The Folk Psychology of Consciousness[*]

Abstract: This paper proposes the 'AGENCY model' of conscious state attribution, according to which an entity's displaying certain relatively simple features (e.g., eyes, distinctive motions, interactive behavior) automatically triggers a disposition to attribute conscious states to that entity. To test the model's predictions, participants completed a speeded object/attribution task, in which they responded positively or negatively to attributions of mental properties (including conscious and non-conscious states) to different sorts of entities (insects, plants, artifacts, etc.). As predicted, participants responded positively to conscious state attributions only for those entities that typically display the simple features identified in the AGENCY model (eyes, distinctive motion, interaction) and took longer to deny conscious states to those entities.

---

[*] All authors made roughly equal contributions to the paper; the order of authors is alphabetical.

We would like to thank the audiences at the Center for Consciousness Studies, Toward a Science of Consciousness 2008 Conference, Arizona's Experimental Philosophy Lab, the Experimental Philosophy Workshop at the 2008 meeting of the Society for Philosophy and Psychology, and the *Integrating Science and the Humanities* conference at the University of British Columbia for their comments. We are also grateful to Mike Bruno, Peter Carruthers, Yali Corea-Levy, Jane Erickson, Ian Evans, Jen Giesselmann, Josh Greene, Terry Horgan, Bryce Huebner, Tony Jack, Chris Kahn, Benji Kozuch, Uriah Kriegel, Trevor Kvaran, Theresa Lopez, John Pollock, Jesse Prinz, Philip Robbins, Deena Skolnik-Weisberg, Justin Sytsma, and especially Josh Knobe and Edouard Machery for suggestions.

Address for correspondence: Dept of Philosophy, University of Arizona, Tucson, Arizona 85721

Email: Arico@email.arizona.edu

1. The problem of other conscious minds

We find it so natural to think that other people and animals are conscious that it is easy to overlook a deep puzzle: *why* do we think anything else is conscious? Consciousness is widely regarded to be a paradigmatically private affair, something that could never be publicly observed. So what prompts us to think that others have conscious minds at all?[1]

The fact that we are inclined to attribute conscious states to other creatures has long been recognized as puzzling by philosophers (Augustine, *De Trinitate* 8.6.9; Reid, 1969[1785]). We can think of this puzzle as the *descriptive* problem of other conscious minds.[2] Philosophers have made various speculations about what leads us to think that others have conscious states. J. S. Mill develops a version of the most famous answer to the puzzle: I come to believe in other minds by drawing an analogy with my own case. He writes:

---

[1] For the purpose of this paper, we focus only on the cognitive process(es) underlying token attributions of conscious experience (e.g., pain) to different entities. We are not here concerned with the folk's metacognitive recognition that one's attribution of (say) pain is an attribution *of consciousness*. For one approach to answering the metacognitive question, see Sytsma and Machery (forthcoming).

[2] We focus on the problem of why we are inclined to attribute conscious states to others. This can be thought of as part of a broader problem, 'the problem of other minds,' which asks why we are inclined to attribute any mental states whatsoever to other beings. But for philosophers like Augustine, Reid, and Mill, the problem of other conscious minds would have been a critical part of the broader problem of other minds.

> [B]y what considerations am I led to believe, that there exist other sentient creatures; that the walking and speaking figures which I see and hear, have sensations and thoughts, or in other words, possess Minds?… I conclude that other human beings have feelings like me, because, first, they have bodies like me, which I know, in my own case, to be the antecedent condition of feelings; and because, secondly, they exhibit the acts, and other outward signs, which in my own case I know by experience to be caused by feelings (Mill, 1865).

Mill's statement suggests the following procedure, which involves analogical reasoning: I know that my feelings come via my body, and I appreciate the analogy between my body and the bodies of others; I also know that certain behaviors of mine are caused by mental states; so when I see other analogous bodies exhibiting similar behaviors, I infer that their behaviors are also caused by mental states. It's easy to see why this view has been historically attractive. Analogical reasoning provides a sensible and familiar explanation for our belief in other minds.

Although the analogical theory has historical weight and a few contemporary advocates (e.g. Hill, 1984; Hyslop, 1995), the currently dominant view is that we believe in other conscious minds because it's the best explanation for what we observe. This approach has been developed in recent years by a number of people. For instance, Robert Pargetter asks, 'What is the nature of the inferences that we all so commonly, and rightly, make from certain behavioural evidence to the mental lives of other people?' He suggests that 'these inferences should best be viewed as being common scientific or hypothetic inferences, or arguments to the best explanation' (Pargetter, 1984, 158). The idea is that we come to believe in other conscious minds by using good inductive techniques – appealing to other conscious minds is the best explanation for the

behavior we observe.[3] This basic approach has been widely adopted in recent philosophy. Indeed, it's promoted in standard textbooks in the philosophy of mind (Churchland, 1988, 71-72; Graham, 1998, 57-63).[4]

The philosophical appeals to analogy and best explanation are based on speculation and informal observation. But over the last several decades there has been an impressive body of empirical work on how children and adults attribute mental states across a broad range of conditions (for reviews see e.g., Goldman, 2006; Gopnik and Meltzoff, 1997; Nichols and Stich, 2003; Perner, 1991; Wellman, 1990). Some of the mental states that have been explored in this literature are, of course, conscious mental states (see e.g., Paul Harris, 1989). This work has taught us a great deal, but very little of this work is focused directly on the traditional question, 'what leads us to think that particular individuals are bearers of *conscious* mental states?'. That is, there is little work aimed at explaining what leads us to regard something as a candidate for having conscious states at all.

---

[3] It should be noted, though, that Inference to the Best Explanation and Analogy need not be seen as mutually exclusive competitors. One could in principle take a particular argument from analogy to be the best explanation for certain kinds of behavior. In other words, one might reason along the following lines: 'well, the best explanation for that person's behavior, which is quite similar to my own, is that they are experiencing the same things as myself.'

[4] Despite the enthusiasm for the best explanation approach, a number of philosophers have worried that it is inadequate for capturing why we attribute *consciousness* to others. The worry is that, while it is easy enough to see why the best explanation for behavior would require attributing information-processing states to others, it's much harder to see why the best explanation for behavior would require attributing *conscious* states (e.g. Melnyk, 1994).

In addition to the intrinsic interest of the descriptive problem, solving the problem may also help us to evaluate whether the processes by which we gauge consciousness are in fact good processes. A familiar view in philosophy is that the way we in fact identify others as conscious is, in general, a rationally sound method (Mill, 1865; Pargetter, 1984). That is, the path by which we identify others as conscious is, roughly, a rationally appropriate path. However, it is an old and familiar philosophical worry that the basis for our thinking that others are conscious might be entirely spurious (e.g. Descartes, 1641/1986). To know whether such a worry is misplaced presumably requires determining the actual basis on which we come to think of others as conscious.

2. The problem of other minds in psychology

The descriptive problem of other minds has a long history in philosophy. Over the last twenty years, cognitive scientists have been intensively exploring how people attribute mental states to others (see e.g., Goldman, 2006; Nichols and Stich, 2003; Perner, 1991; Wellman, 1990). As noted above, little of this work in cognitive science is focused directly at our question of what leads people to attribute *conscious* states—like pain—to others. However, there is one strand of work that provides a very promising framework. This research focuses on what generates the tendency to think that a certain object has mental states *at all.*

One of the earliest investigations on the topic is Heider and Simmel's 1944 study on how adults describe an animation involving geometric objects moving about in distinctive ways. They found overwhelmingly that adults describe the animation by adverting to mental states. For example, at one point in the animation, the big triangle repeatedly bumps up against the

inside edge of a rectangle, and nearly all subjects say that the triangle wants to get out of the box. More recent work shows that children respond in much the same way to these sorts of stimuli. Like the adults in Heider and Simmel's study, when children are asked to describe what they saw, they advert to the goals, beliefs, and intentions of the triangles in a 2D animation (e.g. Bowler and Thommen, 2000; Abell et al., 2000; Dasser, 1989). If you've watched one of these animations, the results will come as no surprise. It's extremely natural to see these objects as having mental states because the motion trajectories of the triangles 'push the right buttons' to trigger mind attribution.[5] This becomes evident when one contrasts Heider-Simmel style animations with an animation of triangles moving about the surface in straight lines at constant speeds. In that case, there is no inclination to start attributing mental states to the triangles. Motion alone is not sufficient. But it remains possible that relatively simple motion cues suffice for agency attributions. For instance, change in speed plus change in direction might be sufficient to generate an attribution of a mind, even if nothing can be discerned about the goals or thoughts of that mind (Scholl and Tremoulet, 2000, 305). Of course, as adults, we don't cave to our first-blush intuitions of mentality here – we know, on slight reflection, that the images don't have minds. Nonetheless, there presumably is a mechanism that generates these powerful, if overridable, inclinations to attribute mental states, and this mechanism likely plays an important role in everyday attributions of mental states.

---

[5] It's possible that people regard the Heider-Simmel stimuli as fictional. But the point can also be made by adverting to natural reactions to the emerging class of robots – Roombas and Rodney Brooks' Mobots – that behave in goal-like ways. People even readily give names to these robots! Watching them just seems to trigger an attribution of mental states. Thanks to Bernie Kobes for these points.

