`

Conceptual Change in Beliefs About Agency: The Joint Roles of Deep Concept Availability and Cognitive Dissonance

Daniel T. Levin[a*], Alexander P. Steger[a], Simon D. Lynn[a], Julie A. Adams[b], and Megan M. Saylor[a]

[a] Department of Psychology and Human Development
Vanderbilt University
Peabody #552
230 Appleton Place
Nashville, TN 37203

[b]Department of Electrical Engineering and Computer Science
Vanderbilt University
359 Jacobs Hall
VU Box 351824 Sta B
Vanderbilt University
Nashville, TN 37235-1824

[*]Corresponding Author
daniel.t.levin@vanderbilt.edu
tel: (615) 322-1518
fax: (615) 343-9494

***Abstract*—We have previously demonstrated that participants strongly distinguish the relative intentionality of computers and humans when making simple behavioral predictions about these entities. In understanding this phenomenon, we have been exploring the degree to which these initial intuitions might change in response to experience or short-term changes in the cognitive framework participants apply to situations in which they consider the intentionality of different agents. We hypothesize that two key events increase the likelihood of this kind of cognitive change. First, change is more likely if deep concepts about the nature of agents are activated, and second, change is still more likely if this activation produces cognitive conflict, or cognitive dissonance. To test this hypothesis, we developed a measure of cognitive dissonance, and in two experiments tested the degree to which conditions that induce deep concept activation and cognitive dissonance produce changes in the behavioral predictions made for different agents. In both experiments, high levels of dissonance predicted relatively intentional behavioral predictions for computers, but only under conditions that activate deep concepts. We argue that conceptual change in beliefs about agency is often a joint product of the availability of basic knowledge-organizing concepts and reframing induced by cognitive conflict.**

Keywords: Human-robot interaction, theory of mind, concepts, cognition.

1 INTRODUCTION

A key element in understanding the knowledge people recruit when interacting with different artificial agents is the need to understand how this knowledge might change with experience. Consider, for example, the basic inference that human behavior is goal directed and intentional, while computers' behavior is not. If we see a person reaching toward a glass of water, we interpret their actions as being driven by a goal (they want the water), and would assume that they would stop reaching if the full glass was suddenly replaced by an empty one. In contrast, if we do not grant computers or robots the same kind of intelligence, then we might assume that these mechanical agents would continue their action no matter what was substituted for the glass. We have been arguing that this difference in attributed motivation is central to a wide range of inferences about agents, and interactions with them. However, perhaps the most important question to ask, especially in a rapidly evolving technological environment, is how these inferences might change with experience. Clearly, today's robots may not really have goals, but tomorrow's might. However, even if some future artificial agents can truly act on human-like goals, not all will (who wants a light switch with its own goals?). So, it becomes important to understand not only the range of agents that people ascribe intentionality to, but also how people will learn about new agents that may, or may not be

`

intentional.

In this paper, we first describe research exploring people's attributions of intentionality to different agents, and then briefly present a framework we are developing to understand the structure of these attributions, and how they might change. A key part of this framework is the hypothesis that the experience of cognitive conflict, or cognitive dissonance, is a key factor mediating experience-driven conceptual change in attributions of intentionality. In two experiments, we asked participants to reason about different agents (a computer, a robot, and a person) in circumstances that were, or were not, designed to induce cognitive dissonance. In two experiments, we asked participants to make a series of behavioral predictions for different agents. In previous research we have demonstrated that these predictions can serve as evidence that participants attribute intentional goal-directed thought to humans, but often do not extend this inference to robots and computers. In the first experiment, we tested whether activating deep concepts about agency would cause dissonance that might be linked with changes to these attributions.

