

The OCC Model Revisited

Bas R. Steunebrink, Mehdi Dastani, and John-Jules Ch. Meyer

Department of Information and Computing Sciences,
Utrecht University, The Netherlands
{bass, mehdi, jj}@cs.uu.nl

Abstract. Although popular among computer scientists, the OCC model of emotions contains a number of ambiguities that stand in the way of a truthful formalization or implementation. This paper aims to identify and clarify several of these ambiguities. Furthermore, a new inheritance-based view of the logical structure of emotions of the OCC model is proposed and discussed.

1 Introduction

In their book “The Cognitive Structure of Emotions” [1], Ortony, Clore & Collins have proposed a very interesting model of emotions that provides a clear and convincing structure of the eliciting conditions of emotions and the variables that affect their intensities. This psychological model is popular among computer scientists that are building systems that reason about emotions or incorporate emotions in artificial characters.

However, although many ad hoc or simplified implementations of “the OCC model” have been made, there have been fewer attempts at formalizing the complete, logical structure of the proposed emotion model (e.g., [2–4]). We are attempting to do so, using a formal logic containing constructs to reason about agents, their beliefs and actions, objects, and events. We are currently formalizing the eliciting conditions of emotions in this logic, trying to stay as close as possible to the book. Unfortunately, we have run into several ambiguities that stand in the way of a truthful formalization. We realize that satisfying logicians was not the primary concern of the book, but given their aim to provide a “computationally tractable account” (quoted from the back cover), it is important that computer scientists have an unambiguous structure of emotions to work with.

The contribution of this paper is to provide clarifications of the logical structure underlying the OCC model. After identifying several issues, we propose an inheritance-based view of the OCC model, supported by a new logical structure (figure 2) and new emotion type specifications (table 2). The structure that we propose aims to resolve several ambiguities in the OCC model, while our emotion type specifications have a stronger correspondence with the logical structure.

It should be emphasized that the current authors are not psychologists. We certainly do not claim to have more knowledge of emotions than OCC. We are computer scientists and logicians, and we have studied the OCC model from this perspective. Our critique only concerns the logical structure underlying the OCC model. Nevertheless, it is entirely possible that our interpretation of the OCC model is colored by our computer scientist’s way of thinking. However, we do not find this possibility problematic, because the aim of our research into emotions is to make emotions *computable* (and make useful systems with them, of course).

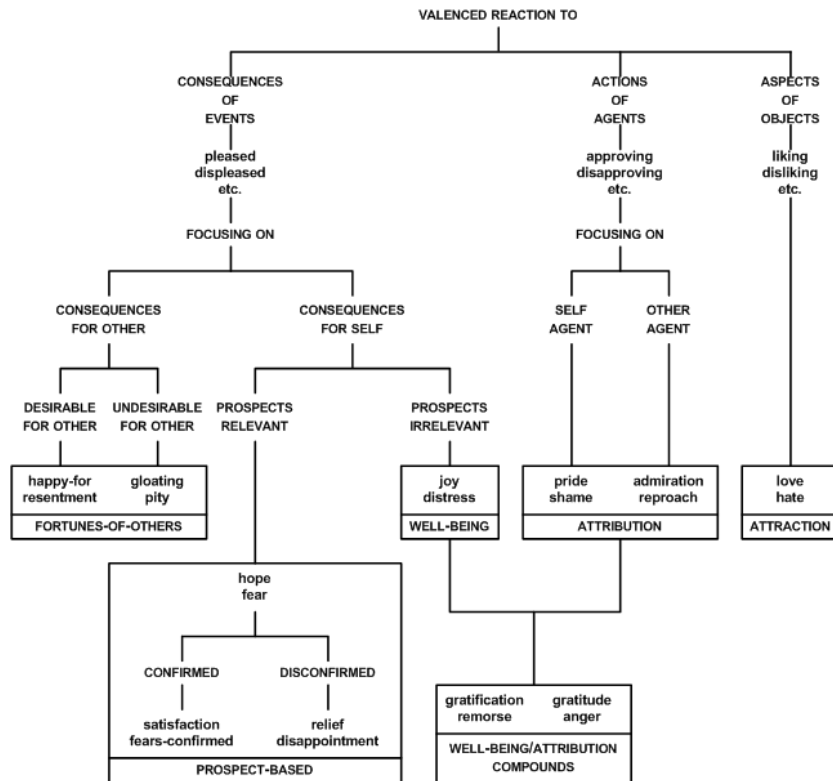


Fig. 1. The original structure of emotions of the OCC model, copied from page 19 [1].