Susan Johnson and colleagues use very different techniques to discern the mechanisms underlying the attribution of a mind to an individual. There are several ways that a baby might reveal that she thinks an object has a mind: she might follow the 'gaze' of the object, try to communicate with the object, imitate the behaviors of the object, or attribute goals to the object. By exploiting this variety of indicators, Johnson provides evidence that infants attribute minds as a result of particular kinds of relatively simple cues. In a representative experiment, 12 month old infants were shown a fuzzy brown object under a variety of different conditions (Johnson, Slaughter and Carey, 1998). In one condition, the fuzzy brown object (with no facial features) interacted contingently with the infant by beeping and flashing when the infant babbled or moved; in another condition, the fuzzy brown object exhibited an equivalent amount of flashing and beeping, but in this condition the activity was not contingent on the infant's behavior. In both conditions, children's looking behavior was measured when the fuzzy brown object 'gazed' at one of two objects by making a smooth, 45 degree turn towards the object and remaining in this orientation for several seconds. What Johnson and colleagues found was that infants were more likely to follow the 'gaze' of the fuzzy brown object when its beeping and flashing were contingent. In another set of conditions the fuzzy brown object did not flash or beep, but Johnson and colleagues found that babies were more likely to follow the 'gaze' of the fuzzy brown object when it had eyes than when it did not. In other experiments, babies were shown a stuffed orangutan that had a face and exhibited contingent interaction. Babies imitated the behavior of the stuffed animal and made communicative gestures toward it, indicating that the babies coded the object as having a mind (Johnson et al., 2001).

In more recent experiments, Johnson and colleagues devised a new object, the blob, a bright green object about the shape of an adult shoe that had no facial features but could beep

and move around on its own. Again they explored contingent interaction, but in this case the contingent interaction was with a confederate rather than the baby herself. In one condition, a confederate engaged the blob in 'small talk', and the blob beeped contingently with the confederate; in the other condition the blob's beeps were not contingent upon the confederate's behavior. Again they found that babies were more likely to follow the 'gaze' of the blob in the contingent interaction condition.

Finally, and perhaps most impressively, the blob design has recently been coupled with Amanda Woodward's goal attribution experiment. Woodward (1998) showed babies an arm moving towards one of two locations, each containing one of two objects. Then the locations of the objects were switched. Babies looked longer when the arm reached to the same location that now held a different object, suggesting that the babies expected the arm to reach for the same goal-object. Shimizu and Johnson (2004) found something similar with the blob – babies looked longer when the blob moved in the same direction but towards a different object. But this effect only occurred when the blob had behaved contingently with the confederate.

Johnson suggests that this broad pattern of results is evidence that by 12 months, the infant has a conceptual representation of *agent*. Characteristics like eyes and contingent interaction trigger this conceptual representation, and this representation then sets up the pattern of behavioral responses that we see in the experiments. As Johnson puts it, 'those characteristics invoke an intermediary representation (intermediary in the processing stream between perception and action) of *intentional agent* that is available to support multiple behaviors' (Johnson, 2005, 254, emphasis added).

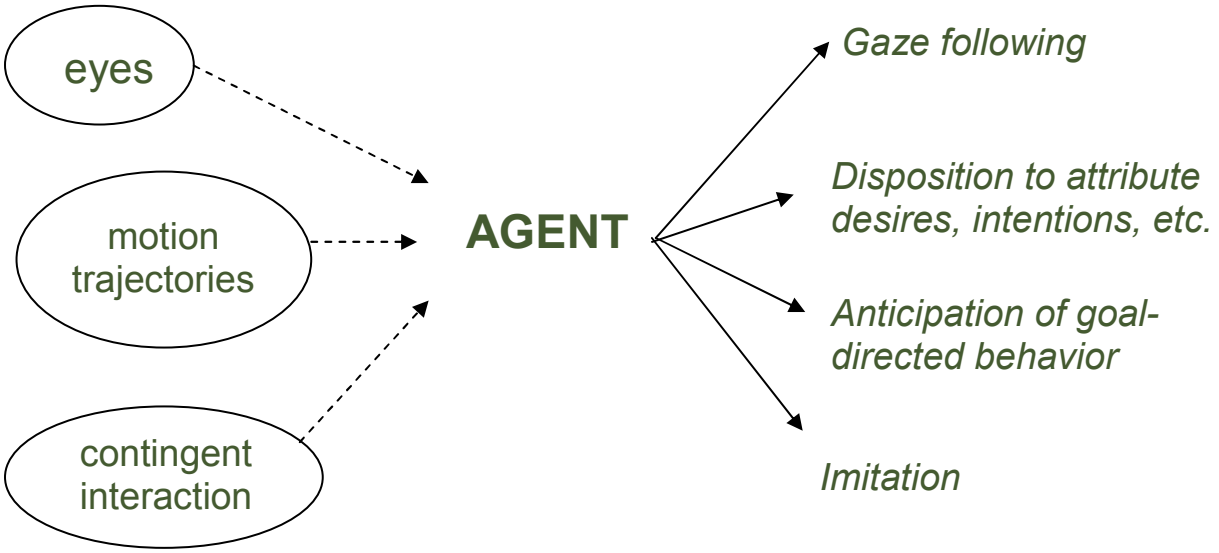

Figure 1: Network of AGENT concept

Johnson's suggestion that the concept of *intentional agent* is a critical intervening factor between the cues and the behavior can easily be expanded to accommodate the results from Heider and Simmel as well. The distinctive motion trajectories of the animation triggers the *intentional agent* concept, which in turn sets up the disposition to attribute desires, intentions, and so forth. Together, this generates the (partial) causal network depicted in figure 1. 'Agent' as it's used in this network is obviously something of a technical notion. It's not clear that it corresponds to any term in vernacular English. We will use *AGENT* to denote this technical category.[6] If an entity exhibits one of the relatively simple features mentioned above, this will

[6] Importantly, this technical category stands in stark contrast to the sophisticated concepts of agency put forth in philosophical theories of agency. (Cf. Korsgaard, 2008) Korsgaard, in the Kantian tradition, takes agents to be the sort of things that base their actions on principles of practical reasoning, including and especially principles of morality. This position is representative of the sophisticated theories alluded to above, in that they define agency in terms of the capacity to deliberate about reasons and principles when deciding how to behave. The

trigger the inclination to categorize the entity as an *AGENT*. Once an individual identifies an entity as an AGENT, this sets up certain dispositions in the person – e.g., the disposition to attribute goals, to imitate, and to anticipate goal directed behavior.

Now we can state our proposal. We maintain that, when it comes to the tendency to attribute mental states to an individual, there is nothing special about *conscious* mental states. Rather, once a person categorizes an individual as an AGENT, she will be inclined to attribute conscious mental states to the individual.[7] This account, which we'll call the 'AGENCY model', can be represented as a minor adjustment to the network characterizing the AGENCY concept,

---

notion of agency at play in the current discussion is significantly simpler and is, at best, a conceptual precursor to the more sophisticated notion under consideration in philosophical theories of practical rationality, free will, moral responsibility, etc. Whereas such theories concern themselves with the sort of agency necessary for moral responsibility, free will, rationality, and so forth, we are here concerned with a much simpler notion. Indeed, the technical category at work in the AGENCY model remains silent on such philosophical questions.

[7] For present purposes, we are neutral on whether the AGENT concept is innate. However, comparative work by Mascalzoni et al. (2010) suggests that a similar mechanism may be in place in baby chickens. If the same mechanism is in place in both humans and chickens, one might argue for innateness on broadly phylogenetic grounds. If the concept is innate, then the view begins to look similar to Thomas Reid's solution to the problem of other minds (1969 [1785]). Our point is that, regardless of how the AGENT concept is acquired, once an entity is categorized as an AGENT, an inclination to attribute conscious states to the entity will be present.

yielding figure 2. The AGENCY model builds naturally on the work in developmental psychology to explain what leads us to think that others have conscious states. But there is a developing literature in philosophy and psychology that seems to challenge it.

