Title:

The Semantic Problem(s) with Research on Animal Mind-Reading

Abstract: Philosophers have worried that research on animal mind-reading faces a 'logical problem': the difficulty of experimentally determining whether animals represent mental states (e.g. *seeing*) or merely the observable evidence (e.g. *line-of-gaze*) for those mental states. The most impressive attempt to confront this problem has been mounted recently by Lurz (2009, 2011). However, Lurz' approach faces its own logical problem, revealing this challenge to be a special case of the more general problem of distal content. Moreover, participants in this debate do not agree on criteria for representation. As such, future debate should either abandon the representational idiom or confront underlying semantic disagreements.

Acknowledgments

I am grateful to Melinda Fagan, Robert Lurz, Daniel Povinelli, Tomoo Ueda, Anna Welpinghus, and an anonymous reviewer for comments and suggestions. This material also benefitted from discussion at University of Houston and the Southern Society for Philosophy and Psychology. I also thank the Alexander von Humboldt Stiftung, which partially supported this research.

Address for Correspondence:
Cameron Buckner
University of Houston
513 Agnes Arnold Hall
Houston, TX 77204-3004
Phone: 713-743-3010
Fax: 713-743-5162
cjbuckner@uh.edu

Word Count: 10840 (including references & footnotes)

1. Background: The 'Logical Problem' with Research on Animal Mind-Reading

The ability to attribute mental states to others to predict and/or explain their behaviour—labelled 'mind-reading' or (interchangeably) 'theory of mind'—has been suggested as a distinctively human capacity that accounts for our ability to learn language, live in complex societies, use tools, and pass on a cultural heritage. While humans are thought able to attribute many types of mental states, a recent debate has concerned whether some animals possess a simpler mind-reading faculty allowing them to attribute perceptual states like *seeing* to other agents. Proponents have been impressed by recent experimental evidence that at least chimpanzees, scrub jays, macaques, and ravens know what conspecifics *see* (Call & Tomasello, 2006, 2008; Emery & Clayton, 2009). A number of sceptics, however—most notably Povinelli and colleagues (Penn & Povinelli, 2007; Povinelli & Vonk, 2003) but also more recently Lurz (2009; 2011)—have here resisted. These sceptics worry that the proponents' experimental evidence can be more parsimoniously explained by supposing that animals track only the situational cues that might be used as evidence for *seeing*, such as *in-direct-line-of-gaze* (i.e. what stands in a direct, unobstructed spatial line from the direction of the animal's gaze). Since the mind-reading interpretation of these data already entail that the animal is tracking the correlations between *direct-line-of-gaze* and behaviour that feature in this behaviour-reading interpretation, these experimental designs are said to face an insurmountable 'logical problem' (Hurley & Nudds, 2006) that renders them unable to provide evidence for a mind-reading hypothesis over its behaviour-reading alternatives.

This trouble might more aptly have been called the 'semantic problem', for the difficulty of distinguishing a representation of proximal evidence from a representation of a distal semantic content—the thing or property the representation is about—is not unique to animal mind-reading experiments. Indeed, this problem—alternatively called the 'problem of distal content' or the problem of 'horizontal indeterminacy' (Dretske, 1986; Neander, 2006; Prinz, 2000)—applies much more widely. Naturalized theories of representation all attempt to derive a representation's content, in one way another, from its patterns of causal covariation with environmental circumstances, so all must overcome the roadblock that any representation more reliably covaries with the proximal evidence it detects than with its distal content. For example, if we detect cows only on the basis of cowish looks, moos, and the smell of manure, then our COW concept will more reliably covary with a conjunction of these cues than with the actual property of being a *cow*. Carefully adorn a suitably-shaped horse in a cow-costume, replay moos from a hidden loudspeaker, and splash on some eau de manure, and we will all be exposed as slaves to

proximal evidence. Animal mind-reading sceptics must be able to point out a disanalogy between their position and that of a '*cow*-sceptic' who here argues that our apparently *cow*-directed behaviour could more parsimoniously be explained in terms of a representation of *cowish-looks* & *moos* & *manure-smell*.

To avoid trapping us in a world of mere appearances, any plausible theory of representation must feature some device that allows us to 'break through' proximal evidence and represent distal contents. Several popular options are currently on offer. Mind-reading sceptics must adopt some such device, lest their arguments reduce to an empirically uninteresting degree of scepticism about representation—as though they were denying experimental evidence for chimpanzee *knowledge* because chimps cannot prove they are not brains in vats. However, several of the most popular psychosemantic theories, when applied to the disputed mind-reading designs, would resolve their more specific logical problems in decidedly inflationary ways.

Mind-reading-sceptics have thus far been coy with their underlying psychosemantics, hoping to obviate the charge of undue scepticism by suggesting alternative designs that they suppose could overcome their behaviour-reading challenge. However, in several articles and a recent book, Lurz has one-upped his fellow sceptics by arguing that the designs offered in this regard by Heyes (1998), Povinelli & Vonk (2003), and Penn & Povinelli (2007) also suffer from logical problems (see Lurz 2009, 312-314, 324-327; 2011, 79-83; and see also Gallagher & Povinelli, 2012, 165-167). Against each of these designs, Lurz applies the following recipe to generate complementary behaviour-reading interpretations of their results (2011, 76-77):

1. A mind-reading hypothesis [MRH] supposes animal *A* to anticipate the behaviour *r* of another agent *B* on the grounds that '*B* is in some mental state *m*, and *B*'s being in *m* is likely to lead it to *r*'.
2. '*A* must apply the mental state concept *m* to *B* on the grounds of [some observable facts or cues], *s*, about *B*'s behaviour or environment.'
3. 'Generate the complementary behaviour-reading hypothesis (CBRH) by replacing *A*'s mental state attributions of *m* to *B* with *A*'s observable grounds for this attribution, [*s*].'

On Lurz' analysis, any experiment for which a CBRH of this form can be constructed that makes the same predictions as (and is at least as plausible as) its MRH fails to overcome the logical problem. In place of these experiments, Lurz has suggested several new designs inspired by his 'Appearance-Reality Mindreading' (ARM) hypothesis, which holds that genuine perceptual state attribution requires the ability to distinguish appearance from

reality and to attribute illusory perceptual states to others. Lurz argues that his new designs can overcome the logical problem, because positive results on these experiments have no plausible CBRH.

In this paper, I will argue for a series of claims, all meant to establish that the fundamental problem in the debate over animal mind-reading is semantic rather than methodological.[1] In Sections 2-3, I argue that the most thoroughly-described design inspired by ARM—Lurz' 'Fake Bananas' (FB) experiment—faces its own logical problems.[2] In particular, I show that in evaluating the FB design, Lurz did not consider its most challenging behaviour-reading interpretation, whereas applying his skeptical formula more thoroughly reveals that a positive result in the FB design also has a viable CBRH. In Section 4, I argue that the FB design can overcome this challenge if we supplement it with a favourable criterion for representation—but since the same can be said about many of the earlier designs, it is not clear that FB wins Lurz (or other sceptics) a dialectical advantage. The issue is rather that the behaviour-reading challenge mongers the much more general problem of distal content; and which experiments can be regarded as sufficiently powerful to distinguish mind-reading from behaviour-reading depends upon our answers to this more general challenge. Finally, in Section 5 I rework Dretske's (1986) approach to the problem of distal content, arguing that it can be used to charitably interpret the claims of animal mind-reading proponents while sketching a productive direction for future research, and in Section 6 I rebut some final objections. The general moral is that until researchers either (1) agree on the conditions under which one agent's mental state should count as representing the mental state of another or (2) abandon the representational idiom in characterizing their dispute, no amount of methodological ingenuity could bring it to rest.