2 The OCC Model

The OCC model describes a hierarchy that classifies 22 emotion types (see figure 1). The hierarchy contains three branches, namely emotions concerning consequences of events (e.g., joy and pity), actions of agents (e.g., pride and reproach), and aspects of objects (e.g., love and hate). Additionally, some branches combine to form a group of compound emotions, namely emotions concerning consequences of events *caused* by actions of agents (e.g., gratitude and anger). Because these notions (i.e. events, actions, and objects) are also commonly used in agent models, this makes the OCC model suitable for use in artificial agents.

Throughout the book, specifications are given for each of the 22 emotion types. For example, below is the specification of the class of emotions labeled as ‘fear’ in the OCC model (copied from page 112 [1]):

<p>FEAR EMOTIONS TYPE SPECIFICATION: (displeased about) the prospect of an undesirable event TOKENS: apprehensive, anxious, cowering, dread, fear, fright, nervous, petrified, scared, terrified, timid, worried, etc. VARIABLES AFFECTING INTENSITY: (1) the degree to which the event is undesirable (2) the likelihood of the event EXAMPLE: The employee, suspecting he was no longer needed, feared that he would be fired.</p>

Joy: (pleased about) a desirable event
Distress: (displeased about) an undesirable event
Happy-for: (pleased about) an event presumed to be desirable for someone else
Pity: (displeased about) an event presumed to be undesirable for someone else
Gloating: (pleased about) an event presumed to be desirable for someone else
Resentment: (displeased about) an event presumed to be desirable for someone else
Hope: (pleased about) the prospect of a desirable event
Fear: (displeased about) the prospect of an undesirable event
Satisfaction: (pleased about) the confirmation of the prospect of a desirable event
Fears-confirmed: (displeased about) the confirmation of the prospect of an undesirable event
Relief: (pleased about) the disconfirmation of the prospect of an undesirable event
Disappointment: (displeased about) the disconfirmation of the prospect of a desirable event
Pride: (approving of) one's own praiseworthy action
Shame: (disapproving of) one's own blameworthy action
Admiration: (approving of) someone else's praiseworthy action
Reproach: (disapproving of) someone else's blameworthy action
Gratification: (approving of) one's own praiseworthy action and (being pleased about) the related desirable event
Remorse: (disapproving of) one's own blameworthy action and (being displeased about) the related undesirable event
Gratitude: (approving of) someone else's praiseworthy action and (being pleased about) the related desirable event
Anger: (disapproving of) someone else's blameworthy action and (being displeased about) the related undesirable event
Love: (liking) an appealing object
Hate: (disliking) an unappealing object

Table 1. The emotion type specifications of the OCC model, copied from the book [1].

Besides an example at the bottom, these specifications have three elements:

(1) The *type specification* provides, in a concise sentence, the conditions that elicit an emotion of the type in question.

(2) A list of *tokens* is provided, showing which emotion words can be classified as belonging to the emotion type in question. For example, ‘fright’, ‘scared’, and ‘terrified’ are all types of fear. (Of course, ‘fear’ is also a type of fear.)

(3) For each emotion type, a list of *variables affecting intensity* is provided. These variables are local to the emotion type in question, i.e., global variables (such as arousal) that affect all emotions are not included. The idea is that higher values for these variables result in higher emotional intensities.

In table 1 we have summarized the type specifications of all 22 emotion types.

3 Ambiguities and Clarifications

The presentation of the structure of the OCC model (see figure 1) is sure to make many computer scientists (including us) think of an inheritance diagram; that is, each emotion type is like its parent type plus some specialization. For example, displeased is a negatively valenced reaction to a consequence of an event; distress is displeased *plus* a focus on the self; regret is distress *plus* the constraint that the consequence in question signals a loss of opportunity, etc. Although it may not have been the authors’ intention to present the structure of emotions as an inheritance hierarchy, we think it is actually useful to think of it that way and pursue this direction further. However, this interpretation reveals several ambiguities in the logical structure underlying the OCC model. In this section, we will list a number of such issues, together with our proposed clarifications.