1.1 Assessing concepts about agents

A number of studies have explored people's concepts about different agents, and many of these have explored a number of contextual factors that affect the degree to which people anthropomorphize agents (for review, see Epley et al., 2007). However, most of these studies have measured anthropomorphism either by simply asking participants to report whether different objects might have "goals" or to have a "mind" (see for example, Morwedge et al., 2007) or by assessing the degree to which people exhibit automatic social behaviors when interacting with different entities (Nass & Moon, 2000). These are both reasonable approaches, but in the former case, it is difficult to know what, exactly, people mean when they attribute a "goal" or a "mind" to something and in the latter, it is difficult to know whether they are invoking any particular concepts about agents when exhibiting nominally automatic social responses. Therefore, we have been exploring concepts about agents by asking participants to make simple behavioral predictions about different agents (Levin et al., in review; Levin et al., 2006) in situations that directly test the entailments of goal-directed intentional reasoning. For example, in one scenario, we present participants with a scene depicting a toy duck and a toy truck sitting in locations A-1 and C-3 on a simple grid (See figure 1). They are asked to imagine that an agent, such as a robot has acted upon the duck (for example, by reaching toward it). Then, participants are asked to imagine that the positions of the duck and truck are switched, and to predict what the robot would do next - would it reach again toward the duck, now in its new location, or would it reach toward the same location and reach to the truck that now occupies it? Following the logic of Woodward (1998), we argue that the answer depends on how participants' interpreted the initial actions. If the actions were interpreted as goal-directed intentional actions, then the reaches were indicative of a preference for the initial object, and so participants would predict that the robot would again reach to the duck, now in its new location. If, on the other hand, the agent's actions were not interpreted as reflecting goals, then one might expect that the action would again be directed toward the same old location, irrespective of the new object. In a similar scenario, we asked participants whether different agents would organize a set of objects by taxonomic category (for example by placing candies together and office supplies together) or by feature (by placing rectangular dark objects together and small colorful objects together). Again, drawing upon arguments that taxonomic classification (as opposed to feature-based classification) is central to intentional understanding (Bloom, 1997), we hypothesize that when participants predict that an agent will organize a group of things taxonomically, they are ascribing intentionality to the agent.

In a range of experiments, we have observed that participants do, indeed, make more intentional behavioral predictions for humans than for computers and robots. Not only is this difference strong, and consistent across experiments, but it has been validated in a number of ways. For example, the putative attributions of intentionality uncovered in the behavioral predictions for computers are significantly correlated with more general attributions of "goal understanding" to computers (even when controlling for attributions of intelligence; Levin et al. in review), and middle-school children's intentional behavioral predictions for people significantly predict their ability to learn from an agent-based teaching system (Hymel et al., in press). Finally, we have observed that adults initially default to making similarly non-intentional behavioral predictions for computers and robots, but begin to make more intentional predictions for robots when they repeatedly observe robots' preferentially looking at objects.

This latter finding is of particular importance because it reveals the role of experience in potentially changing the attributions people make about agents. On one view, these attributions of intentionality stem from relatively long-standing and broadly applied concepts underlying theory of mind (for review see Leslie et al., 2004). If we assume that these skills underlie everyday processing of social action, and that the requisite representational understandings are often invoked implicitly in the service of explicit reasoning, then two important predictions for cognitive change emerge. First, it is possible to hypothesize that changing these deep-seated concepts will require significant motivation - it does not make sense to repeatedly reconfigure the first principles of behavioral understanding unless there is good reason to do so. Second, it is possible that opportunities for conceptual change can be missed if the relevant concepts remain implicit.

Jointly, these two predictions suggest a model of conceptual change whereby change occurs when motivation meets with available concepts. In the social cognition and persuasion literatures, the key factor motivating attitude change is cognitive

`

dissonance. Cognitive dissonance is characterized as a feeling of discomfort that occurs when people detect inconsistencies between their beliefs and their behavior, or among different beliefs (Eliot and Devine, 1994). According to basic theory in persuasion and attitude change, people's desire to lessen this discomfort can lead them to change even firmly held attitudes. A similar idea from the learning literature is that students sometimes experience "cognitive disequilibrium" that causes them to transition from a shallow analysis to a deeper analysis of some problem (Graesser et al., 2005). Based on these ideas, we have developed a simple measure of cognitive dissonance for use in experiments testing concepts about agency with the hypothesis that increased dissonance should predict changes to basic concepts about the intentionality of artificial, and perhaps even human, agents.

Motivated cognition needs a target, and if the relevant concepts about agency remain implicit, conceptual change might not occur, or might occur on a very limited situation-specific basis. For example, a participant who witnesses a computer engage in apparently intentional behavior may simply attribute the behavior to some situation-specific programming instead of any more general intentional skill. Therefore, our second hypothesis is that conceptual change is more likely when participants are forced to activate relatively deep knowledge. This activation could arise in a variety of ways, ranging from a task setting that pushes participants away from default assumptions to direct manipulations that require participants to directly consider deep concepts and assumptions.

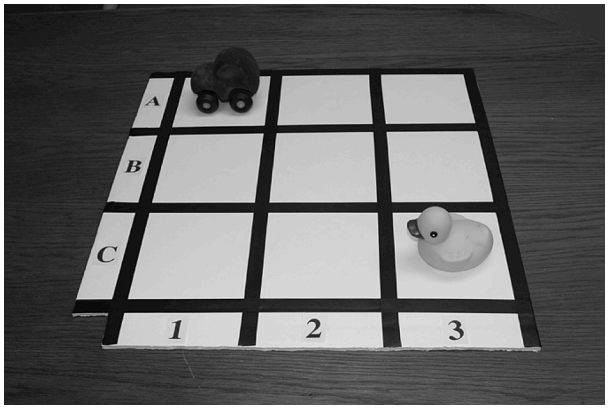

Fig. 1. Illustration accompanying a basic behavioral prediction scenario.