2. The FB Experiment

In the FB design (described in most detail in Lurz 2009), two chimpanzees—a subordinate and a dominant—are housed in separate rooms facing each other across a middle chamber, in which they can compete for food. The FB design calls for two pretraining phases and a test phase. In the first pretraining phase, both chimpanzees are familiarized with two kinds of objects, real yellow bananas and fake orange bananas that from a distance are distinguishable from real bananas only by their colour. Yellow and orange bananas are placed in various locations in the three chambers, and in various combinations of the guillotine doors to their chambers being open and closed.

[1] This is not to say that the problem is *merely* semantic; the question of what might be the 'right' or 'best' psychosemantics to adopt for ToM research will depend in complex ways on other empirical, methodological, and epistemic considerations (see Buckner, forthcoming). The point is rather that attention to methodological and epistemic considerations alone, without confronting the semantic issues, is unlikely to resolve the current debate.

[2] I here focus only on FB, though similar arguments generalize to the other designs inspired by ARM (Lurz, 2011).

As the chimpanzees compete for food in these situations, the subordinate is expected to learn that the dominant will retrieve yellow bananas that are physically and visually accessible to it, but will not retrieve either fake orange bananas (whether or not physically and visually accessible) or yellow bananas that are not physically or visually accessible to it.

In the second pretraining phase, the subordinate alone is familiarized with the effects of two types of barriers: clear transparent barriers that do not change the appearance of objects behind them, and red transparent barriers that make yellow objects behind them look orange. The subordinate is to learn these effects by interacting with a series of yellow and orange objects placed behind such barriers, such as toys—but crucially, as we shall see, neither yellow nor orange bananas.

In the final, test phase of the experiment, both chimpanzees are returned to their rooms, and tested in what Lurz calls the 'alternative barrier test', in which the middle room contains a red transparent and a clear transparent barrier, with a yellow banana on the subordinate's side of the red barrier and an orange banana on the subordinate's side of the clear barrier. Both chimpanzees are then allowed to enter the centre area, with the subordinate given a small head start. Lurz holds that, given the controls of the pretraining phases, if the subordinate retrieves the yellow banana behind the red barrier, then we have evidence that the subordinate can attribute the mental state *seeing-as* to the dominant, and moreover that successful performance in this test cannot be explained by a CBRH.

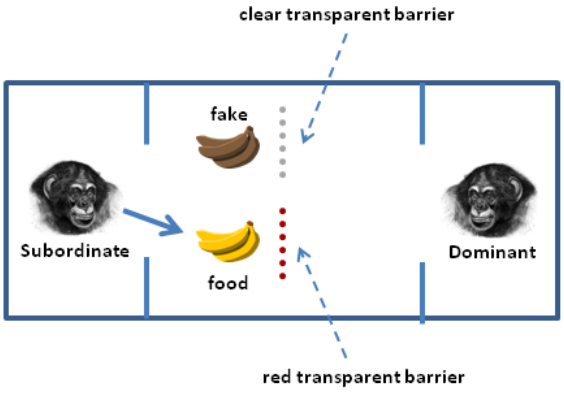


Figure 1. Expected performance of mind-reading chimpanzee in 'alternative barrier test' of Lurz' FB design (diagram reproduced from Lurz 2011, 99).

Lurz claims that this result would provide evidence for a particular three-stage MRH, with the subordinate learning that:

a) 'Yellow bananas that the dominant does not *see* and (hence) does not *see as* yellow can be retrieved, since the dominant does not retrieve them; yellow bananas that the dominant *sees* and *sees as* yellow are not retrievable, since the dominant retrieves them.'

b) 'Red barriers make one *see* yellow objects behind them *as* orange and orange objects *as* dark orange; clear barriers do not change the way the colour of objects behind them *look*.' and

c) 'The yellow banana behind the red barrier is retrievable, since the dominant, *seeing it as orange*, will not attempt to retrieve it.'

There are several peculiarities in Lurz' phrasing of *a-c*. In particular, it is notable that *a* contains no reference to orange bananas, with the transition from *a*,*b* to *c* apparently relying upon on an implicit step from *looks-orange* to *does-not-look-yellow*.[3] Furthermore, in the test phase of the experiment the subordinate must assume that the dominant, unlike himself, is naïve to the barriers' effects on visual appearances, despite the fact that the subordinate has never before observed how the dominant will behave in the presence of the clear and red transparent barriers. Though we might wonder why the subordinate in FB would make this assumption, let us grant all this for the sake of argument. Lurz then takes the CBRH for *a-c* to be the following:

a') 'Yellow bananas with which the dominant does not have direct line of [gaze] can be retrieved, since the dominant does not retrieve them; yellow bananas with which the dominant has direct line of [gaze] are not retrievable, since the dominant retrieves them.'

b') 'Red and clear barriers allow direct line of [gaze] with objects behind them,' and

c') 'The yellow banana behind the red barrier is not retrievable, since the dominant, having direct line of [gaze] with it, will attempt to retrieve it.'[4]

Since *a'-c'* lead to different behavioural predictions than *a-c* in the test phase of FB, Lurz declares the logical problem, at least in principle, solved.

3. A Critique of the FB Design

[3] This peculiarity is absent from the later formulation of *a* in Lurz (2011, 98), which begins instead: 'If it appears to the dominant chimpanzee that he has direct line-of-gaze to an orange banana in a particular location, then he will not attempt to retrieve the banana from that location…'. I focus on the earlier formulation since that work provides a more thorough defense of the FB design, but the arguments presented here apply to both.

[4] Lurz (2009, 317) originally phrases *a'-c'* in terms of 'line-of-sight'; I prefer instead 'line-of-gaze', since 'sight' might be thought to imply *seeing*. Lurz also later rephrases *a'-c'* (2011, 99-100) in terms of 'line-of-gaze' (with other cosmetic changes) but the core worry, that they focus only on the orientation of the dominant's eyes, remains.

While FB perhaps rules out any behaviour-reading hypothesis based solely on the cue *in-direct-line-of-gaze*, this is not the only observational grounds required for the subordinate to attribute *seeing-as*. Lurz' recipe for generating a CBRH requires us to 'replace the mind-reading chimpanzee's judgments of perceptual appearing in a) with their objective, observational grounds' (2011, 77); but evidence for perceptual appearances (*seeing-as*, *looks-orange*, *looks-yellow*) go beyond *line-of-gaze* to include other cues, such as the colour of the transparent barriers through which the line-of-gaze passes and the colours, shapes, and sizes of objects located behind through those barriers. This omission is significant, since these contextual factors provide additional cues to which a behaviour-reading subordinate could associate behavioural consequences. Lurz would appear not to have thoroughly followed his own skeptical recipe for generating a CBRH.

Though Lurz often writes as though each MRH has only one CBRH, a more charitable interpretation would allow that each MRH can have a range of CBRHs focusing on different subsets of the MRH's cues. This interpretation would hold that any hypothesis *Y* counts as a CBRH for an MRH *X* just when

i) *Y* takes one agent to predict the behaviour of another agent using only perceptual or situational cues—possibly a proper subset of the cues—that the former agent uses as evidence for the mental state of the latter agent according to *X*,

ii) *Y* suggests the interpretation that the former agent attributes a mental state to the latter agent because it operates on this evidence,

iii) *Y* does not require this interpretation, as it presumes only behaviour-reading abilities, and,

iv) *X* and *Y* cannot be empirically distinguished by means of further controls or experiments.

However, we might add to Lurz' terminology the notion of a *most challenging* CBRH, which would meet all of the above criteria *i-iv* with the additional constraint *v*:

v) *Y* operates on the full set of observational evidence that *X* takes the former agent to use as evidence for the mental state purportedly attributed to the latter agent, and so the mental state attribution suggested by *Y* involves the same mental state attributed in *X*.

According to this new terminology, though *a'-c'* might count as a CBRH for *a-c*, they are not its most challenging CBRH. In particular, they do not suggest the states of perceptual appearing attributed in *a-c*, for they leave out situational cues required to reliably determine when things *look orange* and *look yellow* in the relevant situations.

