way, while contributing to neuroscientific thinking at the same time. Questions posed from the perspective of emotion theory concern the relation between emotional and cognitive systems. A DS analysis of the psychology of emotion frames these questions in terms of part–whole relations and nonlinear causation, bridging psychological and neuroscientific descriptions. A DS analysis of the neurobiology of emotion then demonstrates, in precise and concrete terms, that cognition and emotion were never two distinct systems at all. They are not distinct at the level of interacting parts, nor at the level of the wholes to which these interactions give rise. Neuroscience relies on psychological definitions of functions such as appraisal and emotion, in order to ensure that the forest is not lost through the trees. Some conceptualization of higher-order functions is necessary to identify the elusive wholes emerging from interactions among relatively well-known parts. But the conceptualization of these functions – their definitions and boundaries – may get transformed in the process, and this is what neuroscience offers psychology in return. Although it remains useful to differentiate cognition and emotion for many research agendas, a neuroscientific analysis finds them to be different aspects of a unitary phenomenon in which interpretation and relevance emerge together.

ACKNOWLEDGMENT

I gratefully acknowledge the financial support provided by grant No.1 R21 MH67357-01 from the Developmental Psychopathology and Prevention Research branch of the National Institute of Mental Health (NIMH), as well as ongoing support from the Natural Sciences and Engineering Research Council of Canada (NSERC). I would also like to thank Nico Frijda, E-man Leung, Phan Luu, Jaak Panksepp, Brian Parkinson, Klaus Scherer, Evan Thompson, Don Tucker, and Phil Zelazo for their thoughtful comments on earlier drafts of this article.

Open Peer Commentary

Why not emotions as motivated behaviors?

George Ainslie[a] and John Monterosso[b]

[a]*Department of Veterans Affairs Medical Center, Coatesville, PA 19320;* [b]*Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, Los Angeles, CA 90024.*

George.Ainslie@med.va.gov **jmont@ucla.edu**

http://www.picoeconomics.com

Abstract: Lewis's dynamic systems approach is a refreshing change from the reflexology of most neuroscience, but it could go a step further: It could include the expected rewardingness of an emotion in the recursive feedback loop that determines whether the emotion will occur. Two possible objections to such a model are discussed: that emotions are not deliberate, and that negative emotions should lose out as instrumental choices.

We suggest that a key element, motivation *for* emotion, should be included in Lewis's rubric for bridging emotion theory and neurobiology. Despite his admirable departure from reflexology, he follows conventional assumptions that emotions and the neural processes that subtend them are unmotivated occurrences, elicited by specific patterns of stimuli without regard to their hedonic consequences. He proposes a thorough integration of these occurrences, both a horizontal one across "[complementary] evaluation, monitoring, and action functions" (sect. 5.1) and a "vertical integration [that] not only maintains but also selects viable EIs" (sect. 6.2). He does mention motivation at various points, including a "motivated action loop" of dynamic process (cf. Figure 3 in sect. 5.1 of the target article), but he is apparently referring to behaviors motivated by emotions: "Anticipation of action and feedback concerning action outcomes are intrinsically motivating" (sect. 6.2). It is well understood that emotions motivate; the interesting possibility is that they *are motivated*.

Many common experiences suggest that emotions can be both cultivated and nipped in the bud. For instance, you can "swallow" your anger or "nurse" it, you can learn to inhibit your phobic anxiety (Marks & Tobena 1990), panic (Clum et al. 1993; Kilic et al. 1997), or grief (Ramsay 1997), and you can refrain from rejoicing or "give yourself over to it." The road rage in Lewis's example (sect. 3.3) does not attack a passive Mr. Smart, but lures him in competition with other available activities. As Lewis points out, angry empowerment vies with shameful submissiveness (sect. 6.2), and, we could add, with enjoying his radio program or worrying about his impending sales conference. The choice first of all is whether to entertain the emotion or not, before (or in conjunction with) the choice of whether particular actions would go well with it. Techniques to deliberately foster or inhibit emotions in everyday life have been described (Parrott 1991); and most schools of acting teach an ability to summon emotion (e.g., Strasberg 1988) because even in actors actual emotion is more convincing than feigned emotion (Gosselin et al. 1998). The frequent philosophical assertion that emotions have a moral quality – good or bad (e.g., Hume, as presented by Baier 1991) – implies motivated participation; some philosophers have gone so far as to call the passions voluntary (e.g., Sartre 1939/1948). Thus, there have been many suggestions that emotion is a motivated activity. Lewis characterizes the competition of an emotion with alternatives as "oscillations between coherent appraisals," which might mean just between appraisals of whether triggering elements are present; but all affective appraisals include hedonic value, positive or negative, and thus could be capable of motivating choice. Mr. Smart is lured, not forced, into rage.

