Emotions as metarepresentational states of mind: Naturalizing the belief–desire theory of emotion

Action Editor: Jonathan Gratch

Rainer Reisenzein

Institute of Psychology, University of Greifswald, Franz-Mehring-Straße 47, 17487 Greifswald, Germany

Abstract

Describes the outlines of a computational explication of the belief–desire theory of emotion, a variant of cognitive emotion theory. According to the proposed explication, a core subset of emotions including surprise are nonconceptual products of hardwired mechanisms whose primary function is to subserve the monitoring and updating of the central representational system of humans, the belief–desire system. The posited emotion-producing mechanisms are analogous to sensory transducers; however, instead of sensing the world, they sense the state of the belief–desire system and signal important changes in this system, in particular the fulfillment and frustration of desires and the confirmation and disconfirmation of beliefs. Because emotions represent this information about the state of the representational system in a nonconceptual format, emotions are nonconceptual metarepresentations. It is argued that this theory of emotions provides for a deepened understanding of the role of emotions in cognitive systems and solves several problems of psychological emotion theory.
© 2008 Elsevier B.V. All rights reserved.

Keywords: Emotion; Belief–desire theory; Metacognition; Affective computing; BDI

What are emotions, and what is their function in the economy of the mind? I propose that at least for a core subset of emotions including surprise, these questions can be answered as follows: Emotions are nonconceptual outputs of hardwired mechanisms whose primary function is to subserve the monitoring and updating of the central representational system of humans, the belief–desire system. This theory of emotions, which closely connects emotions to the monitoring and updating of representations and assigns them important epistemic functions—in holding that they convey to agents important information about their own representational system—may at first appear unusual. In fact, however, it can be argued that this view of emotions is already implicit in the currently dominant theories of emotion, the cognitive emotion theories (for

an overview, see Scherer, Schorr, & Johnstone, 2001). For these theories assume implicitly that emotions are closely tied to changes in beliefs and desires; and at least some of them explicitly attribute to emotions an informational function (e.g., Ortony, Clore, & Collins, 1988). In fact, the computational model of emotions sketched in this article is an attempt to "naturalize" (to integrate into the scientific picture; Dretske, 1995) a particular version of cognitive emotion theory, the belief–desire theory of emotion (*BDTE*).

The motivation for this attempt was the observation that many important issues are still controversial among cognitive emotion theorists (see also, Reisenzein, 2001, 2006a) and the belief that, to resolve these issues, it is indispensable to consider the cognitive architecture that underlies emotions. Although cognitive emotion theorists are agreed that (most) emotions presuppose cognitions, they differ on exactly which cognitions *are* necessary for

E-mail address: rainer.reisenzein@uni-greifswald.de

emotions (e.g., factual beliefs, evaluative beliefs, or both) and in which sense they are necessary (e.g., are they required as causes or components of emotions); whether and in what sense emotions also presuppose desires; how the cognitive processes that generate emotions look like in detail; what the emotion itself is (i.e., how it is to be theoretically defined) and which functions emotion have; how emotions relate to emotional experiences; what accounts for the distinctive quality and the intensity of emotional experiences; and how the object-directedness of emotions can be explained. The theory of emotion proposed in this article gives or at least sketches answers to all these question and thus offers a coherent picture of the emotional mind. At the same time, the theory seeks to preserve the central insights of cognitive emotion theories but to avoid objections that have been raised against them.

The article has three parts. In the first part, I give an overview of *BDTE*. In the second part, I sketch a computational model of *BDTE*. In the third part, I present arguments in favor of the proposed computational explication of *BDTE*.

1. The belief–desire theory of emotion

The belief–desire theory of emotion belongs to the broader class of cognitive emotion theories represented, for example, by the theories of Arnold (1960), Frijda (1986), Lazarus (1991), Oatley and Johnson-Laird (1987), Ortony et al. (1988), and Scherer (2001) in psychology; and those of Kenny (1963), Lyons (1980), Nussbaum (2001), and Solomon (1976) in philosophy (for reviews, see Ellsworth & Scherer, 2003; Goldie, 2007). Cognitive emotion theories have come to dominate psychological and philosophical theorizing on emotions during the past two decades, and they also form the basis of most existing computational models of emotion (e.g., Elliott, 1992; Gratch & Marsella, 2004; Neal Reilly, 1996; Staller & Petta, 2001). As a distinct type of emotion theory within the cognitive approach to emotions, *BDTE* has been primarily promoted by philosophers (e.g., Davis, 1981; Green, 1992; Marks, 1982; Searle, 1983; and for an early version, Meinong, 1894, 1906 [summary in Reisenzein, 2006a]). Proponents of *BDTE* in psychology include Roseman (1979) and Miceli and Castelfranchi (1997, 2007).

1.1. Basic assumptions of BDTE

The conceptual framework of *BDTE* is the same as that of the philosophical belief–desire theory of action (e.g., Bratman, 1987; Mele, 1992) which inspired the *BDI* (belief–desire–intention) approach to artificial agents (Bratman, Israel, & Pollack, 1988). Analogous to the belief–desire theory of action, which holds that actions are the product of cognitive or informational states (beliefs) and motivational states (desires), *BDTE* posits that emotions are the product of cognitions (beliefs) and motives (desires). In this analysis, beliefs and desires are

taken to be basic kinds of mental states—modes of relating to objects—that cannot be reduced to one another. To mention but one difference between the two: Beliefs, but not desires, can be true or false; desires, but not beliefs, can be satisfied or frustrated (e.g., Green, 1992). As discussed in the next section, the distinction between beliefs and desires is ultimately based on the fundamentally different functional roles that these mental states play in the economy of the mind.

BDTE theorists differ, among other issues, on the question of the precise sense in which emotions are "products of" beliefs and desires. Whereas some regard beliefs and desires as *components* of emotions (see Green, 1992, for further discussion), I follow Meinong (1906) in assuming that beliefs and desires are the *causes* of the emotion, which is a mental state of its own. Presupposing this "causalist" interpretation of *BDTE*, its basic assumption can be stated more precisely as follows: At least a core subset of the mental states presystematically subsumed under the category "emotion" are reactions to the cognized actual (e.g., happiness, unhappiness) or potential (e.g., hope, fear) fulfillment or frustration of desires plus, in some cases (e.g., Surprise, disappointment), confirmations or disconfirmations of beliefs.

To illustrate *BDTE*, consider the case of Mary, who feels happy that *Mr. Schroiber was elected chancellor*. (For the time being, I identify emotions with emotional experiences. A distinction between the two is drawn later in the computational explication of the theory.) According to *BDTE*, Mary feels happy about Mr. Schroiber's election if Mary (a) comes to (firmly) believe that Schroiber was elected and (b) desires this state of affairs. In slightly more detail, the process of emotion generation, here illustrated for happiness, typically looks as shown in Fig. 1. First, the person comes to desire some state of affairs or proposition p.[1] For example, Mary acquires the desire that *Schroiber is elected chancellor*. Some time later—as a result of new information acquired through the senses, communication from others, or inference from existing beliefs—the person acquires the belief that p obtains. For example, when watching the news on TV, Mary comes to believe that Schroiber was, indeed, elected chancellor. Thereupon, the emotion occurs: Mary now feels happy or pleased that Schroiber was elected.[2] Note that the belief and desire are connected to the emotion not only as cause to effect, but also semantically: The belief that p, the desire for p and happiness about p concern a common topic, they all refer

---

[1] This is the philosophical usage of "proposition". Psychologists typically use the term to denote a sentence in a language-like mental representation system that represents states of affairs (e.g., Kintsch, 1988; Anderson & Lebiere, 1998).

[2] The converse temporal sequence is possible also. Mary may first learn that Schroiber was elected chancellor, and come to desire that state of affairs only later, when she reads about Schroiber's political program. In this case, the desire is cognized as being fulfilled as soon as it is formed, resulting again in Mary's happiness about Schroiber's election.
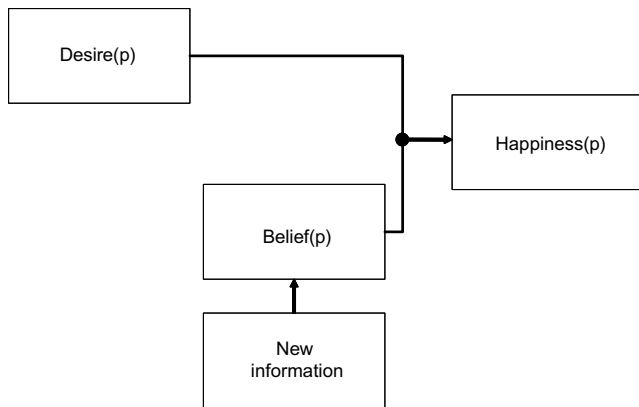
Fig. 1. Basic belief–desire analysis of emotions.

to p (Green, 1992; I will later argue, however, that the object-directedness of the *emotion* is only apparent).

By amending and refining the just-described "basic formula" of the belief–desire analysis of emotions, it is possible to specify the cognitive and motivational preconditions of many emotions distinguished in ordinary language (e.g., Davis, 1981; Meinong, 1894, 1906; Miceli & Castelfranchi, 2007; Searle, 1983). Indeed, there is reason to believe that *all emotions with propositional objects* (all emotions directed at states of affairs) are amenable to a belief–desire analysis. These emotions cover the vast majority of the emotions distinguished in ordinary language (see also, Ortony et al., 1988).