To assess the plausibility of FB's most challenging CBRH, we must carefully replace any reference to *looks yellow* and *looks orange* in *a-c* with their full situational evidence (I only include the parts of the process relevant to determining that the yellow banana behind the red barrier is safe for the subordinate to retrieve):

a") Yellow bananas to which the dominant does not [have non-occluded line-of-gaze] and (hence) [does not have non-occluded line-of-gaze or has line-of-gaze occluded by a clear transparent barrier] can be retrieved, since the dominant does not retrieve them;

b") [Yellow objects of a certain shape to which one has line-of-gaze occluded by a red transparent barrier] have the same behavioural consequences as [orange objects of similar shape and size to which one has non-occluded line-of-gaze],[5]

c") The yellow banana behind the red barrier is retrievable, since the dominant, [having line-of-gaze to it occluded by a red transparent barrier], will not attempt to retrieve it.

Notably, *a"-c"*, though involving no mental state concepts or projections of perceptual experience, make the same predictions as *a-c* in the FB design. Once again, the logical problem threatens a new and more complex design.

However, if *a"-c"* posit abilities that are not plausible or can be empirically distinguished from abilities posited by *a-c*, then they will not pose a logical problem for the FB design. The crucial hurdle is found in *b''*, which takes the subordinate to have grouped perceptual situations into two abstract behavioural equivalence classes, where situations falling under the same grouping have the same behavioural consequences in structurally analogous situations:

Grouping 1: [agent A] has non-occluded line-of-gaze to a yellow object [of a certain size/shape] OR [agent A] has line-of-gaze occluded by a transparent barrier to a yellow object [of a similar size/shape]

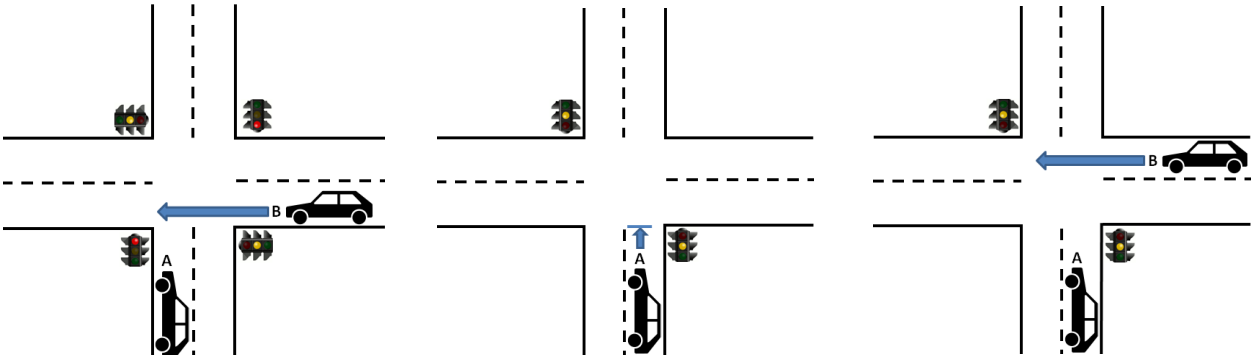
Grouping 2: [agent A] has non-occluded line-of-gaze to an orange object [of a certain size/shape] OR [agent A] has line-of-gaze occluded by a red barrier to a yellow object [of a similar size/shape]

[5] To forestall a misunderstanding, *a"-c"* do not require that any objects from the second pretraining phase need to be of the same shape and size as any objects in the test phase. Indeed, this confound could easily be controlled for in FB. Rather, both *a-c* and *a"-c"* only require that for each trial of the second pretraining phase, the object seen in direct-line-of-gaze and seen through red and clear transparent barriers must be of the same shape and size. (If, for example, a chimpanzee saw something of one shape and size on one side of a barrier and something of a different shape and size on the other side of the barrier in the same trial, it would learn something very different and much stranger about those barriers than is supposed by either *a-c* or *a"-c"*.)

These groupings should be treated as abstract situation-types with free variables (their 'slots' indicated by brackets) that can take agents and object size/shapes as values. Specifically, the subordinate must have associated *Grouping 1*, for the *banana* size/shape and *any agent*, with the behavioural consequences *will-retrieve*, and *Grouping 2*, for the *banana* size/shape and *any (naïve) agent*, with the behavioural consequences *will-not-retrieve*. While the subordinate in a properly-executed FB design will never have observed a conspecific exhibiting the behaviour *does-not-retrieve* in the situation *yellow-banana-behind-red-barrier* prior to the test phase, he will have had experience associating that behavioural consequence with the more abstract *Grouping 2* that contains this situation, for he has been familiarized with his own and the dominant's tendency to not retrieve orange bananas to which they have non-occluded line-of-gaze.

This notion of a 'behavioural equivalence class' might sound abstruse, but can be illustrated by example (Figure 2). Consider the frustrations of a Japanese colleague of mine who recently trained to receive his German driving license. Late at night in Japan after the traffic lights have been taken offline, a flashing red light at a crossroads functions like a stop sign, whereas a flashing yellow light indicates that one is driving on the road that has right-of-way (*Situation 1*). While learning to drive in Germany, my colleague observed a German driver slow when approaching an intersection with flashing yellow lights and no other traffic (*Situation 2*). Since this driver would not be required to slow at a flashing yellow light in Japan, he quickly realized that a flashing yellow light in Germany indicates a behaviourally-equivalent situation to a flashing red light in Japan. After mapping these two perceptually-disparate types of situation, he could then predict—and perhaps thereby avoid a crash!—that when approaching an intersection with flashing yellow lights traffic on the crossroad will not yield right-of-way (*Situation 3*), because this is what one would expect upon approaching a flashing red light in Japan. Similarly, if the subordinate chimpanzee in FB could link the initially perceptually-disparate situations named in *Grouping 2* into a behavioural equivalence class, then he could infer in the test phase, before ever encountering that specific situation, that a naïve dominant would respond to the yellow banana behind the red transparent barrier in the same way he was already observed responding to an orange banana to which he had non-occluded line of gaze.[6]

[6] Note that this inference likely involves the role-based, analogy-making abilities that Penn, Holyoak, and Povinelli (2008) think animals lack (cf. Section 4). Though my own assessment is that animals display significant islands of such ability (though perhaps bound to particular contexts and cues), this concern is not relevant to the critique of FB, for *a-c* require comparably sophisticated analogical reasoning capacities enabling the subordinate to link his own role as a perceiver of illusions to that of the dominant (Lurz 2011, 93). Again, precisely the same perceptual



Situation 1 (Japan)
Learned

Situation 2 (Germany)
Observed

Situation 3 (Germany)
Inferred

Figure 2. In Situation 1 in Japan, *B* (with a flashing yellow light) may proceed through the intersection without slowing, and *A* (with a flashing red light) must stop and give *B* right-of-way. Suppose a Japanese observer sees Driver *A* in Germany slowing before entering an intersection at a flashing yellow light with no other traffic present (Situation 2). Since this behaviour would not be required at a flashing yellow light in Japan but rather only at a flashing red light, this might lead the observer to realize that 'blinking yellow light in Germany' indicates a behaviourally-equivalent situation to 'blinking red light in Japan'. This linkage could then support the prediction in novel Situation 3 that driver *B* would not yield right-of-way to *A* (just as *B* would not yield in the behaviourally-equivalent Situation 1 in Japan). (Note that in Japan cars drive on the left and in Germany on the right, but this is not relevant to the example.)

