in representation.

**Meaning is Relational**

A sign that's not "read" (by a mind) signifies nothing. As Mark Johnson (2007) points out, "No isolated thing, percept, or quality has any meaning in itself. Things, qualities, events, and symbols have meaning for us because of how they connect with other aspects of our actual or possible experience. Meaning is relational . . . ." (p. 268). To put it metaphorically, in the forest of signification, a tree (thought or feeling) that falls without being heard (presented to and interpreted) by a mind falls in vain.

**Sign is Generative**

The intermental formulation of the sign has methodological implications as well. In contrast to the prevalent tendency to interpret signs in terms of what it is at a given moment in time, the Peircean formulation puts an emphasis on the production of new signs and interpretants. In the words of Dewey (1981), "A thing is more significantly what it
makes possible than what it immediately is. ... things in their immediacy are subordinated to what they portend and give evidence of” (p. 105). A cognate idea is expressed by Frijda (2007), who claims that feeling is not qualia but meaning, because it points beyond itself, as “signal for” or “pointer to” further elucidations. This idea can be further elaborated in semiotic terms: “A feeling is a mere sign, awaiting interpretation in its relation with a subsequent thought or feeling before it can have meaning” (Hoopes, 1991, p. 10). That meaning, as a function of the sign, takes time to evolve raises serious questions about data collection methods that operate within the narrow confines of the present moment.

The Triadic Circuitry of the Sign

The sign consists of three terms—object (the signified), sign (the signifier), interpretant (the interpretation). How do they relate to one another? The sign relation constitutes two movements of the sign, one feeding forward and one feeding backward. This can be graphically illustrated:

![Figure 1. The Triadic Circuitry of the Sign.](attachment://triadic_circuitry.png)

In the feeding forward movement, the sign gives rise to the interpretant, which in turn acts like a sign to influence the next interpretant, ad infinitum. The feedback movement is referred to by Wiley (1994) as a “reflexive undertow” (p. 27), which is manifest in the reentrant loops from the interpretant to the sign and the object.

The recursive loop from the interpretant to the sign constitutes a paradoxical fact that the interpretant functions to define (constrain) how the sign is to be interpreted, while at the same time it owes its existence to the sign. This bi-directional flow of influence is applicable to the relationship between the appraisal (the interpretant) and the affective response (the sign)—the latter is interpreted by the former, while it at the same time may have given rise to the former in the first place, according to Frijda & Zeelenberg (2001). This phenomenon is referred to by Lewis (2005) as “circular causality” (p. 187).

To recapitulate, as Figure 1 shows, the triadic circuitry of the sign culminates in a reflexive arc that loops from interpretant back to object. This reflexive arc has its correlate in psychology as the self-
reflexive consciousness, which plays a pivotal role in integrating thought and experience (Philippot, Baeyens, Douilliez, & Francart, 2004). The self-reflexive consciousness is modeled by the two movements of the sign—a forward progression toward further symbolic elaboration, on the one hand; and on the other, a recursive loop that doubles back to the object of representation, which in the present context would be concrete experiences of events. Integration of these two movements of thought results in a match between emotion category and its referent, i.e., personal experience.

The triadic circuitry of the sign has the following implications for emotion theory: First, in its pervasive bi-directional flow of influence, a sign system has no subsystem that can be considered more “basic” than others. Second, if appraisal may be understood as an instance of the interpretant, current appraisal theories fail to model the reflexive undertow of the sign, such that the appraisal process seems to be a one way street, with more advanced appraisals carrying heavier cognitive freight (such as causal considerations, blameworthiness, and so on, see Scherer, 1984). By contrast, an appraisal that has a reflexive undertow may be demonstrated by Focusing techniques (Gendlin, 1981), in which the more one is advanced in Focusing, the less cognitive freight one carries, and the more “embodied” one’s emotion representations tend to be.

Lastly, the mediated (by the mind, or interpretant) relationship between the signifier (sign) and the signified (object) is comparable to Teasdale’s (1999) notion of “buffered” processing, which functions to integrate thought and experience, in contrast to the unmediated reciprocal interaction loop among systems in the “mindless emoting mode” (p. S67). While iteration of cycles of the latter explains the module-like phenomena such as emotional schema known for their impulsivity and cognitive impenetrability, iteration of cycles of the former (mediated or buffered processing) may explain the phenomena that fall under the rubrics of refined emotions (Frijda & Sundararajan, 2007), characterized by cognitive flexibility and better integration between thought and experience.