1.2. Qualitative belief–desire analysis of emotions

A qualitative belief–desire analysis of some emotions is shown in Table 1. To illustrate Table 1, Mary feels *happy* at time t that *Schroiber is elected chancellor* (= proposition p), if Mary desires that Schroiber is elected and comes to firmly believe (i.e., is certain) at t that he was elected. Mary feels *unhappy* at t that Schroiber is elected chancellor if she is aversive to p or "diswants" p to happen (here analyzed as: Mary desires *not-p*, that Schroiber is not elected) and comes to firmly believe at t that p obtains. Mary *hopes* at t that Schroiber is elected chancellor if she wants him to be elected but is uncertain at t whether or not he will be

elected (i.e., believes with uncertainty that p obtains); whereas Mary *fears* at t that Schroiber is elected if she diswants him to be elected but is uncertain whether or not he will be elected. Mary is *surprised* at t that Schroiber is elected chancellor if she up to t believed he would not be elected ($\neg p$) and at t comes to firmly believe that he was elected (p). Mary is *disappointed* at time t that Schroiber is not elected chancellor if she desires him to be elected and up to t believed that this was likely or at least possible, but at t comes to firmly believe that he was not elected ($\neg p$). Finally, Mary is *relieved* that Schroiber is not elected chancellor if she diswants him to be elected and up to t believed that this was at least possible, but at t comes to firmly believe that he was not elected.

Although the emotions listed in Table 1 still comprise only a small subset of the emotions distinguished in ordinary language, from the perspective of *BDTE* they are basic forms. Happiness and unhappiness are the emotional reactions to the cognized *actual* fulfillment versus frustration of desires; hope and fear to their cognized *possible* fulfillment versus frustration. Surprise is the general emotional reaction to belief-disconfirmation; if it co-occurs with desire–fulfillment, it is pleasant surprise; if it co-occurs with desire frustration, it is unpleasant surprise. Disappointment and relief are the emotional reactions to the unexpected (belief-disconfirming) *nonoccurrence* of desire fulfillment versus frustration, respectively.

The diversity and complexity of humans' emotional life gets into the picture primarily, if not exclusively, via the *objects* or *contents* of the beliefs and desires that give rise to emotions. In other words, the diversity and complexity of human emotions is due to the fact that humans have diverse and complex desires and beliefs. In particular, as a highly social species (e.g., Richerson & Boyd, 1998), humans have desires that concern not only their own welfare, but also the welfare of others, as well as their own and others' compliance with social and moral norms. Accordingly, other-regarding emotions, such as *joy for another*, *Schadenfreude, pity*, and *envy* can be analyzed as forms of happiness or unhappiness about, respectively, a desired or undesired state of affairs p that concerns the positive or negative fate of another person (e.g., Meinong, 1894; Ortony et al., 1988). The classical "moral emotions", such as guilt and indignation on the negative side, and pride or moral elevation on the positive side, can be incorporated into the belief–desire framework by assuming that the content of the desire (the desired proposition) in these cases is the compliance of oneself or another person with a social or moral norm (e.g., Ortony et al., 1988; Staller & Petta, 2001). In addition, *BDTE* can be extended to cover "fantasy emotions" (i.e., emotional reactions to anticipated or counterfactually imagined states of affairs), by positing that these emotions are based on *assumptions* rather than *beliefs* (Meinong, 1910; Reisenzein, Meyer, & Schützwohl, 2003). In the computational model of *BDTE* described later, this would mean to add a "pretense store" to the cognitive architecture (see Nichols & Stich, 2000).

Table 1
Belief–desire theory of emotions, qualitative formulation

Emotion	if	Belief at t	Desire at t	Belief at t-1
happy (p, t)		Certain(p, t)	Des(p, t)	
unhappy(p, t)		Certain(p, t)	Des($\neg p$, t)	
hopes(p, t)		Uncertain(p, t)	Des(p, t)	
fears(p, t)		Uncertain(p, t)	Des($\neg p$, t)	
surprised(p, t)		Certain(p, t)	– (*irrelevant*)	Bel($\neg p$, t-1)
disappointed($\neg p$, t)		Certain($\neg p$, t)	Des(p, t)	Bel(p, t-1)
relieved($\neg p$, t)		Certain($\neg p$, t)	Des($\neg p$, t)	Bel(p, t-1)

Notation: Bel(p, t)...believes p at time t; Certain(p, t)...firmly believes p at t
Uncertain(p, t) =: Bel(p, t) & ¬Certain(p, t) & ¬Certain($\neg p$, t). Des(p, t)...
desires p at t; Des($\neg p$, t)...desires *not-p* at t (≈ is aversive against p at t).

1.3. Quantitative formulation of BDTE

Apart from being able to explain the qualitative (i.e., type) differentiation of emotions in a parsimonious way, *BDTE* also allows a straightforward and parsimonious explanation of the intensity aspect of emotion (the fact that emotions vary in intensity). To explain emotional intensity, the qualitative version of *BDTE* is refined to a quantitative theory. In fact, already the first modern formulations of *BDTE* contained proposals for a quantification (e.g., Davis, 1981; Day, 1970). Quantification is achieved by (a) introducing quantitative concepts of belief and desire and (b) proposing quantitative laws, expressed by numerical functions, that connect degrees of belief and desire with the intensity of various emotions. Table 2 shows a quantitative belief–desire analysis of the emotions listed in Table 1 (based mostly on previous proposals; e.g., Davis, 1981; Gratch & Marsella, 2004; Macedo, Reisenzein, & Cardoso, 2004). For example, happiness about p is experienced when $b(p) = 1$ and $d(p) > 0$, and the intensity of happiness about p is a monotonically increasing function of $d(p)$. Fear is experienced whenever $0 < b(p) < 1$ and $d(p) < 0$, and the intensity of fear is a monotonically increasing function of $|d(p) \times b(p)|$.

1.4. BDTE as a psychological background theory for computational models of emotion

As a member of the class of cognitive emotion theories, *BDTE* is supported by the theoretical and empirical argu-

Table 2
Belief-desire theory of emotions, quantitative formulation

Emotion Intensity = function of d and b	for domain subset (else emotion intensity = 0)
Happiness$(p, t) = \Phi_{ha}[d(p, t)]$	$b(p, t) = 1$ & $d(p, t) > 0$
Unhappiness$(p, t) = \Phi_{uh}[d(p, t)]$	$b(p, t) = 1$ & $d(p, t) < 0$
Hope$(p, t) = \Phi_{ho}[b(p, t) \times d(p, t)]$	$0 < b(p, t) < 1$ & $d(p, t) > 0$
Fear$(p, t) = \Phi_{fe}[b(p, t) \times d(p, t)]$	$0 < b(p, t) < 1$ & $d(p, t) < 0$
Surprise$(p, t) = \Phi_{su}[b(p, t\text{-}1)]$	$b(p, t) = 0$ & $b(p, t\text{-}1) > 0$
Disappointment$(\neg p, t) = \Phi_{di}[b(p, t\text{-}1) \times d(p, t\text{-}1)]$	$b(p, t) = 0$ & $b(p, t\text{-}1) > 0$ & $d(p, t\text{-}1) > 0$
Relief$(\neg p, t) = \Phi_{re}[b(p, t\text{-}1) \times d(p, t\text{-}1)]$	$b(p, t) = 0$ & $b(p, t\text{-}1) > 0$ & $d(p, t\text{-}1) < 0$

Notation: b: $\{p_1, p_2, \ldots\} \times \{t_1, t_2, \ldots\} \Rightarrow [0, 1]$ and $d : \{p_1, p_2, \ldots\} \times \{t_1, t_2, \ldots\} \Rightarrow \mathbb{R}$ are the belief and desire functions. $b(p, t)$ represents the strength of belief in p at time t, with 1 denoting certainty that p, 0.5 maximal uncertainty, and 0 certainty that *not-p*. $d(p, t)$ represents the direction and strength of the desire for p at time t, with values > 0 denoting positive desire, 0 indifference, and values < 0 aversion against p. *Happiness*(p, t), *Unhappiness*(p, t) etc. are the emotion intensities, ranging from 0 (absence of the emotion) to some maximum number. For arguments outside the domain subsets indicated in column 2, the value of the emotion functions is 0. Note that the emotion intensity functions (*Happiness, Unhappiness* etc.) are only partly specified in the table, with Φ_{ha}, Φ_{uh} etc. representing their unspecified components. For example, Gratch and Marsella (2004) use $\Phi = ||$ (the absolute value function) to be able to represent all emotion intensities, including those of negative emotions, by positive numbers. If no further specifications are made, Φ is the identity function. The functions for disappointment and relief only capture the case of the simple nonoccurrence of a desired and undesired event, respectively. For a possible treatment of the general case see Mellers, Schwartz, Ho, and Ritov (1997).

ments that have been adduced for cognitive emotion theories generally (e.g., Ellsworth & Scherer, 2003; Goldie, 2007; Reisenzein et al., 2003). Arguments and evidence more specifically supportive of *BDTE* stem from two sources.

First, philosophical proponents of *BDTE* have argued that this theory avoids a number of objections raised against other versions of cognitive emotion theory, especially cognitive–evaluative or appraisal theories of emotion (e.g., Green, 1992; Marks, 1982; see also, Reisenzein, 2006a). A central difference between *BDTE* and the cognitive–evaluative theory of emotion is that the latter regards emotions as products of factual and evaluative beliefs rather than of beliefs and desires. To illustrate, whereas *BDTE* posits that Mary is happy about p if she believes p and desires p, the evaluative theory of emotion assumes that Mary feels happy about p if she believes p and evaluates p as good for herself (i. e., believes that p is good for her; e.g., Arnold, 1960; Lazarus, 1991). Against this assumption, it has been objected that—even in the presence of the necessary factual beliefs—evaluative beliefs are *not sufficient* for emotions. For example, one usually values being healthy highly; nonetheless, the thought that one is healthy (factual belief), and that this is good for oneself (evaluative belief), normally does not cause noticeable joy. *BDTE* can easily explain this: At the time of the evaluative judgment, the desire to be healthy has long been fulfilled. It has also been argued that evaluative beliefs are *not generally necessary* for emotions. *BDTE* can handle this objection, too: To be happy about p, one must desire p and come to believe that p obtains; it is not necessary that one in addition believes that p is good for oneself (Arnold, 1960), congruent with one's motives (Lazarus, 1991), or goal-conducive (Scherer, 2001).

