A Formal Interpretation of the Concept of Emotion

Jonathan Liu

R&D, P&G Co.
1# Bin He Road, HP Econ. Tech. Development District
Guangzhou 510730, China
Liu.jo@pg.com

Abstract

To determine what an emotion is, is a matter of conceptualization—namely, understanding and explicitly encoding the concept of emotion as people use it in everyday life. This is a notoriously difficult problem (Frijda, 1986; Fehr & Russell, 1984). The purpose of this paper is to give a formal representation of the concept of emotion. The view underlying our approach is that primitive concepts are constructed out of the observation of behavioral regularities. The behavioral regularities required for characterizing emotion concepts, together with a level at which the regularities are described, are examined. We argue that the concepts of "emotion" and "desire" are one and the same, and we define the concept of desire (or action tendency) in a formal language. We also address the concept of emotion intensity.

1. Introduction

To approach the problem of emotion, physiologists consider some physiological measures, and some AI researchers propose computational processes (e.g., Sloman, 1981). Yet the question is, are these things emotions? That is, what determines whether the label emotion can be attached to a phenomenon? So first we must determine exactly what we mean by "emotion".

Previous approaches to defining the concept of emotion have a common problem—namely, that such concepts, used to elucidate emotion concepts, as "evaluation", "activity" (Osgood, May & Miron, 1975; Mandler, 1980), "tendency", "satisfaction", "reward", "punishment" (Frijda, 1986), "unexpectness" (Ortony, Clore & Collins, 1988), "want", "good", "bad" (Wierzbicka, 1992) and "motivation", "desire", "goal" or "purpose", etc., have never been defined explicitly; from a scientific point of view what they are supposed to mean is not clear. We may still ask about the meanings of "desire", "want", etc. These so-called indefinables must be defined. This paper shows how this can be accomplished. The semantic primitives we take to serve as the building-blocks from which all others are constructed are "whenever", "situation" (the same notion as in situation calculus; see, McCarthy and Hayes, 1969) and "action". A definition expressed in terms of these semantic primitives may be reckoned as sufficiently formal. On this theory, problems such as "why do I want to eat when I am feeling hungry?" evaporate. The concept of intensity of emotion is elucidated with the theory, and questions about pain, pleasure, feeling of beauty, etc., can be answered.

2. What Emotions Are

On our account such things as feeling hungry, feeling thirst, feeling an itch, even feeling cold or hot, and so forth are also emotions. Whether or not we are right about ordinary usage is less important than whether we can give an account of emotion that shows such cases to be fundamentally different from those that are commonly counted as non-emotions such as thought, etc., as well as to be fundamentally identical with those that we count as emotions.

For the first step, we identify emotion with desire. The emotion is the desire; they are one and the same. To use functionalist terms to illustrate, emotions serve something—they do not represent mere contingencies, merely the way things happen to be. The point we emphasize is that they serve exactly what desires serve. "That I feel an emotion" is identified with "that I desire to do something". In conception, emotion must be replaced with desire. The difference between "emotion" and "desire" lies only in that they play different syntactic role in language. On the other hand, obviously, human beings and other beings must interact with their environment—get information from it and act upon it. Thus, functionally, these two kinds of "mental states" are necessary and sufficient: (a) cognitive mental states which play the role of receiving and processing information, and (b) desires which result in actions. It does seem to be utterly implausible to suppose that emotions serve no function in the interaction of human beings with their environment, but exist merely to provide pleasure or discomfort. Moreover, if emotion and desire are two separate things, we have no explanation of how they interact. Indeed, it really strikes us as odd that one desires to eat when one feels hunger, to drink when feels thirst, and that one does not enjoy agony.