(1) In table 1, the phrase “desirable event” is used many times. However, events are actually always appraised with respect to their *consequences*. For example, an earthquake in itself does not have a valence; only the consequences of this event (e.g., valuable lessons for seismologists, property damage, loss of life) are appraised as being

desirable or undesirable. Because desirability only applies to consequences of events, every instance of the phrase “desirable event” should actually be read as a shorthand for “desirable consequence of an event.” Furthermore, the term “prospect” (used in, e.g., hope and fear) is intentionally ambiguous: it is used to refer to both *future* events and *uncertain* (past or current) events. Many formalizations appear to use OCC’s notion of prospect in only one of these senses. For example, Adam [4] and Gratch & Marsella [2] only used uncertain prospects when formalizing hope and fear, whereas Steunebrink, Dastani & Meyer [3] only used future prospects.

(2) Looking closely at figure 1, we see that each of the three branches is headed by a set of emotional words; namely, pleased/displeased for event-based emotions, approving/disapproving for action-based emotions, and liking/disliking for object-based emotions. These are supposed to be the most *undifferentiated* emotion types. For example, “being pleased” represents the situation where one has appraised a consequence of an event as being desirable, but it says nothing about, e.g., whether that consequence is viewed as prospective or actual (i.e., one cannot say whether it is hope or joy), or whether that consequence is presumed to be desirable or undesirable for someone else (i.e., one cannot say whether it is happy-for or gloating). When the structure as shown in figure 1 is regarded as an inheritance hierarchy, pleased/displeased, approving/disapproving and liking/disliking then become *generalized* emotion types, from which all emotion types below them are derived. Indeed, in OCC’s structure of *variables affecting intensity*, each emotion type inherits all variables from its parents. For example, ‘gratification’ inherits all variables from ‘joy’ and ‘pride’ (this can even be seen as a kind of *multiple inheritance*).

(3) As can be seen in figure 1, ‘joy’ is classified as an emotion type arising from positively appraising the consequences of an event where the focus is on consequences for the self and prospects are irrelevant. However, the type specification of ‘joy’ is given as “(pleased about) a desirable event” (see table 1), with no mention of a focus on the self or the irrelevance of prospects. So either the type specification of ‘joy’ is incomplete, or a focus on the self and a disregard of prospects is implicitly assumed to be the default. (In personal communication, OCC have acknowledged the latter to be the case.) A downside of assuming defaults, however, is that the type specification of ‘joy’ then conflates with ‘pleased’; just as ‘joy’, ‘pleased’ is a valenced reaction to a desirable event. The difference is only implicit; namely, ‘pleased’ is undifferentiated, whereas ‘joy’ has a default differentiation on the self and a disregard of prospects.

(4) As can be seen in table 1, ‘joy’ is specified as “(pleased about) a desirable event” and ‘distress’ is specified as “(displeased about) an undesirable event.” Crossing these specifications, one may wonder whether there is such a thing as “being pleased about an undesirable event” or “being displeased about a desirable event.” The same holds for other pairs of opposite emotion types (i.e., pride/shame, love/hate, etc.): almost all type specifications contain a duplicate positivity or negativity. Often this is not problematic; for example, if one approves of an action, it must be praiseworthy, and vice versa. However, in some cases this duplication can introduce ambiguity; for example, in the type specification of ‘hope’, does “being pleased” enforce the phrase “desirable event,” or does the pleasure apply to having the “prospect” thereof? In other cases, pleased/displeased and desirable/undesirable are indeed crossed; for example, compare

the type specifications of ‘relief’ and ‘disappointment’ in table 1. In the next section we will propose a way to remove these duplications and ambiguities.