If such equivalence classes are coherent, the question is then whether a behaviour-reading chimpanzee could form *Groupings 1* and *2* in FB. Lurz discusses a number of considerations that might be thought to speak against this possibility, so let us consider some rebuttals. First, Lurz insists that the second pretraining phase be structured such that subordinates never manifest a 'differential retrieval bias' between yellow objects behind clear barriers compared to yellow objects behind red barriers (2011, 207-208). However, *a"-c"* do not presume a simple preference for retrieving yellow objects behind red barriers, but rather the recognition that two abstract types of situation can have the same behavioural consequences for a naïve animal across a range of different objects and their associated behaviours. In other words, were there a different orange object that would be retrieved by the dominant

---

situations linked in *a"-c"* must be linked according to *a-c* for the subordinate to correctly track when the yellow banana would *look orange* to the dominant in the test phase of FB.

if left out in the open, then were a yellow object of the same shape located behind the red barrier from the dominant's perspective, *Grouping 2* would support the subordinate's prediction that the dominant would attempt to retrieve it (and thus that the subordinate, to avoid a beating, would not try to retrieve it).

Second, Lurz (2009, 318) considers something close to *b"* offered by Juan Gómez: 'red barriers with yellow objects behind them prevent *line-of-sight* with the yellow object but allow instead *line-of-sight* with a numerically distinct but otherwise analogous orange object.'[7] Lurz questions the coherence of this proposal due to the purportedly factive nature of *sight*: one cannot have *direct-line-of-sight* to an object that does not exist at the space and time that one bears this relation to it. Moreover, the subordinate does not at that time itself have *line-of-sight* to any orange banana at that spot, so it is odd to suppose that the subordinate could think such a banana to be present at that location and time when its own perceptual evidence provides clear evidence against it. However, Gómez' suggestion is distinct from the *b"* offered here, and there is no reason why *b"* need be phrased in terms of two numerically distinct objects. Indeed, *a"-c"* above take the subordinate to infer the dominant's behaviour on the basis not of having *line-of-gaze* to a different object than the subordinate, but rather by recognizing that the dominant is in a type of situation that leads him to have a particular behavioural disposition towards a yellow banana to which they both have *direct-line-of-gaze*.

Third, one might worry that Grouping 2 will only be associated with the behavioural outcome *will-not-retrieve* for the *banana shape* for agents that are naïve to the barriers' effects, but by the test phase, the subordinate is no longer naïve. Ex hypothesi, he should by then have learned to retrieve yellow bananas behind red barriers. However, recall that *a-c* already presume that the subordinate draws a distinction between agents that are naïve and savvy to the perceptual effects of red and clear transparent barriers and classes the dominant as naïve in the testing phase (Lurz 2011, 85)—so this ability, however it might work, could be used to track and attribute naiveté to the dominant in *a"-c"*. This suggestion raises the question as to how a behaviour-reading subordinate would be able to learn that a naïve agent is disposed to produce the same behaviours in the situations included under *Grouping 2* in FB. The experiment's elaborate controls were devised precisely to put the subordinate in the situation where the only evidence the he would have to link these disparate situations is recognition in his own perceptual experience that these two situations initially *look* the same in just the way suggested by *a-c*.

[7] Again, I would prefer that this whole discussion be phrased in terms of '*line-of-gaze*' rather than '*line-of-sight*', but I leave it here in terms of *sight* since it is not clear that *gaze* is factive in the same way.

However, there remain at least two other sources of evidence from which the subordinate in the FB design might induce these groupings. For one, the subordinate might recognize, by observing its own naïve behaviour, that it initially responded in the same way to yellow objects occluded by the red barrier as it does to orange objects of similar shape to which it had non-occluded line of gaze. Perhaps this hypothesis could be ruled out by introducing a further constraint on the second pretraining phase that the subordinate observe the effects of the barriers without being allowed to interact with or respond to the objects behind them in any way. For example, the subordinate might, without being allowed to enter the enclosure, simply observe objects being pulled from one side of the centre area to the other, with their paths passing behind the barriers. Although we might doubt whether chimpanzees could be sufficiently interested in the effects of the barriers in this manner without being allowed to interact with the objects, this control should not be dismissed on a priori grounds alone. Let us grant for the sake of argument that some such control, at least in principle, could rule out the ability to learn the *Groupings* presumed by *a"-c"* by self-observation.

Still, a final source of evidence remains: the affordances for action apparently offered by the various situation-types present in the pretraining phases. I take it to be uncontroversial in philosophy of perception that illusory situations present the viewer with a variety of apparent opportunities for action (many mistaken) from which behaviour might be inferred.[8] For example, when we observe a pencil in a glass of water that appears to be bent (due to light refraction), the pencil presents itself as offering affordances (e.g. where it can be grasped, how it would feel if stroked), some of which are accurate and some of which are mistaken. Behavioural equivalence classes such as *Grouping 1* and *Grouping 2* can be formed by abstracting from the set of situations that (initially) offer identical affordances. This finally is the roadblock that cannot be overcome by additional controls, because perceptual experience of the effects of barriers on appearances is necessarily conflated with experience of their effects on perceived behavioural affordances. Moreover, once one has recognized the correlations between a type of situation and the opportunities for action it appears to offer—based on cues that one must track in any case to determine situations in which things *look-yellow* or *look-orange*—any appeal to representations of another's perceptual states would once again rendered redundant.

[8] There are long-standing controversies about the metaphysical status of affordances—for example, whether they are objective or subjective, dispositional or informational (Caiani, 2013; Scarantino, 2003)—but all participants in these debates agree that we routinely perceive apparent opportunities for action, all that *a"-c"* require. Note also that several other theorists have been exploring even more extensive roles for affordances in mind-reading, especially Gallagher (2008) and Nanay (2013).

Thus, the logical problem again proves its resilience. Since *a"-c"* are a structural mirror of *a-c*, they take the subordinate to learn *Grouping 1* and *Grouping 2* in all the same situations that *a-c* take him to learn to attribute the mental states *looks-yellow* and *looks-orange* (which involve the same free variables) and make the same prediction in the test phase of FB. That such a complex experiment cannot clearly overcome the logical problem ought to renew our suspicion that behaviour-reading critiques presuppose an untenably severe scepticism about representation. Indeed, Lurz' recipe for generating a CBRH (c.f. Section 1 above) would not be out of place as a characterization of the general problem of distal content from its heyday in the late 1980s and early 1990s—and as such must be tempered with a positive account of representation that can overcome this more general problem to be regarded as reasonable and empirically relevant.

4. Is the Problem Logical, or Semantic?

Let us then re-evaluate the role of the logical problem in the debate over animal mind-reading. Lurz attempted to overcome the problem by devising an experiment for which a MRH and its CBRH lead to different predictions, but in evaluating the FB design he did not consider its most challenging CBRH (*a"-c"*). While *a"-c"* are clearly distinct from *a-c*—they presume no general knowledge about the difference between appearance and reality, nor even about other types of situation in which objects would *look-orange* or *look-yellow*—an attractive rebuttal would be to hold that the ability to 'recode perceptually disparate behavioural patterns' required to form *Grouping 1* and *Grouping 2* are already sophisticated enough to count as mind-reading (*sensu* Whiten 1996). The reasoning here is that such abstraction already requires the ability to posit intervening variables uniting sets of situations alike not in virtue of surface perceptual similarity, but rather in terms of underlying psychological relations. Notably, Penn & Povinelli (2007, 733) suggest that such abilities might count as a minimal case of 'bona fide' mind-reading, as they class the ability to learn such hidden similarities as the mind-reading system's core feature. However, deciding what counts as 'bona fide' mind-reading is again a matter of semantic interpretation—this time of theoretical terms in science, rather than of the mental states of chimpanzees—and comparative psychology has no accepted methods for resolving such questions (see Buckner, forthcoming).

