

Are Emotion Metaphors Conceptual or Lexical?

A review of *Metaphors of Anger, Pride and Love: A Lexical Approach to the Structure of Concepts* by Z. Kövecses (1986). Philadelphia/Amsterdam: John Benjamins, pp. vii, 147. ISBN 1-55619-009-3 (US); ISBN 90-272-2558-3 (Europe) \$34

Andrew Ortony

University of Illinois at Urbana-Champaign, U.S.A.

When people *read on your toes*, you sometimes *lose your cool*. Depending on exactly what it is they do, you may *see red*, *be driven wild*, and perhaps, in the *heat of the moment*, you may *blow your top*. Whereas you may think it will help to *get it off your chest*, it can *hurt your pride* when you realise that others see that you cannot *control your anger*...

All this sounds like a parlour game: how many conventionalised expressions about emotions can one string together while retaining a coherent message on the same theme? The answer is a lot, because the language that we use to talk about emotions is replete with conventionalised metaphors and other figurative devices, many of which are systematically related to one another. The fact constitutes the starting point for Zoltan Kövecses' *Metaphors of anger, pride and love*. Using the analytical technique of Lakoff and Johnson (1980), Kövecses presents case studies of the concepts of anger, pride, and love with a view to uncovering the structure and content of the folk models underlying these emotion concepts through a detailed examination of the conventionalised expressions we (naïve speakers of English) use to talk about them. However, there is more on the menu than this. For an appetiser we are presented with the idea that the approach to be adopted has the virtue of focusing on *reference* rather than periphrasing the allegedly methodologically impoverished obsession with the analysis of *sense* (meaning). And for dessert, we are presented with one

Request for reprints should be sent to Professor Andrew Ortony, Center for the Study of Reading, University of Illinois, 51 Gerry Drive, Champaign, Illinois, 61820, U.S.A. This work was supported in part by a grant from the National Science Foundation, BNS83-18077, and by the U.S. Office of Educational Research and Improvement under Cooperative Agreement No. OEG 0087-C1001

© 1988 Lawrence Erlbaum Associates Limited

of cognitive science's most fashionable themes, namely that concepts are organised around prototypes and cannot be analysed in terms of necessary and sufficient conditions. Finally, to ensure that we do not forget how tasty the appetiser was, Kövecses elaborates his objections to traditional componential analysis in linguistics.

Let me first try to lay out the general structure of Kövecses' argument. His central claim is that if we examine the conventionalised expressions—the metaphors and metonymies—speakers use to talk about emotions, we discover a set of underlying metaphors such as CAUSING ANGER IS TRESPASSING (when people *read on our toes*), ANGER IS INSANITY (we may be *driven wild*), ANGER IS THE HEAT OF A FLUID IN A CONTAINER (we may *lose our cool* and *blow our top* in the *heat of the moment*), PRIDE IS A PERSON (what happens may *hurt our pride*), and so on. What these expressions reveal, the argument goes, is that we *understand* concepts such as anger, pride, and love in terms of such metaphors.

This general idea is by no means a new one. Nearly 50 years ago, Stephen Pepper (although in a somewhat different context) wrote: "A man desiring to understand the world looks about for a clue to its comprehension. He pitches upon some area of commonsense fact and tries if he cannot understand other areas in terms of this one. The original area then becomes his *basic analogy* or *root metaphor*. He describes as best he can the characteristics of this area, or if you will, 'discriminates its structure'. A list of its structural characteristics becomes his basic concepts of explanation and description. We call them a set of categories... He undertakes to interpret all facts in terms of these categories" (Pepper, 1942, p. 91).

What Kövecses (and indeed Lakoff & Johnson, 1980) add to this notion is the idea that people's metaphorically based understanding of emotion (and other) concepts is in terms of one or more prototypical "Folk" models—de Sousa (1980, 1987) calls them "paradigm scenarios"—i.e. idealised or typical situations in which emotions are aroused and expressed. Particular instances of one of these emotions may conform more or less closely to the prototypical presentation. For example, Kövecses proposes that the prototype for anger that constitutes our folk model has five stages: (1) an offending event in which a wrongdoer intentionally does something to the experiencer and is at fault for it; (2) anger results with some degree of intensity; (3) the experiencer attempts to control the anger; (4) the experiencer loses control and exhibits angry behaviour; and (5) the experiencer performs an act of retribution comparable in intensity (*sic*) to the experienced anger. The claim is that the metaphors converge on this typical scenario, even though we have other models which are "minimal variants" of it. The first observation one might make about this proposed folk model of anger is that there is nothing very metaphorical about it. This may be because Kövecses' does not present a very sophisticated account of

