

effort required to access it, thereby affecting the relevance of that context. So an explanation of why a particular context was chosen must ultimately refer to a particular organizational structure of EM. In short, to explain context selection, S&W must have a theory about the organization of EM. They do not, however, have such a theory.

S&W suggest that EM is organized into "chunks," which consist of a set of assumptions that enable us to formulate expectations in stereotypical situations. However, they fail to specify the nature of these chunks. This omission enables them to account for the selection of a context through the ad hoc postulation of a particular chunk. Since there are no prior constraints on the individuation of chunks, *any* two pieces of information can belong to the same chunk at their discretion. Thus, through the vacuousness of their claims about EM organization, they can postulate any accessibility relation between contexts, accessibility being a central component in the search for relevance.

S&W's inability to explain context selection is evidenced by their discussion of the course of action an individual should undertake to realize relevance. For example, they consider the case where an assumption has no contextual effects in any of the accessible contexts. In this case, they claim that there is no point in extending the context in order to search for relevance. But how does a person know that an assumption has no contextual effects in any of the accessible contexts without going through each of those contexts? No explanation is offered. Likewise where an assumption has some relevance in the initial context: S&W claim that "an extension of the context will be justified as long as it yields greater contextual effects, and the increase in contextual effects is not outweighed by the increase in processing effort required" (1986, p. 143). But how do people ever know that there is not some other context that will be more relevant than the one they possess at the moment? What prevents people from always searching for a better context? Again, S&W do not tell us. Their injunctions work only by presupposing a solution to the frame problem, only by already assuming that we can realize relevance.

In short, S&W have failed to explain the process of context selection, and therefore beg the question with respect to the frame problem. This is not merely a side issue, it is the *central* issue, and what is at stake is nothing less than an explanation of how our inferences during verbal comprehension manage to be rational. By failing to elucidate the holistic nature of central processes, S&W have not addressed Fodor's central concerns and hence have not given us grounds for casting off the pallor of Fodorian gloom.

Authors' Response

Fodor's frame problem and relevance theory

Dan Sperber^a and Deirdre Wilson^b

^aCentre de Recherche en Épistémologie Appliquée, Ecole Polytechnique, Paris, France

^bDepartment of Linguistics, University College, London, United Kingdom.
sperber@poly.polytechnique.fr

Abstract: Chiappe & Kukla argue that relevance theory fails to solve the frame problem as defined by Fodor. They are right. They are wrong, however, to take Fodor's frame problem too seriously. Fodor's concerns, on the other hand, even though they are wrongly framed, are worth addressing. We argue that relevance theory helps address them.

The AI (artificial intelligence) "frame problem" (McCarthy & Hayes 1969) has been reinterpreted in a variety of ways

(Pylyshyn 1987). Fodor's is the loosest and grandest reinterpretation of all (Fodor 1987). The "frame problem," he writes, is: "Hamlet's problem: when to stop thinking" (p. 140); the problem of formalizing the distinction between "kooky facts" and "computationally relevant ones" (p. 145); "just the problem of nondemonstrative inference" (p. 146); "the problem of formalizing our intuitions about inductive relevance" (p. 148). Fodor concludes that "the frame problem is too important to leave it to the hackers" (p. 148), and Hayes retorts that "Fodor doesn't know the frame problem from a bunch of bananas" (Hayes 1987, p. 132).

Undeterred, Chiappe & Kukla adopt Fodor's interpretation of the problem, and argue that we have not solved it. They are right, of course. Did we ever claim to have solved Fodor's frame problem? Should we have done so? Is it so clear that Fodor's problem is well posed in the first place? What we did claim is that Relevance Theory (1986; 1987), and the study of verbal comprehension in particular, could help us better understand central thought processes. A mere ray of sunshine, obviously not enough to do away with "the pallor of Fodorian gloom."

Fodor seems to know exactly what it takes to be rational, and Chiappe & Kukla seem to have understood him. We don't and we haven't. Fodor argues that "modular cognitive processing is *ipso facto* irrational" in that it arrives at conclusions "by attending to less than all the evidence that is relevant and available" (Fodor 1987; pp. 139–140). By contrast, unencapsulated, central processes of belief fixation are rational, he argues, in that they make use of all the relevant and available evidence. The question then seems to be: How do central processes avoid getting bogged down with all the irrelevant available evidence?