Psychological theories have been prevented from acknowledging the motivated quality of emotions by two considerations: that deliberately emitted emotions seem inauthentic (Frank 1988) and that a motive to experience "negative" emotions such as panic and grief seems contradictory. Ainslie (2001) has indeed argued that emotional processes which are under deliberate control do not behave in the same way as spontaneous ones – that they deteriorate into daydream status through familiarity and consequent premature satiation, leaving those which are occasioned involuntarily by surprising events as the authentic kind (Ainslie 2001, pp. 48–70). However, this is not to say that the latter kind are independent of reward. The class of reward-shaped behaviors includes many examples that arise too quickly, are too strongly motivated, or are just too trivial to be screened by the will. Direction of attention, withdrawal from painful stimuli, and numerous mannerisms and facial expressions can all be determined by reward without being deliberate. Conscious choice is a special case, the tip of a great iceberg of reward-seeking processes that include even behaviors during sleep (Granda & Hammack 1961).

The greater problem is how individuals could be lured into aversive emotions. Ainslie (2001) has argued also that reward of very short duration can motivate acceptance of experiences that are avoided from a distance and are reported as negative. The highly bowed (hyperbolic) shape of the discount curve for delayed rewards can produce temporary preferences for addictive substances, and, cycling more rapidly, for excoriating itches and biting fingernails; the same shape could lead to an urge to panic that is almost irresistible despite instant regret if it is obeyed (Ainslie 2001, pp. 48–70). The key word here is *almost*. With adequate

https://doi.org/10.1017/S0140525X05220046 Published online by Cambridge University Press

training and/or motivation people can learn to resist the urge to give in to negative emotions, as noted above, just as they can learn to resist the temporarily preferred reward in itches and addictions. A rewarding component to such emotions is also evident from the fact that, framed properly, they have market value in horror movies and tearjerkers, and indeed from the fact that there is no line between positive and negative emotions: A continuum between them contains "mixed" emotions like anger, nostalgia, and awe, which are not neutral but which may be either cultivated or avoided, depending on the quality of reward from competing sources. This is to suggest that emotions must all pay off quickly to attract participation, but that wide variance in even slightly longer-range payoffs determines how negative they will seem.

We agree that a dynamic systems model is the best way to approach the highly interactive realm of emotions; and if emotions are at least partially selected by their rewardingness, then the apparent reward center in the nucleus accumbens shell belongs not only in Lewis's "motivated action" loop, but also in his "motivated monitoring" and perhaps in his "object evaluation" loops as well (cf. Fig. 3 in the target article). A key component of the relevant dynamic system will be the rehearsal of an emotion leading to some amount of short-term reward and sometimes an increase in the rewardingness (not necessarily pleasure!) of further rehearsal (the "self-augmenting" recruitment phase in which emotional interpretations [EIs] "grow in strength and scope"; sect. 6.1). The same positive feedback process may occur in "conditioned craving" (Ainslie, in press). A person's opportunity to modify this process by intervening in its recursive self-prediction was described by Darwin (and then James and Lange) in the nineteenth century:

> The free expression by outward signs of an emotion intensifies it. On the other hand, the repression, as far as this is possible, of all outward signs softens our emotions. He who gives way to violent gestures will increase his rage; he who does not control the signs of fear will experience fear in greater degree. (Darwin 1872/1979, p. 366)

This process may proceed with such speed as to seem automatic, triggered reflexively rather than shaped by reward; but the frequent observation that clearly motivated processes can override it, implies that it has to bid for dominance in a common marketplace of motivation. Some neurophysiologists have concluded that such a marketplace integrates all behaviors that are selected by reward (Montague & Berns 2002; Shizgal & Conover 1996). Empirical exploration of how emotions interact with this marketplace will not be a simple matter of correlating activity in the known emotional centers with activity in the known reward centers, since these centers overlap; for instance, stimulation of the emotional centers is often rewarding in its own right (e.g., Touzani & Velley 1998). But at the very least, we should not assume that the emotional processes that Lewis models are selected independently of their hedonic consequences.

