Title
Perspective Taking: Launguage Use in a Visual Context

Author
Greenwood, Michelle Diane

Publication Date
2015

UNIVERSITY OF CALIFORNIA, MERCED

PERSPECTIVE TAKING: LANGUAGE USE IN A VISUAL CONTEXT

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

COGNITIVE AND INFORMATION SCIENCES

by

Michelle Diane Greenwood

Committee in Charge:

Professor Michael J. Spivey
Professor Rick Dale
Professor Paul P. Maglio

2015

The Dissertation of Michelle Diane Greenwood is approved, and it is acceptable in quality and form for publication on microform and electronically:

Professor Rick Dale

Professor Paul P. Maglio

Professor Michael J. Spivey

University of California, Merced

2015

To Jim, Davis, and Dillon

For always being my greatest champion,

may I always be yours in return.

I love you.

Table of Contents

List of Tables

List of Figures

Acknowledgements

Research does not usually happen alone on an island somewhere. Although, thanks to Darwin, I think that image persists and there are times I think some of us wish it did. Thankfully, it happens all around us and with all the people we surround ourselves. It happens in the hallway as you pass a colleague and have a “random” conversation that turns into collaboration down the road. It happens in a coffee shop as you inundate yourself with stimulants of all kinds (coffee, music, conversation, etc.) that sparks some crazy idea that just might work, and eventually does. It happens in a lab where you are surrounded with like minds that want these ideas to work almost as much as you do. And it happens with lots and lots of support. I will try to list the many ways in which I was supported.

First, I want to thank the many professors who drew me into Cognitive Science especially during the pioneer days of UC Merced as an undergrad. It was within their grip that I was taught, inspired, mentored, challenged, and most of all transformed. In the early days of UC Merced when my advisor had not yet arrived on the scene, Teenie Matlock and Jeff Yoshimi were instrumental in my choosing Cognitive Science. If you pitted two majors (and there were not many more to chose from at that time) as perspectives on the world Teenie and Jeff made my options of choices heavily weighted in the direction of Cognitive Science. Once my advisor Michael Spivey arrived, afforded me great opportunities in the lab and in the classroom, I took full advantage and am ever grateful. Spivey was instrumental in all of this research and I can not imagine how different it would have been without his guidance. Rick Dale provided insight into many aspects of the data that I would never have seen without his help. Paul Maglio was always around, from the early days to the very end, he provided insight into the minutiae of how to get through and how to finish.

Thanks to all the research assistants that helped with taking photographs, experiment coding, finding articles, interacting with participants, running copies, keeping logs, and the list goes on: Norma Cardona, Cynthia Cardon, James Greenwood, Caleb Henke, Vandana Koppula, Courtney Griffin-Oliver, Elaine Lai, David Sparks, and Monica Yanez. There was an extra amount of work done by the following research assistants: Yaasha Ephraim, Jesse Falke, Morgan Fleming, Fatima Panes, and Zachary Tosi. Special thanks to Drs. Justin Matthews and Eric Chiu for their tireless help with statistical analysis and helping make sure my assumptions were correct and properly met. In addition, I had statistical help and many conversations regarding the best way to analyze and visualize mouse tracking data with John Bunce. We learned a ton in the process. Also, thanks to Janelle Szary for all her beautiful work of voice recordings on over sixty object references.

This research would not have happened without the support of my family. My husband tolerated a great deal of flexibility in providing a routine for our sons to give me all the time I needed to finish these projects. My sons tolerated pilot versions of these studies, being carted all over campus, and having college

students as their second family. They grew up in so many ways on this campus. My parents (including my in-laws) and siblings were a constant source of encouragement and incentivized the process for me along the way as well. All my family believed in me (even when I didn't myself) and never hesitated saying as much.

