A Transition Model for C cognitions About Agency

Daniel T. Levin
Dept of Psychology and Human Development
Vanderbilt University
Nashville, TN
daniel.t.levin@vanderbilt.edu

Megan M. Saylor
Dept of Psychology and Human Development
Vanderbilt University
Nashville, TN
m.saylor@vanderbilt.edu

Abstract—Recent research in a range of fields has explored people's concepts about agency, and this issue is clearly important for understanding the conceptual basis of human-robot interaction. This research takes a wide range of approaches, but no systematic model of reasoning about agency has combined the concepts and processes involved agency-reasoning comprehensively enough to support research exploring issues such as conceptual change in reasoning about agents, and the interaction between concepts about agents and visual attention. Our goal in this paper is to develop a transition model of reasoning about agency that achieves three important goals. First, we aim to specify the different kinds of knowledge that is likely to be accessed when people reason about agents. Second, we specify the circumstances under which these different kinds of knowledge might be accessed and be changed. Finally, we discuss how this knowledge might affect basic psychological processes of attention and memory. Our approach will be to first describe the transition model, then to discuss how it might be applied in two specific domains: computer interfaces that allow a single operator to track multiple robots, and a teachable agent system currently in use assisting primary and middle school students in learning natural science concepts.

Index Terms—HRI, Concepts, Theory of Mind.

I. INTRODUCTION

As people interact with an increasing variety of intelligent technologies, they are faced with machines that are, in some ways, very similar to people, and in some ways, very different. This makes it important to understand not only how people construe the internal processes inherent to artificial minds, but also how these concepts affect perceptions of, and interactions with, a wide variety of agents. In this paper, we review previous research exploring concepts about agency, and then describe a model of agent-concepts specially adapted to understanding both their structure and their deployment. To understand the need for this type of model, it is useful to imagine an encounter with an intelligent agent of the near future. Imagine that you are walking through the airport during allergy season need to do something about your sniffles and watery eyes. You remember you have some allergy pills in your backpack, so you open it, rummage through it, and have just pulled out a foil pack of allergy pills when a robot rolls up, and asks "Can I help you?" … With what? How does one even begin to understand this interaction? Did the robot come up to you because it is programmed to approach anyone who pauses or because it thinks you are threat, or even a drug dealer? Will it help to show the robot your pills and explain the situation? What if the robot suddenly says "What are those round, white things?" Do you assume its globally stupid, or just that it cannot identify the pills, but could still understand your real intentions. What if the robot suddenly pivots and drives off? Do you assume its seen something else more important, and look ahead to see where its going, or do you assume its just responding to some rote command to move on if it doesn’t get a clear answer? Maybe it's going to get some other robots for backup because something you said made it "worried"…

In this paper we review existing research exploring people's concepts about agency, propose a framework for understanding these concepts, and then discuss an HRI application for our framework. We argue that a new framework is necessary because the default model, proposed in different forms by different researchers, is not complete. In its simple form, the default model assumes that people readily and broadly apply a human model of agency to a wide range of entities as soon as they detect basic signs of anthropomorphism, and only restrict their attributions of agency with further thought. We describe this model and its basis, then describe new research that questions this model. This newer research demonstrates that in many cases people’s first impulse is to strongly differentiate the agency of humans and nonhumans, and only begin to equate them with additional consideration - exactly the opposite of the default model. We argue that a new model of agency concepts should not only incorporate both of these patterns of results, but also explicitly describe factors that can explain why people might shift from their initial impulses to a deeper consideration of the processes that drive a given agent. A key reason to incorporate an explanation of these shifts is that they probably drive conceptual change, and therefore provide a principled explanation for how people might change their thinking about agents as they evidence about their capabilities.

II. A DEVELOPMENTAL BASIS FOR AGENT-CONCEPTS

In trying to understand how people think about different kinds of agents, a good starting point is to consider research
exploring how people think about human cognition. Much of this research has been explored how children learn to reason about the causes of people’s actions. This learning has its foundations in social interactions that take place during the child’s first year of life. For example, a number of studies have demonstrated that during their first year of life, children learn to follow the gaze of a conversation partner. This skill will help them determine the target of others’ attention, and ultimately will help children understand the thoughts that lead people to look at things. ([6][15][22][7]).