what metaphors are or how they work, and he certainly does not relate it to current work in the field or in the crucially important related field of analogy. Furthermore, because most of the metaphorical expressions he does uncover pertain to the *expression* of emotions, the metaphors may simply reflect the fact that when we reason about emotions we often do so analogically on the basis of beliefs about our own angry behaviour and that of animals. In other words, we use analogies as a practical solution to the "other minds" problem.

Kövecses believes that because our concepts of the emotions are represented in terms of prototypical scenarios of this kind, and because the specification of such scenarios is not in terms of a small set of binary necessary and sufficient features, we get a much richer account of conceptual structure than can be obtained by trying to characterise the sense or meaning of the corresponding words with feature descriptions in the tradition of componential analysis. In fact, Kövecses argues, componential analysis provides very impoverished accounts indeed, such as that anger is "a feeling of intense displeasure", which, he rightly observes, tells us almost nothing (although one might note that Kövecses does not specify who it is that characterises anger in this way).

A major theme of the book is that the kind of approach that tries to characterise emotion concepts in terms of idealised cognitive models (prototypical scenarios), as opposed to feature descriptions, enables us to solve all manner of traditional linguistic and philosophical questions (as well as psychological ones) that traditional approaches to lexical semantics cannot handle satisfactorily. For example, Kövecses thinks that it can explain the variety of meanings associated with polysemy without resorting to some unexplained notion of similarity—the different meanings simply draw on different underlying metaphors. He thinks it can explain collocational restrictions (i.e. restrictions governing the co-occurrence of lexical items) in a more satisfactory way than the theoretical machinery of selection restrictions and contextual redundancy rules (Leech, 1974), and he thinks it gives a better account of "semantic fields" (Lehrer, 1974). In this review, I shall not discuss these technical issues further except to say that whereas I share some of Kövecses' concerns about more traditional approaches to these problems, I think he may be over-optimistic in thinking that he has an improved solution. My reasons for this belief flow from my discomfort both with the goals and the methodology of his approach, and it is on these that I shall concentrate.

In introducing the chapter on anger, Kövecses asks: "Are emotions just amorphous 'feelings' or do they have cognitive content?" He then goes on to ask: "If they have cognitive content, how can we find out what it is?" In these two questions, and Kövecses' answers to them, lurk some troublesome misunderstandings and confusions.

Consider the first question: Do emotions have cognitive content? Immediately an obvious issue arises. Is Kövecses talking about emotions *per se*, that is, about emotions as psychological states, or is he talking about people's *concepts* of emotions? If he is talking about emotions *per se*, as the question implies, the answer is clearly "yes" if we take an emotion to include its causal origins, its phenomenal "feel", and its physiological, behavioural, and expressive concomitants. Everybody knows that certain kinds of things *cause* us to become angry, and that when we are angry we have a certain sort of feeling, and that we sometimes express our anger in words and deeds. In this sense, then, at least our beliefs about what has happened and our decisions, if we make them, about how to react, constitute cognitive content. So what is the point of the question?

Presumably, Kövecses would reply that he is not talking about emotions *per se*, but rather that he is talking about our concepts of emotions—our "folk models". And, indeed, he is. But of course, these folk models, because they are mental representations, must have cognitive content by *definition*. However, given Kövecses' insistence that he is giving a *referential* rather than a "sense" account, this raises an immediate problem for him. It surely is the case that the word *anger*, for example, refers to a certain class of psychological states, *not* to a concept. When we attribute anger to someone, we are not attributing a concept to him, just as when I say that Jones has a dog, I am not attributing a concept to him—I am saying he has a dog! Concepts have nothing to do with it. Words do not refer to concepts even though concepts intervene between words and their referents.