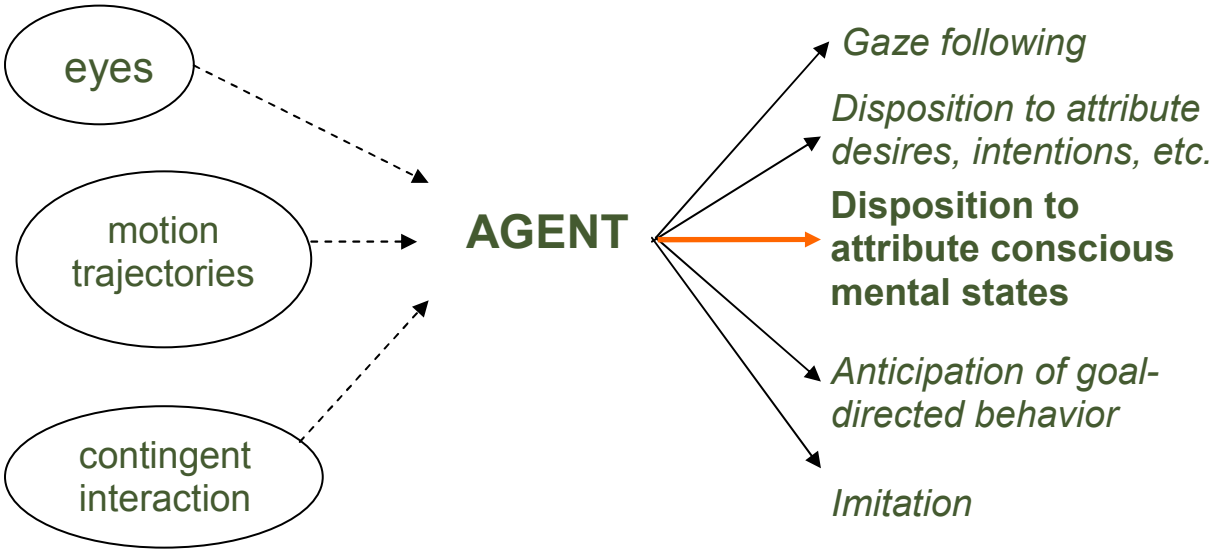

Figure 2: The AGENCY Model

3. Agency and Experience: A Dissociation?

We have suggested that the attribution of agency suffices to dispose one to attribute conscious experiences. In a recent paper in *Science,* Heather Gray, Kurt Gray and Daniel Wegner have argued that agency and experience are independent dimensions of mental attributions. Gray and colleagues collected data from over 2000 respondents to an online survey. Participants were presented with 78 pairings of 13 different characters, and then asked to rate (on a 5-point Likert scale) which character had a greater capacity for possessing a certain particular mental trait (e.g., 'Which character is more capable of experiencing joy?'). The researchers maintain that the responses reveal two quite different dimensions of mind perception: Agency and Experience (Gray et al., 2007, 619; see also Robbins and Jack, 2006). Moreover, Gray and colleagues

conclude from the data that the two features, Agency and Experience, can be dissociated. God, for instance, rates at the top of the Agency scale, but at the absolute bottom of Experience; frogs and fetuses, conversely, rate high on the Experience scale while lacking any degree of Agency.

Although the work of Gray and colleagues might be thought to challenge our AGENT-based approach, there is a natural reply. The capacities that fall under the category of *Agency* in their study are relatively sophisticated: self-control, morality, memory, emotion recognition, planning, communication, and thought. Paradigmatic entities that would trigger AGENT in our sense, like frogs, are judged to lack many of those sophisticated capacities. Thus, while frogs are extremely low on 'agency' by the standards of Gray and colleagues, frogs would certainly trigger the AGENT category on our notion of AGENT. Frogs have it all – eyes, contingent interaction, and distinctive motion trajectories. Furthermore, the traits that fall under Gray and colleague's category of *Experience* also often involve sophisticated capacities, e.g., personality, pride, embarrassment, and joy. So, Gray and colleagues' results leave open the possibility that a minimal notion of agent of the sort we have in mind would produce a tendency to attribute minimal kinds of experience, like pain. The AGENCY model aims to capture the most basic features that generate the disposition to attribute the most basic conscious states. As a result, the model isn't threatened by the findings of Gray and colleagues.

Joshua Knobe and Jesse Prinz present a different kind of evidential argument for dissociating agency and experience, suggesting that people attribute agency while positively *resisting* consciousness attributions. Their studies focus on attributions of mental states to groups. In one study, Knobe and Prinz presented subjects with two sets of sentences; one set attributed non-conscious mental states to a group entity, while another attributed conscious mental states to that same entity. The non-conscious state attributions consisted of sentences like

'Acme Corp intends to release a new product this January' and 'Acme Corp wants to change its corporate image.' The conscious state attributions included such sentences as 'Acme Corp is now experiencing great joy' and 'Acme Corp is feeling excruciating pain.' Knobe and Prinz found that subjects rated sentences about Acme Corp.'s non-conscious mental states as sounding 'natural', while they rated sentences about Acme Corp.'s conscious mental states as sounding 'weird'. The moral of these experiments, they say, is that 'people are unwilling to ascribe to group agents states that require phenomenal consciousness' (Knobe and Prinz, 2008).[8]

Knobe and Prinz's work indicates that people might regard groups as agents in *some* sense, for people seem comfortable attributing desires and plans to groups. But it's possible that groups do not activate the *primitive* notion of AGENCY that we posit. Indeed, although this has

[8] One possible deflationary interpretation of Knobe and Prinz's results is that participants were speaking figuratively when attributing mental states to groups. We conducted pilot studies to investigate this possibility. To screen out those who fail to distinguish between literal and figurative language, we gave subjects a series of sentences, some clearly figurative (e.g., 'Einstein was an egghead') and some clearly literal (e.g., 'Carpenters build houses'), and asked them to rate those sentences on a 7-point Likert scale of literalness (1= 'Figuratively True' 7= 'Literally True'). We also provided those subjects with sentences attributing different mental states to individuals (e.g., 'Some millionaires want tax cuts') and to groups (e.g., 'Some corporations want tax cuts'), and compared their judgments regarding different types of mental states. We found that when subjects rated group attributions, they tended to rate non-conscious state attributions as 'literally true' and gave significantly higher ratings for non-conscious than conscious mental state attributions ($t(66) = 7.735$, $p < .001$). For further discussion of Knobe and Prinz's experiments, see Arico (2010) and Sytsma and Machery (2009).

not been studied yet, we suspect that there is a developmental pattern concerning the attribution of mental states to groups such that young children will be less likely than adults to attribute mental states to groups. If this turns out to be right, it would suggest that children are not identifying groups as AGENTS in our sense.[9]

Moreover, as argued elsewhere (Arico, 2010; Sytsma and Machery, 2009), there are

[9] Bryce Huebner (2010) has suggested that folk psychology functions differently in attributions of beliefs (Agency, loosely speaking) than in attributions of pains and emotions (Experience). Huebner found that people attribute beliefs but not conscious states to robots. Furthermore, other research has found that people are reluctant to attribute pain and emotion to robots even when the robots are described as exhibiting appropriate behavior (Sytsma and Machery, 2010). Once again, this suggests that there is an attribution of AGENCY with no concomitant inclination to attribute conscious states. But alternative explanations are available. One is that, as we suggested for Knobe and Prinz, the AGENCY category is not really triggered by the robots. However, given that Huebner's stimuli included a photo of a robot that possessed one of the relatively simple features that serve as cues mentioned above (eyes), this is not a satisfactory explanation. Another possible explanation is that the AGENT category *is* triggered in Huebner's subjects and does, in fact, generate the disposition to attribute conscious states, but that disposition is subsequently suppressed. Perhaps the disposition is suppressed by some generally-held social schema about robots and sensations/emotions; or perhaps it is suppressed by higher-order deliberations about the robot's physiological makeup. Whatever the reason might be for suppressing the disposition, Huebner's results are problematic for the AGENCY model only in so far as it can be shown that subjects were not initially disposed to attribute phenomenology to the robots.

methodological worries with the Knobe and Prinz experiments that call their conclusion into question. For instance, Knobe and Prinz found that subjects rated sentences attributing phenomenal mental states to groups as sounding 'weirder' (less natural) than sentences attributing non-phenomenal mental states to groups. However, Knobe and Prinz stimuli were not minimal pairs: sentences attributing non-phenomenal states included additional contextual information that phenomenal attributions lacked.[10] Though Knobe and Prinz interpret the difference in sentence ratings as evidence of the folk (tacitly) distinguishing between phenomenal and non-phenomenal states, Arico (2010) provides some evidence that the difference is actually produced by the disparity in contextual information between the two kinds of sentences.

It is important to note that neither Gray and colleagues nor Knobe and Prinz present a positive proposal about the descriptive question with which we started – what features trigger our attributions of conscious states? This is not surprising given that a central purpose of their accounts was to show the independence of *experience* and *agency*. But it does mean that their accounts do not yet give a full answer to the question, 'why do we believe that others have conscious mental states?' And of course that's exactly what the AGENCY model attempts to do. In what follows we will explain the model and its empirical commitments in more detail.

4. The AGENCY Model

We suspect that a more fundamental notion of 'agency' does provide a sufficient basis for attributing conscious states. We have focused on the relatively simple features that serve as

[10] 'Acme Corp is upset about the court's recent ruling' vs. 'Acme Corp is feeling upset.'

cues, which trigger the AGENT concept. But of course these sorts of cues are not the *only* way to activate the AGENT concept. For example, merely thinking about an acquaintance (who is not physically present) may trigger the AGENT concept. Relatedly, if a trusted source of testimony *tells* me that there is a person in the other room, then this will typically lead me to think that there is an AGENT in the other room. For our purposes, the role of the relatively simple features that serve as cues is especially important, since we are particularly interested in identifying a set of minimal sufficient conditions for attributing conscious states. The key idea is that relatively simple features (e.g., motion pattern, facial features, contingent interaction) will suffice to trigger AGENT categorization, and this will in turn produce an inclination to attribute conscious states to the individual. Hence, the core of the AGENCY model is a causal sufficiency thesis:

> *Sufficiency Thesis*. Typically, if an entity is categorized as an AGENT, then there will be an inclination for attributing conscious states to that entity.