There is an unusual bunch of folks that I must also thank. Coffee Bandits coffee shop and 17th Street Public House became a place of refuge for me. They also provided a place for me to become friends with a special group of "townies" (and they know who they are) but to name a few: Ben Davidson, Vanessa Hofmann, and Skip Johnson, and the list goes on and on. Plus a supportive cross-section of grad students who came and met me for coffee and sustenance both of the physical and mental nature: Justin Matthews, Eric Chui, Cyprian Czarnocki, Till Bergmann, Amy Moffat, and Arturo Durazo. There were more but I am unable to list them all and there were others that always shared their table and encouragement on those really busy days.

Although I do not formally have a soundtrack for this dissertation there were a couple of Pandora (http://www.pandora.com/, 2015) stations that were instrumental in providing just the right mix of background noise yet not so overpowering that I wanted to sing along: Sean Hayes Radio, Lana Del Ray Radio, and Ray LaMontagne Radio.

This work was supported by my primary advisor Michael Spivey and a Ronald E. McNair Postbaccalaureate Achievement Program Graduate Student Fellowship.

Curriculum Vitae

Michelle Diane Greenwood

Santa Rosa Junior College

2005	Associate of Arts	Liberal Studies

University of California, Merced

2009	Bachelor's of Arts	Cognitive Science
2015	Doctor of Philosophy	Cognitive and Information Sciences

The author grew up in the small town of Joplin, MO. For as long as she could remember she wanted to leave there and live in a big city. Instead she moved from college town to college town piecing together an Associate of Arts degree, getting married, and raising two boys. Eventually, she focused and found a school and a town that embraced all of her quirks. If you took Joplin and moved it to California, you would end up in Merced. And so it goes, that although she moved to the east coast and then swung back to the west coast, somehow she ended up right back where she started.

Abstract

Perspective Taking: Language Use in a Visual Context

Michelle Diane Greenwood

Doctor of Philosophy in Cognitive and Information Sciences

University of California, Merced 2015

Professor Michael J. Spivey, Chair

It is often argued that the egocentric perspective is the default perspective in natural language use. However, there are many factors that influence which perspective a person takes. In this dissertation I look at various environmental, biological, social, and linguistic cues that point toward another hypothesis for consideration than just the default perspective of egocentricism. It is probable that there is a bias toward the egocentric perspective given we are constrained to a body which provides us with many more experiences of an egocentric perspective but we are able to take other perspectives with relative ease and frequency. These occurrences of an other-centric perspective suggest a graded spatial attraction toward competing perspectives that result in a continuous, rather than a stage-based, account of perspective taking and encourage a more all-inclusive theoretical framework for taking a perspective.

Chapter 1: Perspective Taking: Language Use in a Visual Context

Perspective taking is used in Cognitive Science to study cognition. Perspective taking can be complicated to tease apart due partly to the complexity of the tasks and partly because the tasks do not always provide a clear picture of the perspective taking competition that is unfolding over time. It is usually the case that there are multiple perspectives a person could take during any given perspective-taking task and we are only aware of the finality of their perspective taken but not anything in between. It is the aim to show that there is graded competition that takes place when people are choosing a perspective. We see this in multiple domains of cognitive science. This brief review will take the reader through three of those domains and look at some experiments that test the hypothesis that perspective-taking involves graded competition over time and there are contexts in which the egocentric (taking your own visual perspective) bias can be influenced toward an “other” perspective (non-egocentric, often in this line of research refers to other perspectives specifically i.e., allocentric, bird’s eye view, etc. but for the sake of this dissertation it is simply referring to any other perspective that is not egocentric).

