Extending the computational belief-desire theory of emotions to fantasy emotions

1 author:

Rainer Reisenzein
University of Greifswald
100 PUBLICATIONS   **3,065** CITATIONS

Some of the authors of this publication are also working on these related projects:

Computational Belief Desire Theory of Emotion View project

Extending the Computational Belief-Desire Theory of Emotions to Fantasy Emotions

Rainer Reisenzein (rainer.reisenzein@uni-greifswald.de)
Institute of Psychology, University of Greifswald
Franz-Mehring-Straße 47, 17489 Greifswald, Germany

Keywords: Computational belief-desire theory of emotion, affective computing, fantasy emotions, imaginative emotions, emotional reactions to fiction, simulation theory, paradox of fiction.

Literary and cinematic depictions of events frequently evoke emotions in readers or viewers even when they are fully aware that the portrayed events are fictitious. Likewise, the deliberate imagination of counterfactual events can evoke emotions. These "fantasy emotions", as the Austrian philosopher-psychologist Meinong (1910) called them, pose an explanatory problem for cognitive emotion theories. The reason is that these theories assume—at least in their standard form—that emotions presuppose *beliefs* in the existence of the emotion-elicing events; but such beliefs seem to be lacking in the case of the fantasy emotions (e.g., Green, 1992). In philosophy of art, these considerations have given rise to the much-discussed "paradox of emotional response to ficton" (Radford, 1975; Schneider, 2011). To solve the problem, Meinong (1910) proposed that the fantasy emotions are not based on beliefs but on a different kind of cognitive propositional attitude, called *assumptions* (*Annahmen*). My aim is to explicate Meinong's theory of fantasy emotions in the context of CBDTE, a (sketch of a) computational (C) model of the belief-desire theory of emotion (BDTE) (see Reisenzein, 2009a; 2009b).

The Belief-Desire Theory of Emotion

BDTE is a version of cognitive emotion theory (see e. g., Marsella, Gratch, & Petta, 2010). Its basic assumption is that the core set of the mental states presystematically called "emotions" presuppose, for their existence, both *beliefs* (cognitive or informational states) and *desires* (motivational states) about the emotion-eliciting states of affairs. Thus, the conceptual framework of BDTE is the same as that of the belief-desire theory of action that inspired the BDI (belief-desire-intention) approach to artificial agents (e.g., Bratman, Israel, & Pollack, 1988; Hindriks, 2009). More precisely, emotions are reactions to the cognized actual or potential fulfilment or frustration of desires; plus, in some cases (e.g., relief, disappointment), the confirmation or disconfirmation of beliefs (Reisenzein, 2009a; 2009b). To illustrate, Mary is *happy* that *p* (e.g., that *Mr. Schroiber was elected chancellor*) if she desires *p* and now comes to believe firmly (i.e., is certain) that *p* is the case; whereas Mary is *unhappy* that *p* if she is averse to *p*, and now comes to believe firmly that *p* is the case.

CBDTE: A Computational Explication of BDTE

Following Fodor (1987), CBDTE (Reisenzein, 2009a; 2009b) assumes that beliefs and desires (the causes of emotion according to BDTE) are special modes of processing propositional representations, i.e. sentences in a "language of thought". It is assumed that the central part of this propositional representation system is innate and that its innate components comprise a set of hardwired maintenance and updating mechanisms. At the core of these mechanisms are two comparator devices, the belief-belief comparator (BBC) and the belief-desire comparator (BDC). The BBC compares newly acquired beliefs to pre-existing beliefs, whereas the BDC compares them to existing desires. Computationally speaking, the BBC and BDC compare the "mentalese" sentence tokens s_{new} representing the contents of newly acquired beliefs, with the sentences s_{old} representing the contents of pre-existing beliefs and desires. If either a match (s_{new} is identical to s_{old}) or a mismatch (s_{new} is identical to not-s_{old}) is detected, the comparators generate an output that communicates the detection and degree of the match or mismatch to the rest of the cognitive system. CBDTE assumes that the comparator mechanisms operate automatically (without intention, and preconsciously) and that their outputs are *nonpropositional*: They consist of signals that vary in kind and intensity, but have no internal structure, and hence are analogous to sensations (e.g., of tone or temperature). Output signals that exceed a certain threshold of intensity give rise, directly or indirectly, to unique conscious feeling qualities: the feelings of surprise and expectancy confirmation (BBC), and the feelings of pleasure and displeasure (BDC). According to CBDTE, the BDC and BBC are the basic emotion mechanisms of humans.

Fantasy Emotions in CBDTE

Meinong (1910) proposes that *assuming* is a special mode of cognitively representing states of affairs: the person posits, or hypothetically supposes, that *p* is the case. Furthermore, he suggests that whereas serious emotions are based on beliefs, fantasy emotions are based on assumptions. In the framework of BDTE, this suggestion can be interpreted as follows: One experiences serious joy about *p* if one desires *p* and believes (or more precisely, comes to believe) that *p* is the case; whereas one experiences fantasy joy about *p* if one desires *p* and *assumes* that *p* is the case (Reisenzein, 2012).