Second, *BDTE* is supported by research on the quantitative relations between beliefs, desires, and emotions (Table 2). Although studies of this kind are still few in number and exist only for a few emotions, their results are generally in line with predictions. For example, Reisenzein and Junge (2006) tested several of the "emotion laws" shown in Table 2 in a monetary lottery paradigm (Mellers et al., 1997). Participants experienced occurrences and nonoccurrences of monetary wins and losses that differed in magnitude and probability and rated the intensity of experienced surprise, disappointment and relief caused by each outcome. Support for the predictions of *BDTE* was obtained even on the level of individual participants. Further supporting the self-reports of emotions collected by Reisenzein and Junge (2006), a recent fMRI study (Yacubian et al., 2006) found that unexpected outcomes of a monetary lottery caused changes in activity proportional to the "prediction error" (the difference between the expected and actual outcome value) in brain regions commonly associated with positive and negative emotions (the ventral striatum and the amygdala, respectively) (see also Abler, Walter, Erk, Kammerer, & Spitzer, 2006).

The theoretical and empirical arguments for *BDTE* suggest that *BDTE* may be a good choice as a psychological background theory for computational models of emotion. Other, more pragmatic considerations support this conclusion. First, *BDTE* may be more tractable computationally than are appraisal theories of emotion, as it attributes most of the multidimensionality and complexity of appraisals to the objects of the appraisals. As a consequence, much of the computational labor that appraisal theories of emotion impose on emotion mechanisms can be relegated to ordinary propositional inference processes (Reisenzein, 2001), whereas the emotion-generating mechanism proper can be kept simple. For in contrast to the impression conveyed by appraisal theories, *BDTE* suggests that, behind the apparent complexity of emotions lies a comparatively simple generating mechanism. This suggestion is borne out by the computational model of *BDTE* proposed in the following section. Second, the conceptual closeness of *BDTE* to the *BDI* framework should facilitate the integration of emotions into the *BDI* architecture (e.g., Becker & Wachsmuth, 2006; Rank & Petta, 2005; Staller & Petta, 2001) and into other computational architectures that have affinities to the *BDI* approach (e.g., Gratch & Marsella, 2004; Marinier & Laird, 2006). Finally, *BDTE* connects up naturally with attempts to model emotions in decision theory (e.g., Mellers, 2000; Zeelenberg, van Dijk, Manstead, & van der Pligt, 2000). This suggests straightforward ways of incorporating some effects of emotions on action into computational models of emotion (see also Reisenzein, 1996; for further discussion).

2. Naturalizing the belief–desire theory of emotion

Like most cognitive emotion theories proposed by philosophers and psychologists, *BDTE* is formulated on the "intentional level" of system analysis familiar from common-sense psychology (Dennett, 1987; see also Reisenzein, 2001; Sterelny, 1991). In this section, a proposal is made to naturalize *BDTE*, by moving to the "design level" of the cognitive system, that is, by sketching a computational model of *BDTE* (*CBDTE*; see also, Reisenzein, 1998, 1999, 2001, 2006b). Note that my motivation behind this endeavor was not to develop a worked-out computational model of emotions, but to use computational thinking (thinking in terms of representations and computational mechanisms) as a tool to clarify *BDTE* and thereby—to the degree that *BDTE* is correct as a theory of emotion—to become clearer about several unresolved issues in emotion psychology. Three aspects of *BDTE* in particular needed clarification. First, how (through which cognitive process) is the causal link between factual beliefs and desires on the one hand, and emotions on the other hand (symbolized by the connecting arrows in Fig. 1) mediated? Second, what *exactly* is the emotion in *BDTE*; that is, what is the theoretical definition of emotion in *BDTE* (Reisenzein, 2007)? Third, what are the functions of emotions in agents whose actions and thoughts, like those of humans,

are controlled by a belief–desire representation system? That is, in what ways do these agents benefit from having emotions *in addition to* beliefs and desires?

The naturalization of *BDTE* proceeds in three steps. The first step consists in the modelling of the belief–desire system, for beliefs and desires are the causes of emotions in *BDTE*. Taking recourse to a strategy proposed by Fodor (1987), the belief–desire system is modeled as a propositional representation system. In the second step, this representation system is amended by proposing: (a) The belief–desire system comes equipped with hardwired mechanisms that monitor and, if necessary, update the system in response to newly acquired information. (b) The monitoring mechanisms comprise two submechanisms: One that compares newly acquired beliefs to existing beliefs and another that compares newly acquired beliefs to existing desires. (c) The outputs of these information-processing mechanisms are "nonpropositional signals" (in the sense of Oatley & Johnson-Laird, 1987) that carry information about the degree of match or mismatch of newly acquired beliefs with existing beliefs and desires. In the third step (whose description is interwoven with that of the second) the comparator mechanisms are linked to emotions.

2.1. The representational system

To develop a computational model of *BDTE* requires to embrace the assumption that mental and especially cognitive processes are computations in an internal representation system. I adopt the by now classical representational assumption of cognitive psychology, that the central representation system of humans is symbolic and more precisely, that it is a language-like, or propositional system of representations, a language of thought (e.g., Anderson & Lebiere, 1998; Fodor, 1987; Kintsch, 1988). A main reason for this assumption is that, in contrast to other proposed representation systems, a propositional system lends itself naturally to a transparent and plausible computational analysis of the intentional mental states posited in common-sense psychology, such as beliefs and desires; and thereby, to the naturalization of cognitive emotion theories such as *BDTE* (Aydede, 2004; Fodor, 1987; Sterelny, 1991; see also, Gratch & Marsella, 2004).[3]

How can the naturalization of beliefs and desires be achieved, given a propositional system of mental representation? The answer, suggested by Fodor (1987) and others

---

[3] This is of course not meant to deny the existence of other mental representation systems, such as sensory and image-like representations. In fact, as explained later, I assume that emotions are nonconceptual, sensation-like representations. However, imagistic representations as traditionally conceived of are not suited for representing the contents of beliefs and desires, as they are not capable of capturing the informational selectivity and compositional structure of propositions (e.g., Aydede, 2004). Although recent imagistic representation systems (the perceptual symbol systems described in Barsalou, 1999) are better suited for this purpose, this is so precisely because they incorporate central assumptions of propositional representation systems (see Barsalou, 1999, p. 595).
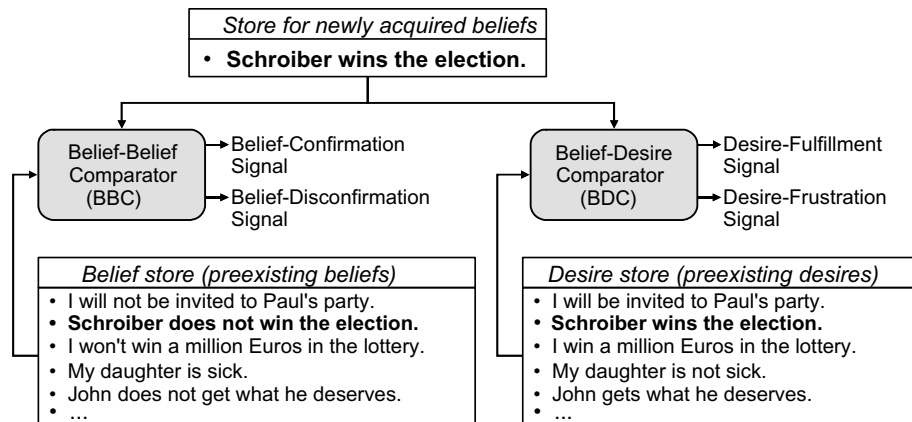
Fig. 2. Illustration of the belief–belief and belief–desire comparators: A moment in Mary's belief–desire system.

(e.g., Schiffer, 1994), is to conceptualize *believing* and *desiring* as special modes of processing propositional representations (sentences in the language of thought). To use Fodor's and Schiffer's metaphor (see also Nichols & Stich, 2000), believing a proposition *p* consists, computationally, of having a token of a sentence *s* representing *p* in a special memory store (which is accordingly called the "belief store"); and desiring *p* consists of having a token of a sentence *s* representing *p* in another memory store (the desire store; note again that I use "proposition" in the philosopher's sense, as a synonym of "state of affairs"; see footnote 1). This mode of speaking is meant to be shorthand for a functional description of beliefs and desires, a description in terms of their causal roles in the system (Fodor, 1987). For example, to say that a sentence is "in the belief store" is shorthand for saying that this sentence is "treated as true" by the system: It is used as a valid premise in inferences, is relied on when planning actions, elicits surprise when it turns out to be false, and so forth.[4] The metaphor of storage bins is thus not essential. What *is* essential, however, is that the system or agent is able to distinguish between propositions that it merely represents, without believing or desiring them; propositions that it believes; and propositions that it desires (see also, Paglieri, 2004).

To illustrate, consider Mary's belief–desire system at the moment when she learns that Schroiber won the election (Fig. 2). At this moment, Mary has the indicated, and many more, "mentalese" sentences in her belief store. These sentences represent the states of affairs she currently believes to obtain. For example, as can be seen, Mary currently believes that Schroiber will not win the election (temporal qualifiers are omitted in Fig. 2 for reasons of simplicity). Likewise, Mary currently has the indicated, and many more, sentences in her desire store. These represent the states of affairs that Mary currently desires. For example, Mary currently wishes that Schroiber wins the election.

