Outlines of a Theory of Emotions as Metarepresentational States of Mind

appeared in Fischer, A. H. (Ed.), *Proceedings of the 10th Conference of the International Society for Research on Emotions (pp. 186-191)*. Amsterdam: ISRE.

Rainer Reisenzein
Department of Psychology, University of Bielefeld
P. O. Box 100131, 33501 Bielefeld, Germany
e-mail: rreisenz@hrz.uni-bielefeld.de

This paper summarizes a theory of emotions as metarepresentational states of mind (for more detail, see Reisenzein, 1998). The basic idea of the theory is that at least a core set of human emotions including surprise are nonconceptual products of hardwired, metarepresentational mechanisms whose main function is to subserve the monitoring and updating of the two basic forms of propositional representations, beliefs and desires.

Origin of the Theory and Relation to Kindred Formulations

The immediate origin of the present theory is an earlier process model of surprise (e.g., Meyer, Reisenzein, & Schützwohl, 1997). The present theory can be regarded as an elaboration of certain aspects of this model (in particular concerning the mechanism responsible for the detection of schema-discrepancy or unexpectedness) and its extension to a broader class of emotions. In making this extension, I exploit an analogy between the appraisal of unexpectedness, which is analyzed as the comparison of newly acquired beliefs with pre-existing beliefs, and the appraisal of desiredness (desire-congruence), which is construed as the comparison of newly acquired beliefs with pre-existing desires.

As to the broader context of the theory, three historical and systematic connections are particularly important. (1) The present theory is a particular version of an *appraisal process theory* (a theory of emotion-relevant appraisal processes; see e.g., Parkinson, 1997, for a recent discussion). In common with most appraisal theories, the present theory takes as one starting point, and accepts as basically correct, core folk-psychological notions about the relations between beliefs and desires on the one hand, and emotions on the other hand—for example, that one is pleased when one believes that one gets what one desires, displeased when one believes that one does not get what one desires, or surprised when a belief is disconfirmed. Another starting point of the present theory is the defining assumption of ("classical") contemporary cognitive psychology, that cognitive processes are computational processes, that is, rule-governed manipulations of internal symbols (e.g., Fodor, 1987). This assumption is taken to hold true equally of emotion-relevant appraisal processes, although the core appraisal processes are thought to have some unique peculiarities (see below). (2) The present theory is also a special version of a "feeling" theory of emotion, in that it assumes that emotional states contain at their core a nonconceptual or nonpropositional component that is essential for the specific phenomenal character (the "feeling") of emotions. More precisely, it is a "mental feelings" theory, that is, it assumes that the nonconceptual states in question are centrally (rather than peripherally) generated (e.g., Oatley & Johnson-Laird, 1987; Reisenzein, 1994). (3) Finally, the present theory connects up to a tradition of emotion psychology that links emotions to *metarepresentations*, that is, representations of one's own mental states and processes (e.g. Ribot, 1897; Clore, 1994): The posited nonconceptual components of emotions are assumed to represent to the experiencer his or her own representational states—more precisely, they represent certain relations and changes in relations between, one the one hand, newly acquired beliefs, and on the other hand, pre-existing beliefs and desires. It is for this reason that emotions are regarded as metarepresentational states of mind.

Main Assumptions of the Theory

I present the main (not all) assumptions of the theory in the form of six (not necessarily logically independent) postulates, and give supportive arguments for each. P1-P5 are not (directly) about emotions, but about humans' representational system, its properties, and postulated mechanisms that service (monitor and update) this system. However, P6 connects these assumptions with human emotions, by identifying basic *emotional mechanisms* with a subpart of the monitor-and-update system, and basic *affective states* with outcomes of these mechanisms.

P1. The representational systems of humans centrally comprise a propositional system, a "language of thought" (e.g., Fodor, 1987).

Although I assume that humans have several representational systems at their disposal, the propositional system is regarded as occupying a central place in the computational architecture: it conceptually interprets and integrates the outputs of other (specifically sensory) modules, and it underlies the strategic control of human action.

My main reason for assuming a propositional representation system as central in the present context coincides with what some authors regard as the most important general reason for making this assumption: Namely, that it promises to allow the "naturalization" (i.e., the integration into the scientific picture) of folk psychology, including its earlier-mentioned assumptions (systematized in structural appraisal theories) about the relation between beliefs, desires, and emotions. The basic strategy is Fodorian (e.g., Fodor, 1987): Propositional attitudes (*believing p, desiring p, being pleased about p, being surprised about p* etc.) are analyzed as computational relations to internally stored sentences in the "language of thought" representing propositions *p*.