As children develop these initial understandings of “mental events” (e.g. thinking), they learn to limit their attributions about the range of things that are capable of thought. These limits probably find support in categorical distinctions children make between living and nonliving things in their first year of life. A key example of research demonstrating these limits comes from Woodward [29]. In these experiments, infants were repeatedly shown a reach to one of two objects (a toy bear, and a ball) sitting next to each other on a stage. For some of the infants, the reach was performed by a human hand, and for some it was performed by a mechanical-looking stick. Then, the positions of the objects were swapped, and the reach went either to the same object in the new position (if the reach was initially directed to the bear on the left, it was now directed to the bear sitting in the new position on the right), or to the new object in the old location (the initial reaches to the bear on the left would now be reaches to the ball on the left). Woodward hypothesized that if infants have some preliminary understanding that reaches by a human agent are goal directed, they would look longer when the human reached to the new object at the old location because it signified interest in a new object. In contrast, if they limit this attribution of interest, they should not show this preference for the mechanical reach. Consistent with these predictions, 9 month-olds looked more when the human hand changed to a new object, but not the machine did so, and even 5-month-olds showed a marginal effect in this direction.

Woodward’s results are important because they demonstrate not only that these young infants understand something akin to the goals that drive behavior, but that they appropriately limit them to human agents. This early understanding about goal-directed behavior is later elaborated into the full-fledged understanding of how thinking causes behavior referred to as “theory of mind” (TOM). TOM refers to the collection of knowledge and skills required to understand how people’s beliefs, desires, and goals underlie behavior. A key idea underlying TOM is that these beliefs, desires and goals are kinds of intentional representation that reflect both reflects a level of analysis for human behavior [8] and a specific property of mental representations that affords a close connection between mental symbols and their referents – this reflects the idea that humans have a semantic system that allows them to truly know what a symbol, such as a word, refers to in the real world, while, according to many philosophers, computers do not [28].

A key insight children need to achieve in order to develop an effective TOM is that intentional representations are predicated upon one’s opportunities to observe the world – one can be in the company of a cookie, and be hungry, but it may not generate a belief if the cookie is hidden from view. Understanding that even incorrect beliefs can drive behavior is an important constituent of TOM, but unlike more basic understandings of goals, it arguably does not develop completely until children are about 4 years old. The most well-known marker of this understanding is the false belief task [30]. In this task, 3- and 4-year old children are first introduced to an agent (usually a puppet). Then, both the child and the puppet witness an object being hidden in one of two locations. The puppet then leaves the room, and the object is switched to the new location. When the puppet returns, the child is asked where the puppet will look for the object. The younger children usually mistakenly predict that the puppet will look in the new location, failing to account for the fact that the puppet has a belief that does not comport with the current state of the world. In contrast, the older children can appreciate this fact, and predict that the puppet will look in the old location where he falsely believes it to be located. The combination of earlier-developing gaze following and simple understandings of goals, and later-developing ability to track false beliefs, has led to several models of TOM that make an important distinction between relatively automatic processes that track basic cues about intentionality (e.g. gaze), and true beliefs, and more resource-intensive processes that track false beliefs [1][21]. So, in many situations, people’s beliefs about the world are correct, and it is possible to engage them without much extra thought. In other situations, a more sophisticated form of TOM is engaged that can effectively track beliefs that diverge from the state of the world.