2.2. The belief–belief comparator and the belief–desire comparator

CBDTE accepts this computational analysis of beliefs and desires but extends it to model emotions. According to *BDTE*, Mary experiences happiness about Schroiber's election as chancellor if she desires this state of affairs and comes to believe that it obtains (Fig. 1). To model this process, let us begin by assuming that newly acquired beliefs are placed into a special memory store, a store reserved for newly acquired beliefs. Computational speaking, then, Mary feels happy that Schroiber won the election when, or very soon after, a sentence representing this state of affairs is deposited in her store for new beliefs (see Fig. 2).

Fig. 1 further suggests that happiness about *p* is the causal effect of the co-occurrence, or joint presence of desiring *p* and believing *p*. However, when looked at from the computational perspective, it is clear that for happiness about *p* to occur, it is not sufficient that the belief that *p* is present in the cognitive system simultaneously with the desire for *p* (i.e., it is not enough that Mary has a sentence token representing *p* in her desire store, and another sentence token representing *p* in her store for new beliefs). In addition, the cognitive system needs to *relate* these two facts: It needs to detect or recognize that the newly believed proposition *is identical to* a desired proposition. To achieve this, a mechanism is needed that compares the newly acquired belief to the preexisting desires, looking for match and mismatch—a *belief–desire comparator* (*BDC*).

Parallel considerations apply to surprise. According to *BDTE*, Mary is surprised about Schroiber's election as

---

[4] In my view, the best explication of the different functional roles of beliefs and desires, that also points to their evolutionary origins, is provided by cybernetic analyses of living and artificial goal-directed systems (e.g., Milsum, 1966; Rosenblueth, Wiener, & Bigelow, 1943). These analyses suggest that negative feedback resulting from the comparison of a representation of an "actual state" with that of a "reference" or "ideal" state is an indispensable control element in even the simplest forms of goal-directed behavior. In my view, human action can be understood as that form of goal-directed behavior that is controlled by propositional representations of actual and ideal states. That is, beliefs and desires are representations of the cyberneticists' actual and ideal states, respectively, in a propositional representation system.

chancellor if she believed that this state of affairs would not occur and then comes to believe that it did, in fact, occur. Again, however, it is not sufficient for the occurrence of surprise that the newly acquired belief that p is present in the system simultaneously with the "old" belief that not-p. In addition, the system must detect or recognize that the content of the newly acquired belief conflicts with that of a preexisting belief. To achieve this, a mechanism is needed that compares the newly acquired belief to the pre-existing beliefs for match versus mismatch—a *belief–belief comparator* (*BBC*).

Importantly, the existence of these belief–desire and belief–belief comparators can be motivated quite independently from any consideration of emotions (Reisenzein, 1999). Presumably, the major evolutionary function of the belief–desire system is to enable adaptive action in an imperfectly known and changing environment. To fulfill this function, the belief–desire system needs to be equipped with mechanisms that, if necessary, update the system in response to newly acquired information. Updating means to add new beliefs and desires to the system as well as, and more important in present context, to abandon existing beliefs (if they turn out to be false) and existing desires (if they are fulfilled). However, for this updating of beliefs and desires to proceed *adaptively*, both the need for and the more precise nature of the updating must first be *diagnosed*, by comparing newly acquired information to the existing beliefs and desires. Accordingly, the updating mechanisms must contain appropriate comparator devices; and again, one can distinguish between two (at least analytically separable) comparators: One that compares newly acquired beliefs to existing beliefs (the *BBC*), and another that compares newly acquired beliefs to existing desires (the *BDC*).

To become clearer about how these comparators might work, let us again consider the moment when Mary comes to believe that Schroiber won the election; or computationally speaking, the moment when a sentence representing this state of affairs is deposited in Mary's store for new beliefs. This belief is now compared to Mary's preexisting beliefs and desires. The *BBC* compares the newly acquired belief to Mary's preexisting beliefs, looking for match versus mismatch (Fig. 2). A match means that a preexisting belief is confirmed by the new information, whereas a mismatch means that a preexisting belief is disconfirmed. Computationally speaking: The sentence token currently in Mary's store for newly acquired beliefs, s_{new}, is compared to the sentence tokens s_{old} currently in her store for preexisting beliefs. If either a match (s_{new} is identical to a sentence s_{old} in the belief store) or a mismatch (s_{new} is identical to the negation of a sentence, $\neg s_{old}$, in the belief store) is detected, the *BBC* generates an output that signals the detection of the match or mismatch. In our example, Mary's *BBC* detects that the content of the newly acquired belief (*Schroiber wins the election*) is inconsistent with (is the negation of) the content of a preexisting belief. Consequently, Mary's *BBC* outputs information about a mismatch, information that one of Mary's beliefs is disconfirmed by the new information (Fig. 2).

The *BDC* compares the newly acquired belief to preexisting desires for match versus mismatch (Fig. 2). A match means that a desire has been (at least subjectively) fulfilled, whereas a mismatch means that a desire has been frustrated. Computationally speaking: The sentence s_{new} in Mary's store for newly acquired beliefs is compared with the sentence tokens s_{old} currently in her desire store. If either a match or a mismatch is detected, the *BDC* generates an output that signals the detection of the match or mismatch. In our example, Mary's *BDC* detects that the content of the newly acquired belief (Schroiber wins the election) is identical to the content of an existing desire. Consequently, Mary's *BDC* outputs information about a match—information that one of Mary's desires has been fulfilled (see Fig. 2).

2.3. Towards a quantitative computational model of BDTE

The foregoing description of the belief–belief and belief–desire comparison mechanisms has been deliberately simplified, to make the basic ideas transparent. A more realistic model of these mechanisms was sketched in Reisenzein (1999). The most important elaborations of the preceding description contained in that more realistic version of *CBDTE* are the following.

1. An explicit distinction is made between long-term memory and working memory and it is assumed that comparisons of newly acquired beliefs with existing beliefs and desires take place in working memory. Hence, propositions in long-term memory need to be retrieved into working memory before they can be compared to other propositions. It is assumed that appropriate retrieval procedures are available for this purpose (e.g., Anderson & Lebiere, 1998). In addition, straightforward, simple inferences are automatically drawn from newly acquired beliefs and are also deposited in working memory as "new" (see Ortony & Partridge, 1987). For example, if Mary believes that if Stoeder is elected chancellor (q), Schroiber is not ($\neg p$), she will normally infer $\neg p$ if she learns q. The assumption of minimal automatic inferences is necessary to explain even the simplest cases of emotion, such as in the described case, Mary's disappointment that Schroiber was not elected chancellor when she learns that Stoeder was elected.

2. To take account of the intensity of emotions, beliefs and desires are modeled as quantitative variables, and the comparators are assumed to compute *degrees* of belief–desire and belief–belief match. To this end, it is assumed that propositions are represented in memory as triples $\langle s_i, b_i, d_i \rangle$, where s_i is a sentence in the mental language expressing proposition p_i, and b_i and d_i are quantities representing, respectively, the agent's degree of belief and desire regarding proposition p_i. To illustrate, assume that $b(p) \in [0, 1]$ and $d(p) \in [-100,$

+100]. Then, prior to the event, Mary's belief and desire referring to Schroiber's election as chancellor could be represented by the triple ⟨"Schroiber wins the election",.05, 50⟩, meaning that Mary was fairly certain that Schroiber would *not* win the election (subjective probability of $p = .05$), but desired Schroiber's election with medium strength (+50). The b_i and d_i are quantitative "tags" of some sort attached to the mentalese sentences s_i, that determine the processing of s_i as believed and desired to a given degree. That is, working memory is constructed in such a way that, if a triple ⟨s_i, b_i, d_i⟩ is entered, its belief tag determines the processing of s_i as believed to degree b_i and its desire tag determines the processing of s_i as desired with degree d_i. These tags thus constitute (together with the assumed architecture of working memory) quantitative analogues to the belief and desire stores assumed in the qualitative model described earlier. The b_i and d_i could be special symbols in the language of thought or (more likely) subsymbolic codes reserved for the representation of belief and desire strengths.[5]

3. Analogous to the qualitative computational model of *BDTE* described earlier, newly acquired beliefs are stored in a functionally separate area of working memory, as triples whose third element (representing desire strength) is undefined. This is meant to reflect that, before a newly believed proposition has made contact with the person's desire system, its motivational relevance is undetermined (Lazarus, 1991). For example, Mary's newly acquired belief that Schroiber won the election would be represented as ⟨"Schroiber wins the election", 1.00, *⟩, with "*" denoting "undefined".

4. Newly-acquired beliefs can, like pre-existing beliefs, be held with different degrees of certainty. As a consequence, degrees of congruence or incongruence to existing beliefs and desires are also computed for newly acquired beliefs held with less than complete certainty. This assumption allows to model *hope* and *fear* (see Table 2).