Drawing the line for 'genuine' mind-reading at the ability to form abstract equivalence classes of this sort is convenient for Penn & Povinelli, given its fit with their broader 'relational reinterpretation hypothesis' (Penn, Holyoak, & Povinelli, 2008)—which holds that the primary psychological difference between humans and animals lies in the ability to posit underlying theoretical relations uniting perceptually dissimilar situations. Other

researchers, however, will demand more or less. Lurz faults Povinelli for drawing the line here on the grounds that 'there is a serious question whether such skills are even required for many ordinary cases of mental state attribution in humans' (2011, 73). Surely, however, the same could be said about Lurz' preferred criterion, a general mastery of the appearance-reality distinction and ability to project illusory perceptual states onto others. The role of perceptual illusions in everyday human social cognition is minor; and even in the rare cases where they are involved, they typically occur when both parties have occurrent access to the illusory stimuli, rather than being attributed to one agent by another that lacks such occurrent access (as the FB design requires). This is so even for the situation described by Lurz in his evolutionary argument for ARM: that the ability to attribute illusory perceptual states might grant fitness benefits to a subordinate during group foraging by allowing him to predict that a dominant will not retrieve a camouflaged food object, such as an insect disguised as a leaf (Lurz 2011, 89-95). The subordinate in such a situation would not actually need to project an illusory perceptual state to the dominant to predict that he will not attempt to eat the insect, since the subordinate has access to his own occurrent behavioural dispositions to treat the insect as a leaf (or, again, affordances from which such dispositions could be derived).[9] Deriving behavioural predictions in this way requires less sophistication than is presumed by Lurz' MRH for the FB design, in which the subordinate does not occurrently experience the perceptual illusion at the time of prediction. Proponents of animal mind-reading (Bugnyar, 2007; Santos, Flombaum, & Phillips, 2007; Tomasello, Call, & Hare, 2003) could thus reasonably regard the criteria for mind-reading favoured by both by Penn & Povinelli and Lurz as unduly complex.

Indeed, this debate has always been framed in terms of the contents of animal psychological states—about whether animals can represent the mental states of others—but sceptics have never paired their critique with a specific theory of representation. This weakens the sceptical critique, for if the right analysis of FB were to say that the capacities presumed by its CBRH are already 'smart enough' to count as mind-reading, then it is not clear why this would not serve as a viable response to sceptical critiques of earlier designs. Penn & Povinelli (2007) importantly gesture at informational theories of representation *sensu* Dretske (1988) to resolve such questions, suggesting that a cognitive state counts as genuinely representing the mental state of another when the two states co-vary 'in a generally reliable manner'. This is a vague specification, however (how reliably, and across what range of

[9] I here assume that the subordinate can access its own naïve and savvy dispositions simultaneously—for becoming familiarized with a perceptual illusion characteristically does not diminish its apparent force. Knowing that the Müller-Lyer lines are the same length does not diminish the appearance that they differ; and knowing where a pencil submerged in a glass of water can actually be touched does not diminish the sense that it could be touched where it appears (falsely) to be bent.

contexts?), and the sceptical refrain—that the logical problem is difficult to overcome because *direct-line-of-gaze* is one of the few cues for *seeing* (Lurz 2011, 45)—has the obvious corollary that a state detecting the former would have highly reliable covariation with the latter.

Indeed, the most popular informational theories of content offer comparatively promiscuous answers as to when a behaviour-reading mechanism would count as representing another's mental states. Many of these theories—a category of views known as *teleosemantics*—answer this question by appealing to a representation's *function*. Under these approaches, an interpretation of current data as evidence for *seeing*-attributions is plausible, for surely *line-of-gaze* is only significant in a chimpanzee's cognitive economy as evidence for *seeing*. In other words, only *seeing* stands at the right place in the dominant's motivational psychology to control the sorts of behaviours that the dominant uses line-of-gaze to track, like pursuit of food items, awareness of a subordinate's sexual advances on females, confrontational eye-gaze, and so on. This requirement is usually put counterfactually: had *line-of-gaze* not been a reliable indicator for *seeing* during an some critical period—if, for example, its conspecifics had been blind, and 'saw' using their hands—then *line-of-gaze* would not possess its current epistemic significance, such as being recruited as evidence to determine which food items are safe to eat. Teleosemantic theories institutionalize this intuition in different ways; for example, even a state that inflexibly responded only to *direct-line-of-gaze* could count as representing *seeing* if that state's ability to track *seeing* explained its fitness benefits (Papineau, 1993), allowed the systems that consume that representation to perform their proper functions (Millikan, 1984), or caused the representation to be recruited to a position of behavioural control in the organism's learning history (Dretske 1988).

There are certainly more restrictive psychosemantics on offer that might legitimize more sceptical attitudes towards animal mind-reading data; the important point at present is that the appearance of 'logical problems' in any experimental design is a function of one's underlying assumptions about the nature of representation. To be fair, animal mind-reading proponents have also not been sufficiently forthcoming about their underlying psychosemantics. Moving forward, empirical researchers thus face a dilemma: either discard the representational idiom and characterize disagreements about social cognition in terms of more precise models, or be explicit about differences in underlying psychosemantics, exposing them to critical scrutiny. Though I will not say more about it here, the former horn of this dilemma remains an attractive option for any theorist up to the challenge. While on this horn the logical problem and psychosemantic differences underlying it would be reduced to distractions, recognizing

them as such—and avoiding any concomitant talking-past and question-begging—would surely constitute theoretical progress. However, in the next section let us evaluate the prospects of a clear, ecumenical, and empirically productive psychosemantics of mindreading.

5. An Informational Proposal

There is an old joke in philosophy[10]: A diner orders lobster in a fancy restaurant and the waiter comes out carrying the silver tray with a lid on top and whisks off the lid, revealing a large, squashed roach. 'But I ordered the lobster!' cries the customer, to which the waiter responds—surely a philosopher at heart—'Well, let's call this "lobster".' Like the perplexed diner in the joke, sceptics at this point in the argument may be feeling a bit disappointed in their dinner, holding that 'mind-reading' (and 'theory of mind') all along meant something more than these deflationary teleosemantic proposals would allow. As such, in this section I briefly sketch an approach to the psychosemantics of mind-reading that is, if not lobster, at least in the more desireable upper ranks of shellfish. To do so, however, we must confront the problem of distal content head on.

To my mind, the most promising attempt to do so was mounted by Dretske (1986).[11] Acknowledging that a representation will always carry more information about a conjunction of the proximal evidence it detects than about its distal content, Dretske nevertheless suggests conditions in which a representation might mean something more than the sum of its proximal indicators. The key, Dretske suggests, is to be found not in a representation's static causal relations at some given time (the evidence it now detects), but rather in how those relations will tend to be modified through learning to better serve its representational needs. In particular, an agent may possess a representation with content that outstrips the sum of its current proximal evidence if its needs and abilities dispose it to learn new evidence for that distal content. On this proposal, at any particular time slice the reliability with which the state of a representation co-varies with its distal content need not be exceedingly high, and also need not hold across all contexts. The number of indicators so recruitable need not be open-ended, and for any species, there may be a wide range of cues that are challenging or impossible for them to master. Even in such cases, the attribution of a distal content to a representation can be legitimate if what unifies the selection of the various proximal cues is their ability to indicate the state of that distal content. Applied to the animal mind-reading debate, this metaphysical solution suggests a direction for future methodology: that we should evaluate candidate mental state representations

[10] I have heard the joke attributed to Barry Loewer.

[11] Dretske later (1988) abandoned this 'forward-looking' style of view because he thought it unable to offer a satisfactory treatment of the problem of mental causation (personal communication). I think it can offer such an account, but this issue need not concern us here.

by seeing whether the animal can revise them to be sensitive to new and additional sources of evidence for the target mental state. From this perspective, a claim of mental state representation is legitimate when better indication of that distal mental state is the limit towards which those revisions converge.