Some might disagree the thesis, for example, that thirst is the desire to drink and assert that thirst causes the desire to drink. The author argues that it is because one desires to drink that one feels thirst, not vice versa. To say that thirst causes the desire to drink is putting the cart before the horse. Our general idea is best illustrated by pursuing an analogy. Consider the case of the notion of "there is a coin in the box". One might say "it is because there is a coin..."
that whenever I look at the place I see the coin". The author would ask, what is meant by "there is a coin"? The argues, in response to this question, that saying that there is a coin (in the box) is equivalent to saying that whenever I look at the place (inside the box) I see the coin. The notion expressed as "there is a coin" is constructed out of the regularity of observable bodily behavior, the regularity that whenever I look at the place I see the coin. It is this regularity which the expression "there is a coin" denotes. In other words, the meaning of the expression "there is a coin (inside the box)" is just: whenever I look at the place (inside the box) I see the coin. Likewise, there is no such a thing as thirst, but just "thirst" which is interpreted as desiring to drink, which in turn denotes a behavioral regularity. When we are talking of thirst we are, by that very token, talking of the desire to drink. Or, it can be said, the desire to drink causes thirst, in the sense in which we speak of one thing explaining, or being explaining of, another thing.

Here we attempt to define some emotions for illustration.

feeling hungry: desiring to eat;
feeling thirst: desiring to drink;
feeling an itch: desiring to scratch; for instance, feeling an itchy toe is the desire to scratch the toe;
feeling cold (hot): desiring to warm (cool) oneself;
feeling of fear: desiring to flee, or escape, etc.; that I feel fear of someone/something means I desire to go away from it.
feeling of love: desiring to be with;
regret: desiring to do what could have made oneself not do what was done by oneself;
hate: desiring to make someone feel pain, die, etc.; desiring to retaliate against someone;
pain (in the general sense of unpleasant feeling): desiring, when P (a fact or a state of affairs) is true, to do something as a result of which P will not hold;
pleasure: desiring to do what is being done.

For instance, desiring to warm myself when I am warming myself (making heat flow into the body, in terms of physics)—for when I am warming myself by, say, sitting near an oven, my body may physically still need heat flowing into it—is a kind of pleasure. Similarly, desiring to drink when one is drinking is a kind of pleasure. A common mistake is to suppose that pleasure results after one's thirst is quenched. Desiring to scratch oneself when one is doing so is another instances of pleasure. (As it is, when one is scratching an itch one is feeling a pleasure.) Where there is no desire there is no pleasure. For more details about pleasure, see section 5.

1 Adopting the same position, Ryle (1949) has accounted for "mind", a more general concept, and Strawson (1985) the concept of causation.

2 One, then, might say that even though one does not see it one can understand the notion of "there is a coin in the box". But how does one first construct the conception of there is something? That is the point.

3. The Concept of Desire

With regard to the idea that emotions are equivalent to desire, much previous work has provided insight on this. Hebb (1949, p. 190) defines hunger as the tendency to eat. Arnold (1960) says that emotion is felt action tendency. Searle (1984, p. 24) says that to be thirsty is to have, among other things, the desire to drink. Frijda (1986) holds that emotions, although not all, are action tendencies. Wierzbicka (1992) characterizes many emotions in terms of want. What impedes this idea's being widely accepted is that there is no further elucidation of what desire is, the problem of the same degree of difficulty, in their theories. Therefore the account of desire is central to the theory.

3.1 The Concept of Desire

Intuitively, it seems that desiring to do conveys the same meaning as tendency to do, or, action tendency (or, a state of readiness to execute a given kind of action). They are virtually synonymous. The goal here is to give a definition of desire, or, tendency.

We now begin our analysis of the concept of desire, and we proceed by cases.

Consider the following cases:
Case 1: a horse tied to a stake, making an effort to move forward, the reins being drawn tight; everything is still.
Case 2: a bow fully drawn with the arrow loaded; everything is still.
Case 3: a stone put on a bridge; everything is still.

Then, compare the following:
Case a: the horse tied to the stake, standing calmly; everything is still.
Case b: the bow not drawn, with the arrow loaded; everything is still.
Case c: the stone put on the bank, beside the bridge; everything is still.