Before we begin, it might prove useful to have an example of perspective taking and its complexity. Imagine yourself visiting a local coffee shop with a friend. You both place the exact same order for black coffee; you are friends for a reason. The employee brews a fresh pot of hot coffee for you both while you head to the bathroom. You return from the restroom to sit at a small square table with two chairs located on either side of it. Your friend is sitting in one of the chairs and you grab the other empty chair. As you sit down there are two identical freshly brewed hot cups of black coffee in white coffee mugs sitting side by side in front of you and your friend. Which one is yours (see Figure 1)? You might assume it is the one on the left because you know your friend is right handed and probably would put his own cup in his right hand while taking a sip and just sat them on the table according to his handedness. In order to do this, you would have to take your friend's perspective though. You might also assume that your friend knows you are right handed as well and wanted you to have easy access to your enticing brew so he arranged the cups according to your handedness. That means that your friend would have used your perspective. In addition to him using your perspective you would have to know your perspective, from his perspective. In this situation, either way, your cup of coffee ends up on the right side of the table as you are looking at your friend. However, you were away momentarily and you also know that your friend had several things in his hand when he arrived and maybe in the entire shuffle somehow your cup ended up on the left side of the table if you are facing your friend. Again, you are back to taking your friend's perspective. Now, in spite of the fact that you two are friends, you really want your coffee but not your friend's germs so you look at your friend and ask, "Which cup is mine?" Are you confused yet? Perspective taking can be perplexing work.

Most of us probably do not routinely go through the aforementioned mental gymnastics when deciding where to place a cup when we are meeting a friend at the coffee shop. Nor do we when trying to figure out which cup on the table is ours. Regardless, it would not be the first time there was confusion about which cup on the table belongs to the actor. Although we may not explicitly run through all those possibilities (only a few were mentioned) there could be some implicit interactions that occur which may involve the type of dynamics involved in the explicit version: vision, cognition, social norms and relationships, and language to name a few. The previous example provides a type of visual perspective taking scenario that exists in the real world and also provides a scenario that involves someone at the table switching their perspective to make sense of the goal at hand.

Visual perspective taking has been used as a way of studying cognition because it is complex and yet also gets at some of the underlying principles that determine many of our contemplative decision-making processes. There are multiple disciplines that provide an excellent framework for allowing us a peek into the intersection of visual perspective taking and cognition: cognitive psychology, social psychology, distributed cognition and linguistics. Embodied cognition or ecological psychology (Chemero, 2009; Gibbs, 2005; & Gibson, 1979) is an overarching principle and a theoretical approach that has weaved its way into each of these disciplines and sheds an appealing light for how we understand the mind and the context in which it finds itself surrounded. Although embodied cognition and ecological psychology are not interchangeable, many of the fundamental principles are similar and will be treated with equal weight and almost interchangeably throughout this dissertation. Although egocentric perspective taking is often thought of as the default perspective in humans (Keysar, Barr, Balin, & Brauner, 2000; Piaget & Inhelder, 1956; Pick & Lockman, 1981; Shelton & McNamara, 1997), there are times when we take other peoples' and objects' perspectives and we seem to do this without difficulty and confusion, ambiguity aside. Previous research has described this other-perspective ability as a two-step process, first using the egocentric perspective, and then adjusting to another perspective when there is an error. For instance, Keysar et al. (2000) describe egocentricism as a way of constraining the possible interpretations in language, and although prone to error in a common ground task, see it as little mental cost and therefore worth it. These multiple instances of taking another's perspective provide us with insight into cognition and give us exemplary case studies.

Through the lens of cognitive psychology there are examples of how the environment feeds embedded cues to the visual system that often determine the specific perspective we adopt. In social psychology there are patterns of visual perspective taking being mediated by our relationships showing that those relationships indeed influence our behaviors and the perspective we choose. Finally, through linguistics we find illustrations of language that also influence which perspective we take and our understanding of the world because of it.

Through each of these lenses we find examples of gradedness that point toward ego-biased perspectives but find attractions to an other-perspective that win out in the perspective taking competition (e.g., Duran, Dale & Kreuz, 2011; Hanna, Tanenhaus, & Trueswell, 2003; Kosslyn, Thomposon & Ganis, 2006). Through this competitive account of perspective taking and the experiments conducted within I will give an explanation of a graded version of perspective taking instead of the two-step process that has dominated the field. We often use language to indicate which perspective we are taking so that seems like a good place to begin.