This approach offers an empirically significant account of mental state representation without reliance on overly complex criteria, based on four planks. First, **integration**: animal performance should tend to be flexible and integrated, in the sense that when an animal learns an additional indicator of a distal mental state, ceteris paribus it should be able to deploy that information and combine it with other cues across some (though not an indefinite number) novel contexts. Second, **flexibility**: animals should be able to learn some (though not an indefinite number) of novel cues to indicate that distal mental state, and their selection should be causally explained by their ability to indicate that distal state. Third, **fallibility**: the list of new cues learnable need not be open-ended; learnability might be limited to salient cues and contexts with ecological validity, and the candidate representational state need not ever achieve perfect or domain-general covariation with the distal mental state. And finally, **graduality**: there need not be any bright line (whether a threshold degree of reliability, or level of conceptual sophistication as brought by relational reinterpretation or mastery of the appearance-reality distinction) after which representation of a distal mental state clearly emerges, so long as the distal content is the limit towards which revisions converge.

From this perspective, to attribute a content *F* to representation *R* is not to say that the possession of *R* grants its bearer anything like a fully abstract, domain-general ability to identify and respond appropriately to *F*'s. While it is often assumed that humans possess such abilities with the concepts they possess, there is very little empirical evidence that this assumption is correct. Where psychologists have looked, they have instead found that even adult humans lack such ability in areas that should bastions of abstract symbolic thought, such as simple arithmetic and algebra. For example, Landy & Goldstone (2007) found that undergraduates' judgments of the validity of mathematical equations (e.g. 'a * b + c * d = a * b + c * d') were sensitive to formally irrelevant manipulations of the stimuli, such as changes of spacing amongst the symbols used in the equation and the names chosen for variables. Notably, this result held even for undergraduates that scored highly on math and algebra tests. In other words, even the abilities of educated humans on simple mathematical tasks depend upon idiosyncratic perceptual scaffolding, and in domains that lack such scaffolding, they show significantly diminished prowess. If representation demands perceptually-invariant, domain-general competence, then the undergraduates tested in

Landy & Goldstone's study must be classed as incompetent with concepts of *validity* and *equivalence*. As Penn, Holyoak, & Povinelli (2008) concede, even adult humans only approximate this form of domain-general mastery with the theoretical concepts they possess; and that humans approximate this ideal to a significantly greater degree than animals should come as no great surprise.

On the Dretskean scheme recommended here, attributing a content *F* to *R* is rather to assert that *R* is in the orbit of an attractor basin *F* in the organism's representational state space, where attractor basins indicate categories in the environment that the organism needs to (and is able to) better track (Figure 3). In order to be sufficiently close to *F* in this state space, *R* must already respond to some specific range of reliable signs for *F*, $s_1,\ldots,s_n$. Again, however, such imperfect covariation alone will be unable to determinately settle whether *R* means *F* or rather a complex configuration of more proximal cues $s_1,\ldots,s_n$. To rule out this latter sceptical interpretation, we require that the organism possess some ability to detect when it is making errors in picking out *F*s—either because in some situations $s_1,\ldots,s_n$ cause *R* to be triggered in the absence of *F* (false alarm), or because there are some cases where *F* is present but $s_1,\ldots,s_n$ are absent (miss)—and to improve the causal contact between *R* and *F* by revising the evidence to which *R* responds to avoid similar errors in the future (see also Allen, 1999). In this sense, all content attributions are idealizations on an imperfect competence—this is what philosophers mean when they say that content is *normative*—but in this case they are idealizations with clear predictive and explanatory value.

A serious worry about the application of teleosemantics to comparative psychology, clearly espoused by Chater & Heyes (1994), is that the counterfactual situations required to evaluate teleosemantic proposals are difficult to empirically assess. Evolutionary teleosemantic proposals, for example, are not beholden to what an animal can now do in the lab, but rather require analysis of the psychological effects of natural selection in counterfactual scenarios—situations to which, beyond crude models fraught with assumptions, psychologists have little epistemic access. By contrast, my modified Dretskean view, by tying representational attributions to the trajectory of learning, takes content attributions to have more empirically-accessible commitments. Experimental designs suggested by this view do not constitute a radical break with what has come before—indeed, on this psychosemantics, these designs are not radically flawed in the way sceptics suppose—but a renewed focus on learning and integration provides an overarching framework to guide future research.

Applied to the perceptual mindreading debate, the first line of investigation suggested by the view concerns whether animals can learn new, 'supra-ocular' sources of evidence for *seeing*. Existing evidence in this regard,

while not conclusive, is already significant; as Call & Tomasello point out (2006), chimpanzees integrate information from body orientations, whether the line of gaze does or does not terminate in a plausible target, presence or absence of occluders, and the type of occluders involved. The behaviour-reading interpretations of these results must concede that animals possess sophisticated, flexible learning abilities that are 'not limited to…statistical contingencies between behavioural cues', can 'represent the special causal structure amongst cues', and even 'generalize this top-down causal knowledge in an inferentially flexible and ecologically rational (i.e. adaptive) fashion' (Penn & Povinelli, 2013). If sceptics concede such learning abilities, they must also concede that the list of cues that can be learned in this way has not yet been exhausted, and we should continue to pursue alternative ecologically-valid indicators of what is or could be seen, such as gestures, acoustic signals, and lighting conditions. Second, based on a suggestion of Heyes & Dickinson (1990), we should investigate whether animals can revise their representations to be less sensitive to cues such as eye-gaze or body orientation across some range of 'looking glass' environments where contingencies between those cues and mental states have been reversed. Can animals learn that *line-of-gaze* is not threatening in the case of 'blind' agents, or adjust to situations where only the food standing in *line-of-gaze* is safe to retrieve? Third are 'integration designs', which would explicitly test whether animals can transfer information learned from one of these cues to situations involving only the others. For example, experimenters could design an enclosure in which subordinates can see the position of occluders for gaze, or routes to access its possible targets (e.g. food), but not both at the same time (see Figure 4). Can they learn to integrate information from the two perspectives to conclude which food item would be safe to eat? No one experiment would be conclusive, and these are only sketches; but rather than flesh them out with appropriate controls here, I move to consider a final rebuttal that these designs would still miss the point of the sceptical challenge.

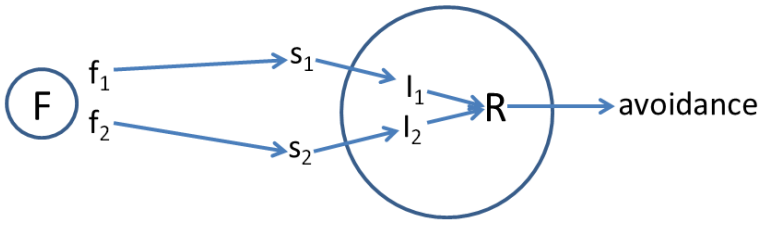


Figure 3. Reproduced sketch from Dretske's (1986) proposed solution to the problem of distal content. *R* can represent distal state *F* by tracking proximal evidence s_1 and s_2 (using perceptual indicators I_1 and I_2) if what diverse

s_1 and s_2 have in common—what explains why they were selected for behavioural control (in this diagram, avoidance)—is their ability to indicate properties f_1 and f_2 possessed by F.

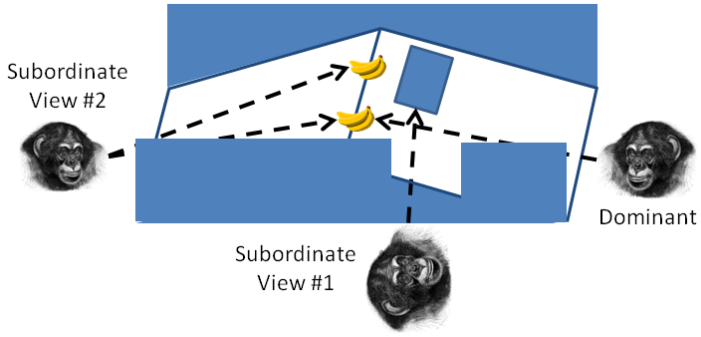


Figure 4. A proposed cue-integration design. In this modification of the Hare design, the subordinate has two views on an enclosure with an inclined surface that peaks at the middle, where the food objects will be placed. View #1 (where the subordinate would begin) reveals the location of the occluders, whereas the food items are only accessible from View #2, which reveals only the locations of the food items (and not the location of occluders). From neither view does the subordinate see the dominant's eyes. The design could be applied as a transfer test on the original Hare design, once subordinates have been familiarized with competing for food when occluders are present. (Dotted lines indicate what can be seen from each location.)