In Experiment 1, we tested our hypotheses about conceptual change by asking participants to view a context-setting video in which a researcher described a robot. After viewing this video, participants watched a series of events in which the described robot drove among objects, "looking at" a subset of the objects. After exposure to the robot, participants completed our behavioral prediction scenarios, and then completed our cognitive dissonance scale. There were two conditions. In the intentional condition, the context video described the robot in heavily intentional terms, giving it a name, and referring to its actions using intentional terms such as "goals". In the other condition, the mechanical condition, the robot was referred to in much less intentional terms. As reviewed above, in previous research we have observed that adults initially assume that robots are no more intentional than computers, but begin to diverge from this default with strong manipulations of the apparent intentionality of the robot. We predicted that the intentional condition would depart from participants' default tendency to classify robots as non-intentional and would therefore not only produce cognitive dissonance, but also dissonance-moderated conceptual change. That is, we predicted a positive correlation between dissonance and agency concepts, as measured by our behavioral prediction scenarios.

2 Experiment 1

2.1 Subjects

A total of 37 Vanderbilt University undergraduates completed Experiment 1 in exchange for course credit (mean age=18.8, 27 female). Nineteen completed the intentional condition, and eighteen completed the mechanical condition.

2.2 Materials

Introductory videos and a subsequent set of still images depicting robot actions were displayed full-screen on a 17" Apple eMac computer. Participants completed paper-based questionnaires after viewing the videos.

`

2.3 Procedure

2.3.1 Context video

After completing the informed consent, participants viewed a ninety-second documentary-style video in which an actor (posing as a researcher) described a robot's creation and training to perform certain tasks. In the intentional condition, the researcher described the robot using anthropomorphic terminology: the robot's name was OSCAR, he was described as learning to "explore his environment and look for specific things or things that don't belong", "learning about his world visually", and "developing his own goals and interests so that he can work with people as a partner". During the narration, a human trainer was shown presenting pairs of objects, while OSCAR inspected them using its video camera. The researcher underscored OSCAR's intentionality by explaining that, "the real key is that OSCAR decides what to look at without us telling him what to do--we're developing a sense of curiosity for him. He's developing his own way of looking at the world, and wanting to find out more about what he can see, and what he can't see. He's coming to understand that there are things in his environment that he doesn't know and doesn't see, and he's driven to find out more about those things."

In the mechanical condition, the researcher described an identical robot named V5, but using mechanical terminology: V5 was described as "being programmed to process information from its video cameras to analyze and identify objects," as having an onboard computer, and as "mov[ing] through its environment based on the data it extracts from its surroundings." In contrast to OSCAR, V5 was not described as having goals, or as "choosing" what to look at on its own. The narrator did not use anthropomorphic language and used "it" in place of "he". To emphasize the V5's lack of intentionality, the researcher narrated in technical detail how V5 analyzed its data; meanwhile a programmer was shown writing code and updating V5 as it was attached to a computer. For both videos, views of the researcher were alternated with views of a simple 4-wheeled robot (made from a Vex robotics kit, with a single video camera attached to a stalk in front of the robot) driving about and being worked with by a person.

The two conditions were replicated using a female actor and a different style robot (a Pioneer P3-DX wheeled robot base with a color video camera and laser range finder) while the narrated scripts remained the same. Participants were counterbalanced between the two alternate versions of the two conditions.

After watching the context video, participants were instructed to "describe the most important aspects of the robot from the introduction video" and were given five minutes to write.

2.3.2 Robot event sequences

Next, participants were shown a set of four different still image sequences, each consisting of a comic book style depiction of one of two different robots driving among a set of objects and stopping to look at some of them. Within each sequence, the robot drove to and "looked at" four of eight objects. For each of the four looked-at objects, the sequence included one still of the robot driving toward the object, and one with the robot stopped at the object. In the first sequence, sports equipment and containers were arranged on the ground, each item resting on one of two different colors of paper (see figure 2 for a sample still). The robot featured in the introduction video (either OSCAR or V5) drove among the objects to inspect them.