It cannot be denied that there is a thing which case (1), case (2) and case (3) have in common, while case (a), case (b) and case (c) do not. What is it, or, how do we express it? Allow us to say, it is a kind of tendency. But what is a tendency? One might say that that thing is just a physical force. Admittedly, it is force. What then is force? One might say that "force" is a clearly explained physical no-
tion which stands for something which has both magnitude and direction, and if applied to a free body results chiefly in motion or a change in the direction or speed of motion of the body, and so on. But, actually, the concept of force has never been defined, just as time, although people have acquired intuitive concepts of them. Moreover, even if you have no knowledge of physics, couldn't you judge that there exists a similarity among case (1), case (2) and case (3), and a difference between any one of them and the corresponding case in the second group? Certainly, you can. It follows that people do possess the intuitive concept of tendency, or desire, or force, which is not in physics terms though, and which we will formalize in a moment.

Let us now attempt to generalize and abstract. Letting \( p \) be a circumstance or fact, which can be described by a propositional formula, and \( a \) be an event, we use \( p \rightarrow a \) to represent a tendency; that is to say, a tendency involves two parameters, namely circumstance (may be called "condition" in this case) and event; not event only. The operator " \( \rightarrow \) " is interpreted as this:

**Definition** \( (\rightarrow) \): \( p \rightarrow a \) iff whenever \( p \) holds true, \( a \) occurs.

Let || \( t \) || denote the meaning of "\( t \) is a tendency". Thus we have

**Definition** (tendency): || \( t \) || iff \( p \rightarrow a \).

It cannot be said that tendency causes \( p \rightarrow a \), but should be said that we mean \( p \rightarrow a \) by the word "tendency".

Both force and desire are tendencies.

To make this clearer, let us make an examination of how the concept of force is acquired by people. Suppose you had no concept of force; for example, imagine that you were a man before the time of Galileo. You would have observed that whenever an apple is cut off from the branch, a certain event occurs, that is, the apple moves in a certain direction (downward). In general, whenever \( p \) (an apple is cut off from branch, for instance) is true, \( a \) (the apple moves in a certain direction, for instance) occurs. Having encountered many times the facts of such sort, you might be aware of a regularity and indicate it by a word "force", and thus possess the concept— but everyday concept—of force. In physics, however, only properties of force such as magnitude and direction, rather than the concept itself, are dealt with. Neither has the semantic interpretation of force ever been defined in linguistics. The same can be said with respect to the case of desire; where event \( a \) is commonly called "action".

Returning to the cases described above, in case (1), case (2) and case (3), \( p \) is respectively: the reins are cut, the bow string is released, the bridge is broken. And \( a \) is respectively: the horse running forward, the arrow shooting, the stone dropping down. In case (1), for example, when the reins are cut, the horse moves, while in case (a), it won't.

In the case of feeling hungry, \( p \) may be this: food is at hand, or food is in my mouth, or food is at hand and at the same time I am not on a diet, and so on, and the \( a \) is eating. If the case of thirst, \( p \) may be this: water, or beer, and the like is at hand, etc. The action part \( a \) characterizes the desire's quality, which determines a desire's characteristic feel or quale. For example, if \( a \) is eating, the desire \( p \rightarrow a \) counts as feeling hungry, and if scratching, counts as feeling an itch, and so forth. The condition part \( p \) determines a desire's intensity, of which we will go into details in the next section.

### 3.2 Physical Realization

It might be questioned that how we could know that an agent—if it really has a desire but the condition \( p \) has never held, and even will in no way hold, so that no action is performed—has the desire; for example, if it is true that I am itching but in practice take no action to scratch, how the desire could be said to exist. It amounts to this how to realize of an agent an internal state which can be described as \( p \rightarrow a \).

Let, to reply to this question, \( P \) be a proposition represented as a propositional formula, which determines that \( p \) must be the case in order to make it true; and that means:

\[(\text{the proposition } P) = (\text{the proposition that fact } p \text{ makes true}).\]

We can construct whatever mechanism, an internal physical state, which realizes the function that whenever proposition \( P \) holds true a certain type of action, for which we use \( A \) instead of \( a \) in this case, is performed. Thus, it may be claimed that an agent in whose brain this mechanism is encoded has a desire. Consequently, we rewrite the preceding definitions as follows:

**Definition** \( (\rightarrow) \): \( P \rightarrow A \) iff whenever \( P \) holds true \( A \) is performed.

The capital letter \( A \) denotes the performances of actuators rather than actual events performed. Noticing that proposition \( P \) can be represented as an internal state of whatever

---

3. In practice, a proposition naturally determines that a type of circumstance, rather than an individual circumstance, must be the case in order to make it true; for example, if \( P = \text{a cup of liquid is at hand and it is drinkable} \) then all such circumstances as a cup of beer is at hand, a cup of tea is at hand, etc. would make \( P \) hold true. For the sake of simplicity, here we would not distinguish between the notions of "proposition" and "propositional formula".