III. IMPLICATIONS OF EARLY-DEVELOPING TOM FOR ADULT UNDERSTANDINGS OF AGENTS

One interesting fact about research exploring TOM is the majority of it has been on young children. Recently, however, this has been changing, as researchers begin to ask how adults deploy their (arguably) fully developed TOM. Particularly important for present purposes is the range of agents to which TOM is applied. Based on the developmental work, one would think that adults have an especially firm basis to limit generalizing the attribution of goals to humans, as this would seem to be at the core of the early-developing, automatic TOM-subprocess. However, recent research demonstrates that adults sometimes make a variety of anthropomorphic attributions of clearly mechanical agents. Much of these findings come from the field of social cognition, where researchers are asking what makes for a social agent? For example, work by Nass and colleagues demonstrates that adults will apply social norms to computers they are interacting with [24]. More recent research has explored the degree to which adults attribute some form of agency to robots. Attributes of agency to robots were affected by simple variations of synchrony between the robot’s dancing movements and the rhythm of music [36], and robots that
mediate conversations using human derived gaze patterns induced increased rapport [35]. In addition, researchers in social cognition have demonstrated that adults often rely on simple heuristics instead of fully deploying their TOM skills. For example, adults often fail to inform their partner of important facts that the partner plainly was not aware of [3].

Findings such as these, led Epley and colleagues to develop a psychological model of anthropomorphism to account for the range of circumstances under which people are likely to anthropomorphize an agent [10]. The model assumes that people have ready access to a self-model that is effectively applied automatically when interacting with others who share a common framework of understanding. Thus, it is often safe to assume that a conversation partner not only can see the same things as you see, and knows many of the same things you know, but also thinks in fundamentally the same way as you do. However, sometimes this is clearly not true so one must correct the bias to apply a basic self-model. The key to the Epley model is that this correction requires mental effort, which will not be applied without some motivation.

To provide an explanation for this correction, Epley's model specifies three broad factors that can affect motivation to either correct an initial bias to anthropomorphize, or prevent correction. The first factor is "elicited agent knowledge". This refers to the degree to which knowledge about other forms of nonhuman agency is present. However, having this knowledge is not sufficient if it remains unaccessed so Epley et al hypothesize that a range of known psychological phenomena may affect access to alternative models. For examples, they suggest that individuals high in "Need for cognition will more likely access this knowledge in any given situation because these individuals like to think deeply about things. The second factor, "Effectance motivation" represents the degree to which an individual perceives the need to be successful in some task such as interaction with an agent. If someone believes that they are about to actually interact with an ambiguous agent, they will be less likely to anthropomorphize it. Finally, high "sociality motivation" results from social isolation or loneliness, and increases anthropomorphization.

A similar approach to agent-reasoning can be seen in recent work connecting concepts about TOM and agency with religious concepts. For example, Barrett and Leman review data supporting the theory that religious belief can, in part, be attributed to a heightened sensitivity to detect agency based on simple perceptual cues [4]. On this view a cognitive/perceptual subsystem they refer to as the "hyperactive agency detection device" searches out minimal signs of agency (such as self movement), and produces the percept of an agent even when other information suggests that the stimulus is not an agent. Similar to Epley, Barrett argues that the initial beliefs caused by HADD is sometimes overridden by subsequent more explicit processes.

Both of these models make an excellent start at a broad psychological account of agency attributions. However, one limit to the basic distinction between an initial TOM-generalizing/HADD process, and a later correction process is that neither naturally account for at least some of the knowledge that children and adults clearly have. As reviewed above, even infants clearly distinguish between living and nonliving things, and between thinking and nonthinking things, and if we assume that these lay at the core of their conceptual apparatus, one would think that they should be readily available, even for automatic nondeliberative processes. Even Nass and Moon's very basic observation that adults explicitly insist that computers are different from humans, and are undeserving of the same social niceties as humans can serve as evidence that something in addition overgeneralization is occurring even as part of adults' initial judgments about agency. Thus, we believe that adults may have a better understanding of the differences between intelligent agents than these models give them credit for. Of course, it is possible to dismiss these verbal reports as mere spandrels of thought that really only cover for the deeper, and more important impulse for more automatic systems to detect agency broadly. We would argue that this view is shortsighted, and that it stems from an overly literal interpretation of the view that only modular systems are an appropriate target for psychological investigation [11]. Not only are these beliefs likely to be important as inputs to a wide range of cognitive problem-solving processes, but they may also provide an important guide for perceptual selection. More important, we review evidence below that explicit reports distinguishing agents can satisfy many key criteria necessary to establish that any measured phenomenon is scientifically useful. As we will review below, these reports reveal replicable phenomena, they can be validated against interesting external variables, and they seem to predict interesting phenomena.