It seems to me that the only course open to Kövecses at this point is to abandon his claim that he is offering a referential account. He is not. Although, inexplicably, he never uses the term himself, he is offering an account of what in cognitive science is usually referred to as *mental models* (e.g. Gentner & Stevens, 1983; Johnson-Laird, 1983). There is nothing wrong with that, except that a mental model is not the same thing as that of which it is a model. I am sympathetic to the view that we might be able to learn something about emotions *per se* by understanding how people conceptualise them, perhaps because the status of emotions as experiences endows our folk theories of them with more built-in validity than, say, our folk models of meteorology. On the other hand, Kövecses offers no discussion of the relation between our folk models of emotions and the emotions themselves. Interestingly, Kövecses' neglect of the difference between emotions and emotion concepts is mirrored by another conflation. He treats lexical items and their corresponding concepts as the same thing [cf. for example, "In other words, *lexical items* (*concepts*) may differ in the number and kinds of secondary metonymies they allow" p. 135, italics added]. One might view this as being particularly surprising coming from a linguist.

It is Kövescses' failure to distinguish clearly between words, their referents, and the concepts that mediate the two, what best explains what I take to be his misplaced attack on componential analysis. Because componential analysis is concerned with words and their meaning constituents, and because its goal is to specify meaning constituents (features) that can explain how the meaning of higher-order structural constituents can be derived from lower order ones, it cannot be faulted for ignoring the question of the structure of people's concepts. That is not the name of the game. Thus, linguists who engage in componential analysis, whatever its weaknesses, cannot be blamed for not analysing the referents of the lexical items they study. A linguist studying a lexical field like cooking need not be a chef or a chemist or an expert in convection and radiation. Similarly, a linguist studying the emotion lexicon does not have to be a psychologist, which perhaps is just as well.

We come now to Kövescses' second question. If emotions have cognitive content, how can we find out what it is? The answer here is that if we are talking about emotions *per se*, we have to do empirical psychology, which is certainly not what Kövescses is doing. If the question really means "If people's *concepts* of emotions have cognitive content, how can we find out what it is?", the answer is surely, "Ask them!". But this is not what Kövescses does. Rather, what he does is to use an indirect and, in my opinion, rather unreliable method. He consults his intuitions about what people do or could say about different emotions, and shows that these intuitions are consistent with some fairly homely models of what those emotions typically are. I certainly do not to deny that we can discover something interesting by looking at the figurative language that is (sometimes) used to talk about emotions. However, as I shall argue shortly, I do not think we discover what Kövescses thinks we discover. But first, I need to discuss some methodological points.

I started this review with some fairly obvious cases of conventionalised metaphorical expressions used to talk about emotions. Certainly, the list was not exhaustive. If one were to attempt to draw any serious generalisations from such a list, a reasonable first challenge would be "How complete is the list?". One of the most troubling problems about the methodology employed by Kövescses (and indeed, by Lakoff & Johnson, 1980) is that we have no idea what the answer to this question is. Perhaps the failure to address this issue (except to specify the actual number of examples discussed) has to do with the fact that the method puts the cart before the horse (which phrase, incidentally, in no way establishes that I have conceptualised my notion of a theoretical argument in terms of horse-drawn carriages). The method is advertised as a discovery procedure, but in reality a hypothesis confirmation procedure. Several observations invite this conclusion. First, the chapters on anger and pride *start* by presenting (parts of) the proposed folk model. One would expect the method to start

with a corpus of conventionalised expressions and to move from there to an organisation of the corpus into the hypothesised underlying metaphorical concepts. In Kövescses' defence, here, it might be argued that the presentation of the model before the data could just be an exegetical move, so that my objection, while perhaps suggestive, is of limited force. Secondly, and more troublesome, is the problem of the selective use of data. Many of the examples are forced, artificial, and (conveniently) limited to cases that are consistent with the proposed hypothesis. Thus, one is left with the impression that their purpose is to substantiate a preconceived idea rather than that their existence inexorably leads to the proposed conclusion.

A third, related, problem is actually the converse of the one I have just raised. Some of the examples are fabricated cases that are not in fact conventionalised expressions at all. In other words, they are cases of creating data to fit a hypothesis, rather than explaining data with a generalisation. For example, people do not really speak of anger as being "insatiable" (p. 23), or of people "overflowing with love" (p. 82) (at least, not in the sense of experiencing, as opposed to giving it). Nor do people speak of being "chesty" to indicate pride (p. 42). The inclusion of these and many other strange or marginal examples ("to cast off/discard one's dignity", "to be overcome by concert", and so on) appears to increase the pervasiveness of the underlying metaphors, but in fact it only serves to increase one's scepticism. There are other suspect inferences too: for example, the claim (p. 26) that expressions like "get out of my sight", "leave me alone", "drawing the line", and "stepping on someone's toes" are evidence for an underlying THE CAUSE OF ANGER IS TRESPASSING metaphor is quite gratuitous.