4. Perhaps it is not necessary to examine the everyday concept of force for the purpose of physics. But it is not always unnecessary to examine the everyday concepts for the purpose; for example, such concepts as "simultaneity".

---

8. In practice, a proposition naturally determines that a type of circumstance, rather than an individual circumstance, must be the case in order to make it true; for example, if \( P = \text{a cup of liquid is at hand and it is drinkable} \) then all such circumstances as a cup of beer is at hand, a cup of tea is at hand, etc. would make \( P \) hold true. For the sake of simplicity, here we would not distinguish between the notions of "proposition" and "propositional formula".

6. The fact that \( P \) holds true means that the agent believes that \( p \). For reasons of space, we shall not address the issue of "belief" here. Instead, we just give a definition: The agent believes that \( p \) iff this very \( P \) holds true.
form, so $P \rightarrow A$ denotes an internal state. Letting $d$ be an internal state, we have

Definition (desire): $d$ is a desire if $d$ is represented as $P \rightarrow A$.

Here, "$\rightarrow$" is interpreted as a kind of real mechanism which we construe as "whenever". If one programmed a robot to have a mechanism of $P \rightarrow A$, one would be right to claim that one had produced an emotion, because one had produced what is fundamentally identical to what people designate as "emotion".

4. Intensity of Emotions

Our daily language and experience recognize that emotions may exist in varying degrees of intensity. For example, "I am feeling a little bit hungry", "I am feeling very hungry", etc., do reflect the conception of emotional intensity. We often distinguish between apprehension, anxiety, fear, and panic or between contentment, pleasure, happiness, and ecstasy. Furthermore, there seems to be a kind of order relationship among emotions. People are clearly aware that if emotion $e_1$ is "stronger" than $e_2$, and $e_2$ is "stronger" than $e_3$, then there is the same relationship between $e_1$ and $e_3$—transitivity. And people do not think that emotion $e_1$ is "stronger" than $e_2$ that, for example, a bit hungry is "stronger" than very hungry—antisymmetry. Also, people do not think that there is such a relationship between an emotion and itself as that between, say, very hungry and a bit hungry—antireflexivity. Yet strangely enough, psychologists and philosophers have rarely examined this dimension of emotion. The key to the problem of emotion intensity, again, is to elucidate the concept of emotional intensity, to determine what the meaning of "stronger" is.

To begin with, it is instructive to reflect on some important facts which might be easily overlooked:

- when extremely hungry, one would eat whatever edibles are available; while if only a bit hungry, one would not until some "delicious" food is available,
- when feeling a mild need to urinate, one might urinate only when a convenient opportunity presents itself; while if feeling an urgent need to urinate, one might urinate without so much (or any) attention to the place and its convenience,
- with feeling a little bit hungry and at the same time very thirsty, if both food and drink are available one would begin to drink rather than eat,
- et cetera.

Further consideration of what it means for an emotion to be stronger than another lead to the following definition: Suppose $e_1 = P_1 \rightarrow A_1$ and $e_2 = P_2 \rightarrow A_2$. Using $e_1 \geq_d e_2$ for "$e_1$ is stronger than $e_2$", we have

Definition (stronger): $e_1 \geq_d e_2$ if $e_2$ gives rise to action $A_2$ in some circumstances then $e_1$ will necessarily give rise to action $A_1$ under the same circumstances, and not vice versa; in other words, there cannot be such a circumstance that $A_2$ is triggered under it while $A_1$ is not, but there are some circumstances under which $A_1$ will be triggered whereas $A_2$ is not.