We take the sufficiency thesis to be central to the AGENCY model. But we also want to promote a more specific view about the nature of the process that generates these attributions. The research on animations (e.g. Heider and Simmel, 1944) suggests that the distinctive motion trajectories lead to attributions of goals and intentions by a process that is *fast*, *automatic*, and *unavailable to introspection* (see, e.g., Scholl and Tremoulet, 2000). Following Johnson (2005) we've suggested that this process is mediated by triggering the concept AGENT. On our AGENCY model, the disposition to attribute conscious states follows from triggering this concept. As a result, it's natural to maintain that the attribution of conscious states mediated by the AGENT concept is generated by a process that is fast, automatic, and unavailable to introspection. This is not, of course, the *only* way that an attribution of consciousness can come

about. Our model allows for attributions of consciousness that do not involve these quick, automatic inclinations. The model thus aligns with various dual-process pictures, according to which, in addition to quick, automatic, low-level cognitive processes there are slow, deliberative, controlled, high-level reasoning processes capable of operating on the same domain (see, e.g., Chaiken and Trope, 1999; Sloman, 1996; Stanovich and West, 2002).

With the dual process framework in the background, we can now state a further, more ambitious thesis of the AGENCY model. We suggest that typically the only way to get the fast, automatic, gut-level inclination to attribute conscious states to an entity is by triggering the AGENT concept. That is, categorizing an object as an AGENT is causally necessary for a certain sort of inclination to attribute conscious states:

> *Necessity Thesis*. Typically, there will be a quick, automatic inclination for attributing conscious states to an entity *only if* that entity is categorized as an AGENT.

The claim is qualified to allow for the possibility that we might attribute consciousness to an entity on the basis of more deliberate, controlled processes. But to get the gut-level immediate inclination to attribute consciousness requires categorizing the object as an AGENT.

The *sufficiency thesis* suggests that if a person categorizes an object as an AGENT, then she will be disposed to attribute conscious states to that object. This produces the interesting prediction that even relatively simple features will generate an inclination to attribute conscious states to an object. The *necessity thesis* predicts that if a person rejects the categorization of AGENT for a given object, then she will typically not have the automatic inclination to attribute conscious states to the object.

Insects provide an intriguing real-world test case here. Insects are widely regarded among neuroscientists as incapable of experiences like pain, since they lack the relevant neural

structures. Nonetheless, insects exhibit all of the simple features reviewed above – eyes, distinctive motion trajectories, and contingent interaction—that serve as cues for categorizing a thing as an AGENT. As a result, the AGENCY model predicts that insects should be categorized as AGENTS and, as a result, people should have a greater tendency to attribute conscious states to insects than to objects like clouds and rivers that lack the central cues of AGENCY and presumably do not get categorized as AGENTS.

5. Experiment

In order to test the AGENCY Model, we ran a reaction-time study in which subjects were presented with a sequence of Object/Attribution pairs. In the present study, we were interested to learn about the processes underlying folk attributions of states that psychologists and philosophers typically consider conscious or phenomenal, rather than the folk's categorizations of certain states as 'conscious'. As such, the attributions we examined specified states or properties that we take to be paradigmatic of phenomenally conscious experience, such as 'Feels Happy' and 'Feels Pain'. [11] Objects were drawn from several categories, including Vehicles, Insects, and Plants.

[11] We follow philosophical tradition in taking pain to be a paradigmatic example of a phenomenally conscious state. For an alternative view, see Carruthers, 2004. AlsoSytsma and Machery (2010) maintain that the folk conception of subjective experience is importantly different than the corresponding philosophical conception. Still, they recognize that a state like 'feeling pain' is traditionally taken to be a phenomenally conscious state. Whether the folk conceptualize it as such (tacitly or explicitly) has no bearing on the discussion at hand (Cf. fn 1).

Because the model we are proposing describes a low-level cognitive process that, we suggest, reacts automatically to simple features, we wanted to measure the mental chronometry for attributions of conscious states to various objects. Thus, for each Object/Attribution pair, subjects were asked to respond as quickly as possible (Yes or No) whether the object has the attribute. The AGENCY model makes predictions both about overt responses and about reaction times. For overt responses, the core of the AGENCY model, the sufficiency thesis, predicts that subjects should exhibit a tendency to attribute conscious states to objects that are likely to trigger AGENT categorization. Since insects exhibit the simple features that trigger the AGENT concept, and so are likely to be categorized as AGENTS, they are of particular interest. What will be especially telling will be to contrast the responses to insects with responses to objects that seem unlikely to trigger the AGENT concept, including individuals in the categories Vehicles, Natural Moving Objects, and Plants. The *necessity thesis* generates the further prediction that under speeded conditions, participants should not be inclined to make overt attributions of conscious states to objects that are not categorized as AGENT.

In a reaction time paradigm, faster response times suggest that the responses are dictated (largely, if not exclusively) by lower-level, automatic processes, while longer reaction times suggest that the responses are influenced by higher-level considerations, deliberations, associations, etc. Given this (standard) interpretation of response times, the *sufficiency thesis* predicts slower reaction times when participants *deny* conscious state attributions to objects that are typically classified as AGENTS (as compared broadly to non-AGENTS). The idea is that even if someone were to overtly express the belief that insects don't feel pain (e.g. because they lack appropriate neural structures), she would still have an strong, automatic inclination to think that they do feel pain; she thus has to overcome that automatic inclination to get out her answer

of 'no', and this will take some extra time.[12] Or, to put it more colorfully, when I am asked whether ants feel pain, there is a little guy inside of me saying 'yes!' And I need to repress that guy before I can get out my 'no' response.[13] By contrast, there's no little guy inside of me saying that trucks or rivers feel pain, so my denial in those cases should not be delayed in this way. Finally, the *necessity thesis* predicts that, for *anything* that is not counted as an AGENT, there will be no immediate inclination to attribute conscious states, so there should be no hesitation to respond 'no' for anything that does not trigger the AGENT concept.[14]

---

[12] A bit more fully, the prediction goes as follows. The presence of the cues biases the subject toward categorizing the individual as an AGENT with the consequent inclination to attribute conscious states to the individual; but competing processes defy these attributions. This creates an uncertainty that takes time to resolve, driving up RT times as a result. Since the whole issue here is about processing time, the way to investigate these matters is by testing performance under speeded conditions.

[13] This tendency may be so strong as to persist even after extensive instruction and experiences in the domain. Goldberg and Thompson-Schill (2009) found that even biology professors made predictable, childish mistakes and slower reaction times on a speeded classification task of living/nonliving. Likewise, although neuroscientists might deny that insects feel pain, our bet is that they too would be significantly slower in those denials and be more likely to make mistakes (by their own lights) in speeded conditions.

[14] The differences between the RT predictions flowing from the sufficiency and necessity theses is that the sufficiency thesis only says that we should expect faster denials (of conscious states) for at least some individuals that don't get categorized as AGENTs. The necessity thesis makes

Method

Participants

Thirty-four participants (14 male, 20 female, mean age = 19.2) from the University of Pittsburgh volunteered for this study to fulfill course requirements.

Materials

Subjects performed a timed property-attribution task, in which they were asked to respond positively or negatively to a series of questions attributing different sorts of properties to different sorts of entities. Properties included 'Feels Anger', 'Feels Happy', 'Hunts', 'Made of Metal', 'Feels Pain', 'Feels Pride', 'Is A Living Thing', and 'Is Colored White'. Entities included fifteen word items for each of eight categories: Mammals, Birds, Insects, Plants, Artifacts, Vehicles, Inanimate Natural Objects (e.g., stone, mountain), and Moving Natural Objects (e.g., cloud, blizzard). All category items were matched for letter length, number of syllables, and lexical familiarity.

Procedure

The experiment used a within-subjects design. All stimuli were presented using E-Prime software (Psychology Software Tools, Pittsburgh PA), which measured participants' response time for each trial. Participants were required to make a speeded response within two seconds before the next item was automatically presented and with a response the next item was

---

the stronger claim that we should typically expect fast denials for *any* individual that doesn't get categorized as AGENT.

presented. For each category (e.g., Insect, Plant), all subjects responded to 120 stimulus items (8 property attributions [e.g., feels pain, feels pride] x 15 entities [e.g., for Insect: bee, wasp]). Since our primary interest was in the attribution of simple conscious states, we collapsed responses to 'feels anger', 'feels happy' and 'feels pain' for our analyses. We preplanned comparisons for overt responses to attributions of these simple conscious states between the Insect category and the categories of Vehicles, Moving Natural Objects, and Plants. We also preplanned comparisons of reaction times of negative responses to simple conscious states between the category of insect and the categories of Vehicle, Moving Natural Objects, and Plants.