In yet other cases, we get implausible interpretations of examples. For instance, we are told (p. 13) that the expression "to see red" is an example of the fact that redness metonymically indicates anger through its (anger's) interference with accurate perception. But it seems just as likely that the expression gains its currency through its allusion to cultural beliefs about the behaviour of bulls in response to the colour red. Similarly, the ANGER IS INSANITY metaphor is supposed to be supported by the expression "to go bananas" (p. 20), but no explanation is offered as to how or why this should be so.

What is lacking are constraints on examples and constraints on the metaphors from which they allegedly derive. The enterprise would have been more interesting and more convincing had it been based on a real corpus derived from a representative sample of English speakers. As it is, we simply do not know how many missing cases there might be, and whether missing cases would merely reflect incompleteness, or whether they would actually constitute counter-evidence. Nor do we know whether the suspicious examples really are spurious, or whether they are manifesta-

tions of regional and dialect differences (although the existence of such differences might prove to be an embarrassment to Köveses' position).

So what should we make of all this? Certainly there do exist some systematic metaphors underlying the way we talk about emotions, and notwithstanding the reservations I have expressed, Köveses had done a good job in identifying a number of them. But what conclusions can be drawn from the existence of such metaphors? The conclusion that Köveses would have us draw is that these metaphors "structure" the emotion concepts. This is a strong interpretation of the data in that it maintains that people actually represent their emotion concepts *in terms of* the various metaphorical ones. On this interpretation, people represent their concept of, say, anger, at least partly by using their already existing representations of, say, hot liquids. This I shall refer to as the *conceptualisation* view. An alternative, weaker, interpretation of the data is that we simply borrow the language from one lexically rich domain to talk about a lexically less rich domain. This view I shall call the *lexicalisation* view.

Köveses writes: "[the concept of] ANGER is understood as being, or is created to be, an ENTITY. ANGER and LOVE are understood as having, or are created as having, a CONTROL aspect: the PROUD PERSON is understood as having, or is created to have a VALUE, etc. Let us now examine these *constitutive metaphors*. Their source domains—ENTITY, INTENSITY, LIMIT, FORCE, CONTROL, VALUE, UNITY, etc.—all seem to be superordinate concepts, that is, concepts that are fairly abstract. By contrast, the principal metaphors that map onto the anger, pride, and love ontologies—HOT FLUID, FIRE, INSANITY, BURDEN, STRUGGLE, ECONOMIC VALUE, PHYSICAL UNITY (like a BOND), etc.—appear to be basic-level concepts that are linked more directly to experience" (p. 116, italics added). The implication of the underlying metaphors being constitutive and experientially based is that we would not (could not?) have the corresponding concepts without them. However, this seems an improbable conclusion. Let us consider what ought to be true if this strong, conceptualisation, view were to be correct.

Taking, first, a psychological developmental perspective, if concepts such as anger and love only exist because they are represented in terms of other concepts such as HOT FLUID, INSANITY, STRUGGLE, etc., one would expect that it would not be possible for a child to possess the emotion concepts until and unless he or she already possessed the concepts that constitute their metaphoric, of better, analogical, "gist". As far as I know, there is absolutely no evidence to this effect, bearing in mind that the re-namings and overgeneralisations that are so ubiquitous in language acquisition are not generally regarded as metaphors. If anything, it seems more likely that the emotion concepts precede these more sophisticated

concepts upon which they are alleged to depend. For example, it seems more reasonable to suppose that children come to acquire concepts of anger and love *before* acquiring an understanding of the behaviour of hot fluids, or of the concepts of insanity or a unity. Indeed, this seems to be true on the basis of the very criterion that Köveses himself claims to be crucial, namely, that there be an experiential basis for the concepts. It is much easier to imagine a child growing up with an experientially based concept of anger without an experientially based concept of insanity than it is to imagine a child growing up with an experientially based concept of insanity without one for anger.

