A Tale of Two Froggies

Colin Allen, Department of Philosophy, Texas A&M University, College Station, TX 77843-4237 <colin-allen@tamu.edu>

There once was an ugly duckling. Except he wasn't a duckling at all, and once he realized his error he lived happily ever after. And there you have an early primer from the animal literature on the issue of misrepresentation -- perhaps one of the few on this topic to have a happy ending.

Philosophers interested in misrepresentation have turned their attention to a different fairy tale animal: the frog. No one gets kissed in this story and the controversial issue of self-recognition is avoided. There are simply some scientifically established facts about ways to get a frog to stick out its tongue. (Who would want to kiss a frog under those conditions, anyway?) Some frogs, it seems, are fairly indiscriminate about sticking out their tongues. Not just flies, but a whole slew of other things will go down the hatch if propelled at just the right velocity and range through a frog's visual field, provoking a tongue-flicking response. Fortunately for us all, frogs seem to be a bit more discriminating about whom they will kiss.

At first sight, the frog's tongue-flicking response seems like an ideal starting point for those who wish to promote evolutionary or "teleological" theories of intentional content. The signals passed from the frog's retina to the frog's brain were undoubtedly honed by the deaths of untold millions of insects snagged by countless generations of amphibians. Those amphibian ancestors whose eyes generated signals that were more

reliable guides to the location of food in the environment did better at propagating their genes, all other things being equal, than their cohorts whose eye to brain signals were less reliable. The teleosemanticist identifies the content of frogs' intracranial signals in terms of the environmental conditions that historically corresponded to successful tongue-flicking, namely the presence of frog food -- typically flies -- in tongue-flicking range. And their descendants live happily ever after.

But this would not be a fairy tale unless there were something to pose a credible threat to this happy ending. Enter villains, armed with BBs (but not shotguns), who find that their amphibian subjects have as much of an eye for lead pellets as for more nutritious fare. Paradise lost, for no longer can frogs flick their tongues with impunity. Too heavy a dose of BBs could very literally sink a frog. But the real moral of the story, according to those whose target is teleosemantics not frogs, is that natural selection doesn't care what the frog's eye tells the frog's brain so long as the environment serves up the right kind of thing with sufficient frequency. Whether the signal sent to the frog's brain says "food" or whether it says "food or BB" doesn't matter. When there are no BBs around, both messages provoke tongue-flicking in exactly the same circumstances and are thus equally adaptive. So, it seems, there is no evolutionary reason to prefer one phrase over the other if you want to specify the content of the frog's eye-to-brain signal and thus there are no grounds for claiming that the frog who swallows a BB is the victim of a misrepresentation.

There are, of course, gaps in this argument against teleosemantics and these gaps have been fruitfully exploited by several defenders of teleosemantics. This is neither the time nor the place to survey all of that fine work, some of it conducted by other contributors to this issue. Instead, I wish to reorient the discussion by examining the role of the English expressions that purport to give the content of the frogs' neural signals. When examined, it will be seen that the so-called disjunction problem collapses into a choice between alternatives that cannot be adjudicated in a practical vacuum.

We begin by focusing on the internal representations of two specific frogs. The first frog whom I will call "frog-then" lived at a time before BBs existed. The second frog, "frog-now", we will suppose to be a direct descendant of frog-then, and has the misfortune of being the first in her lineage to be spotted by scientists hell-bent on feeding a frog some BBs. For the example, let us imagine that frog-then's visual system has been faithfully replicated through the generations to frog-now, and that some signal token `s-then` that passed from frog-then's eye to brain is neurologically indistinguishable from `s-now` that passes from frog-now's eye to brain on her first exposure to a BB. They are both tokens of neurobiological type S.