5. On the background of these representational and computational assumptions, the *BDC* and *BBC* are implemented as a *modified pattern matcher* (cf. Anderson & Lebiere, 1998) that works as follows: Newly acquired beliefs are compared in parallel with all preexisting beliefs and desires in working memory for identity versus opposition of content. In the simplest case, this means that a "new" triple ⟨r_j, b_j, *⟩ is compared to all preexisting triples ⟨s_i, b_i, d_i⟩ in working memory with respect to identity or contrariness of the first elements (r_j, s_i). For instance, the newly acquired belief ⟨"Schroi-

ber wins the election", 1.00, *⟩ is compared, among others, with the pre-existing representation ⟨"Schroiber wins the election", .05, 50⟩. If a match ($r_j = s_i$) or a mismatch or conflict ($r_j = \neg s_i$) between the two sentences r_j and s_i is detected, then the belief strength b_j of the newly acquired proposition is computationally integrated with the belief and desire values b_i and d_i by means of a procedure that generates one or more nonpropositional output signals. These signals reflect, in particular, the degree of "expectedness versus unexpectedness" of r_j (computed as a function of b_j and b_i) and the degree of "desiredness versus undesiredness" of r_j (computed as a function of b_j and d_i). However, as the *BBC* and *BDC* are assumed to operate in parallel, their joint outputs could well be integrated at a subconscious level into more complex output signals. For example, co-occurring signals of unexpectedness and desire–frustration might be combined into a disappointment signal. The majority of the interesting functions that can be computed, separately or jointly, by the *BDC* and the *BBC* can be gleaned from Table 2. (Although Table 2 refers to emotional *experience*, as explained later the experience of emotions is assumed to be closely linked to the outputs of the *BDC* and the *BBC*). To illustrate, in Mary's case the comparison of ⟨"Schroiber wins the election", 1.00, *⟩ with ⟨"Schroiber wins the election", .05, 50⟩ yields a match (identity of the first component), which in this case produces an unexpectedness signal of (1-.05) and a desiredness signal of (1 × 50).

2.4. Functional consequences of the detection of belief– and desire (in-) congruence

Having discussed how belief– or desire congruence and incongruence are detected, I turn to the functional consequences of detected (in-)congruence (note that these consequences are not shown in Fig. 2). According to *CBDTE*, the most important of these consequences—which are particularly pronounced in the case of belief or desire *discrepancy*—are the following three.

First, attention is automatically focused on the contents of the newly acquired beliefs that gave rise to match or mismatch—the (un-)expected or (un-)desired propositions. For example, in Mary's case, attention is directed to Schroiber's unexpected election victory. This reallocation of processing resources is a precondition for the further conscious analysis of the event in question; regarding, for example, the analysis of its causes, implications, and relevance for ongoing action (Meyer, Reisenzein, & Schützwohl, 1997).

Second, some minimal, immediate updating of the belief–desire system takes place automatically: Mentalese sentences representing disconfirmed beliefs are deleted from the belief store, and sentences representing state of affairs that are now believed to obtain are deleted from the desire store. Note that this does not mean that the belief or desire contents are *forgotten*; it only means that they are

---

[5] In the latter case, the belief–desire representation system is a hybrid of a symbolic and a (localist) subsymbolic system (Kelley, 2003; Sun, Zhang, & Mathews, 2006): Propositions are represented in a symbolic code, whereas their associated belief and desire strengths are represented subsymbolically, analogously to the activation levels of nodes in a propositional spreading activation network (Anderson & Lebiere, 1998).

no longer believed or desired, respectively (see Paglieri, 2004). These automatic, minimal local updates of the belief–desire system prevent the persistence of at least the most blatant inconsistencies in the system.

Third, BBC and BDC output signals that exceed a certain threshold of intensity give rise, directly or indirectly (e.g., by activating subcortical structures; see Yacubian et al., 2006) to unique, conscious feeling qualities. In the case of the BBC, these are the feeling of surprise and the feeling of "expectancy confirmation" (see also, Ortony et al., 1988); in the case of the BDC, they are the feelings of pleasure and displeasure. However, as mentioned, additional distinct feelings (e.g., of disappointment and relief, hope and fear) could result from subconscious integrations of the BDC and BBC outputs. The general function of these feelings is assumed to be the same as that attributed to other conscious experiences: To make information system-wide available and thereby poised for exerting global control (e.g., Baars, 1988; Chalmers, 1995; Oatley & Johnson-Laird, 1987). Specifically, the function of the feeling of surprise is to communicate system-wide that a belief–discrepant state of affairs has been detected. This communication typically elicits curiosity and may thereby motivate epistemic search (Reisenzein, 2000). The function of the feelings of pleasure and displeasure is to communicate system-wide that something desire–congruent (pleasure) or desire–discrepant (displeasure) has happened. These communications may, depending on circumstances, motivate the person to continue with a course of action (Fredrickson, 2001) or to "coast" (Carver, 2003) in the case of pleasure; or to deal with the problematic state of affairs in the case of displeasure (e.g., Schwarz & Clore, 1996). It needs to be emphasized, however, that to be available for further processing by the belief–desire system (the propositional representation system), the emotional feelings must first be conceptually interpreted. That is, the information carried by these feelings must be extracted and represented as the content of beliefs, analogous to the case of sensory feelings (e.g., Jacob, 1997).

In addition, displeasure signals seem to automatically create a desire to reduce them, whereas pleasure signals seem to create a desire to maintain them (e.g., Baumeister, Vohs, DeWall, & Zhang, 2007; Mellers, 2000). In my view, the most plausible way to incorporate this assumption into CBDTE is to posit the existence of a hardwired procedure that generates concrete hedonic desires under appropriate circumstances (e.g., the desire that the currently present unpleasant feeling be abolished). The hedonic motive (the desire to maintain pleasure and avoid pain) is thus not represented explicitly (i. e., as a desire with a general content), but is implemented as a procedure. From the perspective of CBDTE, the hedonic motive is best viewed as a motivational support mechanism (Reisenzein, 1996). In particular, it aids the satisfaction of the original desire that p, when it is threatened or frustrated, by creating an auxiliary desire to reduce or abolish the displeasure caused by the threat to or frustration of the primary desire. In this way, the secondary, hedonic desire reinforces the primary desire even though it is, in and of itself, blind to the aim of the primary desire.

2.5. Implementation, mode of operation, and output format

So far, the BBC and the BDC were described in terms of their inputs, outputs, and immediate functional effects. I now come to what are, in one sense, the most important assumptions of the present theory. They concern the implementation, mode of operation, and output format of the proposed comparator mechanisms.

First, the BBC and the BDC are not learned procedures, but are "hardwired" into the brain.

Second, partly as a consequence, these mechanisms (a) operate on a preconscious level, without and even against the agent's intentions; (b) they operate continuously, that is, they work on every single, newly acquired belief; and (c) they compare each belief in parallel to the preexisting beliefs and desire contents in working memory. In short, every newly acquired belief is automatically (without intention, and preconsciously) compared simultaneously to all belief and desire contents currently in working memory.

Third, as was already suggested when describing the output of the BBC and BDC, I assume that this output is not propositional in nature—it is not just another sentence in the language of thought (representing the fact that a match or mismatch was detected). Rather, the outputs of the BBC and BDC are nonpropositional and nonconceptual: They consist of signals that vary in kind and intensity, but have no internal structure (see Oatley & Johnson-Laird, 1987; Picard, 1997), analogous to simple sensations of tone or temperature. These signals carry information about the degree of expectedness versus unexpectedness, and the degree of desiredness versus undesiredness of the propositional contents of newly acquired beliefs; but they do not represent these contents themselves.

In sum, I propose that the belief–desire system comes equipped with a set of hardwired monitoring- and updating mechanisms, the BBC and the BDC. These mechanisms are, in a sense, similar to sensory transducers (i.e., sense organs for color, sound, touch, or bodily changes). In particular, their immediate outputs are nonpropositional signals. However, instead of sensing the world (at least directly), these "internal transducers" sense the state of the belief–desire-system and signal important states and state changes in this system as it deals with new information.

2.6. The belief–belief and belief–desire comparators and the emotions

According to BDTE, happiness about a state of affairs p occurs if one desires p and comes to believe p; whereas surprise about p is felt if one previously believed not-p and now comes to believe p. The computational analysis of BDTE described in the last section suggested that the

causal link between newly acquired beliefs and preexisting beliefs and desires on the one hand, and emotions on the other hand, is mediated by the *BBC* and the *BDC*: Emotions result when the comparator mechanisms detect a match or mismatch of a newly acquired belief with preexisting beliefs (*BBC*) or desires (*BDC*). It follows that emotions are intimately related to the updating of the belief–desire system. In fact, according to *CBDTE*, the hardwired comparator mechanisms that service the belief–desire system, the *BBC* and the *BDC*, *are simultaneously the basic emotion-producing mechanisms*. This answers the first question about *BDTE* raised at the beginning of this section, the question of which process mediates the causal link between beliefs and desires on the one hand, and emotions on the other hand.

However, if this answer is accepted, then the other two questions raised, concerning the nature (theoretical definition) and function of emotions in *BDTE*, can also be answered.

1. Emotions are the nonpropositional signals of congruence and incongruence produced by the two basic, hardwired comparator mechanisms that service the belief–desire system, the *BDC* and the *BBC*. The *kind* of the signals indicates the *type* of the detected congruence or incongruence (e.g., desire fulfillment, or belief disconfirmation); their *intensity* indicates the *degree* of congruence or incongruence. *CBDTE* suggests these signals as the best candidates for the scientific referents of emotions because of their central causal role: They are caused by beliefs and desires (the inputs to the emotion mechanisms), and they in turn cause emotional experience, attentional focusing, updating of the belief–desire system, and adaptive action. By identifying emotions with the nonpropositional signals produced by the *BDC* and the *BBC*, *CBDTE* allows for the possibility of unconscious emotions. For the output signals of the *BDC* and *BBC* need not necessarily give rise to conscious experiences (e.g., when they are below a threshold of intensity). Nevertheless, because one function of the *BDC* and *BBC* is precisely to make the experiencer aware that incoming information matches or mismatches existing beliefs or desires, emotions should normally be conscious.