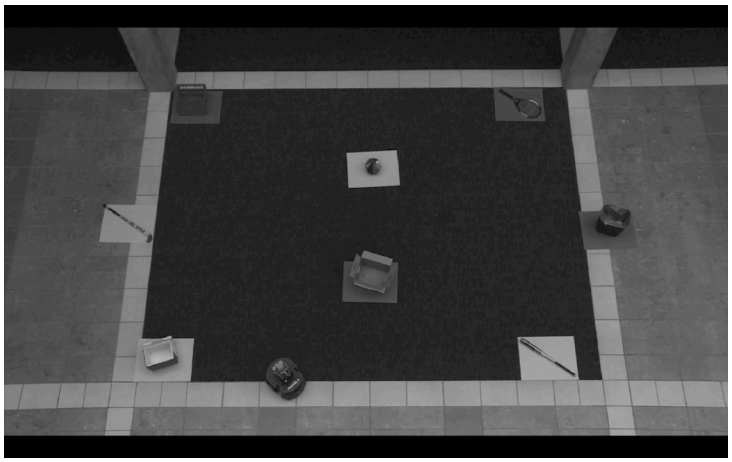

Fig. 2. Example image from robot event sequence.

In the second sequence, the robot from the alternate introduction video (not watched by the participant) moves through a collection of tools and clothing items. After watching the two sequences, they were repeated, but with the robots taking different paths and visiting different objects.

Between each sequence, participants viewed a series of images and indicated whether or not each image was shown in the prior sequence. They also completed a paper questionnaire that tested their ability to recall the different categories of objects in the scene before beginning the next scenario.

`

After completing four of these sequence-test cycles, participants completed the object/location questionnaire, responding on the same computers used to view the sequences. Participants read general instructions concerning three entities: a human named John, a computer named Yd3, and a robot named either OSCAR or V5 (whichever was robot was featured in the introduction video). Four scenarios asked participants to predict the behavior of each entity in each situation, and the directions emphasized that were “no right or wrong answers”, and that predictions should be “based on your judgments or intuitions."

2.3.3 Behavioral prediction scenarios

The four behavioral prediction scenarios were presented to the participants. The first scenario, the “Object vs. Location” scenario (see Figure 1), described two trials in which an entity acted upon one of two objects (a toy duck and toy truck) in two locations on a 3x3 grid. Then, the object locations were switched and the subject was asked whether the entity would reach for the previously reached-for object in its new location, or for the original object in the previously reached-for location. The intentional response was indicated by a prediction that the entity would act upon the original object in its new location, thereby demonstrating a belief that the entity was engaged in goal-directed action toward that object and would reach for it again.

In the next scenario, the “Feature vs. Category” scenario, six objects (a dry-erase eraser, electronic organizer, push pin, small piece of hard candy, gummy bear, and full size Hershey’s candy bar) were pictured in an array. Participants were asked to choose how each entity would organize the objects, as illustrated by two possible arrangements. One arrangement, the intentional response, showed the objects grouped according to taxonomic category with candy grouped separately from office supplies. The alternate arrangement showed the objects grouped according to size and color with the Hershey's candy bar, dry-erase eraser, and electronic organizer grouped separately from the gummy bear, push pin, and small piece of hard candy.

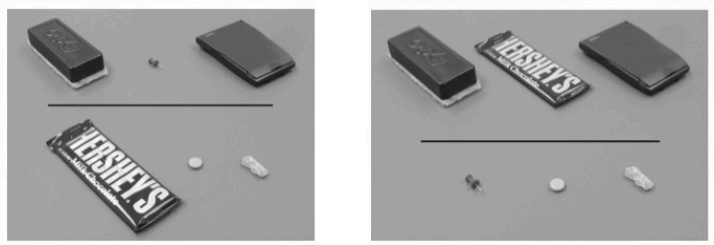

Figure 3. Choice screen for “feature vs. category” scenario.

The third scenario, another “Object vs. Location” scenario, involved seven items (two ink pens, a pencil, dry-erase marker, spoon, pair of scissors, and screwdriver) arranged in a row. The first three actions were directed toward every other item, one pen (in the first position in the row), another pen (in the third position), and the pencil (in the 5th position). Participants were asked to choose a final action that either continued the spatial “every other object” pattern of responding (a screw driver in the 7th position), or a category-based response that violated the spatial pattern (a dry erase marker in the 6th position).