A corollary to P1 is **P1C**: The propositional representation system is the most important for human emotions: it is indispensable for all emotions but "sensory feelings" (i.e., not conceptually mediated, direct affective reactions to sensations, such as the pleasantness elicited by some odors or sounds; see Reisenzein & Schönpflug, 1992). The justification of P1C parallels that of P1. Without going into the details, consider a simple everyday case of an emotion, such as Irina's happiness about Yeltsin's winning of the Russian presidential elections. To have this emotion, Irina must be aware of the state of affairs *Yeltsin won the Russian presidential elections*, which presupposes that she has available the concepts "Yeltsin", "Russia", "presidential election", "win" etc. and combines them appropriately. According to the classical cognitive approach, cognition presupposes the presence of an internal representation medium; this accepted, the example given suggests that this medium must be suited for the representation of concepts and propositions.

P2. The two basic, mutually irreducible forms of mental states involving propositional representations are *beliefs* and *desires*.

One reason for making assumption P2 is that it accords with both common sense (folk psychology), and with the conclusions of numerous philosophers and psychologists who have analyzed the human mind (see e.g., Reisenzein, 1996). A second reason is the following: Cybernetic (control-theoretic) analyses of living and artificial systems showing adaptive behavior strongly suggest that representations of "actual states" and "ideal states" of the world, and their comparison, are fundamental to adaptive behavior at even the simplest level; beliefs and desires are however just the representations of actual and ideal states at the *propositional* level of representation.

A corollary to P2 is **P2C**: Belief and desires are not all-or-none affairs, but come in degrees: Beliefs range from complete uncertainty that *p* to compete certainty that *p*; desires range from extreme aversion against *p* through indifference to extreme (positive) desire for *p*.

P3. The belief-desire representation system comes equipped with mechanisms that update the system in response to newly acquired information—more precisely, in response to newly acquired beliefs. Updating the system means (a) to add new beliefs and desires and (b)—more important in the present context—to abandon old beliefs (if false) and old desires (if fulfilled, and perhaps also when regarded as impossible to fulfill).

The justification for P3 is simply that, without such mechanisms, the belief-desire system would not be able to fulfill its major presumed function—to enable adaptive action in an imperfectly known and changing environment.

P4. The mechanisms that update the belief-desire representation system comprise at core two *comparator devices*: One device that compares newly acquired beliefs with pre-existing beliefs (belief-belief comparator, *BBC*); and another device that compares newly acquired beliefs with pre-existing desires (belief-desire comparator, *BDC*).

The reasons for P4 are straightforward: (a) To affect an *adaptive* modification of existing beliefs and desires (see P3), the need for a modification must first be diagnosed. (b) Corresponding to the two basic forms of propositional representations (P2) there are two basic kinds of modifications of the belief-desire system, modifications of beliefs, and modifications of desires. For each, the need for modification must be detected. (c) Accordingly, there are two kinds of at least analytically separable comparator mechanisms, the *BBC* and the *BDC*. Note that these processes correspond (roughly) to two appraisal processes commonly assumed in appraisal theories of emotion: The appraisal of unexpectedness *(BBC*) and the appraisal of valence (*BDC*).

Conceptual Note: The *BBC* and the *BDC* are appropriately called *metarepresentational* comparator mechanisms (*MRCMs*), because the comparisons which they make concern not only the *contents* of beliefs and desires (the believed or desired propositions), but also their *mode*—the fact that the propositions are, respectively, believed or desired, or considered quantitatively, the degree or strength of the beliefs and desires regarding the propositions. In other words, the *MRCMs* take as input information about both propositions and of the associated belief and desire strengths (see P5). Information that a proposition is believed or desired (with a certain strength) is however not information only about the proposition, but also about the representing system—it is metarepresentational.

P5. The *MRCMs* have the following properties. (1) They have as inputs (a) a sentence in the language of thought representing a newly-believed proposition p_n (an appraised object in the terminology of appraisal theories); (b) sentences in the language of thought that represent already-believed and desired states of affairs $p_1...p_m$; (c) information about the belief strength attached to p_n, and of the belief and desire strengths attached to $p_1...p_m$. (2) The possible outputs of the *MRCMs* are up to four (meta-)representations that carry information about, respectively, the degree of unexpectedness, expectedness, desiredness, or undesiredness of p_n. That is, these representations reflect the degree of congruence or incongruence (discrepancy) between the newly acquired belief (concerning p_n) on the one hand, and the pre-existing beliefs (*BBC*) or pre-existing desires (*BDC*) on the other hand. (3) The *MRCMs* operate continuously: they analyze each newly-believed proposition p_n. (4) The *MRCMs* operate in parallel rather than serially, in two senses: (a) the *BDC* operates in parallel with the *BBC*; (b) each newly believed proposition p_n is compared in parallel with all activated, pre-existing beliefs and desires. (5) The *MRCMs* operate outside of consciousness. This is so because (6) the *MRCMs* are hardwired (rather than stored, and hence presumably learned) procedures. (7) The *MRCM* outputs are not propositions, but analog signals whose intensity represents degrees of (un)expectedness and (un)desiredness. (8) Certain outputs of the *MRCMs* have characteristic hardwired consequences. These include in particular, the focussing of attention on unexpected/undesired propositions, and the automatic resetting of belief and desire strength values (for more detail, see Reisenzein, 1998). Hence, there is a hardwired belief- and desire update mechanism that avoids at least blatant inconsistencies in the belief-desire system. This mechanism is another component of the hardwired machinery that "services" the belief-desire system.