This is all very puzzling. Is it really the case that central processes of belief fixation use *all* the relevant and available evidence? Wouldn't *that* be enough to bog them down? You have invited Granny for dinner and you wonder what main course would most please her. Osso bucco, you decide, remembering that she likes Italian food, raves about the Capri restaurant whose specialty is osso bucco, and has complained that you always serve her kedgerree. Reasonable enough, but don't you have, after all these years, much more evidence of Granny's likes and dislikes? Didn't she, for instance, once say that you couldn't find good veal any more? And yet here you are, processing the veal shanks, and not all these further bits of relevant information. The truth of the matter is that central processes consider *some* of the available relevant evidence, never all of it.

If it were crucial to rational belief fixation to consider all the available relevant evidence, why shouldn't evidence available in the environment (in libraries, for instance, or in other people's memories) be exploited too? Given Fodor's criterion of rational belief fixation, why should the way in which you access the relevant evidence – by remembering or by consulting – matter to whether rationality demands its use? Actually, we often do consider some of the environmentally relevant information (you did check with Grandpa that Granny had not eaten veal this week, didn't you?), but never all of it.

By Fodor's criterion of rationality, since we fail to consider all the relevant evidence, we are, in any case, irrational. Come to think of it, would you *want* to be rational in his sense? Do you want to consider all the (internally and externally) available evidence every time you fix a belief – which still would not guarantee that all your beliefs were

true, but it would guarantee that you would fix far fewer of them? Fodor's rationality is a purely epistemic matter: the only utility is truth, and no price is too high to pay to increase the chances that your beliefs are true. Fodor's frame problem is: How do we manage to pay the exorbitant price of such rationality? The short answer is that we don't.

A kind of rationality worth having is one based on sound accounting principles, where not only benefits, but also costs are weighed. This, incidentally, is also the only kind of rationality that is at all likely to be found in evolved wetware like ourselves. To be rational in this sense is to maximize the expected cognitive utility of the information one attends to, be it information picked from the environment or information retrieved from memory.

We use "relevance" as a theoretical term to refer to the cognitive utility of a piece of information in a context, or for an individual at a given time (Sperber & Wilson 1986; 1987). Relevance so understood has two aspects, cognitive effect (the benefit) and processing effort (the cost). The cognitive effect, if any, of processing a piece of information is to allow fixation or revision of beliefs. Effort is a matter of greater or lesser mobilization of brain resources in order to achieve this effect. *Ceteris paribus*, the greater the effect of processing a given piece of information, the greater its relevance. *Ceteris paribus*, the greater the effort involved in processing a given piece of information, the lower its relevance.

Here is a toy illustration. You have bought a ticket for a lottery and you know the prizes are \$10, \$500, and \$1000. Suppose you are informed of one of three things:

- (a) You have won \$500.
- (b) You have won \$10, \$500, or \$1000.
- (c) Either you have not won \$500, or the square root of 2207 is not 49, but not both.

Information (a) is more relevant than information (b) because (a) implies (b) and therefore has all the effects of (b) plus some of its own, without greater cost in effort. Information (a) is also more relevant than information (c), although (a) and (c) are logically equivalent and therefore carry exactly the same effects. However, in the case of (c), achieving these effects involves greater effort. This may not correspond to your favorite meaning for the vague English word "relevance." If so, we would want to argue that either relevance in your sense plays no distinct role in cognitive processes, or else relevance in your sense is a special case of our more general theoretical notion.

At any given moment in one's cognitive life, there is a wide range of new information being monitored in the environment, and there is an even wider range of information in memory, bits of which might be activated and would provide a context in which to process the information from the environment (or other pieces of information from memory). Only some of the possible combinations of new and contextual information would yield relevance, and this to a greater or lesser degree. There is no way for the mind to review all possible combinations of new and contextual information in order to find out which would maximize relevance. Even if there were a way, the effort involved in such a review would so lower the overall cognitive utility of the process as to defeat the whole enterprise. So how should the mind proceed? Since it cannot have foreknowledge of relevance, how can the mind have, at least, nonarbitrary *expectations* of relevance?