Results

As predicted, participants were significantly more likely (all p's <.001) to attribute simple conscious states (pain, happy, anger) to insects (70% of trials) than to Plants (10%) or items that exhibit motion, including Vehicles (6%) and Natural Moving Objects (6 %). (See chart 1). Also as predicted, participants were also significantly *slower* to reject the attributions of simple conscious states to insects (670 ms) than to natural moving objects (610 ms) ($t(33) = 2.17$, p <.05) or vehicles (616 ms) ($t(33) = 2.39$, p <.05). Contrary to our predictions, no significant difference in response times was found between insects and plants (651 ms).

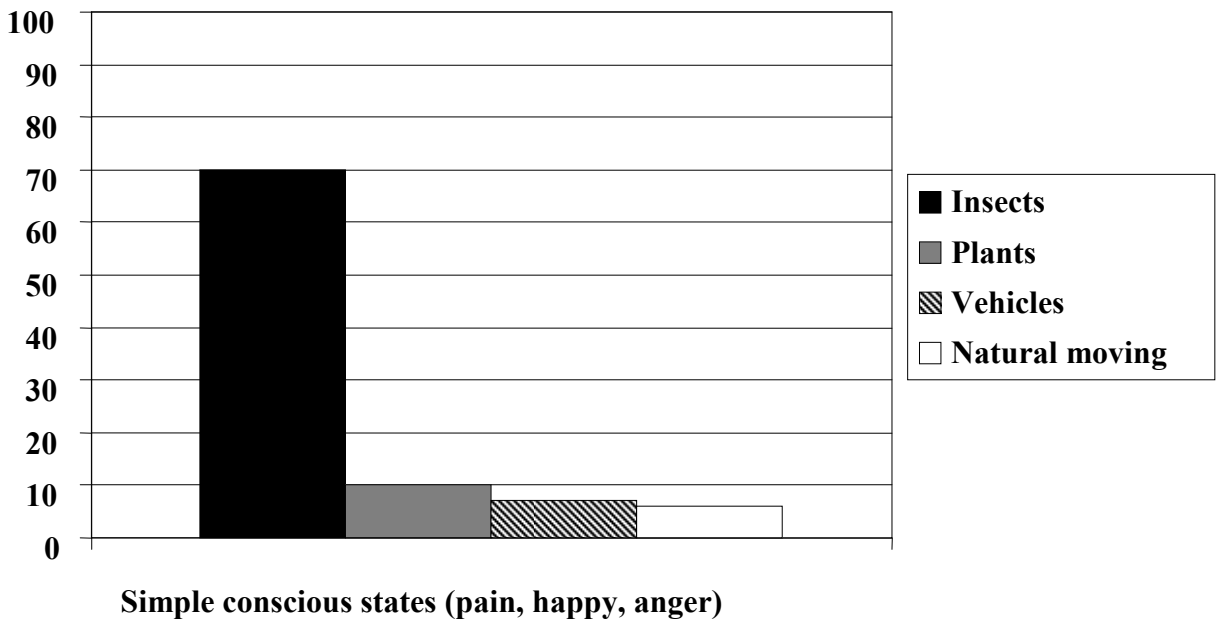

Chart 1

we conducted additional analyses comparing Plants to Vehicles and Natural Moving Objects. To our surprise, we found that participants were also more likely to ascribe simple conscious states to Plants than to Vehicles ($t(33) = 3.28$, $p < .01$) and Natural Moving Objects ($t(33) = 4.79$, $p < .001$). Participants also demonstrated significantly slower RTs for denying simple conscious states to plants (651 ms) as compared to natural moving objects ($t(33) = 4.26$, $p < .001$) and to vehicles ($t(33) = 4.73$, $p < .001$). (See chart 2).

Chart 2

(averaging across subjects, treating each item as a random variable in the analysis) replicated each of these effects (p's < .001). As a control on whether the effect with insect RTs was simply a general feature of how people respond to insects, we looked at the reactions times for denying a superficial property (e.g., *colored white*) to an object. No significant differences in response time were found between insects, plants, vehicles, and natural moving objects for these cases.[15]

[15] One might take a deflationary stance toward the results of the experiment, holding that they do not call for the postulation of any specialized mental mechanism. Instead, the results might be explicable in terms of prior beliefs about the relevant categories. It might be that subjects take longer to respond to PAIN/INSECT stimuli because they do not have any prior beliefs about the presence or absence of conscious states in insects, and so they have to deliberate longer. By contrast, it might be that subjects *do* have prior knowledge that mammals, birds and so forth *do* have conscious states, and that artifacts, natural objects and so forth *do not* have conscious states. So responding to the insect stimuli requires some on-the-spot deliberation, whereas responding

6. Discussion

The experiment supports the central hypothesis of the AGENCY model: that categorizing an entity as AGENT is typically sufficient for generating an inclination to attribute conscious states. The overt responses suggest that being identified as an AGENT generates strong inclination to attribute simple conscious states. In particular, insects, but not vehicles or clouds or rivers, were judged to have conscious states.[16] The reaction time data tell a similar story.

---

to the other categories does not. (Dave Schmitz and an anonymous referee have raised objections along these lines.)

There are a couple of reasons to doubt that this deflationary interpretation is correct. First, although the RTs for a No response to insects is significantly greater than RTs for other categories, the difference is on the order of 50 milliseconds. If the differential response reflected the necessity of conscious deliberation in the case of insects, one would expect a much larger difference. Second, we see no obvious reason for expecting people to have existing beliefs about the mental states of mammals and birds and plants, but not insects. In the absence of evidence for thinking that subjects lack beliefs specifically and uniquely about insect consciousness, there seems little positive reason to adopt this interpretation of the data. Nonetheless, it is an empirically-testable hypothesis that calls for further research on the matter.

[16] The virtue of using familiar objects like insects, birds, and vehicles, is that the examples are ecologically realistic. We aren't using some completely artificial psychological construct. However, when using familiar objects, it's hard to exclude all effects of training and education even while the word items used were extensively controlled to minimize these differences. As a

Again, our prediction was that there would be a tendency to attribute simple conscious states like pain to insects, so denying that insects feel pain would require one to override an initial natural inclination.[17] As predicted, we found that denials of simple conscious states to insects were indeed much slower than denials of such states to vehicles and natural moving objects.[18]

---

result, to make a persuasive case for the AGENCY model, it will be important to supplement our results. And perhaps the most important way to supplement the results is by using unfamiliar, novel objects. That, of course, is a central virtue of the work by Heider and Simmel and Johnson. To further explore the AGENCY model, it will be important to follow up that work by looking directly at attributions of conscious states. So, for instance, in the standard method developed by Johnson, infants will come to regard the blob (a very un-AGENT like entity) as an AGENT if it beeps contingently with a person's speech. This raises an obvious question about attribution of conscious states. Would infants in the contingent interaction condition be more likely to regard the blob as suffering pain if it were visibly damaged (as compared to a non-contingently interacting condition)? Our model predicts that they would, but the relevant experiments remain to be conducted.

[17] There is a further question about why we attribute the particular states that we do (in this case, pain, anger, and happiness). Then there is the broader question of exactly what's included within the range of states that we are sometimes willing to attribute to AGENTS. We leave this as an important area for future research.

[18] Because participants only had two seconds to respond to each item, it's unlikely that differences were caused by extended deliberations on each item. Rather, these response time differences are likely to reflect more immediate decision processes as shown in seminal and

The *necessity thesis* – the claim that AGENCY is necessary for the immediate inclination to attribute conscious states – also looks to get some support from the overt responses. Overall, items that lacked AGENCY cues were unlikely to be afforded the simple conscious states. Insects were overwhelmingly more likely to be granted simple conscious states than were the items that did not have paradigmatic AGENCY features. But the situation with RTs is more complex. Participants showed slower response times when denying simple conscious states to plants as compared with similar responses to vehicles and natural moving objects, and this seems to run contrary to the prediction generated by the *necessity thesis:* that *only* items categorized as AGENTS will generate the automatic inclination to attribute conscious states to the item.

As noted, the RT evidence on plants contravened our predictions, and this demands further consideration of how to model the role of AGENCY in the attribution of conscious states. There are a number of different models that can accommodate the data. We will not try to be exhaustive. Instead, we will review three models that we find especially interesting. First, though, we would like to note that in this experiment we also collected responses for each item on whether or not the object 'is a living thing'. In contrast to vehicles and natural moving objects, participants overwhelmingly judged insects and plants to be living things. So, throughout our discussion we will rely on the idea that plants (and insects) are coded as ALIVE.[19] The relevance of this fact will become more salient in the following discussion of the three explanatory models.

---

contemporary accounts of semantic processing based on these types of response time differences (e.g. McCloskey and Glucksberg, 1979; Smith et al., 1974).

[19] Still, as noted above for the AGENT concept, we want to allow for the possibility that the ALIVE concept is not a perfect match for any word in English.