Ignoring the indexical components (i.e. here and now) of the contents of `s-then` and `s-now`, either `s-then` and `s-now` have the same intentional content, or they do not. (I shall not include the qualification about indexicality anywhere else in this paper; it should be taken as implicit throughout.) If these two token signals have different contents, then it is possible for the content of tokens of type S to change without any

structural changes in the organisms producing those tokens. While some content-externalist views might in principle allow this possibility, it is not an option for backward-looking teleosemantic views. There has been no time for a changed selection regime to take effect because, by hypothesis, frog-now is the first of her lineage to have been exposed to a BB. With respect to the environment of selection that matters to teleosemantic accounts of content, frog-then and frog-now are in the same moat.

Proponents of alternative, non-teleological externalist accounts of content have also put forward theories which entail that S tokens have the same content then and now. Thus, for example, Fodor's (1990) principle of *asymmetric dependence* maintains constancy of content because the frogs' tendency to respond to BBs is counterfactually dependent upon the tendency to respond to flies but not conversely; i.e., the frogs' tendency to respond to BBs is causally dependent on their tendency to respond to flies, and not vice versa. Nevertheless, the environment external to frog-now is different from that encountered by frog-then, so one can imagine constructing an externalist theory of content that is sensitive to this fact. While it might be fun to explore the construction of theories that have the consequence that the contents of S tokens are different then and now, this is not the way in which the philosophical discussion of what the frog's eye tells the frog's brain has proceeded. Accordingly, I will limit the present investigation to the consequences of assuming that the contents of `s-then` and `s-now` are the same.

Let us, then, call the putative shared content of `s-then` and `s-now` "FETFB" (an acronym for Frog's Eye To Frog's Brain; note that FETFB is intended to indicate the

signals' content, not the signals themselves). How might one specify FETFB using the English language? An obvious suggestion is this: provide an English expression L such that the content of L is identical to FETFB.

One obvious problem with this proposal relates to the requirement that FETFB, the content of a neural signal, be *identical* to the content of some fragment of English. English and the signals of frog-neuralese are tools with rather different functions and it would be surprising indeed if any simple phrase of English had exactly the same content as a frog's neural signal. What, then, are we to make of the choice between "food" and "food or BB" for expressing the content FETFB? Sticking to the search for identical contents one could regard "food" and "food or BB" as shorthand for two longer expressions of English that are the genuine candidates for having the content FETFB. To settle which, if either, of the indicated longer expressions has a content that is identical with FETFB, one must determine the difference between their contents; in the absence of the longer expressions themselves, one must attempt to determine what is added to the content of the expression indicated by "food" by attaching the words "or BB".

First, a relatively superficial objection to the use of "BB" in specifying FETFB. Recall that we are operating under the lemma that `s-then` and `s-now` have the same content. At the time of frog-then's existence BBs had not been invented and there was no necessity, metaphysical or otherwise, that BBs would be invented. The planet could have been destroyed that year by a huge comet and no BBs would have ever come to be. If the term "BB" makes a contribution to the intentional content of `s-then`, then so must

names for many other entities that failed to exist then or now. But there are no good empirical reasons to refer to merely possible entities (if entities they be) in specifying the contents of the representations of actual organisms. One does not need to have much of a Quinean streak to draw the line at this point. But just in case the reader lacks a Quinean streak altogether, there is another line of argument available, viz., that if mere possibilia are relevant to the specification of content then we can never know the content of any actual signal for we can never know or specify the full range of objects that would cause the signal to be tokened (except in this unspecific, circular fashion). The utility of content specification for scientific purposes would be undermined if one cannot draw the line so as to exclude the merely possible, for it is of no interest at all to be told that a signal's content is given in terms of its possible causes, and its possible causes are just those things that might cause the signal. Drawing the line so as to exclude the merely potential causes of `s-then` eliminates reference to BBs as such from the expression of the content of `s-then`, and hence too of `s-now`. So, if the choice is directly between the contents indicated by "food" and "food or BBs", then "food" wins going away. But there is a deeper point lurking.