If one takes a diachronic linguistic perspective, one comes to similar conclusions about the implications of the conceptualisation view. If the view was right, one would expect that the emotion words would have to appear later in the language than words referring to those aspects of the world (HOT FLUID, INSANITY, etc.) that are allegedly required for the emotion concepts to exist. Again, I am not aware of any evidence to this effect.

Another, and perhaps more telling consequence of the conceptualisation view, is that if emotion concepts were actually represented in terms of these other concepts, one would expect much greater freedom in the allowable metaphoric expressions. If emotion concepts were really metaphorical mental models one would expect that the metaphor that permits (or "creates") expressions such as being "filled with" love, love "welling up inside", and love "pouring out" (p. 32), would also permit expressions such as being "immersed" in it, which it does not (despite claims to contrary on pp. 82 and 88). Nor can we be "submerged" in love "drown" in it, or "swim" in it. Similarly, there is nothing in Köveses' account that can explain why we speak of people being "blind with rage", but not "blind with fear". For Köveses, the folk model of anger includes the idea that anger interferes with accurate perception. But this is equally true of our folk model of fear and panic—if we are very frightened we often behave irrationally because, to modify Köveses' claim about anger (p. 12): THERE IS A LIMIT BEYOND WHICH THE PHYSIOLOGICAL EFFECTS OF FEAR IMPAIR NORMAL FUNCTIONING. In general, the problem is that Köveses pays too little attention to what one *cannot* say, focusing instead only on what one can say.

Although the improbable consequences of the strong conceptualisation view pose a serious threat to it, they are not problems at all for the less radical lexicalisation explanation. The lexicalisation view can account for the linguistic data without claiming that the representational meat of emotion concepts (or any concepts for that matter) is comprised of the underlying metaphorical constructs. In its starkest form, the lexicalisation view would treat conventionalised figurative expressions as sort of intra-

language loan words. Such expressions simply involve borrowing words from one (source) domain to express relations between conceptual elements which are not lexicalised in the other (target) domain—the more, and the more systematic, the better. In other words, the lexicalisation view explains the data by proposing that there exists (or from a historical perspective, existed) source domains in which one or more already lexicalised relations could be mapped onto similar relations in the target (emotion) domain and that in this latter domain those relations were not already independently lexicalised.

No doubt one could massage the conceptualisation view and the lexicalisation view in such a way as to make them meet in the middle—for example, one might modify the lexicalisation view to allow for some “conceptual leakage”. Nevertheless, on balance, the lexicalisation view, which Kovecses does not even consider, strikes me as a more parsimonious and a more credible explanation not only of the facts that Kovecses concentrates on, but also of some that he does not. One of these is that there can be inter- and intra-language differences in the conventionalised metaphors used to talk about different emotions while one might want to deny that these differences reflect any radical differences in the corresponding concepts. Furthermore, our concepts change over time even when the conventionalised language we use may be frozen. Thus, although we speak of the sun “going behind the clouds”, “setting”, “rising”, “going down”, and “coming up”, it would be foolish to argue that the systematicity of these expressions indicates that people *currently* hold a heliocentric view. Today’s conventionalised expressions do not necessarily reflect today’s folk models. They might reflect yesterday’s, or they might just represent convenient ways of talking about things, and not reflect anything at all about conceptual structure.

I think Kovecses is wrong to suppose that the existence conventionalised analogies for emotion concepts entails that these concepts depend on the analogies for their existence. I think he is right to identify the linguistic evidence for such analogies as an interesting and possibly fruitful area for systematic study, albeit an area ultimately more likely to be of interest to linguists than to psychologists.

Manuscript received 19 January, 1988

REFERENCES

- De Sousa, R. (1987). *The rationality of emotions*. Cambridge, Mass: MIT Press.
 De Sousa, R. (1980). The rationality of emotions. In A. O. Rorty (Ed.), *Explaining emotions*. Berkeley: University of California Press.

104 ORTONY

- Gentner, D. & A. Stevens (Eds) (1983). *Mental models*. Hillsdale, N.J.: Lawrence Erlbaum Associates Inc.
 Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge University Press.
 Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
 Leech, G. (1974). *Semantics*. Harmondsworth: Penguin.
 Lehrer, A. (1970). *Semantic fields and lexical structure*. Amsterdam: North Holland.
 Pepperc, S. (1942). *World hypotheses*. Berkeley: University of California Press.