IV. MEASURING AGENT CONCEPTS

To further explore the basis for adults' apparently strong differentiation between human agents and mechanical agents such as computers and robots, we developed a measure of agency attribution that tests specific predictions people make about the actions and thought processes of different agents [18]. This project was particularly mindful of the idea that there ought to be continuity between distinctions infants make between living and nonliving agents and distinctions made by adults. In our experiments, adults made predictions in a series of scenarios, some of which were modeled on Woodward's infant paradigm. In the scenario most similar to Woodward's, participants were shown a pair of objects (a small toy duck and a truck) sitting next to each other on a table and asked to imagine that each of three entities (a computer, a person, or a robot) had "picked up" the duck, while ignoring the truck. Then, participants were shown a new image with the locations of the duck and truck switched, and were asked what each entity would do on a third trial. Following Woodward's logic, if participants are treating the behavior in a goal-directed fashion, they should predict that the entity will retain the same
goal on the third trial, and pick up the duck again, even though it was in the location that had formerly been occupied by the truck. On the other hand, if the entity was acting in a more rote (and possibly location-oriented) fashion, participants should predict that it would reach again to the old location, despite the presence of the new object.

In addition to goal-oriented action, we tested attributions about agents’ ability to organize the world using knowledge-based categories, vs. feature-based categories. This scenario is based on the idea that artifact categories are inherently intentional in that the commonalities binding these categories reflect the goals of the humans who created the objects [5]. For example, a “tool” is a tool because all members of the category fulfill a common set of human goals. Accordingly, if participants believe that an agent uses these knowledge-based categories, they are concretely attributing an important form of intentional understanding to the agent. In contrast, if participants do not attribute intentionality to an agent they might assume that the agent would group objects that share simple perceptual features (Figure 1).

![Knowledge-based categories: Office Supplies vs. Food](image1)

![Feature-Based Categories: Large dark vs small colorful objects](image2)

Figure 1. Choice examples for attributions about the kind of categories an entity might use.

Results of several experiments consistently demonstrated that adults do, indeed, strongly distinguish computers and people by making far fewer intentional responses for computers than for humans [18]. More important, participant predict that robots will be just as nonintentional as computers, even when the robot is given a human name (“OSCAR”), is describing it as having goals, and is shown in a walking in a video. In two additional experiments, however, we asked participants to track a robot’s focus of attention as it was shown pairs of objects. The robot was shown (in a video) looking to one of the objects, and then the other, and participants were told that they would need to remember which of the two the robot “preferred”. After tracking the robot's looking for 10 object pairs, participants made more intentional predictions for the robot than the computer.

It is important to note that participants’ nonintentional responses for computers and robots were not simply the result of their belief that current computers are inherently unintelligent. In all of these experiments, participants rated how intelligent they thought computers were and how effectively computers can infer human goals based on visual information. Participants who believed that computers are good at understanding human goals showed less of a contrast between humans and computers in behavioral predictions, while there was no such relationship for beliefs about overall intelligence. We have also validated our prediction measure by showing how it is linked with success in using computer-aided learning systems. We have shown that junior high school students who make relatively intentional behavioral predictions for people learn more effectively from an agent-based on-line tutoring system (the “Betty’s Brain” system) [34].

In this context, a key question is whether people's strong differentiation of human and machine thought is characteristic of a secondary problem-solving process that “corrects” an initial impulse to equate human and superficially intentional nonhuman agency. This would be consistent with both the Epley model and the hyper-active agency detection device. On the other hand, it is possible that people do have readily available concepts that differentiate human and machine thought, and that they sometimes apply this as an initial judgment. To test for this possibility, we timed participants as they made their predictions, and found the strongest contrasts between human and non-human predictions in those who decided the most quickly. In a second experiment, we confirmed this effect by asking some participants to decide quickly, and by asking others to think deeply about their decisions. Again, those who decided quickly differentiated human and mechanical predictions, while those who thought more did not [17]. These findings are precisely the opposite of predictions made by “equate-then-correct” models of agent concepts. In the next section, we argue that both of these patterns are characteristic of important processes, and propose a model broad enough to encompass this diversity.