To begin with, when expectations of effect are wholly indeterminate, the mind should base itself on considerations of effort – picking up from the environment the most easily attended stimulus and processing it in the context of what comes most readily to mind. *Ceteris paribus*, what is easier is more relevant, if it is relevant at all. But what are the chances that what comes more easily to mind is, in fact, relevant? They would be close to nil, if salience in the environment and accessibility in memory were both random, and, moreover, uncorrelated. But humans are evolved organisms with learning capacities of sorts, so it is not too surprising to find that they spontaneously pay more attention to moving objects than to still objects, to looming objects than to receding objects, to sudden noises than to constant noises, to other people's faces than to other people's feet, to their own children than to others', and so on, that is, to objects and events that, on average, are more likely to be relevant to them.

For the same reason, it is not surprising that the perceptual categorization of a distal stimulus should tend to activate related information in memory. Thus, having your attention attracted by a snake tends to make your beliefs about snakes, at that moment, more accessible than your beliefs about the frame problem. Nor is it surprising that memory is so organized that pieces of information that are likely to be simultaneously relevant tend to be coaccessed or coactivated in chunks variously described in the literature as "concepts," "schemas," "scripts," "dossiers," and so forth.

Chiappe & Kukla might want to follow Fodor and argue that such suggestions are a way to beg, rather than to begin answering, the question. Consider the concept of a fridgeon: "x is a fridgeon at t iff x is a particle at t and [Fodor's] fridge is on at t" (Fodor 1987; p. 144). Were you to learn that Fodor's fridge has just been turned on, you could infer of every particle in the Universe that it is now a fridgeon. How is that for cognitive effect! Why, then, don't we have such kooky concepts, and why don't we keep inferring such kooky facts? Because, contrary to appearances, such cognitive effect is of the weakest kind. Once you have inferred that a given particle is a fridgeon, or that all particles are fridgeons, nothing further follows. Such deadend inferences are not worth the effort. Compare inferring that some food is refrigerated: from this you can infer that it will keep longer, that it will taste different, and these facts in turn have further consequences. Why the difference? Because we have a "theory" of refrigeration, not one of fridgeonization. Relevance considerations will favor concepts with rich inferential potential, typically concepts embodying some kind of causal theory. But why couldn't we have inferentially rich kooky concepts? We can and we do: astrology is an example. However, the biological function of cognition is served mostly through roughly true theories that give the organism some control over specific aspects of its environment (there is a much longer story to be told here: see Sperber 1994; Tooby & Cosmides 1992).

Chiappe & Kukla object to our claim that memory is organized in chunks, as if it were some controversial posit of ours, and not the most common presupposition of all the memory literature. They also object that, since we don't say much about what goes in a given chunk, we leave open the possibility of tailoring particular chunks so as to confirm – vacuously – our relevance-based predictions. They are quite right: we don't have a theory of memory of our own,

nor do we claim to have one. Our concern has been, rather, to develop an account of some cognitive processes that relates in a mutually beneficial way with what is known, or will be discovered, about the organization of human memory. In general, relevance theory predicts that memory will tend (both from a phylogenetic and an ontogenetic point of view) to be organized in a relevance-boosting manner. Relevance-theoretic analyses of a particular cognitive process, say, of the retrieval of implicature from a given utterance, imply that some particular pieces of information are chunked and tend to be activated together, thus making the analysis vulnerable to experimental techniques (e.g., priming) used in memory and categorization research.

Adopting independently motivated and testable assumptions about attention and memory, we argue that relevant evidence is likely to be found by following a path of least effort. Minimizing effort, then, is not just reasonable thrift, it is an *epistemically* sound strategy.

But once on the path of least effort, how far should you go? Fodor's Hamlet problem, "when to stop thinking," would have no general answer if you had a single, open-ended thought process active in your mind (e.g., comparing the merits of being vs. not being). Unlike Fodor's Hamlet, however, humans have in mind, at any given time, several active or near-active conceptual processes competing for cognitive resources. In such conditions, Fodor's Hamlet problem has a simple in-principle answer. Let the processes with greater expected relevance win. But, of course, this time, we want expectations of effect to be a determinant factor, for least effort by itself would end up favoring no effort at all.