The fourth and final scenario, another “Object vs. Location” scenario, involved six items (a computer cable, computer video adapter, electrical plug, and three different types of candy) from two categories organized in an array. Three were highlighted in their location by an underlying piece of paper and three were not. The first arrow indicated action upon the wrapped single piece of candy on paper and the second upon the unwrapped orange candy, also on paper. A subsequent pair of arrows offered a choice between the wrapped Smarties candy, not on paper, and the computer video adapter, on paper. An intentional or goal-directed response would have been indicated by a predicted action toward the Smarties candy, as it was similar to previous objects in its utility, though set apart in its location. The computer video adapter, different in utility, was similarly located on paper.

For each scenario, participants made predicted how a human, a computer, and a robot would respond.

2.3.4 Cognitive dissonance questionnaire.

Finally, participants answered 6 Likert-like items to measure cognitive dissonance created for the purpose of this investigation. Items include: “Sometimes I was uncomfortable answering these questions”, “At times I worried that some of my answers were inconsistent with my other answers”, “If I were allowed to, I would go back and change some of my responses”, “Some of the answers I gave in this experiment were inconsistent with my previous beliefs about the subject”, “I was always certain about my responses” (reverse scored) and “I never had difficulty putting together all of the facts in this experiment” (reverse scored). Participants were instructed to base their answers on their reactions while responding to the

`

behavioral prediction questionnaire.

2.4 Results

2.4.1 Behavioral predictions

As in previous research, participants made many more intentional predictions for the human (90%) than for the robot (44%; $t(36)=10.737$, $p<0.001$), or for the computer (34%; $t(36)=10.737$, $p<0.001$). The difference in predictions for the computer and the robot was not significant, $t(36)=1.587$, $p=0.121$. The pattern of predictions was globally similar between the conditions (see figure 4), although the difference between the computer predictions (28%) and robot predictions (47%) was barely not significant in the intentional condition ($t(18)=2.002$, $p=0.061$, while there was no difference at all in the mechanical condition (computer: 40%, robot: 40%).

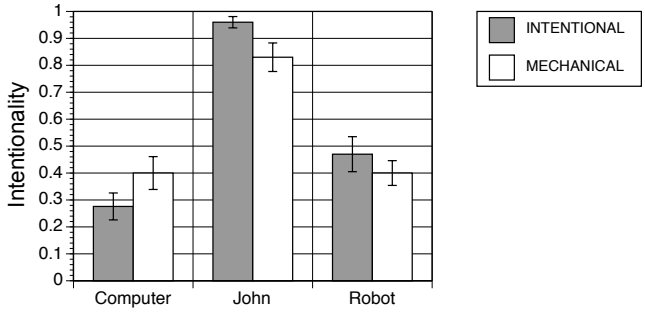

Figure 4. Behavioral predictions in the intentional and mechanical conditions.

2.4.2 Cognitive dissonance

Overall, levels of cognitive dissonance were similar in the intentional (4.07) and mechanical conditions (4.15; $t(35)=0.197$, ns). However, there was a strong positive correlation between intentionality of computer behavioral predictions and cognitive dissonance in the anthropomorphic condition ($r=0.644$, $p=0.003$), while there was no such link for the mechanical condition ($r=-0.094$, $p=0.711$; see figure 5). There were no correlations between cognitive dissonance and behavioral predictions for the robot or the human in either condition (r's<0.24, p's>0.33).

2.5 Discussion

In Experiment 1, we observed an interaction between dissonance and the intentionality manipulation such that high-dissonance participants in the intentional condition were more likely to give intentional behavioral predictions for computers, while low-dissonance participants were more likely to give non-intentional predictions for computers. Although globally consistent with our hypothesis linking dissonance with conceptual change, the specific form of these findings deserves some comment. Most important, there was no overall increase in dissonance in the intentional condition. Rather, both conditions caused similar amounts of dissonance, but in the intentional condition, this dissonance was linked with the behavioral predictions. Although we cannot be certain why this was observed, one possibility is that the complexity of the experiment caused dissonance about a range of different parts of the experiment, and only in the intentional condition did this dissonance cause motivation to change concepts about computer agency.