V. THE TRANSITION MODEL OF AGENCY

Based on the empirical findings described above, we are developing the transition model of agency concepts. Our goal in creating this model is not only to understand the empirical patterns described above, but also to develop a useful framework for understanding processes that are central in HRI. The most important of these is conceptual change. How does experience with novel agents such as robots affect people’s concepts, both about the agents themselves, and about agency more generally?
The transition model hypothesizes two kinds of agent-conceptual reasoning: First-line default reasoning, and second-line conceptualizing (see Figure 2), each of which represents a diverse collection of implicit and explicit processes. Therefore, what unifies first-line and second-line processing is not processing style—there is no neat implicit/explicit or automatic/controlled divide here. Rather, the model assumes that it is important to understand how a range of processes might be initially applied to an agent, and then how different processes might be applied when revisions to initial conclusions are deemed necessary. A key element of the model is that it places strong emphasis on accounting for factors that are responsible for shifts from first-line defaults to second line reasoning, and for situations where second-line reasoning can affect first-line reasoning.

**First-line Defaults.** First-line defaults include four key processes. The most basic of these include the act of perceptually detecting agents, mid-level classifications of entities as intentional or nonintentional, and a few basic visual-cognitive entailments of that classification. Core conceptual systems studied by developmentalists that distinguish living and nonliving things are clearly part of this basic process, but it is also important in adults. For example, recent work exploring adults’ perceptual detection of inter-agent relationships such as chasing, would be part of this subprocess [12].

Another key set of first-line processes support understandings of agents as they participate in basic goal-driven actions and events. Basic agency and reference supports detection of self-initiated behavior, movement patterns, and anthropomorphic physical features such as faces and eyes, as well as a basic understanding of the presence of true (e.g. correct) beliefs [2][21] that are probably automatic [1]. We suspect that this process will be drawn upon heavily both when people track simple active agents, and engage in relatively automatic social cognitions when dealing with other people. Under this model, participants' relatively automatic application of social norms to computers may reflect the operation of these basic agency processes.

First-line knowledge activation is meant to encompass the initial cognitive responses to agents. This includes basic heuristics about the kind of thinking an agent typically engages in (such as the inferences measured by our behavioral prediction scenarios), and the results of default problem-solving strategies. In addition, local givens constitute details and contingencies about the current setting that enter into behavioral predictions (for example, stimuli in the agent’s environment, and situational constraints on the agent in the current environment). These are particularly important when participants reason about technology, which is likely to be associated with application-specific functional constraints [20].

**Second-line conceptualization.** According to the transition model, the first-line default characterizes most everyday cognition about entities in much the same way that basic intentionality is assumed to represent the most common, and usually-correct entailments of a TOM reasoning. However, these basic entailments are not always sufficient to reason accurately about entities. The TOM literature makes this point by emphasizing the cognitive control inherent to correcting for untrue beliefs [21], and the present context includes this need, but goes beyond it because we also want to understand what happens if initial classifications of entities are incorrect or must be revised. Accordingly, the transition model hypothesizes a broad set of processes and knowledge, which we will refer to as second-line conceptualization. These processes may involve explicit intentional reasoning (similar to selection processing in the Leslie et al., TOM model), and appeals to and potentially modifications of broad concepts about the entailments of agent categories. Second-line reasoning could be considered the cognitive “heavy-lifting” necessary when expectations are disconfirmed, when first-line defaults run out, and especially when basic concepts and classifications might require revision.

**The Transition Model for Agency Cognition**

Figure 2. The transition model.