BBs provide a nice, real-world example of the kinds of non-food items at which frogs will stick out their tongues. The point of the disjunction problem is surely not tied to BBs per se, but to a more general class of non-food items whose ingestion by frogs was not nutritious for them. The more interesting distinction to be made in trying to specify the content FETFB is between contents that designate a class F of objects that have both the

property of causing the signal and whose ingestion has, on the whole, been good for frogs, and a class G of objects that have the former property but which lack the latter. Construed this way, G is not a class that should raise any Quinean hackles. Frogs being as indiscriminate tongue-flickers as they are, it is virtually certain that members of this class were a part of frog-then's environment, even in the distant past when BBs were not even a gleam in Fodor's eye. The disjunction problem for FETFB may then be construed as the problem of selecting between two ways of specifying the content indicated by the expressions "F" and "F or G".

Given a choice between these ways of expressing the content FETFB, we must still ask what role the word "or" is playing in the second alternative. Given that this topic is known as "the disjunction problem" one could be forgiven for thinking that it is crucial. But it is not. Philosophers' discussions of what the frog's eye tells the frog's brain have (intentionally) been carried on in the absence of any commitment to a functional role for FETFB beyond its involvement in the tongue-flicking response. FETFB is not, for example, required to play any role in inferences such as disjunctive syllogism. (The simplicity of the frog example is considered one of its main virtues for thinking about semantic theories. Its simplicity also underlies the muttered and footnoted comments of many philosophers that teleosemantics might be fine for frogs but is unlikely to scale up to princes.) If disjunction is assumed to play no structural role in the content FETFB, then the content of the English "food or BB" cannot be identical to FETFB (although their extensions may, of course, be the same). Neither can the content of "F or G" be identical

to FETFB insofar as "F or G" is an abbreviation for a disjunction of English. If the word "or" is to feature in the expression L of the schema above, we must give up on the idea that L and FETFB will have identical contents. In other words, we must give up on the proposal as it was formulated above.

Once we give up on identity between FETFB and the content of whatever expression of English we choose to represent it, the discussion is stalled unless criteria are provided for determining whether a given expression of English adequately captures FETFB. To my knowledge, no one has provided such criteria. There is an old, sizable, and largely pessimistic literature on the difficulties of using expressions of English to express the mental contents of nonhuman animals, particularly for the purposes of precise prediction of animal behavior (e.g. Dennett 1969, Stich 1983). Optimists (e.g., Allen 1992) have done little more than attempt to shoot down the arguments of the pessimists and gesture in the direction of how one might come up with qualifications upon qualifications of English phrases to express the contents of animal minds. The arguments against giving precise contents for animal thoughts aren't very good, but the theories that would make content attributions precise aren't available either. My assessment is that there is presently stalemate on this point.

I do not intend to remedy the situation here. (Indeed I am not sure how to remedy it.) However, some progress might be made by investigating a weaker condition on L, that it have the same reference or extension as FETFB. With respect to the original question of whether to gloss FETFB as "food" or "food or BB" what we need to know is

this: Do BBs belong in the extension of those signals, or don't they? But focusing on extension renders the question of choosing between "food" and "food or BB" entirely moot. For, as Fodor is fond of pointing out, in frog-then's environment the extensions of these two descriptions are identical. If the only criterion for selecting between these expressions is what they designate in current environment, then there is nothing to choose between them as expressions of what `s-then` refers to.

This result is, of course, an artifact of the non-existence of BBs in frog-then's environment. Presumably, though, it did not require the intervention of scientists to get frogs occasionally to stick out their tongues at things that aren't so good to eat, so, even discounting the relevance of BBs to frog-then, there is a real distinction between the alternatives previously labeled "F" and "F or G". Notice, however, that the move to consider these alternatives further reinforces the irrelevance of disjunction to the so-called disjunction problem. The disjunction "F or G", which stands for "signal-causing and nutritious or signal-causing and not nutritious", is logically equivalent, and thus extensionally equivalent, to the simpler phrase "signal-causing". Thus, if extensional equivalence is the only criterion applied, there is nothing to choose between these phrases. That is, from the standpoint of extensionality, the choice between "F" and "F or G" to express the extension of the frogs' neural signal is equivalent to the choice between "nutritious and signal-causing" on the one hand, and simply "signal-causing" on the other. Disjunction has disappeared from the disjunction problem, along with the not-so-ubiquitous BBs, and what is left is the question of whether teleosemantic theories