`

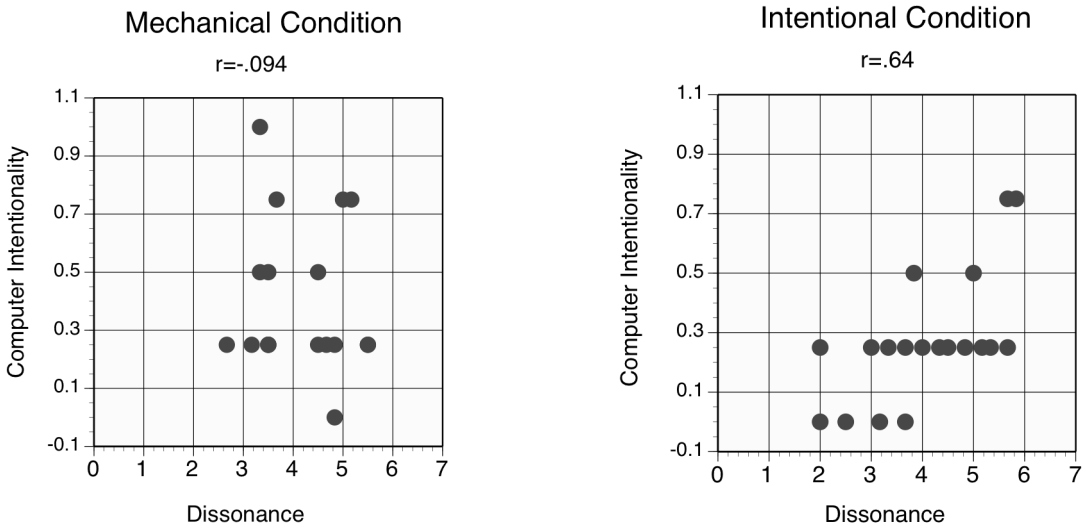

Fig. 5. Correlations between cognitive dissonance and intentionality of computer behavioral predictions.

A second element of the findings that requires explanation is that the dissonance-agency link was present for computers and not the robots that were the focus of the context videos. We suspect this occurred because the computer predictions were for a generic computer, whereas the robot predictions were for one of the specific robots used in the experiment. Previous research in social cognition suggests that motivated heuristic reasoning is much less likely for specific referents (such as a personality trait that has a very narrow definition like "punctuality"), than a less specific referent (for example, a trait such as sophistication) (Dunning et al., 1989). The prominence of heuristic reasoning for ambiguous targets occurs because a motivated reasoner can search a range of instances for agreeable examples in the less specific case. In a similar way, dissonance-induced motivated reasoning (e.g., reasoning that is motivated to lessen the discomfort at cognitive conflict) could more easily affect judgments of a large class such as computers. Thus, the increased search space of knowledge instances for the computer allows high-dissonance participants to isolate facts consistent with a high-agency interpretation of computer behavior that could reduce their dissonance. In contrast, there is little such flexibility in concepts about a specific robot, leaving little room for dissonance-motivated changes in concepts.

Both of these issues make clear that dissonance can be caused by a range of elements in a given experiment, and that its effects depend on a range of factors inherent to participants' knowledge base. Therefore, in Experiment 2, we lessened the complexity of the experiment to eliminate possible ambiguities in the source of cognitive dissonance.

3 Experiment 2

In Experiment 2, we more directly tested the hypothesis that deep concept availability interacts with dissonance to produce conceptual change. Instead of assuming that a highly anthropomorphic description would lead to deep concept availability, we directly manipulated deep concept availability by asking participants to complete an exercise in which they a) explicitly compare the kinds of intelligence inherent to people and computers, and b) listed ways in which the thinking inherent to computers might be different from that inherent to people. If cognitive dissonance causes conceptual change in concert with deep concept availability, then the positive correlation between dissonance and computer behavioral predictions should be limited to the deep-concept availability condition, just as it was for the intentional condition in Experiment 1.

3.1 Participants

36 Vanderbilt University undergraduates completed Experiment 2 in exchange for course credit (mean age=19.4, 27 female). Eighteen completed the anthropomorphic condition, and eighteen completed the mechanical condition.

3.2 Materials

Paper-based behavioral prediction questionnaires were administered with questions and images identical to those used in Experiment 1.

3.3 Procedure

In this experiment, the training videos and the robot event sequences were eliminated. After giving informed consent, participants in the experimental group were asked to respond to two written essay questions intended to invoke deep concepts about the nature of machine and human intelligence and about the difference between living and nonliving things. The first

`

question asked, “How might the intelligence of a person and a computer be the same and/or different?” and the second asked, “List all the ways living things are different from non-living things.” Participants were given five minutes to answer each question; if the participant finished his or her answer before five minutes, he or she was asked to “continue to think about the question for the remainder of the time and to write anything else you may think of.” Participants in the control group did not complete this questionnaire. The behavioral prediction questionnaire was then administered, followed by the cognitive dissonance questionnaire.

3.4 Results

3.4.1 Behavioral predictions

Overall, participants gave much more intentional predictions for humans (83%) than for computers (25%; $t(35)=8.367$, $p<0.001$) or for robots (27%; $t(35)=7.730$, $p<0.001$). The difference between computers and robots did not approach significance, $t(35)=0.325$, ns. This pattern was very similar between the control and deep concept conditions - see Figure 6.