**Transitions: Cognitive dissonance and deep-concept availability.** This initial sketch of the two parts of the transition model makes clear the need to explain when and how individuals switch between reasoning modes. These switches are important to understand, especially for reasoning about novel agents such as robots, because switches to second-line conceptualization determine whether an individual simply uses existing categories and defaults, or considers the entity more deeply. This second-line processing has the potential to change people’s understanding about either a specific entity, or about entire entity categories. For example, in findings described above, we appear to have induced second-line processes, and conceptual change when we asked participants to repeatedly attend to a robot looking at objects. The question is, what caused this? We hypothesize that participants attributed more agency to robots because participants they focused on the robot “looking at” objects and having preferences for them, which may have built up cognitive dissonance with a simple mechanical classification of the robot. The idea of dissonance, which we define as discomfort or cognitive conflict caused by the detection of mismatches among beliefs and observations, comes from several sources. These include research exploring conceptual change that documents a learner’s dissatisfaction with current concepts emphasized in the conceptual change literature (for example, [27]), the idea that students will rely on relatively shallow analyses until they experience "cognitive
disequilibrium” which leads them to deeper analysis [13], the “theory-theory” explanation of TOM [40], and long standing theory in social cognition, in that it reflects some amount of unease at violations of basic understandings [9]. However, our notion of dissonance lacks the threat to self image that characterizes the social-cognitive dissonance theory. This idea resonates with the concept of the “uncanny valley” [23], which is the observation that people are uneasy with robots that are very similar to, but nonetheless discriminable from humans.

In addition to dissonance, we hypothesize that the current availability of the second-line concepts facilitate transitions. This idea is based both on Epley et al’s idea of elicited agent knowledge and on developmental research demonstrating that inductions based on the broad living-nonliving distinction (which we would assume to be a characteristic second-line concept) require that children be reminded of the status of specific things as living or nonliving [14]. The effectiveness of this simple reminder not only suggests the importance of activating second-line concepts, but also suggests that one plausible developmental achievement would be to obtain the metacognitive skills necessary to self-generate these reminders when the situation calls for second-line reasoning.

We have collected data in support of both of these predictions. To do this, we developed a simple 6-item response scale to measure dissonance. It includes items such as “Sometimes I was uncomfortable answering these questions”, and “Some of the answers I gave in this experiment were inconsistent with my previous beliefs about the subject”. In an initial experiment, we had one group of participants complete the dissonance scale after having completed our behavioral prediction measure of agency attributions, and another group of participants completed the same measures after completing a deep-concept availability manipulation in which they were asked to describe ways in which computer and human intelligence might be similar or different, and to list properties of living and nonliving things [37]. Therefore, just as children can use reminders to increase the availability of knowledge about differences between living and nonliving things, the participants in this experiment self-generated reminders of their understanding about this basic categorical contrast. We found that the deep-concept availability manipulation produced a correlation between dissonance and agency attributed to computers such that increased dissonance was associated with high assignment of intentionality to computers. In contrast, for participants who did not receive the deep-concept manipulation, there was no link between dissonance and attributions of intentionality to computers. In another experiment we found that asking participants to watch a video that heavily anthropomorphized robots induced more dissonance than a similar video that did not anthropomorphize the robot. Similar to the first experiment, participants in the anthropomorphic condition again exhibited a positive correlation between dissonance and attributions of intentionality to computers, while there was no such correlation in participants who got the standard mechanical description.

We have also explored the links between dissonance and agency attributions in more realistic HRI settings [38]. In this experiment, participants completed a realistic triage scenario with either a human partner or a robot partner in which the Human-Human or Human-Robot team was tasked with assessing the state of a series of simulated disaster victims (Figure 3). In the Human-Robot condition, participants interacted with a Pioneer 3-DX which guided participants through the steps necessary to triage the victim. In the comparable human-human interaction, the participant worked with the same simulated victims, but was guided by a human helper who communicated with the participant via a radio. The interactions were scripted so as to make the contents of the interaction as comparable as possible. Similar to the video-based experiment described above, we found that the human-robot condition both produced more dissonance overall, and produced another dissonance - concepts link such that increased dissonance was again associated with increased attributions of intentionality to computers. It is, of course, important to consider why interacting with a robot changed influenced concepts about computers. We believe that this occurred both because participants often see close links between the two, and because the robots we asked about in the agency-attribute questionnaires were the specific ones used in the experiments whereas the computers were generic. Based on social-cognitive research suggesting that heuristics have a stronger impact on reasoning when the target broadly defined because broad categories afford the opportunity to focus attention on exemplars that might match the new concept [39], we suspect it was easier for participants to adjust their concepts about the broad category of computers rather than the specific robot from the experiment. In any case, these experiments have repeatedly demonstrated a clear link between dissonance and agent-concepts, and have further demonstrated that deep concept activation can facilitate this link.