Three Types of Signs

Charles Peirce has distinguished three types of signs—icon, index, and symbol, based on their respective referential competence (Deacon, 1997):

1. Icon refers to the type of reference making that defines a sign relation in terms of spatiotemporal contiguity between the sign and the object of its representation (Object). For example, sensation of pain that accompanies tissue damage. This type of sign capitalizes on physical representations at the expense of conceptualization.
2. The index is the type of reference making that calls attention to the object of representation, like the finger that points at the moon, rather than representing it. The function of index is closely related to that of icon, for instance, pain also functions as index to call attention to physical injury.

3. The symbol refers to the type of reference making that defines a sign relation in terms of arbitrary conventions, for instance affective lexicon, the mode of representation characteristic of language and culture. The symbol is one step removed from experience, a price it pays to capitalize on conceptualization as preferred mode of representation.

This framework can explain both association and dissociation among the multiple components of emotions. Tight coupling of elements is characteristic of iconic representations, whereas lose coupling or even dissociation is characteristic of symbolic representations. Since a fully developed sign has both iconic and symbolic dimensions, both association and dissociation are to be expected. However, from the perspective of pragmatism, a more important question than association and dissociation is the capacity of the sign for integration. Integration is a function of the full fledged sign which, according to Charles Peirce, embodies all three—icon, index, and symbol—in one. An optimal functioning of such a sign has a reflexive arc that integrates the different levels of reference making, from iconic to symbolic and back again to the ground of experience. Consistent with a computer model of emotion integration that is empirically tested (Teasdale & Barnard, 1993), the Peircean model of integration can be illustrated graphically:

```
    Iconic          Symbolic
  Object -------> Sign -------> Interpretant
       \                        /          (Indexical)
        v                        v
Object (updated) ------->
```

Figure 2. A Full Fledged Sign Incorporates Three Modes of Representation

The potentially endless reiteration of this process may result in an interdigitation of the concrete and the abstract—concrete iconic expression of the event, on the one hand; and abstract symbolic elaboration, on the other—that renders experiences meaningful. Thanks also to a “reflexive undertow” (Wiley, 1994) capitalized by the indexical function of the sign, new interpretations in the symbolic mode can
feedback and update the original experience (the Object), resulting in self-modification. A concrete example would help.

Consider the scenario of a child touching the hot stove: Peirce claims that when the child feels the pain, "he becomes aware of ignorance . . ." (cited in Fisch, 1982, Vol.2, p.202). The feeling of pain is an iconic representation, which is interpreted at the symbolic level as ignorance; the interpretation of pain as error feeds back to form a new sense of self as one capable of ignorance. In other words, learning has taken place.

This framework gains a new perspective on emotion schemas and other alleged relics of our evolutionary past. These seemingly pre-packaged emotion syndromes may be understood in terms of fragmentary and unanchored mechanisms that are not integrated with a wider array of patterns of information. It is in this vein that Stern (2004) suggests that traumatic memories are the "temporally unanchored past" (p. 218).

Applications to the Basic Emotions Debate

The basic emotions debate can be understood in terms of a debate over the relationship between two fundamentally different systems, A and B.

A refers to biology—the discourse of Nature's joints.

B refers to emotion concepts and categorizations.

What is the relationship between A and B? The answer from the basic emotions theories (BET) is inherently contradictory: on the one hand, there is the assumption of an ontological gap between A and B—the former being the basic building blocks, whereas the latter secondary development or icing on the emotion cake; on the other hand, there is the notion of a direct translation from A to B—a discrete emotion is a direct readout (Oatley & Johnson-Laird, 1987) from a specific neural circuit. This contradiction is exposed by the core affect theory (CAT), which casts serious doubts (Barrett, 2006) upon the possibility of a direct translation between terms across the chasm of an ontological gap that separates nature from culture, lower animals from humans, and so on. A very different formulation of the relationship between A and B is proposed by Peircean semiotics, which denies any ontological gap between levels of functioning that are evolutionarily continuous (Deacon, 1997), on the one hand; and rejects the possibility of a direct translation between systems, on the other.