Again, it is, in the same sense as before, not because $e_1$ is stronger than $e_2$, that there is no such circumstances that $A_2$ is triggered under them while $A_1$ is not, and that under some circumstances $A_1$ will be performed whereas $A_2$ will not; but rather, the conception that $e_1$ is stronger than $e_2$ is constructed from the latter. It is a misconception that emotions have magnitudes in the sense in which physical quantities have. This misconception leads people to view some connected physical parameters as variables modulating or indicating, directly or indirectly, the intensity of emotion.

We now turn to the explanation of why the condition part of a desire determines the intensity of the desire. As it is mentioned earlier, the condition part $P$, a formula of proposition, refers to a circumstance type of which a certain circumstance will make it true. For example, given that $P=it$ rains, $P$ refers to all the circumstances where it rains; a morning when it is raining, an evening when it is raining, for example, both make $P$ true. We use $P(c)$ for "$P$ holds true in circumstance $c$"., then $P$ specifies a set of situations:

$$S_p = \{x \mid P(x) = \text{Truth}\}$$

Rewriting $P$ as $P(c)$, and thus $e= P \rightarrow A$ as $e=P(c) \rightarrow A$, we have

$$P(c) \text{ holds iff } x \in S_p$$

Then the preceding definition of desire is rewritten as

Definition: $e$ is a desire iff $\forall x(\text{whenever } P(x) = \text{Truth}, A \text{ is performed}).$

Let $e_1 = P_1(c) \rightarrow A_1$ and $e_2 = P_2(c) \rightarrow A_2$. Then, the definitions of desire and "stronger" immediately lead to the following theorem:

Theorem 1: $e_1 \geq_d e_2$ iff

$$\forall x(P_2(c) \rightarrow P_1(x)) \land \neg \forall x(P_1(x) \rightarrow P_2(x)) \text{ holds.}$$

Corollary: Given

$$e_1 = \bigwedge_{c \in C_1} e_1 \rightarrow A_1, e_2 = \bigwedge_{c \in C_2} e_2 \rightarrow A_2,$$

where $C_1$ and $C_2$ are two sets of propositions, then,

$$C_1 \subseteq C_2 \Rightarrow e_1 \geq_d e_2.$$

That is, the fewer the conditions appear in the condition part, the stronger the emotion is. Theorem 1 implies:

Theorem 2: $\geq_d$ is an order relation. Namely, it satisfies

(1) antireflexivity: for all $e_i, e \geq_d e$ does not hold,
(2) antisymmetry: if $e_1 \geq_d e_2$, then $e_2 \geq_d e_1$ does not hold,
(3) transitivity: if $e_1 \geq_d e_2$ and $e_2 \geq_d e_3$, then $e_1 \geq_d e_3$.

Theorem 2 accounts for the intuitions mentioned at the beginning of this section.

Since the intensity of emotion or desire only involves the condition parts of emotions or desires, two emotions of
different quality may be comparable with each other with respect to intensity. This is also reflected by our intuitions. For example, it seems meaningful to say that feeling a little bit hungry is weaker than feeling extremely thirsty. Nevertheless, not every pair of emotions are comparable with one another. For example, it seems that, intuitively, some feelings of fear and some of thirst are, to some extent, not comparable with one another with respect to intensity. This is because, suppose $e_1$ and $e_2$ are not comparable with one another, neither

$$\forall x(P_2(x) \rightarrow P_1(x)) \land \forall x(P_1(x) \rightarrow P_2(x))$$

nor

$$\forall x(P_1(x) \rightarrow P_2(x)) \land \forall x(P_2(x) \rightarrow P_1(x))$$

holds true and therefore neither $e_1 \succ d e_2$ nor $e_2 \succ d e_1$ holds, in this case.

The conditions specified by the condition parts not only refer to the external world states but also may include the internal states of an agent. For example, when one feels hungry, one might take no action to eat even if some food is at hand, if one is on a diet; with the hunger becoming strong enough, one would eat the food.