One possibility is to retain the AGENCY model and explain away the unexpected reaction times for plants. For instance, since many things coded as ALIVE are also AGENTS, ALIVE might have high cue validity – it might provide a good heuristic for thinking that an object has mental states. That is, if something is coded as ALIVE then that's a good cue that it's an AGENT, even though it remains possible for something to be coded as ALIVE but not as AGENT.[20] As a result, it takes some time to process the fact that plants fall into the class of things that are ALIVE but not AGENTS (this proposal is illustrated in figure 3). The key point of this model is that people do not actually have a disposition to go all the way with attribution of conscious states to plants. Rather, they just need to do a bit more processing in order to exclude plants than they do for vehicles. Note that this would also explain why there were comparatively few overt responses in favor of plants having simple conscious states.

[20] As we discuss in the following section, there is some evidence in the developmental psychology literature for thinking that categorizing plants as AGENTs is, in fact, a prerequisite for categorizing plants as ALIVE. That is, rather than merely seeing ALIVE as a highly-valid cue for AGENCY, we might see ALIVE categorizations as being routed through AGENT categorizations.

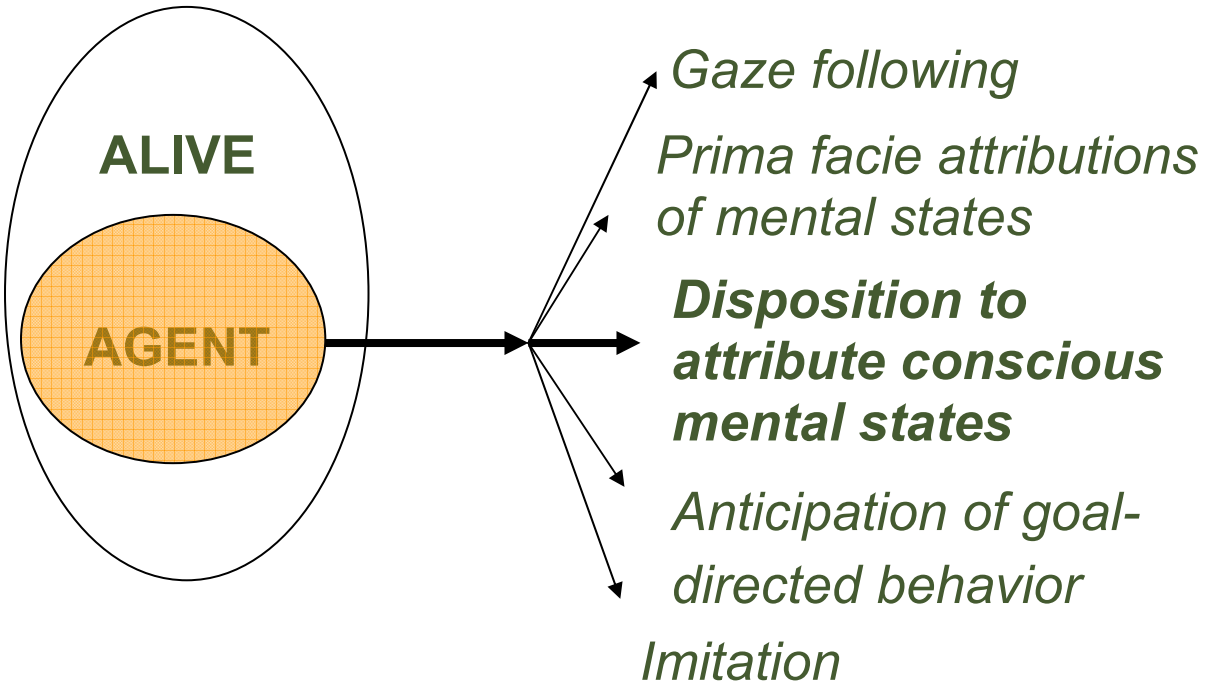

Figure 3: The Agency Model, with ALIVE as cue

A second possible model holds that categorizing an object as ALIVE is actually a (previously unnoticed) cue, which then triggers the individual to categorize that object as AGENT. That is, perhaps activating the notion that a thing is ALIVE would also activate the notion of AGENT.[21] We'll call this the LIFE-to-AGENCY model (see figure 4). On such an account, the presumed tendency to attribute conscious states to plants is mediated by the AGENT attribution.

[21] It is also, of course, possible that categorizing an individual as an AGENT would bring with it the categorization of the individual as ALIVE. For instance, it might be that once the child identifies Johnson's blob as an AGENT, the child would also think of the blob as ALIVE. We set this issue aside for present purposes.

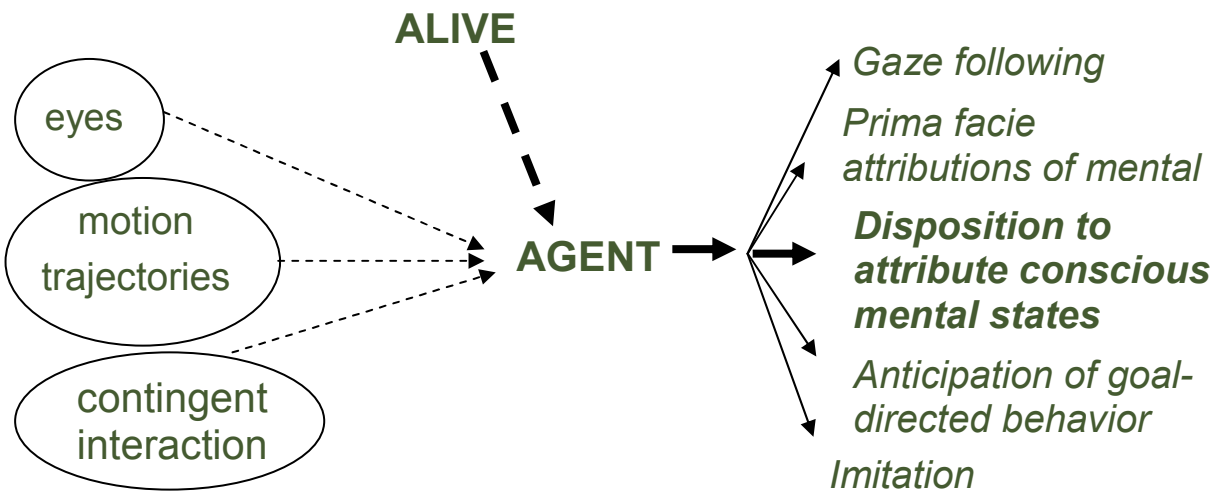

Figure 4: Life-to-Agency Model

A third and much more dramatic possibility is that AGENCY is not the important category at all. Perhaps what really matters for attributing conscious states is categorizing an object as ALIVE (as depicted in figure 5). This explanatory model provides a new way to defend the view that agency and experience are independent dimensions. An object, according to this explanation, can be regarded as something that is capable of having experiences even if it is not regarded as an AGENT. This approach would suggest that the deep conceptual link is not between AGENCY and consciousness, but rather between LIFE and consciousness. As such, we call this the LIFE model.

Figure 5: Life Model

7. Developmental questions

The evidence from our experiment does not decide between the three positive models we've just sketched. Nonetheless, we can begin to adjudicate between them by considering data from related areas of inquiry. One critical source of evidence for deciding between the models is developmental. We know that in our experiment, adults tended to overtly deny that plants have conscious states. But given their reaction times, the question is whether there is an automatic inclination in favor of attributing conscious states to plants which then gets suppressed. If there is an automatic inclination to make such attributions, then we might find fully overt attributions of conscious states to plants if we look earlier in development, before children have acquired whatever it is that suppresses overt attributions in adults.

In a pair of classic studies, Inagaki and Hatano (1987, 1991) found that Japanese children attributed conscious states to tulips. Strikingly, they found that most children (72%) in their study said that tulips feel pain, but most (80%) deny that stones feel pain. The children also overwhelmingly (98%) denied that tulips *think* (1987, 1016). This is a significant data point in favor of the idea that attributions of LIFE are sufficient to generate the tendency to attribute simple conscious states. However, one (deflationary) explanation for the Inagaki and Hatano results is that the results are culturally specific to Japanese children.

Recent work by Jane Erickson and Frank Keil (2008) suggests that the attribution of mental states to plants is not isolated to Japanese children. Erickson and Keil offered American children different kinds of explanations for a variety of entities, including humans, plants, and non-living natural kinds. In one experiment, the child's task was to evaluate the quality of a psychological explanation for a behavior exhibited by different kinds of entities. The

experimenters showed a picture of a plant growing toward the sun and offered the following psychological explanation: 'The flower grew towards the sun because it felt tired and hungry and knew that the sunlight would give it energy to make food.' Children regarded such a psychological explanation as apt for plants, but they rejected parallel psychological explanations for artifacts and non-living natural kinds. In addition, Erickson and Keil found a developmental progression – older children were less likely to allow psychological explanations for plants (though they continued to regard plants as exhibiting goal-directed behavior). All of this seems to fit with the possibility that the AGENCY model is completely misguided and should be rejected in favor of the LIFE model. For the data leaves open the possibility that LIFE is the operative categorization behind the immediate inclination to attribute conscious mental states.