(like other naturalistic semantic theories) have the resources to limit the function of these signals to the representation of anything less extensive than their causes. Surely not even the most ardent critic of teleosemantics can doubt that natural selection does care about this distinction; animals have evolved to learn from their mistakes, after all.

If something beyond mere reference is at stake for `s-then`, then we are owed an account of how our English expressions of content are supposed to map onto Diplasicoelan content. But it doesn't seem possible to do this without begging the question against rival theories of content, for one must produce a theory of content to show how the contents of English and frog-neuralese are related. There is no theory-independent way of settling this matter. Consequently it is my view that selection of the best theory of content is not a matter for mere philosophical reflection on the consequences of each theory for our intuitive judgements about content. Rather, the theories must be judged in a different way that is based on the (putative) roles of content attribution in the behavioral sciences. The ultimate test of any theory of content will be the success of the sciences that adopt it (see Allen 1995a). Such a "metaempirical" test means that the viability of a theory of content cannot be subjected directly to experiment, but is determined by the empirical successes or failures of entire research programs. Thus the competition between theories cannot be adjudicated in advance by philosophers, although philosophers certainly play a role in distinguishing the possible forms that naturalistic semantic theories can take.

The tale of froggie neurosemantics is perhaps a useful nursery-primer for philosophers seeking to develop competence in naturalistic semantics. But despite its origins in the scientific literature, it is not a good place to pursue the metaempirical evaluation of naturalistic semantic theories. The real scientific action lies elsewhere. For whatever reasons, among them perhaps that there is little interest in employing intentional notions to describe such simple systems, there are no scientific controversies about the proper way to gloss the neural signals of frogs and no particular explanatory use to which scientists attempt to put such descriptions. So it is elsewhere that we might expect to get useful feedback about the utility of competing semantic theories, and consequently elsewhere that philosophers might more fruitfully devote their energies.

In my opinion, the much more interesting discussions are those taking place in comparative psychology and ethology about how to gloss the various communicative behaviors of animals, within both natural and artificial systems of communication. There are those comparative psychologists (particularly among chimpanzee researchers) who apply individualistic behavioral criteria -- such as whether animals learn from their mistakes and generalize what they learn -- to the attribution of content to communicative acts of animals (e.g. Savage-Rumbaugh, 1990, who labels language a "cause-effect system"). Kanzi (a bonobo) gets up and puts the hammer in the refrigerator having heard "put the hammer in the refrigerator" even though he has never heard that exact sequence of words before, because he apparently understands the content of the instruction by having generalized from other examples such as "put the hammer in the box" and "put the

cup in the refrigerator". When Kanzi later uses the lexigram keyboard to produce a series of words, the content of that sequence is interpreted in light of his individual history of learning to use similar sequences of lexigrams in particular ways, for example to signal his desire for a particular food item or to go to a particular location outdoors. The patterns of explanation in this genre are very familiar to philosophers because they closely follow the pattern of folk-psychological explanations of human behavior.