Cast into the Peircean framework, A and B are terms that differ in representational structure:

A's representational structure is iconic, characterized by diffused global state and tight coupling between components.
B's representational structure is symbolic, characterized by loose coupling and sharp differentiations.

The relationship between A and B is a dynamic one, characterized by potential conflict and integration. The configuration of their relations is neither pre-packaged (pace BET) nor fortuitous (pace CAT), but structured by two moments—feedforward and feedback—of a recursive movement of the sign system.

**A Neuroscience Illustration**

A cognate idea is found in Don Tucker's (2007) core to shell formulation of the neural structure, which is summed up succinctly by Johnson (2007):

> The limbic core, with its dense interconnections and emotional valences, would present us with a holistic, feeling-rich, emotionally nuanced grasp of a situation. The more modular and highly differentiated sensory and motor regions of the shell (cortical) structure would permit the discrimination and differentiation that we call conceptualization. (pp. 100-101)

This formulation translates readily into the Peircean framework:

The limbic networks correspond to Term A, the iconic representational structure of which is well captured by the densely interconnected patterns and syncretic holism found at the limbic core.

The neocortical networks correspond to Term B, characterized by sparsely connected networks, and greater specificity in representation.

The relation between A and B, between the subsymbolic and the symbolic-cortical systems, is referred to by Tucker (2007) as a "vertical integration" which is defined as a "recursive processing" (p. 223) that consists of movements in two opposite directions, limbifugal and limbipetal:

1. Limbifugal movement refers to Core to Shell connection: This is the feedforward movement toward increasing differentiation into specific and concrete forms.

2. Limbipetal movement refers to Shell to Core connection: This is the feedback, reentrant loop toward integration and self-modification.

Together, limbifugal and limbipetal movements constitute one cycle of the recursive processing referred to as vertical integration: The result of neural network patterns traversing in both directions is the emergence of meaning, which is not surprising—and abstract concepts, which may be a surprise to some. Tucker is emphatic in his claim that abstract concepts
emerge through recursive, bidirectional exchange across the linked, hierarchic, and embedded networks. At the core, concepts may take the form of feelings and intuitive, preconscious constructs. . . . The progression from diffuse core to differentiated shell is a process in which each concept becomes represented in reentrant fashion. (2007, p. 211)

This claim has two implications for the relationship between nature’s joints (A) and emotion concepts (B):

1. Emotion concepts are not icing on the emotion cake, but rather intimately connected with what lies at the biophysiological core.
2. On the other hand, the connection between the two systems is not necessarily smooth and automatic as a pre-packaged deal would be. As Tucker points out, the relationship between the two systems is dialectical:

The consolidation process across the linked networks from shell to core is dialectical in that an inherent opposition of structural forms—fused versus separated—exists between the core and shell. . . . Each wave in the cycle of abstraction traverses this conflict in some way. In those rare optimal instances of the human mind, the dialectic is extended, recursive, and progressive. (2007, pp. 224-225)

In Freudian terms, to invoke an earlier incarnation of the “core and shell” metaphor, where Id was there shall Ego be. In postmodern parlance, the terms of the sign system, A and B for instance, are rolling beads of glass—it is the continuous eliding/transformation of one into the other that gives rise to meaning. In either scenario, Freudian or postmodern, when the ongoing connectedness and interanimation of representations across systems break down, when an outcome defaults toward either A or B, we have what Tucker (2007) calls “a dialectical failure” (p. 228), or what Charles Taylor calls a “paradox”:

The paradox of human emotions is that although only an articulated emotional life is properly human, all our articulations are open to challenge from our inarticulate sense of what is important, that is, we recognize that they ought to be faithful articulations of something of which we have as yet only fragmentary intimations. If one focuses only on the first point, one can believe that human beings are formed arbitrarily by the language they have accepted. If we focus only on the second, one can think that we ought to be able to isolate scientifically the pure, uninterpreted basis of human emotion that all these languages are about. But neither of these is true. (1985, p. 75)
Different Recipes of Coherence

After a comprehensive review of the literature, James Russell (2003) concludes that the so-called emotion is perceived pattern of configuration out of multiple ingredients—brain modes, instrumental action, action tendencies, reflexes, attitudes, cognitive structures, motives, sensations, feelings, facial, vocal and autonomic changes—none of which have any intrinsic connection with one another. Concerning this lack of intrinsic connection between components of emotion, two solutions can be found along the divide between representation in and to the mind.