The concept of circular causality should be discarded

Bram Bakker

Informatics Institute, University of Amsterdam, 1098 SJ Amsterdam, The Netherlands. **bram@science.uva.nl** **http://www.science.uva.nl/~bram/**

Abstract: This commentary argues that one specific but central concept in Lewis's theory, circular causality, is fundamentally flawed and should be discarded – first, because it does not make theoretical sense, and, second, because it leads to problems in practice, such as confounding the interaction between different systems with the relationship between different levels of analysis of a single system.

In recent years, the dynamic systems framework has become increasingly popular with cognitive scientists and neuroscientists. Lewis's target article is a valuable contribution to this area, and it extends the reach of the framework to the psychology and neurobiology of emotions. Lewis makes a convincing case for the claim that in emotional interpretations (EIs), the strong interaction between emotions and appraisals may be understood best in terms of the dynamic systems concepts of positive and negative feedback, and the rapid unfolding of an EI may be understood as a phase transition ending in convergence to one of a limited set of attractors. Furthermore, Lewis argues convincingly that such explanations can be plausibly connected to the underlying neural machinery. Even if not all details have been worked out, his comprehensive account of emotions and appraisals provides significant insights as well as valuable guidance to further work.

This commentary focuses on one specific but central element in Lewis's article, the concept of "circular causality." This concept is used in much dynamic systems theorizing and is defined in section 3.2.4 of the target article as follows:

> Feedback is one form of nonlinear causation. A second form, termed *circular causality* (Haken 1977), describes bidirectional causation between different *levels* of a system. A coherent, higher-order form or function *causes* a particular pattern of coupling among lower-order elements, while this pattern simultaneously *causes* the higher-order form. The top-down flow of causation may be considered an emergent constraint (by the system as a whole) on the actions of the parts. (emphasis in original)

One typical example cited by theorists using the term *circular causality* is the phenomenon of emerging orderly patterns in boiling water, called "convection rolls" (e.g., Kelso 1995). In this example, circular causality refers to the relationship between the water molecules and the global convection roll pattern. As is apparent in Lewis's definition, and also in other literature (e.g., Haken 1977; Kelso 1995), circular causality is construed as a kind of "nonlinear" causality, to be contrasted with ordinary, "linear" causation, which is predominant in theorizing about emotions (sects. 1, 2.2–2.4, 3.1, 7). Throughout the target article, Lewis develops the argument that emotion theories based on linear causal processes should be replaced by theories based on nonlinear, circular causality.

This commentary's main argument is that the concept of circular causality is fundamentally flawed and should be discarded. Circular causality, as construed by Lewis and other authors, is about the relationship between different levels of analysis of a single system, between *parts* and *wholes*: the behavior of the parts "causes" the whole form, and the whole form in turn "causes" the behavior of the individual parts. However, since we are considering a single system, the whole form simply *is* the collection of individual parts interacting in a particular way. The word "correspondence" seems more appropriate than "causality" to describe this relationship. The relationship between a *single* constituent part and the whole form is the relationship of being one of multiple components making up the whole; again it is not properly conceived as a causal relationship. In the example of self-organizing convection rolls in boiling water, the convection roll pattern *corresponds* to the combined behavior of all constituent water molecules; and a single water molecule is one *component* of the whole system. In fact, this relationship between different levels of analysis of the same system is not fundamentally different for self-organizing systems than for other types of systems. A car corresponds to the collection of its interacting parts, and a wheel is one component of the whole system.

One of the hallmarks of circular causality (see the definition above) is the "top-down flow of causation" in a self-organizing system, which is considered a special kind of *constraint* by the whole system on the actions of the parts. However, such "constraints" are also there in other kinds of systems. The behavior of a wheel of a car is constrained, in the same sense, by the behavior of the whole car and vice versa: the wheel is moving when the whole car is moving and vice versa. Such a "constraint" is only natural as the wheel is part of the car; but one would not normally refer to it as "circular causality." In other words, these types of constraints are always

https://doi.org/10.1017/S0140525X05220046 Published online by Cambridge University Press