Less familiar to philosophers is the pattern of explanation among those ethologists who adopt a "functional" approach to signal content, that glosses the content of signals according to their evolutionary and ecological significance (e.g. Evans & Marler 1995). Such "functional" approaches are typically more concerned with explaining the evolution of patterns of behavior in a population than with moment-to-moment predictions of the behavior of individual organisms. In vervet monkeys, individual learning does play a role in the entrainment of specific vocalizations to specific avian predators (Cheney & Seyfarth, 1990) whereas the vocalizations of chickens to their aerial predators may be more fixed (Evans & Marler, 1995). Both, however, are functionally characterized by ethologists as alarm calls and glossed as conveying the message "avian predator present" to conspecifics. In both species there has evolved a kind of "audience effect" (Marler et al., 1991) such that the signal is emitted only in the presence of conspecifics who are likely to benefit from it. Despite the fact that the presence of a conspecific is part of the normal condition for these signals to be produced, as well as a condition of their efficacity, ethologists have not shown any inclination to gloss these signals having the

conjunctive meaning "avian predator and conspecific present". Here, then, we have a clear case of the signals' content being distinguished from both causes and the conditions under which they enhance fitness. A proper understanding of this practice will require philosophers to pay more attention to the real explanatory goals in the sciences that make use of content attributions.

A further step away from the nursery may be achieved by coming to see alternative semantic theories as complementary rather than antagonistic. Different approaches to content specification may well be compatible, reflecting different but perhaps equally legitimate explanatory goals. I believe that functional approaches correspond best to a "metacausal" explanatory role for intentional properties in the behavioral sciences, meaning that the attribution of such properties figures in explanations of how particular causal relationships between signals and behaviors are established (Allen, 1995b; see also Dretske, 1988). Other semantic theories may be more suitable for attempts to show that intentional properties are directly causally involved in the moment-to-moment behavior of organisms. However, the approaches are independent in that it is possible that the metacausal project can be sustained even if, contrary to the wishes of many, more direct kinds of intentional causation cannot be defended (Allen, 1995b).

Perhaps this is another story about misrepresentation that lacks a thoroughly satisfactory ending. But I never was much of a fan of fairy tales. Here's an ending with realism. Ethologists can choose between the range of theories of intentional content on

offer to see which best fits their explanatory projects. Philosophers can continue to search for ingenious new ways of thinking about content without the burden of having to show that any particular theory defeats allcomers. And frogs can get on with the business of catching a meal, free, it is to be hoped, from the clutches of people with government grants to flick BBs in front of their noses.

Acknowledgement

I am grateful to Scott Kimbrough, Chris Menzel, Eric Saidel and especially to Jill McIntosh for comments on earlier drafts of this essay.

References

Allen, C. 1992. "Mental content." *British Journal for the Philosophy of Science* 43: 537-553.

---. 1995a. "Intentionality: natural and artificial." In *Comparative Approaches to Cognitive Science* ed. J.-A. Meyer and H.L. Roitblat. Cambridge, MA: MIT Press, pp. 93-110.

---. 1995b. "It isn't what you think: a new idea about intentional causation." *Noûs* 29:115-126.

Cheney, D., and Seyfarth, R. 1990. *How Monkeys See the World: Inside the Mind of Another Species*. Chicago, IL: University of Chicago Press.

Dennett, D. C. 1969. *Content and Consciousness*. London: Routledge and Kegan Paul.

Dretske, F. 1988. *Explaining Behavior.* Cambridge, MA: MIT Press.

Evans, C., and Marler, P. 1995. "Language and animal communication: Parallels and contrasts." In *Comparative Approaches to Cognitive Science.* ed. J.-A. Meyer and H.L. Roitblat. Cambridge, MA: MIT Press, pp. 341-382.

Fodor, J. 1990. *A Theory of Content and Other Essays.* Cambridge, MA: MIT Press.

Marler, P., Karakashian, S., and Gyger, M. 1991. "Do animals have the option of withholding signals when communication is inappropriate? The audience effect." In *Cognitive ethology: The minds of other animals. Essays in honor of Donald R. Griffin.* ed. C. A. Ristau. Hillsdale, New Jersey: Lawrence Erlbaum, 187-208.

Savage-Rumbaugh, E. S. 1990. "Language as a cause-effect communication system." *Philosophical Psychology* 3: 55-76.

Stich, S. 1983. *From Folk Psychology to Cognitive Science: the case against belief.* Cambridge, MA: M.I.T. Press.