Combined with our own evidence with respect to reaction times for plants, one might construe the evidence from Inagaki and Hatano and Erickson and Keil as thoroughly undercutting the AGENCY model. After all, one might take these studies to show that children are willing to attribute psychological states to plants based solely on the fact that plants are alive; and if LIFE suffices to generate the inclination to attribute conscious states to an individual, then there seems to be no need to appeal to AGENCY to explain the inclination to attribute conscious states to people and insects. Although this looks very bleak indeed for the AGENCY model, recent work by John Opfer and colleagues motives a much more favorable interpretation of the foregoing results. Opfer (2002) showed subjects—both children and adults—a series of videos in which irregularly-shaped blobs moved along various paths. Among other things, the task required subjects to determine what the object was. In one condition, the videos included goal objects that corresponded to the blob's movement, while in the other condition, the videos did not include goal objects; they then presented subjects with a battery of attributions that included

both psychological traits and biological traits. According to these studies, attributions of life themselves depend on the individual seeing the entity as a goal-directed agent. Even children as young as 5 years old, they say, relied on the blob's motion being goal-directed in order to attribute to it either biological states (like being alive) or mental states (like being happy or feeling pain). Opfer observes, 'the goal-directedness appeared to be the decisive factor in these displays: identical blobs that moved identically but towards no goal failed to convince children and adults that they were living things.'(116)

In a separate study, Opfer and Siegel (2004) found that 5-year-olds did not initially take plants either to be alive or to behave in goal-directed manners. Indeed, Opfer and Siegel found that most children only take animals to be alive. Yet when told that plants move towards sources of nutrition, such as sunlight and water,[22] 5-year-olds inferred that these motions were goal-directed. Importantly, Opfer and Siegel also found that the attributions of *life* seem to depend on attributing goal-directedness to the entity:

> The changes in categorization by children in the teleology feedback group suggest that although 5-year-olds initially categorized only animals as living things, the preschoolers' concept of living things included capacity for goal-directed movement as an important property of life. After the preschoolers learned that plants, like animals, were capable of goal-directed movement, they quickly inferred that plants, like animals, are alive. Conversely, children who failed to revise their judgments about teleological agency also failed to revise their life judgments. (321)

While most of the children in the study initially categorized animals--but not plants--as living things, the data from this study suggests that the exclusion of plants arises from the failure to

[22] Opfer and Siegel label this group the 'teleology feedback' condition.

recognize that plants display goal-directed behavior. Opfer and Siegal write, 'once children conclude that plants can act in goal-directed, self-sustaining ways, they also conclude that plants are alive.'(329) This casts an entirely new light on the results concerning attribution of consciousness to plants. Although one might initially be inclined to interpret the data from Inagaki and Hatano or Erickson and Keil as undercutting the AGENCY models, this interpretation seems less plausible in light of the fact that the attribution of life itself seems to depend on the recognition of an entity *as a goal-directed entity*.

Moreover, considering the work by Opfer and colleagues, it now seems natural to interpret the Erickson and Keil data as providing some *support* for the AGENCY model. Erickson and Keil offered an explanation of plant behavior that was explicitly goal directed. To accept these explanations (as the children do) is to accept a goal-directed understanding of plants. According to the AGENCY model, AGENTs are, *inter alia*, the sorts of things that we are inclined to attribute goals to. That children approve of psychological explanations of goal-directed behavior attributed to plants just seems to show that, once children see plants as the sorts of entities that have goals and behave according to those goals, they are inclined to say that plants also can have psychological states. As such, Erickson and Keil's data seem to provide evidence that, once they are primed to categorize plants as AGENTs, children are inclined to attribute psychological states to plants. And that just is what the AGENCY model predicts.[23]

[23] If categorizing plants as ALIVE requires categorizing them as AGENTs, then the *necessity thesis* is also preserved: the delayed responses for plants are no longer an obvious counter-example, since the initial inclination to attribute simple conscious states can plausibly be explained by attributions of LIFE tacitly depending upon attributions of AGENT.

8. Philosophical Implications

Our proposed account has some relatively clear implications regarding the descriptive problem of other conscious minds. The picture on offer suggests that our cognitive systems are differentially sensitive to very specific features – eyes, distinctive motion trajectories and contingent interaction –the detection of which is normally causally sufficient to bias the subject toward attribution of conscious states. Importantly, the cognitive processes that operate on these low-level cues appear to operate even when global considerations militate against the idea that the target has mental states. For example, my belief that a triangle from Heider and Simmel's animation is not a mentalistic agent does not stop me from seeing the triangle as 'wanting to get out of the box'. Similarly, my belief that insects are not conscious does not prevent me from having the inclination to attribute pain and anger to them. The simple features suffice to trigger the inclination to attribute conscious states. It is natural to think of the relevant cognitive process as exhibiting two central features of modularity: domain specificity and informational encapsulation (Fodor, 1983). The process is domain specific to the extent that a limited number of simple features are able to trigger that process. In the normal case, things like clouds, blizzards, cars and mountains do not exhibit the right sorts of features, and hence fail to trigger the process.[24] Since the relevant process fails to incorporate important information from the subject's broader set of beliefs (for example, the belief that that triangle is *literally* a two-dimensional figure), the process also seems, to that extent, informationally encapsulated. These points apply equally to the various models under consideration (whether they give a central role to AGENCY, LIFE, or both), for all models give special weight to a small set of relatively

[24] However, such objects may *seem* to exhibit some of these features in unusual circumstances.

simple featural cues.

Consider again the two traditional answers to the descriptive problem of other conscious minds: analogy and inference to the best explanation. Both analogical reasoning and inference to the best explanation are thought to be paradigm cases of non-modular processing (Fodor 1983, 88-89, 107). First, neither process is domain specific: analogies can in principle relate anything to anything, and explanations can in principle target arbitrary explanatory domains. And second, neither analogical reasoning nor inference-to-the-best-explanation is informationally encapsulated; both are highly sensitive to information from all over the cognitive system (Fodor, 1983, 104-107). If this assessment of the processes is correct, then there is a tension between the traditional answers to the descriptive problem of other conscious minds and the view that we are promoting. For unlike traditional accounts, on our account it is natural to view some of the mental processes responsible for the attribution of conscious states as modular.

The modularity hypothesis provides a familiar framework for articulating the difference between the AGENCY model and the traditional analogy and best-explanation models. But the important differences can still be drawn without adopting the modularity framework. The important point is that on the AGENCY model, the way we attribute consciousness to other things often involves simple cues (like contingent behavior) that quickly, automatically, and effortlessly trigger the AGENT concept, which in turn leads directly to an inclination to attribute conscious states. These operations differ radically from traditional intellectualist processes like analogical reasoning and inference to the best explanation. This point again resonates with various dual-process models that describe two distinct pathways for reaching the attributions. Kahneman provides a nice summary of dual-process architecture:

The operations of System 1 are typically fast, automatic, effortless, associative, implicit

> (not available to introspection), and often emotionally charged; they are also governed by habit and are therefore difficult to control or modify. The operations of System 2 are slower, serial, effortful, more likely to be consciously monitored and deliberately controlled (Kahneman, 2003, 698).

Regarding attributions of consciousness, we have essentially been developing an account of a System 1 or *low road* process involving automatic processing of relatively simple featural cues. But we also believe that people have another pathway for attributing conscious states, a *high road* process involving relatively slow, controlled and deliberative reasoning (see Fiala et al. forthcoming).

In addition to the implications for the traditional descriptive problem of other minds, the models we've sketched here might also have import for the status of particular intuitions about the consciousness (or not) of other entities. A number of prominent philosophers have built explicit theories of mind partly on the basis of our intuitions about what is conscious and what is not. One such case is Ned Block's famous example in which we are to imagine that all the residents of China are rigged up with radio transmitters so as to functionally mimic a living brain (Block, 1980). The intuitive reaction to this case is presumably that the nation of China (as a whole) does not have any mental states, and surely not any conscious states! This is then supposed to support the conclusion that merely getting a system to functionally mimic a brain does not make that system conscious. For intuitively, it seems clear that functionally organizing the residents of China to behave like a brain would not make the collective nation have conscious states.

Notice, however, that if our proposal is correct, there is a potential explanation for these intuitions that does not involve the denial that the nation of China enjoys conscious states.

Instead, it may be that the example tends to provoke these intuitions because the sorts of featural cues that typically incline a subject toward attributions of consciousness are not salient with respect to the nation of China.[25] Each Chinese person, on the other hand, does possess the relevant cues (which are readily imaginable when considering Block's scenario). Hence, we have an inclination to attribute conscious states to individual Chinese people, but not to the nation of China. This is just one example, but it illustrates how the answer to the descriptive problem might influence the way the relevant intuitions are used in philosophical debate. For depending on what one thinks about the epistemic status of the relevant psychological processes, one might be led either to dismiss the intuitions, or to give them special weight. [26]

An answer to the descriptive problem might also bear on the idea that people intuitively embrace a 'folk dualism' (Bloom, 2004), according to which the mind is radically different than the body.[27] It's plausible that one aspect of such a dualism is the apparent gulf between

---

[25] This is not to say that the *only* path to attributions of consciousness is via simple cues. Rather, the point is that if an entity fails to manifest the cues that trigger the AGENCY category, it will be significantly less natural to attribute conscious states to that entity. Instead, attributions of consciousness in such cases are likely to be the result of deliberate high-level reasoning.