P6. The *MRCMs* are the basic emotion mechanisms for "propositional emotions" (emotions that have propositional objects). Their outputs—the analog signals of expectedness, unexpectedness desiredness, and undesiredness—make up together the core set of "basic emotions".

P5 and P6 are the most important postulates of the present theory, P5 because it contains the comparatively most distinctive assumptions of the theory concerning humans' representation system, and P6 because it connects these assumptions with emotional states. In contrast, P1 and P2 are shared by many psychological theories; and although P3-P4 are somewhat less frequently assumed (at least explicitly), I believe they follow with "evolutionary necessity" from P1-P2.

P5-1 and P5-2 are (partial) computational explications of folk-psychological notions concerning the relations between beliefs and desires on the one hand, and surprise/expectedness versus pleasure and displeasure on the other hand (i.e., of the appraisals of unexpectedness and desiredness). They and the additional explications of these appraisal processes, most importantly P5-6 and P5-7—claiming that the postulated *MRCMs* are hardwired and output not propositions but analog signals—are the assumptions that distinguish the present theory most strongly from other appraisal process theories. It is these assumptions, then, that need most support, and I give several arguments for them in Reisenzein (1998). In particular, I argue that the only visible alternative to the proposed theory of *MRCMs*—the theory that the *MRCMs* are ordinary propositional inference processes—is highly implausible.

P6 constitutes a *theoretical definition* of (propositional) emotions. It says that emotions are at core centrally generated, analog metarepresentations. Simultaneously, these signals are also appraisals (i.e., the immediate outcomes of appraisal outcomes)—the appraisals of (un)expectedness and (un)desiredness. Hence, the present theory can be said to, in one sense at least, *resolve* the much-discussed problem of the relation between cognitions (I mean appraisals) and emotions (for a recent discussion of this problem, see Parkinson, 1997). Concerning the relation between the *MRCM* output signals and conscious feelings, it could be assumed that, if the *MRCM* outputs cross a certain threshold, they become conscious and then *are* the feelings responsible for the characteristic phenomenal character of surprise, belief confirmation, pleasure and displeasure. Alternatively, these conscious feelings could be generated by some additional mediating central process (Reisenzein, 1998).

The range of propositional emotions that the present theory covers cannot be precisely delineated a priori. Certainly the theory covers surprise and being pleased or displeased about a state of affairs; and it appears able to comparatively easily accommodate a number of other propositional emotions, such as disappointment, relief, hope, and fear. The analysis of still other propositional emotions is possible if one assumes that the cognitive antecedents and actional or other consequences of the metarepresentational feelings contribute essentially to their distinction, including their phenomenal differences (cf. Reisenzein, 1994; Reisenzein & Schönpflug, 1992). In any case, the present theory is a priori no worse off in this respect (the explanation of the diversity of emotional qualities) than are most other emotion theories. Furthermore, theoretically principled extensions of the theory to include other metarepresentational signals are possible (Reisenzein, 1998).

References

Clore, G. L. (1994). Why emotions are felt. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion* (pp. 103-111). Oxford: Oxford University Press.

Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind.* Cambridge, MA: MIT Press.

Meyer, W.-U., Reisenzein, R., & Schützwohl, A. (1997). Towards a process analysis of emotions: The case of surprise. *Motivation and Emotion*, *21*, 251-274.

Oatley, K., & Johnson-Laird, P. N. (1987). Towards a cognitive theory of emotions. *Cognition and Emotion*, *1*, 29-50.

Parkinson, B. (1997). Untangling the appraisal-emotion connection. *Personality and Social Psychology Review, 1,* 62-79.

Reisenzein, R. (1994). Pleasure-arousal theory and the intensity of emotions. *Journal of Personality and Social Psychology, 67,* 525-539.

Reisenzein, R. (1996). Emotional action generation. In W. Battmann & S. Dutke (Eds.*), Processes of the molar regulation of behavior* (pp. 151-165). Lengerich: Pabst Science Publishers.

Reisenzein, R. (1998). *A theory of emotions as metarepresentational states of mind.* Paper under review. (A copy of this paper is available from the author on request).

Reisenzein, R., & Schönpflug, W. (1992). Stumpf's cognitive-evaluative theory of emotion. *American Psychologist, 47,* 34-45.

Ribot, T. (1896). *Psychologie des sentiments* [The psychology of emotions]. Paris.