Where representation in the mind is the reigning paradigm, coherence is achieved by capitalizing on the tight coupling of systems characteristic of the iconic discourse, such as hard wired modules, mechanisms, causal relations, dedicated brain circuits (Panksepp, 1998), etc. CAT repudiates much of BET’s claims to modularity (Russell, 2006), only to reserve the honor for itself—core affects are modules, says Russell (2008). But why privileging the iconic representations as the yardstick of coherence?

The icon is the form of representation found at the bottom of the interpretative hierarchy, a sign with an impetus for interpretation at its lowest ebb: “Iconism is where the referential buck stops when nothing more is added. . . . the production of new interpretants stops” (Deacon, 1997, p. 76). Why privileging a sign deficient in interpretant? This seems to be the price that a purity-based notion of coherence—one that drives the quest for the raw, un-contaminated nature, such as the basic building blocks of emotion—has to pay.

An alternative definition of truth/coherence, one based on integration rather than purity, was entertained by William James (1907/1955): “ideas (which themselves are but parts of our experience) become true just in so far as they help us to get into satisfactory relation with other parts of our experience” (p. 49, emphasis in original). This integrative definition of truth is consistent with the Peircean approach to coherence that capitalizes on the “referential competence” of the mind. What constitutes the referential competence is “the ability to produce an interpretative response that provides the necessary infrastructure of more basic iconic and/or indexical interpretations” (Deacon, 1997, p. 74). Thus the more referential competence there is to support a fully developed sign, one that embodies multiple modes of interpretation at once—the icon, the index, and the symbol, the more likely it is for multiple and diverse systems—body and mind; experience and concepts; nature and culture—to be brought into mutual illumination and integration.

These two approaches to coherence make the opposite recommendations: one aspires to what mystics call “poverty of the spirit,” as evidenced by the strip down approach to emotions practiced by CAT; the
other celebrates the wealth of complexity and creativity as hallmarks of the human mind.

Concluding Observations

The prototype can be misleading. For instance, honeybees are the prototype of the bees, but they are actually the minority in nature—about 75% of the bee species have no hives, no honey, and live solitary lives, according to a report in *Science News* (Milius, 2007). In a similar vein, Russell (2003) has shown that prototypes of emotions, such as the fear reaction of meeting a bear in the woods, are actually infrequent, and violation of the prototype actually common in our emotional lives. To model the full diapason of emotions beyond the fold of a handful basic emotions, this paper proposes a comprehensive model derived from the semiotics of Charles Perice. With integration as a measure of optimal functioning of the sign, the Peircean framework can explain a wide spectrum of affective phenomena, ranging from the aesthetic to the pathological, from the creative to the inflexible manifestations of emotions. It also gives a more nuanced account of the causal chain, besides explaining in a systematic way association and dissociation among components of emotions. When these bones of contention—causal chain, association, dissociation, etc.—between BET and CAT are dissolved in a more comprehensive framework, what remains is a lacuna rendered the more glaring by the neglect by both BET and CAT—the intermental dimension of emotions. What Frith (1995) says in the context of consciousness studies can be extended to emotion research: “the major mistake of most theories of consciousness is to try to develop an explanation in terms of an isolated organism” (p. 683).

The issues raised in this paper offer multiple points of agreement as well as disagreement with BET and CAT. Hopefully this will help BET and CAT theorists to get beyond the nuts and bolts questions, and reflect instead on the larger picture, so as to articulate more clearly their respective research “decision” (Zachar, 2006). Inspired by the self-reflexive ethos of the Peircean perspective, I believe that the self-reflection of eminent researchers on their own basic assumptions has the potential to move the field forward (Sundararajan, 2008b).

References


**Author Note**

Correspondence concerning this article should be addressed to Louise Sundararajan, 691 French Road, Rochester, NY 14618. Email: louiselu@frontiernet.net