(5) In figure 1, the six prospect-based emotion types are grouped in a framed structure with an inner hierarchy. How should this nesting be interpreted? On page 19 [1], OCC explicitly state that the figure represents a *logical* structure of emotions, not a *temporal* one. This would mean that both satisfaction/fears-confirmed and relief/disappointment are more differentiated types of hope/fear, or, to continue our inheritance-based perspective, they are *specializations* of hope/fear. In particular, disappointment would then have to be a specialization of fear. However, this cannot be, because disappointment does not inherit anything from fear; it inherits its *variables affecting intensity* from hope, but its valence is reversed (hope is positive and disappointment is negative). The relation between hope and disappointment appears to be more of a temporal kind. For example, first Bob hopes Alice will show up for their date, but when she does not, his hope turns into disappointment. Thus satisfaction, fears-confirmed, relief, and disappointment are not special kinds of hope or fear, but more like continuations of hope or fear, counting from the point when an event has been perceived that signals the confirmation or disconfirmation of the thing hoped for or feared. In other words, these four types are emotions in response to *actual* consequences of events, namely consequences signaling the confirmation or disconfirmation of a *previously prospective* consequence. In the next section we will propose to move these four emotion types from under hope/fear to become specializations of joy/distress.

(6) Consider the group of fortunes-of-others emotion types (happy-for, resentment, gloating, pity) and their type specifications in table 1. For there to be a ‘happy-for’ emotion, the consequence that is desirable for the other must also be desirable for oneself to some degree (probably because it satisfies an interest goal; page 94 [1]). But if a consequence of an event is appraised as being desirable for oneself, the conditions for ‘joy’ are satisfied (see table 1). So logically speaking, happy-for/gloating implies joy, and resentment/pity implies distress. If the specifications of joy and distress are subsets of those of the fortunes-of-others emotions, then joy and distress are generalizations of happy-for, resentment, gloating, and pity and should thus be their parents in the hierarchy.

(7) The structure of emotions as illustrated in figure 1 contains the phrases “focusing on” and “prospects (ir)relevant” several times. We believe this terminology weakens the logical structure, or at least the inheritance view of it. It may indeed be intuitive to say that ‘joy’ is an emotion type where one is “focusing on” a consequence for the self and that ‘happy-for’ is an emotion type where one is “focusing on” a consequence for someone else, but as we have just shown above, ‘happy-for’ logically implies ‘joy’. Moreover, we have also shown in (5) that the connections from hope/fear to the emotion types below them are not purely logical but also temporal. So there are relations between emotion types in the structure that are not captured by presenting it as in figure 1. In the next section, this will be resolved by abandoning the phrases “focusing on” and “prospects (ir)relevant” and adopting a stricter inheritance-based view.

(8) Finally, the branch of attraction-based emotions (i.e., liking/disliking and love/hate) looks a bit awkward to us, because there are no conditions to distinguish love/hate from its generalization liking/disliking. In other words, love/hate does not seem to ex-

tend its parent type, at least not judging from figure 1. However, whereas liking/disliking has appealingness as its only *variable affecting intensity*, love/hate extends this by adding the variable *familiarity*. According to OCC, the more familiar one is with an appealing object, the more one will love it, and the more familiar one is with an unappealing object, the more one will hate it. Interestingly, OCC have chosen not to differentiate based on familiarity. In contrast, below approving/disapproving, the variable *strength of cognitive unit* is used to differentiate between pride/shame on the one hand (i.e., the acting agent is in a cognitive unit with the experiencing agent) and admiration/reproach on the other hand (i.e., the acting agent is distinct from the experiencing agent). Likewise, below pleased/displeased, the variable *likelihood* is used to differentiate between hope/fear on the one hand (i.e., an event is possible but not certain) and joy/distress on the other hand (i.e., an event has actually happened). Analogously, one could use the variable *familiarity* to differentiate between love/hate on the one hand (i.e., for familiar objects) and interest/disgust on the other hand (i.e., for unfamiliar objects).

4 The OCC Model Revisited

The structure of emotions as proposed by OCC and illustrated in figure 1 may be intuitive, but it falls short of capturing the logical structure underlying the OCC model. Taking the points of the previous section into consideration, we have constructed a new hierarchy, as shown in figure 2. It may not be as beautiful as the original figure; e.g., the naming of groups of emotion types (“attraction,” “attribution,” etc.) has been lost. Although it does not contain anything really new (we are not psychologists, after all), the structure has been altered in several ways to bring it more in line with our computer scientists’ inheritance-based way of thinking. Below we elaborate on the differences.

(1) Figure 2 represents the inheritance structure explicitly, because it has labels at every point in the hierarchy, and the conditions of every child node are a superset of those of its parent node(s).