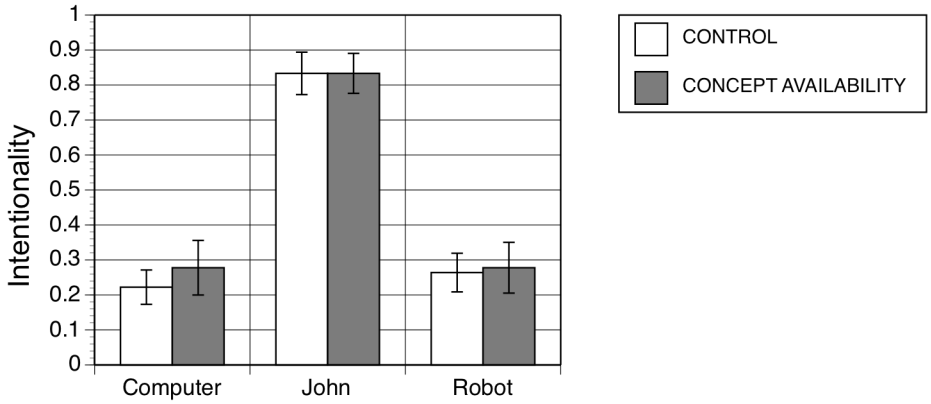

Fig. 6. Behavioral prediction scenario results for control and deep concept availability condition.

3.4.2 Cognitive dissonance

The deep-concept condition produced significantly more cognitive dissonance (mean=3.54) than the control condition (mean=2.86; $t(34)=2.163$; $p=0.038$). In addition, the correlation between dissonance and intentionality of computer predictions was significantly positive in the deep-concept condition ($r=0.58$, $p=0.012$), and not significant in the control condition ($r=0.17$, *ns*; see figure 7). The correlations between cognitive dissonance and behavioral predictions for the human and the robot were not significant in both conditions (r’s<0.21, p’s>0.40).

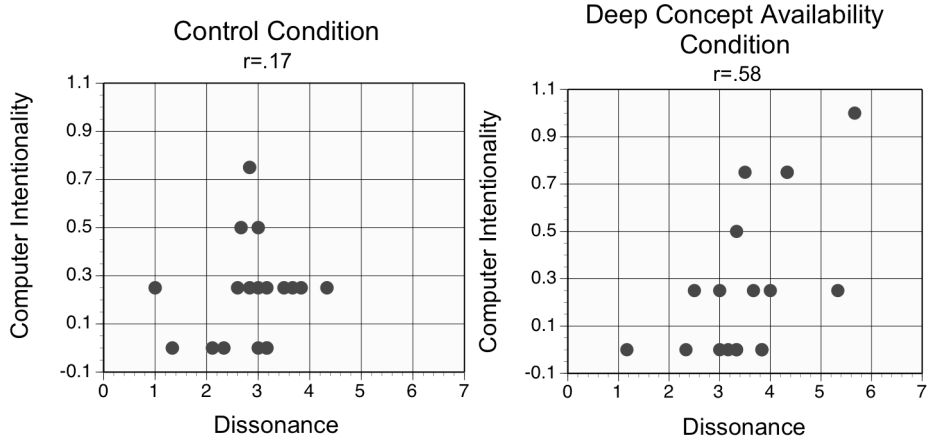

Fig. 7. Correlations between cognitive dissonance and computer intentionality in control and deep-concept availability conditions.

E. Discussion

`

In Experiment 2 we observed that making basic concepts about living versus nonliving things, and about the nature of machine vs. human intelligence not only caused increased dissonance, but also afforded a correlation between dissonance and intentional behavioral predictions for computers. It is important to point out that this manipulation did not require participants to take any specific position regarding the relative natures of machine and human intelligence. Instead, they were merely asked to report their own opinion on the matter, and then were asked to describe differences between living and nonliving things. Accordingly, we would argue that in this case the resulting dissonance occurred not because participants were confronted with new facts about the agents, but rather because they activated their existing knowledge about intelligence and differences between living and nonliving things.

4 General Discussion

Two basic findings are replicated across the two experiments. First, increased dissonance is associated with increased intentionality of behavioral predictions for computers. This basic finding suggests that increased cognitive conflict is associated with a relatively intentional view of computer behavior. Second, this relationship occurred in situations that involve invocation of deep concepts, either directly (in experiment 2), or indirectly via an intentional description of a robot.