[26] Huebner, Bruno, and Sarkissian, 2010 present some evidence that intuitions regarding the conscious states of group entities (like the Nation of China) are culturally diverse. They found that English-speaking students in Hong Kong treated consciousness ascriptions to groups and to individuals more similarly then did their U.S. counterparts. This cultural variance, Huebner et al. argue, undermines the epistemic status of (some) intuitions about whether groups can have conscious states.

[27] Cf. Chalmers, 2002 for a review of the gulf between consciousness and physical objects.

consciousness and physical objects. For instance, when we think about a brain as a massive collection of neurons that has various chemical and physical characteristics, it is not at all intuitive that this mass has consciousness. Something similar can, of course, be said for other bodily organs. Even after we are told that the brain is the part of the body responsible for consciousness, this does not render it *intuitive* that the brain is where conscious experience occurs. We suggest that part of the reason for this is that when we consider brains as hunks of physical stuff, we are considering descriptions that exclude the sort of cues that tend to activate the low-level processes that generate the intuitive sense that an entity is conscious.[28] Hence, it's not surprising if we find some initial resistance to the idea that the physical brain is conscious.[29] In this light, it's somewhat ironic that, while people have difficulty thinking of the brain as conscious, they have no trouble at all thinking that *ants* are conscious. On the contrary, our experiments indicate that people have trouble thinking that ants are *not* conscious.

Adam Arico†, Brian Fiala†, Robert F. Goldberg‡, and Shaun Nichols†

†Department of Philosophy

University of Arizona

‡ Department of Brain and Cognitive Sciences

Massachusetts Institute of Technology

Word Count: 10,737

---

[28] Though, as noted above, there might be other pathways that generate the judgment that an entity has conscious states.

[29] For a more detailed treatment of dualist intuitions, especially 'explanatory gap' intuitions, see Fiala, et al. (in press).

References:

Abell, F., Happé, F. and Frith, U. 2000: Do triangles play tricks? attribution of mental states to animated shapes in normal and abnormal development. *Journal of Cognitive Development*, 15, 1-20.

Arico, A.J. 2010: Folk psychology, consciousness, and context effects. *Review of Philosophy and Psychology*, 1(3), 371-393.

Block, N. 1978: Troubles with functionalism. In W. Savage (ed.), *Perception and Cognition: Issues in the Foundations of Psychology*. Minneapolis: University of Minnesota Press.

Bloom, P. 2004: *Descartes' Baby*. New York, NY: Basic Books.

Bowler, D. and Thommen, E. 2000: Attribution of mechanical and social causality to animated displays by children with autism. *Autism*, 4, 147–71.

Carruthers, P. 2004: Suffering without subjectivity. *Philosophical Studies*, 121, 99-125.

Chaiken, S. and Trope, Y. (eds). 1999: *Dual Process Theories in Social Psychology*. New York, NY: Guilford.

Chisholm, R. 1977: *Theory of Knowledge.* 2nd Edition. Englewood Cliffs: Prentice-Hall.

Cohen, S. 1984: Justification and truth. *Philosophical Studies* 46, 279-295.

Dasser, V., Ulbaek, I. and Premack, D. 1989: The perception of intention. *Science,* 243, 365-7.

Descartes, R. 1641/1986: *Meditations on First Philosophy*. Translated by John Cottingham. Cambridge: Cambridge University Press.

Erickson, J. and Keil, F. 2008: The role of intention in teleological explanations. Poster at 2008 meeting of Society for Philosophy and Psychology.

Fiala, B., Arico, A. and Nichols, S. in press: On the psychological origins of dualism: Dual-process cognition and the explanatory gap. In E. Slingerland & M. Collard (eds.)

Creating Consilience: Issues and Case Studies in the Integration of the Sciences and Humanities. Oxford University Press.

Fodor, J. 1983: *The Modularity of Mind*. Cambridge, MA: MIT Press.

Goldberg, R. and Thompson-Schill, S. 2009: Developmental 'roots' in mature biological knowledge. *Psychological Science, 20,* 480-487.

Goldman, A. 1979: What is justified belief? In G. Pappas (ed) *Justification and Knowledge*. Cambridge, MA: MIT Press.

Goldman A. 2006: *Simulating Minds*. Oxford: Oxford University Press.

Gopnik, A. and Meltzoff, A. 1997: *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press.

Gray, H., Gray, K. and Wegner, D. 2007: Dimensions of mind perception. *Science*, 315, 619.

Harris, P. 1989: *Children and Emotion*. Oxford: Blackwell.

Heider, F. and Simmel, M. 1944: An experimental study of apparent behavior. *American Journal of Psychology*, 57, 243-259.

Huebner, B. 2010: Commonsense concepts of phenomenal consciousness: Does anyone care about functional zombies? *Phenomenology and the Cognitive Sciences*, 9(1), 133-155.

Huebner, B., Bruno, M., and Sarkissian, H. 2010: What does the nation of China think about phenomenal states? *Review of Philosophy and Psychology,* 1(2), 225-243.

Hyslop, A. 2005: Other minds. *The Stanford Encyclopedia of Philosophy (Winter 2005 Edition)*, ed. E. Zalta. From http://plato.stanford.edu/archives/win2005/entries/other-minds/.

Inagaki, K. and Hatano, G. 1987: Young children's spontaneous personification as analogy. *Child Development*, 58, 1013-1020.

Inagaki, K. and Hatano, G. 1991: Constrained person analogy in young children's biological inference. *Cognitive development*, 6, 219-31.

Johnson, S. 2005: Reasoning about intentionality in preverbal infants. In P. Carruthers, S. Laurence, and S. Stich (eds.), *The Innate Mind: Structure and Content.* Oxford University Press, 254-271.

Johnson, S., Booth, A. and O'Hearn, K. 2001: Inferring the goals of non-human agents. *Cognitive Development*, 16, 637–656.

Johnson, S., Slaughter, V. and Carey, S. 1998: Whose gaze will infants follow? features that elicit gaze-following in 12-month-olds. *Developmental Science*, 1, 233-238.

Kahneman, D. 2003: A perspective on judgment and choice: mapping bounded rationality. *American Psychologist*, 58, 697–720.

Knobe, J. and Prinz, J. 2008: Intuitions about consciousness: experimental studies. *Phenomenology and the Cognitive Sciences*, 7(1), 67-83.

Korsgaard, C. 2008: *The Constitution of Agency: Essays on Practical Reason and Moral Psychology.* Oxford University Press.

Mascalzoni, E., Regolin, L. and Vallortigara, G. 2010: Innate sensitivity for self-propelled causal agency in newly hatched chicks. PNAS, 107, 4483-4485.

McCloskey, M. and Glucksberg, S. 1979: Decision processes in verifying category membership statements: Implications for models of semantic memory. *Cognitive Psychology*, 11, 1-37.

Melnyk, A. 1994: Inference to the best explanation and other minds. *Australasian Journal of Philosophy*, 72(4), 221-241.

Mill, J. 1865: *An Examination of Sir William Hamilton's Philosophy*. London: Longmans.

Nichols, S. and Stich, S. 2003: *Mindreading.* Oxford: Oxford University Press.

Opfer, J. E. 2002: Identifying living and sentient kinds from dynamic information: The case of

goal-directed versus aimless autonomous movement in conceptual change. *Cognition*, 86, 97.

Opfer, J. E., and Siegler, R. S. 2004: Revisiting preschoolers' living things concept: A microgenetic analysis of conceptual change in basic biology. *Cognitive Psychology*, 49, 301-332.

Perner, J. 1991: *Understanding the Representational Mind*. Cambridge, MA: Bradford Books/MIT Press.

Reid, T. 1969/1785: *Essays on the Intellectual Powers of Man*. Cambridge, MA: MIT Press.

Robbins, P. and Jack, A. 2006: The phenomenal stance. *Philosophical Studies*,127, 1, 59-85.

Scholl, B. and Tremoulet, P. 2000: Perceptual causality and animacy. *Trends in Cognitive Sciences*, 4, 299-309.

Searle, J. 1980: Minds, brains, and programs. *Behavioral and Brain Sciences*, 3, 417-424.

Shimizu, Y. and Johnson, S. 2004: Infants' attribution of a goal to a morphologically unfamiliar agent. *Developmental Science*, 7, 425-30.

Sloman, S. 1996: The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3-22.

Smith, E.E., Shoben, E.J., and Rips, L.J. 1974: Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, 81, 214-241.

Stanovich, K. and West, R. 2000: Individual differences in reasoning: Implications for the rationality debate. *Behavioural and Brain Sciences*, 23, 645–726.

Sytsma, J and Machery, E. 2009: How to study folk intuitions about consciousness. *Philosophical Psychology*, 22(1), 21-35.

Sytsma, J. and Machery, E. 2010: Two conceptions of subjective experience. *Philosophical*

Studies, 151(2), 299-327.

Wellman, H. 1990: *The Child's Theory of Mind*. Cambridge, MA: Bradford Books/MIT Press.

Woodward, A. 1998: Infants selectively encode the goal object of an actor's reach. *Cognition*, 69, 1-34.