2. The function of emotions (the output signals of the *BDC* and *BBC*) is, generally speaking, to assist the updating of the fundamental action-guiding representational system of humans, the belief–desire system. Emotions do this by focusing attention on detected belief–belief and belief–desire (in-)congruence, by causing some minimal local updating of beliefs and desires, and by giving rise to conscious experiences that make information about belief–belief and belief–desire (in-)congruence system-wide available. Thereby, emotions prepare the system to respond adaptively to detected belief– and desire (mis-) matches. This answers the question of the function of emotions as *occurrent emotional states*. The question of the function of emotions can also be asked for the *mechanisms* that produce emotional states. To the latter question, *CBDTE* suggests the following answer: The function of the emotion mechanisms is to detect matches and mismatches of newly acquired beliefs with existing beliefs and desires, and to prepare the cognitive system to deal with them once they have been detected. This conclusion agrees well with Frijda's (1986, 1994) proposal that the emotion mechanisms are at core "concern relevance detectors", but it extends Frijda's proposal to the detection of "epistemic relevance".

Finally, note that emotions, as conceived here, are representations; specifically, they are *nonconceptual metarepresentations*. This can be seen as follows. (1) The information carried by the *BBC* and the *BDC* output signals—for example, "a belief has been disconfirmed" or "a desire has been fulfilled"—is information *about* beliefs and desires; hence, it is metarepresentational information (I use "representation" here as shorthand for "mental representation). A *belief* with this content—e.g., Mary's belief "one of my beliefs has been disconfirmed" would be a *conceptual* metarepresentation, specifically, a higher-order thought.[6] (2) Although the outputs of the *BBC* and *BDC* are not beliefs, but nonpropositional signals, I follow traditional (e.g., Brentano, 1874/1955) and recent philosophers of mind (e.g., Crane, 1998; Dretske, 1995; Tye, 1995) in assuming that not only propositional attitudes, but also nonconceptual mental states such as sensations can legitimately be regarded as representations. Specifically, I follow Dretske (1995) in assuming that nonconceptual mental states are representations if their evolutionary function is to convey information to a (sub-)system of the mind. (3) This is the case with regard to emotions, the signals produced by the *BBC* and *BDC*: These signals do not just happen to carry metarepresentational information; rather, it is their (evolutionary) function to convey that information to other mental systems. (4) Hence, I conclude that emotions are *nonconceptual metarepresentations*. They belong to a category of mental states that Flavell (1979), in a classic paper on metacognition, listed as one subtype of metacognition termed "metacognitive experiences" (see also,

---

[6] Adherents of the definition of "metarepresentation" as "representation of a representation" (e.g., Jacob, 1997) might object that this belief is not a genuine metarepresentation because it does not represent the *contents* of the disconfirmed or disconfirming first-order beliefs, at least not exactly. This objection is based on a restrictive exegesis of the definition. In my use of "metarepresentation" which is meant to be a generalization of "metacognition" that avoids unwanted implications of this term (Reisenzein, 1999), metarepresentations need not represent lower-order representations *in detail*. Accordingly, not only Mary's belief that she believes *p* is a metarepresentation (in this case, a higher-order thought), but so is Mary's belief that she believes many contradictory things, the belief that she has just acquired a new belief, and the belief that one of her beliefs has been disconfirmed. In any case, the important point is that emotions are mental states that nonconceptually represent certain states and state changes in the belief–desire system, regardless of whether or not one wants to call these mental states "metarepresentations".

Nelson, 1996). As such, emotional experiences belong to the same category as, for examples, feelings of familiarity and novelty (e.g., Koriat, 2000; Metcalfe, 1994). Specifically, emotional feelings represent to the experiencer, in a nonconceptual format, current states and state changes of his or her belief–desire system (e.g., surprise: "a belief has just been disconfirmed"; pleasure: "a desire has just been fulfilled"; fear "a desire may now be frustrated").

3. Arguments for the computational explication of *BDTE*

As Green (1992, p. 104) noted in his exposition of *BDTE*, "the value of a theory is a function of the problems it resolves and the phenomena it explains". What, then, are the phenomena explained, and the problems resolved, by the proposed *computational explication* of *BDTE*?

To begin with, *CBDTE* reproduces the assumptions of *BDTE* (e.g., Fig. 1 and Tables 1 and 2); it therefore accounts for the same common-sense intuitions and systematic data that support *BDTE*. However, the explanatory capacity of *CBDTE* goes farther than this: The theory also suggests answers to several moot questions of emotion psychology. It is to these that I now turn.

3.1. The nature and function of emotions, basic emotions, and the distinctiveness of emotional experience

As already explained, *CBDTE* suggests precise answers to the questions of what emotions are and which functions they have in a belief–desire system. Beyond this, *CBDTE* provides for a *principled* demarcation of basic emotions: The basic emotions are exactly the outputs of the two comparator mechanisms, the *BBC* and the *BDC*. Furthermore, *CBDTE* explains what is *distinctive* about emotional experiences, what sets these experiences apart from others: Namely, that they are at core *unique* (i.e., specific to emotions) sensation-like experiences produced (directly or indirectly) by dedicated mechanisms. Finally, as argued in more detail in Reisenzein (2008), *CBDTE* is able to account for the intensity aspect of emotional experiences, as well as for their qualitative differences.

3.2. Resolving the puzzle of the object-directedness of emotions

CBDTE suggests a solution to a traditional puzzle of the object-directedness of emotions. The puzzle is this: On the one hand, emotions such as happiness, unhappiness, fear, hope and so on usually present themselves to the experiencer as being directed at objects. For example, Mary feels happy *that Schroiber was elected chancellor*. On the other hand, sometimes emotions seem to lack objects, as in the case of moods: Sometimes Mary just feels happy, without feeling happy about anything in particular. This is puzzling, particularly because emotions differ in this respect from the paradigmatic propositional attitudes, beliefs and desires, which have objects *necessarily*: It is not possible to believe or desire without believing or desiring *something*, even though that something is sometimes only very vaguely specified.

CBDTE suggests that the appearance of object-directedness of emotions is just that: an appearance. As mentioned, emotions (the nonpropositional signals produced by the *BDC* and *BBC*) represent only the congruence or incongruence of newly acquired beliefs with existing beliefs or desires; they do not represent the *propositional contents* of these beliefs and desires. This implies that emotions cannot literally be directed at the contents, for "to be directed at" here means "to represent" (e.g., Searle, 1983). Inasmuch as emotions, from the first-person perspective, *appear* to be directed at these objects, this appearance must therefore be an illusion. The strong point of *CBDTE* is that it is able to explain, at least up to a point, how this illusion occurs: It is due to the automatic focusing of attention on propositions that (mis-) match existing representations. For example, when an existing belief is disconfirmed by a newly acquired belief $Bel(p)$, the person experiences a feeling of surprise and has her attention near-simultaneously drawn to the "offending" proposition. It then appears to the person that she is surprised *about p*, that her feeling *represents* *p* in a particular way (namely, as a *surprising* fact). Conceivably, under the described temporal and causal circumstances, the feeling of surprise gets "bound" to the mental representation of *p* by a process analogous to the binding of different features of an object (e.g., shape and color) in object perception (Roskies, 1999); by an implicit causal attribution (Clore, 1994; Costall, 1991); or by the implicit categorization of the experience (Barrett, 2006; Mandler, 1984; Reisenzein, 1994) as an instance of "being surprised by".

3.3. Resolving the cognition–emotion debate

CBDTE resolves, to a degree, the "cognition–emotion debate" in psychology. Started by Zajonc (1980) and Lazarus (1982), the cognition–emotion debate revolves around the question of whether or not cognitions are necessary for emotions, and if yes in which sense they are necessary. Despite much discussion, the question continues to be controversial (see e.g., Kappas, 2006; Parkinson, 1997). *CBDTE* in a sense resolves this question, by proposing the following, differentiated answer to it. On the one hand, cognitions *are* necessary for the emotions explained by *BDTE*. First, these emotions are the results of operations on propositional representations including beliefs; therefore, they presuppose these beliefs (e.g., to feel happy about Schroiber's election as chancellor, Mary must believe that Schroiber was elected). Second, because *BDTE*-emotions are representations whose main function is to provide information, they are themselves cognitions in a wide sense of the term.

On the other hand, cognitions are *not* necessary for emotions. First, the beliefs that are the inputs to the *BBC* and *BDC* are in the typical case factual beliefs, not evaluative

ones. For example, to feel happy about *p*, Mary must believe that *p*, but not that *p* is good for her, or that *p* is motive-congruent. Evaluative beliefs are needed for an emotion only if that emotion is a reaction to the fulfillment or frustration of a desire for an explicitly evaluative state of affairs (e.g., Mary wishes not to have done wrong; Green, 1992). Second, the mechanisms that produce emotions according to *CBDTE*, the *BDC* and the *BBC*, are not propositional inference procedures, and their outputs are nonpropositional signals. Because of these features, the emotion-producing mechanisms are not cognitive in the classical, narrow sense of the term.

Still, these mechanisms fulfill similar functions as two central cognitive appraisal processes postulated in appraisal theories of emotion: the process of appraising motive–congruence (the *BDC*, Lazarus, 1991), and the process of appraising unexpectedness (the *BBC*; e.g., Roseman, Antoniou, & Jose, 1996). *CBDTE* claims that (analogues of) these appraisal processes are implemented as hardwired procedures, and that their "appraisal outputs"—the nonconceptual representations of (un)expectedness and (un)desiredness—*are* emotions. According to *CBDTE*, the distinction between these cognitive appraisals and emotions is therefore futile: The outcomes of the *BDC* and *BBC* are identical to emotions; emotions *are* appraisals.

3.4. Why humans need hardwired belief–desire and belief–belief comparators

A final consideration that speaks in favor of *CBDTE* is that alternative attempts to explicate *BDTE* face difficulties. As argued before, belief–desire and belief–belief comparators of some kind are needed to explain the generation of emotions in *BDTE* (see Fig. 1), and are indispensable components of the machinery that updates the belief–desire system. If this is accepted, then what is at issue is not the *existence* of the *BDC* and *BBC*, but only their *special form*.