![Figure 3. Four of the mannequins in the triage environment](image)

**VI. APPLYING THE TRANSITION MODEL**

We believe that an effective understanding of concepts about agency can guide research exploring a wide variety of agents, especially in an HRI context. Therefore, to give a sense of how the transition model might be of use in HRI, we would like to briefly discuss how recent research exploring computer interfaces for unmanned vehicle (UV) operators might benefit from applying the transition model.
A key component of any human-computer interaction is situation awareness, and when interfaces must be designed to allow users to monitor multiple robots, concepts about agency become crucial in determining how users construe the situations presented to them. Endsley [42][43], defines three levels of situation awareness. Level one is the perception of the relevant events in the environment. This perception occurs based upon the sensory feedback from the remote system via auditory, visual, and tactile sensors. Level two is the comprehension of the perceived information as it relates to the current tasks or goals. Level three requires the human to predict what will happen in the near future in order to plan appropriate actions and responses.

A multilevel model of situation awareness such as this makes clear the diversity of cognitions necessary to monitor multiple UVs. The deployment of robotic teams into remote environments that do not permit direct line of sight supervision is heavily reliant on the UV Specialist’s ability to understand the robot teams’ actions and ability to achieve the mission goals. Accordingly, the UV Specialist must not only understand broader mission goals, but must also understand how a more local set of “goals” nested within individual UV’s that may or may not have sophisticated decision making and perceptual capabilities. This likely requires a considerable amount of attentional switching not only between different display locations about also between different levels of analysis within locations.

We would argue that the transition model can be useful in situations like this for two basic reasons. The most simple is that the model provides a useful checklist for analyzing the range of first and second-line cognitions that may be relevant to the specialists’ understanding of the range of agents in the situation. For example, first-line processes may include both a simple anthropomorphism derived from social heuristics that guides the specialists time-pressure verbal responses, while also including a more differentiated consideration of the UV’s decision making capabilities that guide explicit level three predictions. According to the transition model both of these apparently contradictory responses might coexist and be readily available as first-line responses.

However, many of the more specific details of the transition model may provide guidance that extends beyond a basic check-list aid. Most important, the model may help explain the process whereby UV specialists change their approach to the agents, either as they transition from being novices to experts, or as more expert users encounter novel UV systems. In either case, the transition model specifies measurable processes that can predict conceptual change. One could easily measure cognitive dissonance in UV specialists, and use this to predict not only when they might be changing their ideas about the UV’s but also to detect situations where specialists are ignoring evidence that is crucial to understanding a system. Thus, the activation of second-line cognitions via measurable dissonance may moderate success in adapting to changing conditions, and can therefore serve as a diagnostic for UV specialists’ awareness of, and response to, the deep structure of their cognitive environment.

VII. CONCLUSIONS

Although the above application is clearly speculative, a key insight of the transition model is that a range of situations activate agency concepts, and each may serve as a stimulus for first-line processes. So, in some cases, concepts must be activated in the context of time-pressure social interaction, while in others the concepts are activated in a less pressured situation where more explicit problem-solving is required. We would argue that BOTH should be considered first-line processes because in both cases, people produce initial interactive behaviors, or develop initial hypotheses that can sometimes be changed with a second stage of deeper thought. Thus, this model is different from many psychological models of problem solving in that the two stages are defined not by the nature of the processes that characterize them (as the typical implicit/automatic -> explicit/controlled model might), but rather by the status a range of processes as initial defaults or secondary conceptualization in typical situations where the reasoning is applied. Thus, the transition model can serve as a link between more basic research on psychological processes, and more applied research exploring how these processes might be deployed in specific situations. In addition, the central emphasis on factors producing transitions between modes of reasoning makes the transition model uniquely suited to studying the dynamics of conceptual change in concepts about agency, clearly an important topic as computers start turning into robots that look like people, act like then, and even think like them.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant No. 0826701 to DTL, MMS, JAA and GB.

REFERENCES

of multiple robots in the RoboFlag simulation environment.