6. Simplicity, Perceptual Similarity, and Needs

A final objection to the Dretskean approach sketched above is that it is merely a form of a 'unifying hypothesis' strategy that Lurz (2011, 44-49) and other sceptics have already dismissed. On the unifying hypothesis strategy, proponents claim that there is already significant evidence for animal mind-reading because sceptics must posit a disorderly hodge-podge of behavioural rules to explain existing experimental results, compared to a single (and thus more parsimonious) MRH that explains all them all. Lurz dismisses this strategy for perceptual state attributions on the grounds that 'it does not work well for mental state attributions that possess a fairly circumscribed observable grounds' (45). In other words, since subordinates might perceive all these various situations as similar in virtue of

sharing a single cue, *in-direct-line-of-gaze*, there is not now—nor is there likely to be in the future—a multiplicity of other cues that require unification by a MRH.[12]

Three replies should be made to this potential objection. First, it is important to point out that while the Dretskean approach sketched above is close kin to the unifying hypothesis strategy, it goes beyond this strategy in the way it is typically understood. The unifying hypothesis strategy justifies a mind-reading interpretation of current results on the basis of parsimony. While there is certainly something to this line of thought, numerous commentators have remarked that reflection on parsimony alone is unlikely to resolve this particular debate. The problem is that 'parsimony' conflates several more precise notions—higher or lower levels of description, cognitive difficulty for the subject or the researcher, presence of hidden variables, number of free parameters, and compatibility with prior theories—that pull in different directions (Heyes, 1998; Shettleworth, 2009; Sober, 2005). The Dretskean approach offered here, on the other hand, justifies a mind-reading hypothesis not just on the basis of parsimony, but rather by expanding the explanandum to include the process of learning itself. While the Dretskean picture explains why any design will have a CBRH predicting that the animal responds only to proximal cues, explanations phrased in terms of specific proximal cues are not even of the right form to causally explain why those cues were acquired. Though sceptics such as Penn & Povinelli (Penn & Povinelli, 2012) have repeatedly bemoaned the insufficiency of 'behaviouristic' and 'statistical' models of learning to explain the acquisition and organization of such cues, they have not offered a more accurate model in their place.

At minimum, however, the Dretskean approach sketched above requires multiple cues in order to form an educated guess about the direction of the learning trajectory, and Lurz alleges that current results can be explained by appeal to a single cue, *in-direct-line-of-gaze*. This brings us to our second rebuttal: that whether *in-direct-line-of-gaze* counts as a single cue or a complex configuration of numerous simpler cues depends upon our theoretical perspective. As noted above, existing experiments already demonstrate that chimpanzees track *line-of-gaze* across a variety of perceptually disparate situations (Tomasello & Call, 2006). Lurz, inspired by recent-work on the theory- and concept-ladenness of perception, points out that subjects in these experiments might simply perceive each of these situations as similar in terms of a single common relation, *in-direct-line-of-gaze*, rather than needing to

[12] Note that this would be a tortured interpretation of a successful result in the design depicted in Figure 4, since neither View #1 nor View #2 offer all the cues relevant to computing *in-direct-line-of-gaze*.

explicitly compute this relation by inference from simpler perceivable features.[13] A similar line of reasoning could be used to challenge the idea, canvassed above, that the situations united in *Grouping 1* and *Grouping 2* are really so perceptually disparate—for they might be perceived as similar in virtue of initially offering the same affordances. However, as far as the metaphysics of representation are concerned, this may all just be sleight of hand; it is likely irrelevant whether the subject consciously perceives the cues as one or many, the situations as similar or dissimilar. Indeed, many proponents of theory of mind in humans have long supposed that much of the relevant processing would be tacit (e.g. Stich & Nichols, 1992), and even tacit cue unification merits explanation.

Finally, sceptics may simply adopt the Dretskean semantics offered here, but argue that we can still explain all current results with the hypothesis that the learning trajectories of animals terminate at *in-direct-line-of-gaze* rather than *seeing*. From this perspective, we might suppose the methodological worries of Povinelli and Lurz to be generally on the right track, if justified by an overly-strong behaviour-reading challenge. In particular, we might worry that the inspiration behind Lurz' ARM designs—to (finally) get past the eyes and inside the head—is still not satisfied by the modified Dretskean framework.

This response, however, again relies on an oversimplification of the Dretskean solution to the problem of distal content, because it leaves out a crucial component of nearly every naturalized theory of content: the organism's needs. The short answer to this revived sceptical challenge is that we still have no reason to believe that animals need to track *line-of-gaze* except as evidence for what conspecifics *see*. It may be an enduring consequence of the anatomy of perception that *eye-gaze* is causally downstream from *seeing*, but this should not force us to conclude that chimpanzees (or other non-verbal animals) need to represent *eye-gaze* for its own sake. For example, if we want to explain why the subordinate avoids food standing in the line of gaze of a dominant, his motivation is to avoid a beating, and only *seeing* has the right causal connection to pummelling (as well as other relevant behaviours) in the dominant's motivational psychology. If sceptics concede that performance is driven by a powerful, flexible ability to learn and integrate new cues, then they should predict that animals would learn new cues to better track the state that actually controls these relevant behavioural outcomes. So with the caveat that current evidence is not conclusive and we should continue to pursue the additional lines of investigation suggested in Section 4, if asked to

[13] This kind of response is probably not viable for the experiment suggested in Figure 4. While one could argue that the chimpanzee is somehow assembling the distant perspectives to perceive the dominant's line-of-gaze from View #2, this goes well beyond plausible phenomenology.

extrapolate now about this trajectory, it is much less plausible that it terminates at *in-direct-line-of-gaze* rather than at *seeing*.

To forestall a final objection, it is important to note that this reasoning is not so powerful that it could rebut any arbitrary sceptical challenge to the contents of animals' representational states. Consider, for example, Povinelli's similar sceptical challenge concerning chimpanzees' understanding of *weight*. Chimpanzees appear to select stone tools for nut-cracking on the basis of their weights, and their choices of tools reflect trade-offs between the distance they must carry the stone against the hardness of the nuts to be cracked (Sakura & Matsuzawa, 1991; Schrauf & Call, 2011). Whereas proponents have suggested that these experiments and observations provide evidence that chimpanzees are capable of representing an object's *weight*, Povinelli argues that these results can be explained more parsimoniously by holding that chimpanzee tool choices are instead driven by a simpler sensorimotor representation of *effort-to-lift*. The modified Dretskean strategy could not so easily be applied to rebut this sceptical challenge over *weight*—for whereas *line-of-gaze* has no intrinsic significance in the chimpanzees' life except as a cue for *seeing*, *effort-to-lift* comes with its own primitive motivational significance: a subjective cost that chimpanzees would prefer to avoid if possible.

7. Conclusion

> 'Another reason for withdrawing to a semantical plane is to find common ground on which to argue…In so far as our basic controversy over ontology can be translated upward into a semantical controversy about words and what to do with them, the collapse of the controversy into question-begging may be delayed.'
>
> -Quine (1948)

In summary, the only construal of the animal mindreading debate acceptable to both proponents and sceptics is cast in representational terms: they disagree as to whether current evidence supports the hypothesis that animals can represent at least some perceptual states of conspecifics. Whether current evidence supports the representational claim, however, depends upon one's underlying psychosemantics. If sceptics require fully domain-general, abstract competence, they appear to beg the question against proponents. Moreover, the sceptics' behaviour-reading challenge threatens to reduce to a much more thoroughgoing scepticism about representation derived from the problem of distal content. Any naturalized semantics must adopt some device to avoid this extreme scepticism, and the most popular such devices appeal in one way or another to the organism's needs. It is, however, implausible that chimpanzees (or other animals) need to represent *in-direct-line-of-gaze* (or other proximal behavioural cues) except

as evidence for *seeing*, and current experimental evidence provides strong evidence that animals can flexibly recruit a number of different cues as proximal evidence to better serve this need.