It seems that the only alternative to the present proposal—that the *BDC* and *BBC* are hardwired procedures with nonconceptual outputs—that has been entertained in the literature is that the comparators are ordinary propositional inference procedures. According to this view, the language of thought is used in a self-reflexive manner for metacognitive monitoring purposes. To illustrate, in the case of surprise about Schroiber's election as chancellor (*p*), Mary reasons: "Up to now, I believed *not-p*; just now, I came to believe *p*; thus, my previous belief that *not-p* is disconfirmed." This metacognition then presumably causes Mary's experience of surprise. Analogously, in the case of happiness about Schroiber's election, Mary reasons: "I desired *p*; just now I came to believe *p*; thus one of my desires is fulfilled", and this metacognition then presumably causes her feeling of happiness. This "metapropositional" theory was explicitly proposed by Davidson (1982) for the belief–belief comparison process, and by Miceli and Castelfranchi (1997) as a special case of the belief–desire comparison process (required, in their view,

for emotions resulting from the frustration of self-related desires). Furthermore, as mentioned, the *BBC* and *BDC* are similar to the appraisals of unexpectedness and motive–congruence posited by appraisal theorists (e.g., Lazarus, 1991); appraisals are however paradigmatically regarded as (evaluative or nonevaluative) beliefs (see Reisenzein, 2001, 2006a).

This "metapropositional" theory of the *BBC* and *BDC* strikes me as extremely implausible.

1. Introspection provides no evidence that metacognitions of the described kind precede emotions. As Becher (1916) noted long ago for the case of surprise, the thought "something unexpected happened" occurs to the surprised person, if at all, only as an afterthought.

2. As mentioned, a realistic account of the comparison of newly acquired beliefs with existing beliefs and desires requires the comparison and integration of *degrees* of belief and desire. If the *BBC* and the *BDC* were implemented as propositional inference procedures, people would therefore have to form metacognitive beliefs such as "the strength of the newly acquired belief concerning *p* differs *n* units from the strength of the pre-existing belief concerning *p*". Even assuming that only few categories of belief strength are distinguished, this scenario looks implausible. It becomes still more implausible if one adds the assumptions that newly acquired beliefs are compared with pre-existing beliefs and desires in parallel, and continuously.

3. Already small children seem to experience pleasure in response to goal success, displeasure in response to goal failure, and surprise when their beliefs are disconfirmed (e.g., Lewis, 2000). It is doubtful, however, whether small children are capable of forming higher-order thoughts about their beliefs and desires. To form such thoughts, they would have to possess the concepts of belief and desire. Developmental research suggests that children acquire these concepts only between 3 and 5 years (e.g., Flavell, 2004; Wellman, 2002). A parallel argument can be made for emotions and higher-order beliefs in animals (see also, Carruthers, 2007; Proust, 2006).

4. The assumption that the *BBC* and *BDC* are hardwired procedures explains both the universality of the links between beliefs/desires and emotions and the involuntary aspects of emotion. To achieve the same explanatory success, the "metapropositional" theory of the *BDC* and *BBC* would have to assume that there are hardwired connections between emotional reactions and particular kinds of beliefs, that differ from other beliefs only with respect to their *content*. For example, this theory is committed to assumptions such as that the belief "one of my desires was frustrated" is an innate activator of physiological arousal.

5. The "metapropositional theory" of the *BBC* and *BDC* does not square well with the idea that emotional experiences have informational functions (e.g., that the

function of surprise is to inform the conscious self that something unexpected has happened). If appraisal processes are metapropositional inference processes and their outcomes are, accordingly, occurrent beliefs, it is unclear why the feelings presumably caused by them are needed for informational purposes, and why they should be heeded: If the feelings agree with the beliefs, the feelings are uninformative; if they disagree with the beliefs, the feelings constitute misinformation.

6. The "metapropositional theory" of the *BBC* and *BDC* gets the object-directedness of emotions wrong. If, as this theory assumes, the proximate causes of emotions are higher-order beliefs with contents such as "*p* is unexpected" and "*p* is desire–congruent", then one should expect that the emotions are (or phenomenally appear to be) directed at the objects of these proximate cognitions. This is not so, however: One is surprised that *p* is the case, not that *p* is unexpected; and one is happy that *p*, not that *p* is desire–congruent.

CBDTE avoids all these problems. It does not require that to have emotions, humans must form higher-order thoughts about their beliefs and desires, much less thoughts about the *degree* of belief–belief and belief–desire congruence or discrepancy. They need not even possess the concepts of belief and desire to have emotions. *CBDTE* also obviates the need to assume hardwired links between certain (metacognitive) beliefs and emotions. At the same time, the posited emotion mechanisms, the *BDC* and *BCC*, are first-rate candidates for hardwired procedures. As such, they could plausibly operate in parallel (and thus quickly) and continuously (Logan, 1992), and they can be present already in infants, as well as in animals. Furthermore, *CBDTE* explains how feelings can provide reliable, nonredundant information about belief- and desire congruence and incongruence. Finally, *CBDTE* restores to emotions their correct (apparent) objects.

References

Abler, B., Walter, H., Erk, S., Kammerer, H., & Spitzer, M. (2006). Prediction error as a linear function of reward probability is coded in human nucleus accumbens. *NeuroImage, 31*, 790–795.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Arnold, M. B. (1960). *Emotion and personality* (Vol. 1 & 2). New York: Columbia University Press.

Aydede, M. (2004). The language of thought hypothesis. Stanford encyclopedia of philosophy. <http://plato.stanford.edu/entries/language-thought/>.

Baars, B. (1988). *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.

Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review, 10*, 20–46.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–660.

Baumeister, R. F., Vohs, K. D., DeWall, C. N., & Zhang, L. (2007). How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation. *Personality and Social Psychology Review, 11*, 167–203.

Becher, E. (1916). Gefühlsbegriff und Lust-Unlustelemente [Concept of emotion and hedonic elements]. *Zeitschrift für Psychologie, 74*, 128–154.

Becker, C., & Wachsmuth, I. (2006). Modeling primary and secondary emotions for a believable communication agent. In D. Reichardt, P. Levi, & J.-J. C. Meyer (Eds.), *Proceedings of the first international workshop on emotion and computing (in conj. with KI2006)* (pp. 31–34). Universität Bremen.

Bratman, M. E. (1987). *Intentions, plans, and practical reason*. Cambridge, MA: Harvard University Press.

Bratman, M., Israel, D., & Pollack, M. (1988). Plans and resource-bounded practical reasoning. *Computational Intelligence, 4*, 349–355.

Brentano, F. (1874/1955). In O. Kraus (Ed.), *Psychologie vom empirischen Standpunkt [Psychology from the empirical point of view]* (Vol. 1). Hamburg: Meiner.

Carruthers, P. (2007). Meta-cognition in animals: A skeptical look. *Mind and Language, 23*, 58–89.

Carver, C. S. (2003). Pleasure as a sign you can attend to something else: Placing positive feelings within a general model of affect. *Cognition and Emotion, 17*, 241–261.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2*, 200–219.

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.

Costall, A. (1991). Phenomenal causality. In G. Thinès, A. Costall, & G. Butterworth (Eds.), *Michotte's experimental phenomenology of perception* (pp. 51–64). Hove, East Sussex: Erlbaum.

Crane, T. (1998). Intentionality as the mark of the mental. In A. O'Hear (Ed.), *Current issues in philosophy of mind* (pp. 229–251). Cambridge: Cambridge University Press.

Davidson, D. (1982). Rational animals. *Dialectica, 36*, 317–327.

Davis, W. (1981). A theory of happiness. *Philosophical Studies, 39*, 305–317.

Day, J. P. (1970). The anatomy of hope and fear. *Mind, 79*, 369–384.

Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: MIT Press.

Dretske, F. (1995). *Naturalizing the mind*. Cambridge, MA: MIT Press.

Elliott, C. (1992). *The affective reasoner: A process model of emotions in a multi-agent system*. Ph.D. Thesis. Evanston, IL: Northwestern University.

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 572–595). Oxford: Oxford University Press.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*, 906–911.

Flavell, J. H. (2004). Theory-of-mind development: Retrospect and prospect. *Merrill-Palmer Quarterly, 50*, 274–290.

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

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist, 56*, 218–226.

Frijda, N. H. (1986). *The emotions*. Cambridge, England: Cambridge University Press.

Frijda, N. H. (1994). Emotions are functional, most of the time. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion* (pp. 112–136). NY: Oxford University Press.

Goldie, P. (2007). Emotion. *Philosophy Compass, 2*, 928–938.

Gratch, J., & Marsella, S. (2004). A domain-independent framework for modeling emotion. *Cognitive Systems Research, 5*, 269–306.

Green, O. H. (1992). *The emotions: A philosophical theory*. Dordrecht: Kluwer.

Jacob, P. (1997). *What minds can do. Intentionality in a non-intentional world*. Cambridge: Cambridge University Press.

Kappas, A. (2006). Appraisals are direct, immediate, intuitive, and unwitting... and some are reflective. *Cognition and Emotion, 20*, 952–975.

Kelley, T. D. (2003). Symbolic and sub-symbolic representations in computational models of human cognition: What can be learned from biology? *Theory & Psychology, 13*, 847–860.

Kenny, A. (1963). *Action, emotion, and will*. London: Routledge and Kegan Paul.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction–integration model. *Psychological Review, 95*, 163–182.

Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness and Cognition, 9*, 149–171.

Lazarus, R. S. (1982). Thoughts on the relations between emotion and cognition. *American Psychologist, 37*, 1019–1024.