5. More about Pleasure and Pain

5.1 Pleasure

As has been seen, the concept of emotion is stood for by $p \rightarrow d a$, and thereby we mean that an emotion, or desire, is uniquely determined by a proposition and an action. But pleasure has not this characteristic form of expression. In fact, if we call those that are characterized as the form $p \rightarrow d a$, such as fear, itch, etc., "basic" or "fundamental" emotions, pleasures are not basic emotions; they are "forms" of desires, rather than desires themselves. As mentioned earlier, the concept of pleasure is defined as (5.1) desire to do $a$ when $a$ is being performed; for instance, one desires to scratch (that is, one is itchy) when one is scratching.

This is best illustrated by examples. Suppose the desire now is thirst, namely desire to drink water. Let us write $P(\cdot)(w)$ to represent the condition part of the thirst, and $A$ to represent the action—drinking. The notation of $P(\cdot)(w)$ is used to indicate that the condition under which $A$ occurs depends on the level of fluid content of one's body or in one's stomach, or throat. The lower the level is, the weaker the condition is and therefore the stronger the desire to drink is, and vice versa. In short, the condition part $P(\cdot)$ is a function of the level of fluid content. Let the desire be written

$$P(\cdot)(w_0) \rightarrow d A.$$

The occurrence of $A$ can have affects on $P(\cdot)(w)$, leading to the increase of the level of the fluid content and this in turn causing $P(\cdot)(w)$ to vary. As it is, after one's having taken a swallow of water, the thirst would be weakened. The goal of $A$, and also of the desire, is, say, "water is flowing through the throat". At the starting point, the desire is

$$P(\cdot)(w_0) \rightarrow d A.$$
where something holds repeatedly as opposed to "continuously" during an interval.

5.2 Pain

Let \(-g\), the negative of \(g\), be the goal of desire \(P(\cdot) \rightarrow A\). The concept of pain is defined as the conjunction of

1. \(g\) holds,
2. \(P(\cdot) \rightarrow A\) holds,
3. \(P()\) does not hold.

That is to say, when one is having a pain, then one has a desire to do an action which will cause a certain fact not to hold; without taking into account the difference between the two forms of goals—one is a proposition, another is in the form of the negative of a proposition—we can simply say that pain is a desire to do an action which has not taken place. This explains why feeling thirsty is a pain, love is a sort of painful, and so on.

6. Conclusion

To determine what an emotion is, is to find a scientifically defined concept of emotion, which accords with our everyday concept of what an emotion is. Hence what we should do is to make clear the everyday concept -- to express it explicitly. The existing literature, however, has rarely seriously done this, since conceptualization is difficult to make when attempting to express knowledge about the everyday, "commonsense" world. We advocate that concepts are constructed out of regularities of observable behavior. This view underlies our approach. We have argued in this paper that the concept of emotion is identified with that of desire. While several researchers have made the similar claim that emotion is desire (or action tendency), they have never formulated a systematic theory based on this view, and none of them holds that every emotion is a desire; in particular, they have never developed accounts of some crucial instances of emotion, such as pleasure, pain, aesthetic feeling, etc. We have presented definitions for some emotions for illustration; among them are pleasure and pain. On this view, we can define every emotion by showing what one desires to do when having an emotion. We have revealed the behavioral regularities required for characterizing the concept of desire and presented a formal definition of desire; and in effect have formally defined the synonyms thereof such as "tendency", "want", and so on. This is central to the theory. In so far as we know, no one has attempted to define the concept of "desire". Within our theory of emotion, emotion intensity is naturally elucidated. This problem has also rarely addressed in previous literature.

As has been widely viewed, there are two main questions concerning emotions. First: What, in general, distinguishes emotional experience from other modes of experience and enables the subject to identify his or her experience as "emotional"? Second: What, specifically, distinguishes the experience of one emotion from that of another and enables the subject to identify his or her experience as one of fear, love, hunger, thirst, and so forth? We add the third: What, more specifically, distinguishes the experience of one (type of) emotion of some intensity from that of another intensity and enables the subject to identify his or her experience as one of weak itch, strong itch, unbearable itch, and so forth? In brief, writing \(p \rightarrow A\) for the mechanism of emotion, we propose the answer to the three questions as follows: \(\sim p\) distinguishes emotional experience from non-emotional experience, \(a\) distinguishes the experience of one emotion from that of another, and \(p\) accounts for differentiation between the experiences of emotional intensity.

References