These findings converge to suggest that in some situations, a relatively large amount of dissonance leads to increased attributions of agency to computers. However, overall levels of intentionality attributed to computers did not significantly increase in the deep-concept conditions. Therefore, the increased intentionality attribution in the high-dissonance subjects must have been offset by relatively lowered attributions of intentionality in participants who experienced relatively low levels of dissonance. One means of accommodating this pattern of results would be to hypothesize that some participants readily activated a range of concepts about the potential agency of computers, some of them consistent with an intentional hypothesis and some consistent with a non-intentional hypothesis. These participants therefore experienced dissonance, and changed their concepts about computers to incorporate the possibility for intentional action. In contrast, the low dissonance participants either activated fewer concepts about computers, or activated a set of concepts that more consistently pointed them toward a non-intentional view of computers.

Whatever the specific pattern of cognition that produces the dissonance-intentionality link, these experiments not only converge to support the hypothesis that dissonance is a key element of conceptual change, but they also validate our measure of dissonance. This is especially true of Experiment 2, for which both the group level manipulation of dissonance and individual differences in dissonance were consistent with the hypothesis that deep-concept activation would produce dissonance, and that particularly high levels of dissonance would push participants away from their presumption that computers are non-intentional. Experiment 1 is somewhat more difficult to interpret because although it did produce a significant (and strong) link between dissonance and attributions of intentionality to computers, this link appeared only as an individual difference, and there was no overall increase in dissonance in the intentional condition. One potential hypothesis that might be explored in future research is that no between-condition dissonance effect occurred in experiment 1 because a sizeable minority of participants already hold a relatively intentional view of computers, and therefore would experience no particular conflict with an intentional description of a robot. In experiment 1, this may have been particularly true because the event sequences shown in both conditions conspired to increase the levels of attributed intentionality to over 40% in some cases, as compared with Experiment 2 where the highest mean level of intentionality in any specific condition was 27%.

5 Conclusion

In summary, these experiments demonstrate that cognitive dissonance predicts increased attribution of intentionality to computers. These results therefore suggest that measuring dissonance is an excellent starting point for understanding conceptual change in people's understanding of agency in machines. Further research might extend these findings by assessing what specific facts about a given experimental situation cause the measured dissonance. We believe that a well-delineated understanding of conceptual change will not only specify the effects of dissonance but will also describe the link between dissonance and shifts from default reasoning to deeper problem solving that people engage in when their biases are challenged, as surely they will be by a rapidly changing technological environment that gradually incorporates machines that (arguably) think.

6 Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 0826701.

7 References

Bloom, P., 1997. Intentionality and word learning. Trends in Cognitive Sci. 1, 9-12.

`

D. Dunning, J.A. Meyerowtiz, A.D. Holtzberg, 1989. Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. Journal of Personality and Social Psychology, 67, 1082-1090.

Hymel, A., Levin, D.T. , Barrett, J., Saylor, M., & Biswas, G. in-press. The interaction of children's concepts about agents and their ability to use an agent-based tutoring system. Proceedings of the 2011 HCI International Conference.

Eliot, A.J. and Devine, 1994. On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort. Journal of Personality and Social Psychology, 67, pp. 1082-1090.

Epley, N., Waytz, A., & Cacioppo, J. T., 2007. On seeing human: A three-factor theory of anthropomorphism. Psychological Review, 114, 864-886.

Graesser, A.C., McNamara, D.S., & VanLehn, K., 2005. Scaffolding deep comprehension strategies through Point&Query, AutoTutor, and iSTART. Educational Psychologist*, 40*, 225-234.

Leslie, A.M., Friedman, O., German, T.P., 2004. Core mechanisms of theory of mind. Trends in Cognitive Sci. 8, 528-533.

Levin, D.T., Saylor, M.M., Varakin, D.M., Gordon, S.M., Kawamura, K., Wilkes, D.M., 2006. Thinking about thinking in computers, robots, and people. Proceedings of the 5th Annual International Conference on Development and Learning, pp. 49-54.

Levin, D.T., Saylor, M.M., Killingsworth, S.S., Gordon, S., Kawamura, K., in review. Testing the scope of intentional theory of mind in adults: Predictions about the behavior of computers, robots, and people.

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

Leslie, A.M., Friedman, O., and German, T.P., 2004. Core mechanisms in "theory of mind". Trends in Cognitive Sciences, 8, 528-533.

Morewedge, C.K., Preston, J., and Wegner, D.M., 2007. Timescale bias in the attribution of mind. Journal of Personality and Social Psychology. 93, 1-11.

Nass, C. and Moon, Y., 2000. Machines and mindlessness: Social responses to computers. Journal of Social Issues, *56*, 81-103.