Thus, if we are capable of representing distal contents generally, then we already have significant evidence that animals can represent the perceptual states of others. Rather than accepting the consequent of this conditional, anti-representationalists may instead deny the antecedent, celebrating the discovery that the 'logical problem' in mind-reading research is at bottom the problem of distal content. Sceptical debates over other cognitive capacities might similarly be seen to depend upon the vagaries of representation, as other disputed capacities such as episodic memory, metacognition, transitive inference, and cognitive mapping are also characterized in terms of their representational contents. If we lack a consensus solution to the problem of distal content, then we might also lack a way to decide when these other mental states count as 'genuinely' representing their purported contents. Alternatively, this insight might inspire philosophers to set aside Swampman and Twin Earth to revisit old debates about content with a series of new practical applications and test cases in mind.

Before closing, it is worth noting that none of these observations impugn the relevance of Lurz' or other sceptics' experimental designs. On any plausible psychosemantics, a positive result in these designs would provide yet more impressive evidence for animal mind-reading (albeit, perhaps, in the sense that calculus would provide more impressive evidence of mathematical ability than long division). However, a perplexing feature of this debate is that most of these experiments have not yet been performed. Because they focus on increasingly sophisticated inferential abilities, and require elaborate materials, pretraining conditions, and controls, there are considerable practical barriers to their proper execution and reproduction. Simplicity is one of the greatest virtues of an experimental design, for every additional complication introduces an opportunity for unexpected challenges: uncooperative or confused animals, failures of apparatus, complexity of data analysis, and the introduction of confounds. In short, while any lab up to the task should be encouraged to perform the experiments suggested by Povinelli, Penn, Vonk, and Lurz, in the meantime the positive view sketched here justifies a more modest approach. In particular, experimentalists should continue to explore the full range of cues and contexts that animals can use in a flexible and integrated way to track the mental states of others, without letting the dream of the perfect experiment become the enemy of the good.

Department of Philosophy
University of Houston

References

Allen, C. 1999: Animal Concepts Revisited: The use of self-monitoring as an empirical approach. *Erkenntnis*, 51(1), 537–544.

Buckner, C. Forthcoming: Morgan's Canon, meet Hume's Dictum: Avoiding anthropofabulation in cross-species comparisons. *Biology & Philosophy*.

Bugnyar, T. 2007: An integrative approach to the study of 'theory-of-mind'-like abilities in ravens. *Japanese Journal of Animal Psychology*, 57(1), 15–27.

Caiani, S. 2013: Extending the notion of affordance. *Phenomenology and the Cognitive Sciences*, 1–19.

Call, J. and Tomasello, M. 2008: Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, 12(5), 187–92.

Chater, N. and Heyes, C. 1994: Animal concepts: Content and Discontent. *Mind & Language* 9(3), 209-246.

Dretske, F. 1986: Misrepresentation. In R. Bogan (Ed.), *Belief: Form, content and function* (pp. 17–36). New York: Oxford.

Dretske, F. 1988: *Explaining Behavior: Reasons in a World of Causes.* Cambridge: The MIT Press.

Emery, N. and Clayton, N. 2009: Comparative social cognition. *Annual Review of Psychology*, 60(1), 87–113.

Gallagher, S. 2008: Direct perception in the intersubjective context. *Consciousness and Cognition*, 17(2), 535–543.

Heyes, C. 1998: Theory of mind in nonhuman primates. *Behavioral and Brain Sciences*, 21, 101–148.

Heyes, C., and Dickinson, A. 1990: The intentionality of animal action. *Mind & Language*, 5(1), 87-104.

Hurley, S., and Nudds, M. 2006: The questions of animal rationality: Theory and evidence. In S. Hurley & M. Nudds (Eds.), *Rational Animals?* (pp. 1–43). Oxford: Oxford University Press.

Landy, D., and Goldstone, R. 2007: How abstract is symbolic thought? *Journal of Experimental Psychology: Leaning, Memory, and Cognition*, 33(4), 720–733.

Lurz, R. 2009: If chimpanzees are mindreaders, could behavioral science tell? Toward a solution of the logical problem. *Philosophical Psychology*, 22(3), 305–328. doi:10.1080/09515080902970673

Lurz, R. 2011: *Mindreading Animals: The Debate Over What Animals Know about Other Minds*. Cambridge: The MIT Press.

Millikan, R. 1984: *Language, Thought, and Other Biological Categories*. *Philosophical Books* (Vol. 40, pp. 441–8). MIT Press.

Nanay, B. 2013: *Between Perception and Action*. Oxford: Oxford University Press.

Neander, K. 2006: Naturalistic theories of reference. In M. Devitt & R. Hanley (Eds.), *The Blackwell Guide to the Philosophy of Language*. Oxford: Blackwell.

Papineau, D. 1993: *Philosophical Naturalism*. Oxford: Blackwell.

Penn, D., Holyoak, K., and Povinelli, D. 2008: Darwin's mistake: explaining the discontinuity between human and nonhuman minds. (B. Smith & D. Woodruff Smith, Eds.) *Behavioral and Brain Sciences*, 31(2), 109–130; discussion 130–178.

Penn, D. and Povinelli, D. 2007: On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences*, 362(1480), 731–744.

Penn, D. and Povinelli, D. 2013: The comparative delusion: the behaviorstic/mentalistic dichotomy in comparative theory of mind research. *Oxford Handbook of Philosophy & Cognitive Science*.

Povinelli, D. and Vonk, J. 2003: Chimpanzee minds: suspiciously human? *Trends in Cognitive Sciences*, *7*(4), 157–160.

Prinz, J. 2000: The duality of content. *Philosophical Studies*, 100(1), 34.

Quine, W. 1948: On what there is. *Review of Metaphysics*, 2, 21–38.

Sakura, O. and Matsuzawa, T. 1991: Flexibility of wild chimpanzee nut-cracking behavior using stone hammers and anvils: an experimental analysis. *Ethology*, 87, 237–238.

Santos, L., Flombaum, J., and Phillips, W. 2007: The evolution of human mindreading: How non-human primates can inform social cognitive neuroscience. In S. Platek, J. Keenan, & T. Shackleford (Eds.), *Evolutionary Cognitive Neuroscience* (pp. 433–456). Cambridge: MIT Press.

Scarantino, A. 2003: Affordances explained. *Philosophy of Science*, 71(5), 949–961.

Schrauf, C. and Call, J. 2011: Great apes use weight as a cue to find hidden food. *American Journal of Primatology*, 73, 323–334.

Shettleworth, S. 2009: The evolution of comparative cognition: is the snark still a boojum? *Behavioural Processes*, 80(3), 210–217.

Sober, E. 2005: Comparative psychology meets evolutionary biology: Morgan's canon and cladistic parsimony. In L. Dalston & G. Mitman (Eds.), *Thinking with Animals: New Perspectives on Anthropomorphism*. New York: Columbia University Press.

Stich, S. and Nichols, S. 1992: Folk psychology: Simulation or tacit theory? *Mind & Language*, 7(1-2), 35–71.

Tomasello, M, Call, J., and Hare, B. 2003: Chimpanzees versus humans: it's not that simple. *Trends in Cognitive Sciences*, 7(8), 239–240.

Tomasello, Michael, and Call, J. 2006: Do chimpanzees know what others see--or only what they are looking at? In S. Hurley & M. Nudds (Eds.), *Rational Animals?* Oxford: Oxford University Press.