Lazarus, R. S. (1991). *Emotion and adaptation*. New York: Oxford University Press.

Lewis, M. (2000). The emergence of human emotions. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of Emotions* (2nd ed., pp. 265–280). New York: Guilford Press.

Logan, G. D. (1992). Attention and preattention in theories of automaticity. *American Journal of Psychology, 105*, 317–339.

Lyons, W. (1980). *Emotion*. Cambridge: Cambridge University Press.

Macedo, L., Reisenzein, R., & Cardoso, A. (2004). Modeling forms of surprise in artificial agents: Empirical and theoretical study of surprise functions. In K. Forbus, D. Gentner, & T. Reiger (Eds.), *Proceedings of the 26th annual conference of the cognitive science society*, August 2004. Mahwah, NJ: Erlbaum.

Mandler, G. (1984). *Mind and body*. New York: Norton.

Marinier, R., & Laird, J. (2006). A cognitive architecture theory of comprehension and appraisal. In R. Trappl (Ed.), *Cybernetics and systems 2006* (Vol. 2). *Proceedings of the 18th European meeting on cybernetics and systems research*. Vienna: Austrian Society for Cybernetic Studies.

Marks, J. (1982). A theory of emotion. *Philosophical Studies, 42*, 227–242.

Meinong, A. (1894). *Psychologisch-ethische Untersuchungen zur Werttheorie [Psychological-ethical investigations concerning the theory of value]*. Graz, Austria: Leuschner & Lubensky. Reprinted in R. Haller, & R. Kindinger (Eds.) (1968). *Alexius Meinong Gesamtausgabe (Band III*, pp. 3–244). Graz, Austria: Akademische Druck- und Verlagsanstalt.

Meinong, A. (1906). Über Urteilsgefühle: Was sie sind und was sie nicht sind [On judgment-based feelings: What they are and what they are not]. *Archiv für die gesamte Psychologie, 6*, 22–58.

Meinong, A. (1910). *Über Annahmen [On assumptions]* (2nd ed.). Leipzig, Austria: Barth. Reprinted in R. Haller, & R. Kindinger (Eds.) (1977). *Alexius Meinong Gesamtausgabe (Band IV)*. Graz, Austria: Akademische Druck- und Verlagsanstalt.

Mele, A. R. (1992). *Springs of action: Understanding intentional behavior*. NY: Oxford University Press.

Mellers, B. A. (2000). Choice and the relative pleasure of consequences. *Psychological Bulletin, 126*, 910–924.

Mellers, B. A., Schwartz, A., Ho, K., & Ritov, I. (1997). Decision affect theory: Emotional reactions to the outcomes of risky options. *Psychological Science, 8*, 423–429.

Metcalfe, J. (1994). A computational modeling approach to novelty monitoring, metacognition, and frontal lobe dysfunction. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 137–156). 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.

Miceli, M., & Castelfranchi, C. (1997). Basic principles of psychic suffering: A preliminary account. *Theory & Psychology, 7*, 769–798.

Miceli, M., & Castelfranchi, C. (2007). The envious mind. *Cognition and Emotion, 21*, 449–479.

Milsum, J. H. (1966). *Biological control systems analysis*. New York: McGraw-Hill.

Neal Reilly, W. S. (1996). *Believable social and emotional agents* (Ph.D. Thesis No. CMU-CS-96-138). Pittsburgh, PA: Carnegie Mellon University.

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist, 51*, 102–116.

Nichols, S., & Stich, S. (2000). A cognitive theory of pretense. *Cognition, 74*, 115–147.

Nussbaum, M. C. (2001). *Upheavals of thought: The intelligence of emotions*. Cambridge: Cambridge University Press.

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

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge: Cambridge University Press.

Ortony, A., & Partridge, D. (1987). Surprisingness and expectation failure: What's the difference? In *Proceedings of the 10th International Joint Conference on Artificial Intelligence* (pp. 106–108). Los Altos: Morgan Kaufman.

Paglieri, F. (2004). Data-oriented belief revision: Towards a unified theory of epistemic processing. In E. Onaindia & S. Staab (Eds.), *Proceedings of STAIRS 2004* (pp. 179–190). Amsterdam: IOS Press.

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

Picard, R. W. (1997). *Affective computing*. Cambridge, MA: MIT Press.

Proust, J. (2006). Rationality and metacognition in non-human animals. In S. Hurley & M. Nudds (Eds.), *Rational animals?* (pp 247–274). Oxford: Oxford University Press.

Rank, S., & Petta, P. (2005). Appraisal for a character-based story-world. In T. Panayiotopoulos et al. (Eds.), *Intelligent virtual agents 2005* (pp. 495–496). Berlin: Springer.

Reisenzein, R. (1994). Kausalattribution und Emotion [Causal attribution and emotion]. In F. Försterling, & J. Stiensmeier-Pelster (Eds.), *Attributionstheorie: Grundlagen und Anwendungen* (pp. 123–161). Göttingen: Hogrefe.

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). Outlines of a theory of emotions as metarepresentational states of mind. In A. H. Fischer (Ed.), *ISRE'98. Proceedings of the 10th conference of the international society for research on emotions, 1998* (pp. 181–191). Amsterdam: Faculty of Psychology.

Reisenzein, R. (1999). A theory of emotions as metarepresentational states of mind. Unpublished manuscript, University of Bielefeld [available from the author on request].

Reisenzein, R. (2000). The subjective experience of surprise. In H. Bless & J. P. Forgas (Eds.), *The message within: The role of subjective experience in social cognition and behavior* (pp. 262–279). Philadelphia, PA: Psychology Press.

Reisenzein, R. (2001). Appraisal processes conceptualized from a schema-theoretic perspective: Contributions to a process analysis of emotions. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 187–201). Oxford: Oxford University Press.

Reisenzein, R. (2006a). Arnold's theory of emotion in historical perspective. *Cognition and Emotion, 20*, 920–951.

Reisenzein, R. (2006b). Emotions as metarepresentational states of mind. In R. Trappl (Ed.), *Cybernetics and systems 2006. Proceedings of the 18th European meeting on cybernetics and systems research* (pp. 649–653). Vienna: Austrian Society for Cybernetic Studies.

Reisenzein, R. (2007). What is a definition of emotion? And are emotions mental-behavioral processes? *Social Science Information, 46*, 424–428.

Reisenzein, R. (2008). Emotional experience in the computational belief–desire theory of emotion. *Emotion Review*, in press.

Reisenzein, R., & Junge, M. (2006). Überraschung, Enttäuschung und Erleichterung: Emotionsintensität als Funktion von subjektiver Wahrscheinlichkeit und Erwünschtheit [Surprise, disappointment and relief: Emotion intensity as a function of subjective probability and desire strength]. Paper presented at the 45th Congress of the German Psychological Association (DGPs) 2006 in Nürnberg.

Reisenzein, R., Meyer, W.-U., & Schützwohl, A. (2003). Einführung in die Emotionspsychologie, Band III: Kognitive Emotionstheorien. *Introduction to the psychology of emotions Cognitive emotion theories* (Vol. 3). Bern: Huber.

Richerson, P., & Boyd, R. (1998). The evolution of ultrasociality. In I. Eibl-Eibesfeldt & F. K. Salter (Eds.), *Indoctrinability, ideology and warfare* (pp. 71–96). New York: Berghahn Books.

Roseman, I. J. (1979). Cognitive aspects of emotions and emotional behavior. *Paper presented at the 87th Annual Convention of the APA*, New York City.

Roseman, I. J., Antoniou, A. A., & Jose, P. E. (1996). Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory. *Cognition and Emotion, 10*, 241–277.

Rosenblueth, A., Wiener, N., & Bigelow, J. (1943). Behavior, purpose, and teleology. *Philosophy of Science, 10*, 18–24.

Roskies, A. L. (1999). The binding problem. *Neuron, 24*, 7–9.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92–129). Oxford: Oxford University Press.

Scherer, K. R., Schorr, A., & Johnstone, T. (2001). *Appraisal processes in emotion: Theory, methods, research*. Oxford: Oxford University Press.

Schiffer, S. (1994). The language-of-thought relation and its implications. *Philosophical Studies, 76*, 263–285.

Schwarz, N., & Clore, G. L. (1996). Feelings and phenomenal experiences. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social Psychology: Handbook of basic principles* (pp. 433–465). New York, NY: Guilford Press.

Searle, J. (1983). *Intentionality*. Cambridge: Cambridge University Press.

Solomon, R. C. (1976). *The passions*. Garden City, NY: Anchor Press.

Staller, A. & Petta, P. (2001). Introducing emotions into the computational study of social norms: A first evaluation. *Journal of Artificial Societies and Social Simulation,* 4(1). <http://www.soc.surrey.ac.uk/JASSS/4/1/2.html>.

Sterelny, K. (1991). *The representational theory of mind: an introduction*. Oxford: Blackwell.

Sun, R., Zhang, X., & Mathews, R. (2006). Modeling meta-cognition in a cognitive architecture. *Cognitive Systems Research, 7*, 327–338.

Tye, M. (1995). *Ten problems of consciousness. A representational theory of the phenomenal mind*. Cambridge, MA: MIT Press.

Wellman, H. M. (2002). Understanding the psychological world: Developing a theory of mind. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 167–187). Malden, MA: Blackwell.

Yacubian, J., Gläscher, J., Schroeder, K., Sommer, T., Braus, D. F., & Büchel, C. (2006). Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. *Journal of Neuroscience, 26*, 9530–9537.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35*, 151–175.

Zeelenberg, M., van Dijk, W. W., Manstead, A. S. R., & van der Pligt, J. (2000). On bad decisions and disconfirmed expectancies: The psychology of regret and disappointment. *Cognition and Emotion, 14*, 521–541.