(2) Along each connection are written (in small caps) the additional condition(s) that make an emotion type a specialization of its parent type(s). Ambiguous terms are avoided; for example, we have omitted or replaced the phrases “focusing on” and “prospects (ir)relevant.” In distinguishing pride/shame from admiration/reproach, the action in question must have been performed by either the self or another agent, so the phrase “focusing on” is redundant. In distinguishing hope/fear from joy/distress, we consider the consequence in question to be perceived as either prospective or actual. In contrast, the phrase “prospects irrelevant,” as used in figure 1, suggests that there is no differentiation on prospects for joy/distress, which would mean that joy/distress is a generalization of hope/fear, which would contradict figure 1 because joy/distress is depicted next to hope/fear instead of above it. To avoid such ambiguities we have strived to use clearer terms.

(3) Satisfaction, fears-confirmed, relief, and disappointment have been moved from under hope/fear to become specializations of joy/distress. This is because we regard a confirmation or disconfirmation to be an *actual* consequence (of an event). According to OCC, these four emotion types require an “attendant” hope or fear emotion; however, we interpret this as meaning that one emotion requires the *presence* of another emotion

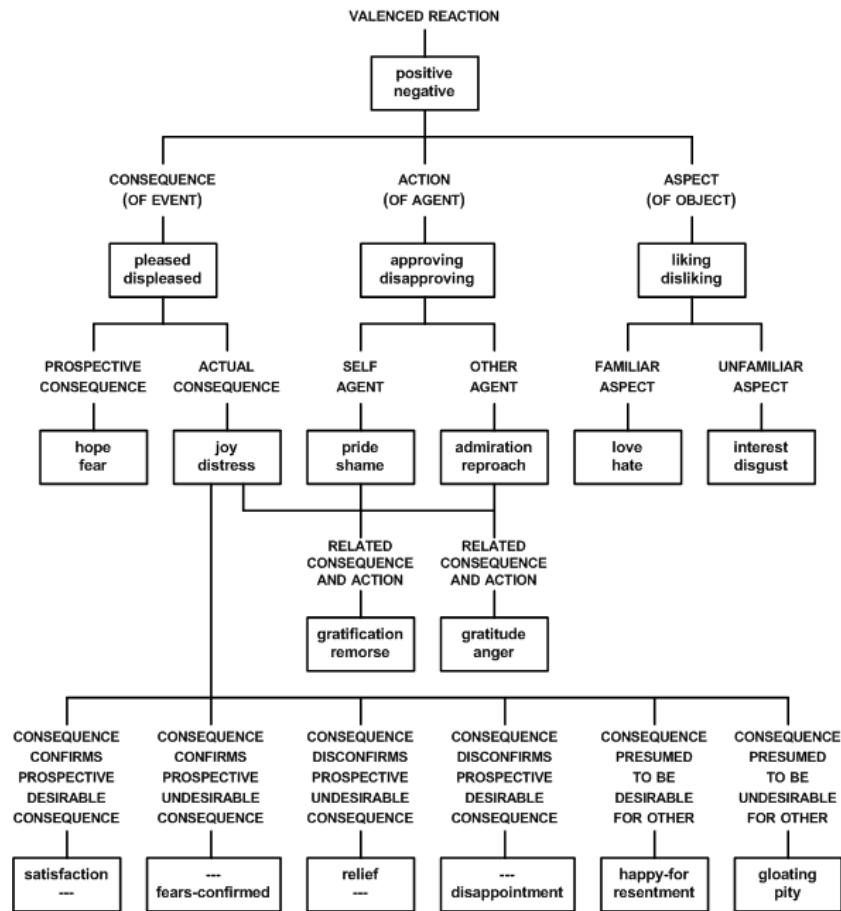


Fig. 2. A disambiguated, inheritance-based hierarchy of emotions of the OCC model.

to be elicited (e.g., relief requires that something is or was feared), not that one emotion is a specialization of another emotion (e.g., relief is not a special type of fear). The conditions that we have put on the connections in figure 2 capture this interpretation as follows: the phrase “prospective desirable consequence” matches the type specification for hope, and “prospective undesirable consequence” matches with fear. Thus the conditions of satisfaction, fears-confirmed, relief, and disappointment require the existence of an attendant hope or fear emotion, in line with OCC’s specifications.

(4) Happy-for, resentment, gloating, and pity have been moved from the left of figure 1 to the bottom right of figure 2. This is due to the reasoning of point (6) in the previous section.