To incorporate fantasy feelings into CBDTE, I begin by assuming that, like believing *p* and desiring *p*, *assuming p* is a special mode of processing propositional representations.

An elaboration of this idea has been proposed by Nichols and Stich (2003) in their theory of mental simulation. However, to explain fantasy emotions, important extensions of this model are needed. These extensions are directly suggested by CBDTE's assumptions about serious emotions. Specifically, I assume that the updating mechanisms for assumptions include hardwired comparator mechanism analogous to the BBC and BDC: An *assumption-assumption comparator* (AAC), and an *assumption-desire comparator* (ADC). The AAC compares newly made assumptions with existing assumptions, whereas the ADC compares newly made assumptions with existing desires. Fantasy emotions arise when the AAC or the ADC detect an agreement or a conflict between (a) a newly made assumption and (b) an existing assumption or desire, respectively.

To illustrate, Mary experiences fantasy joy about Schroiber's election victory (= *p*) if she desires *p* and assumes *p* to be the case. On the computational level, this corresponds to: Mary's ADC discovers that the mental sentence representing the content of an existing desire is identical to that of a newly made assumption; as a consequence, it generates a nonpropositional signal that communicates the detection of this agreement to the rest of the cognitive system, and that is subjectively experienced as a feeling of fantasy pleasure. Analogously, Mary experiences *fantasy displeasure* about *p* if she is averse against *p* and assumes *p* to be the case. On the computational level, this corresponds to: Mary's ADC discovers that there is a contradiction between the content of an existing desire and the newly made assumption *p*; it then generates a signal which communicates the detection of this incongruence to the rest of the cognitive system, and which is subjectively experienced as fantasy displeasure. Mary can also experience *fantasy surprise*—namely, if she first assumed that Schroiber did not win the election (not-*p*) and then makes the new assumption that Schroiber did, after all, win the election (*p*). In this case, Mary's AAC detects a contradiction between an assumption that is part of a current simulation and a newly made assumption, and as a consequence generates a signal that is experienced as fantasy surprise.

Explanatory Capacity of the Theory

CBDTE can explain the thorough-going *parallelism* between fantasy feelings and serious feelings. Each serious emotion (joy, sorrow, fear, hopeetc.) can also occur in a fantasy form (as fantasy joy, fantasy sorrow, and so on). Likewise, both serious and fantasy emotions can be experienced in different intensities and both can be directed at the same state of affairs. According to CBDTE, this parallelism between serious and fantasy emotions is the consequence of the parallel construction of their generating mechanisms. CBDTE can also account for the different motivational effects of serious and fantasy emotions (see Schneider, 2011): Whereas serious emotions often motivate coping actions, the corresponding fantasy emotions usually do not have such effects. CBDTE can explain this difference, at least in part, by the assumption that an immediate update of beliefs and desires takes place only in the case of serious emotions, but not in the case of the fantasy emotions. Fantasy emo-

tions can influence actions only indirectly; in particular, by generating beliefs *about* fantasy emotions. Finally, the CBDTE theory of fantasy emotions throws new light on the question of whether or not fantasy emotions qualify as "genuine" emotions (Reisenzein, 2012).

References

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

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

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

Hindriks, K. V. (2009). Programming rational agents in GOAL. In R. H. Bordini, M. Dastani, J. Dix, & A. El Fallah Seghrouchni (Eds.), *Multiagent Programming: Languages, tools and applications* (pp. 119-157). Berlin: Springer.

Marsella, S., Gratch, J., & Petta, P. (2010). Computational models of emotion. In K. R. Scherer, T. Bänziger, & E. Roesch, Eds.) *A blueprint for a affective computing: A sourcebook and manual*. Oxford: Oxford University Press.

Meinong, A. (1910). *Über Annahmen* [*On Assumptions*]. (2nd Ed.) Leipzig, Germany: Barth. Reprinted in: R. Haller & R. Kindinger (Eds.) (1977). *Alexius Meinong Gesamtausgabe* [*Alexius Meinong's complete works*]. (Vol. 4). Graz, Austria: Akademische Druck- und Verlagsanstalt.

Nichols, S., & Stich, S. (2003). *Mindreading: An integrated account of pretense, self-awareness and understanding other minds*. Oxford: Oxford University Press.

Radford, C. (1975). How can we be moved by the fate of Anna Karenina? *Proceedings of the Aristotelian Society*, Supplemental Vol. *49*, 67-80.

Reisenzein, R. (2009a). Emotions as metarepresentational states of mind: Naturalizing the belief-desire theory of emotion. *Cognitive Systems Research*, *10*, 6-20.

Reisenzein, R. (2009b). Emotional experience in the computational belief-desire theory of emotion. *Emotion Review*, *1*, 214-222.

Reisenzein, R. (2012). Fantasiegefühle aus der Sicht der kognitiv-motivationalen Theorie der Emotion [Fantasy feelings from the perspective of the cognitive-motivational theory of emotion]. In S. Poppe (Ed.), *Emotionen in Literatur und Film* (pp. 31-63). Würzburg: Königshausen & Neumann.

Schneider, S. (2011). The paradox of fiction. *Internet Encyclopedia of philosophy*. http://www.iep.utm.edu/fict-par/ (retrieved June 17, 2011).