We assume that cognitive processes proceed in a way that is sensitive to the level of effect they achieve, and to the level of effort they expend (just as bodily movement proceeds in a way that is sensitive to the effect achieved and to muscular effort expended). This does not mean that the mind computes representations of effect and effort, let alone absolute values. All that might be involved is a sensitivity to marginal changes in levels of effect and effort and, for example, an automatic increase of effort for processes where effect is on the increase, and, after an initial grace period, a decrease of effort or a deactivation for processes where effect is on the decrease or is nil. Of course, automatic allocation of cognitive resources based on such a very rough implicit evaluation of expected relevance would allow many unproductive processes to carry on for too long, and would terminate too early some processes with great hidden potential. Your chances of ever making a true scientific discovery would be extremely slim. Well, actually, they are.

Relevance theory makes claims about cognition in general and about communication in particular. Chiappe & Kukla show no understanding of our claims about communication. Communication, we argue, raises and exploits

definite expectations of relevance. Whereas individual spontaneous cognitive activity aims at maximal relevance and may have no better way of doing so than a form of blind hill climbing (feel the terrain, choose a path that goes up but is not too rough), comprehension aims at a specific level of relevance indicated by the act of communication itself. (How? Read *Relevance*.)

Fodor asks: "[w]hat is a nonarbitrary strategy for delimiting the evidence that should be searched in rational belief fixation?" (Fodor 1987, p. 140). We have just hinted at how to answer this question: a nonarbitrary strategy available to cognitively endowed evolved organisms consists of trying to maximize the expected effect/effort ratio. This will be effective even if the organism has nothing better to base its choices on than a sensitivity to immediate increments and decrements in levels of effect and effort. Of course, such an organism would not be rational in Fodor's sense, but we claim that no individual organism ever is. Enduring collective cognitive enterprises where, through communication, relevance can be better targeted may begin to display shades of the kind of rationality that Fodor attributes to individual human cognition. Scientific thinking is a case in point. The chances of an isolated individual cognizer making a true scientific discovery are not slim – they are nonexistent. So are the chances of understanding the cognitive basis of scientific achievements without understanding the more modest cognitive feat that each of us performs thousand times every day in understanding what someone else is saying.

References

- Fodor, J. A. (1983) *The modularity of mind*. MIT Press. [DLC]
- Fodor, J. A. (1987) Modules, frames, fridgeons, sleeping dogs and the music of the spheres. In: *The robot's dilemma: The frame problem in artificial intelligence*. Ed. Z. Pylyshyn. Ablex. [DS, DLC]
- Hayes, P. J. (1987) What the frame problem is and isn't. In: *The robot's dilemma: The frame problem in artificial intelligence*. Ed. Z. Pylyshyn. Ablex. [DS]
- McCarthy, J. & Hayes, P. J. (1969) Some philosophical problems from the standpoint of artificial intelligence. In: *Machine intelligence 4*. Ed. B. Meltzer & D. Michie, Edinburgh University Press. [DS]
- Pylyshyn, Z. (ed) (1987) *The robot's dilemma: The frame problem in artificial intelligence*. Ablex. [DS]
- Sperber, D. (1994) The modularity of thought and the epidemiology of representations. In: L. A. Hirschfeld & S. A. Gelman (eds), *Mapping the mind: Domain specificity in cognition and culture*. Cambridge University Press. [DS]
- Sperber, D. & Wilson, D. (1986, revised 1995) *Relevance: Communication and cognition*. Blackwell. [DS, DLC]
- Sperber, D. & Wilson, D. (1987) Précis of *Relevance: Communication and cognition*. *Behavioral and Brain Sciences*, 10:697–754. [DS, DLC]
- Tooby, J. & Cosmides, L. (1992) The psychological foundations of culture. In: J. Barkow, L. Cosmides & J. Tooby (eds.) *The adapted mind: Evolutionary psychology and the generation of culture*. Oxford University Press. [DS]