(5) In line with point (8) in the previous section, we have added interest/digust as an additional (i.e., besides love/hate) specialization of liking/disliking, based on the familiarity with the object in question.

(6) One thing that we find particularly attractive about figure 2 is that type specifications now follow *immediately* from the diagram. Descriptions can easily be formed by

<i>positive</i> and <i>negative</i> are valenced reactions (to “something”)
<i>pleased</i> is being <i>positive</i> about a consequence (of an event)
<i>displeased</i> is being <i>negative</i> about a consequence (of an event)
<i>hope</i> is being <i>pleased</i> about a prospective consequence (of an event)
<i>fear</i> is being <i>displeased</i> about a prospective consequence (of an event)
<i>joy</i> is being <i>pleased</i> about an actual consequence (of an event)
<i>distress</i> is being <i>displeased</i> about an actual consequence (of an event)
<i>satisfaction</i> is <i>joy</i> about the confirmation of a prospective desirable consequence
<i>fears-confirmed</i> is <i>distress</i> about the confirmation of a prospective undesirable consequence
<i>relief</i> is <i>joy</i> about the disconfirmation of a prospective undesirable consequence
<i>disappointment</i> is <i>distress</i> about the disconfirmation of a prospective desirable consequence
<i>happy-for</i> is <i>joy</i> about a consequence (of an event) presumed to be desirable for someone else
<i>resentment</i> is <i>distress</i> about a consequence (of an event) presumed to be desirable for someone else
<i>loathing</i> is <i>joy</i> about a consequence (of an event) presumed to be undesirable for someone else
<i>pity</i> is <i>distress</i> about a consequence (of an event) presumed to be undesirable for someone else
<i>approving</i> is being <i>positive</i> about an action (of an agent)
<i>disapproving</i> is being <i>negative</i> about an action (of an agent)
<i>pride</i> is <i>approving</i> of one’s own action
<i>shame</i> is <i>disapproving</i> of one’s own action
<i>admiration</i> is <i>approving</i> of someone else’s action
<i>reproach</i> is <i>disapproving</i> of someone else’s action
<i>ratification</i> is <i>pride</i> about an action and <i>joy</i> about a related consequence
<i>remorse</i> is <i>shame</i> about an action and <i>distress</i> about a related consequence
<i>gratitude</i> is <i>admiration</i> about an action and <i>joy</i> about a related consequence
<i>anger</i> is <i>reproach</i> about an action and <i>distress</i> about a related consequence
<i>liking</i> is being <i>positive</i> about an aspect (of an object)
<i>disliking</i> is being <i>negative</i> about an aspect (of an object)
<i>love</i> is <i>liking</i> a familiar aspect (of an object)
<i>hate</i> is <i>disliking</i> a familiar aspect (of an object)
<i>interest</i> is <i>liking</i> an unfamiliar aspect (of an object)
<i>disgust</i> is <i>disliking</i> an unfamiliar aspect (of an object)

Table 2. These emotion type specifications correspond directly to figure 2.

following any link from child to parent node, inserting the text on the link. The resulting type specifications are displayed in table 2 (cf. table 1).

5 Conclusion and Future Work

In this paper we have identified a number of ambiguities in the logical structure underlying the popular OCC model of emotions [1]. If these ambiguities are not resolved, computer scientists wishing to formalize or implement emotions may come up with conflicting interpretations of the psychological model. Therefore we have proposed a new inheritance-based view of the structure of emotions, which addresses the ambiguities that we have identified. This effort is part of our goal to formalize emotions in logic in a way which is both truthful to the OCC model and useful for Artificial Intelligence.

References

1. Ortony, A., Clore, G.L., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press, Cambridge, UK (1988)
2. Gratch, J., Marsella, S.: A domain-independent framework for modeling emotions. Journal of Cognitive Systems Research **5**(4) (2004) 269–306
3. Steunebrink, B.R., Dastani, M., Meyer, J.-J.Ch.: A logic of emotions for intelligent agents. In: Proceedings of the 22nd Conference on Artificial Intelligence (AAAI’07), AAAI Press (2007)
4. Adam, C.: The Emotions: From Psychological Theories to Logical Formalization and Implementation in a BDI Agent. PhD thesis, Institut National Polytechnique de Toulouse (2007)