Perspective Taking: Linguistics

We often use language as a way to signal which point of view we are taking when communicating with someone (Talmy, 2000). The phrases we use can be literal, "*from my point of view* I can see the car" or figurative, "I wish you understood what it was like *from my point of view*." In both of these examples, the words *from my point of view* are used but they take on very different meanings. One is explaining what is visually available to the individual while the other is about things that are not visually seen at all, but represent a state of mind. However, we often use the word perspective to describe both of these situations: literal & metaphorical.

When we talk about perspective metaphorically we map the abstract perspective (point of view) to visual and spatial perspective taking and the principles that accompany them. So we use spatial mappings such that our mind is the space that one needs to traverse to understand the context of our thoughts, feelings and understanding (Lakoff & Johnson, 1980; Lakoff, 1987). It is in this realm of visual and spatial perspective that we will now journey (metaphorically speaking, of course).

Language is a powerful and well-studied mode of cognition. And because perspective taking intersects with language we see many ways in which perspective taking is influenced in that intersection where they meet. Anytime one does an experiment, instructions mediate out interpretation of the activities that unfold. Many experiments have been carefully constructed and fine-tuned to make sure that the words we used and the instructions that were given are carefully coordinated and orchestrated as not to interfere with the true task at hand. However, there are many times when that same language is used as a way to influence the task at hand and perspective taking seems to be a place where that is evident. One area where we find evidence of this is in common-ground constraints (Clark & Krych, 2004). These common-ground constraints help us interpret the world around us. Adults are not the only ones who use common-ground constraints, children as young as five or six are also able to make rapid use of common-ground constraints (Nadig & Sedivy, 2002). There is also a great deal of support for how signaling is used in language through describing-as, indicating and demonstrating (Clark, 1996) and the importance of gesture (i.e. exhibiting, pointing at, eye gaze, etc.) in language (Clark, 2004). All of these things help us constrain our interpretation of the discourse. Neglecting this part of language in modeling can lead to false assumptions. Not everyone subscribes to this way of thinking. Horton and Keysar (1996) interpret their results as evidence against a model that assumes that common ground is part of an initial utterance design and instead support the monitoring and adjustment model. However, this model does not seem to be computationally tenable, nor do they show evidence that common-ground constraints are not being used initially. Whereas, goal-based domain restrictions, especially those signaled by relatively simple non-verbal cues are likely to provide an important source of information that allows participants to coordinate their language without requiring resource demanding models of each other's mutual benefits (Hanna, Tanenhaus, & Trueswell, 2003).

Metaphor is also used to conceptualize space and the way we think about it (Gibbs, 2008; Lakoff & Johnson, 1980). People's conceptual experience of information space is largely structured metaphorically and based on embodied experience in physical space (Maglio & Matlock, 1999). There is also a relationship between thought about motion and language about motion, as is the case in fictive motion. This presents a challenge to models that favor purely static representations and demands instead some kind of dynamic representation. These claims of fictive motion support the underlying conceptual nature of language (Matlock, 2004). Figurative uses of motion verbs appear to evoke conceptual structure that is dynamic and reflective of the way we perceive and enact motion in the world (Matlock, 2006). Indeed language affects the way we think about the world. Some would even say that the structure of a language affects the way speakers conceptualize their world, including spatial inference through linguistic relativity (Levinson, 1996). Although it is evident that language influences our thinking about spatial matters, motion, common-ground, gesture and the like, it is not as clear that people are not able to develop specific conceptualizations of the world just because of the language they learn. In fact, after giving subjects a few intelligence tests (i.e. general intelligence, verbal, spatial, etc.), self report sense-of-direction independently predicts learning assessed in a real environment, whereas verbal ability and general ability are not independent predictors of large-scale spatial abilities. Small-scale spatial abilities are highly predictive of learning from direct experience (Hegarty, Montello, Richardson, Ishikawa, & Lovelace, 2006). However, that direct experience translates into our linguistic understanding of things through metaphor. Not only does language affect how we think about the world but those linguistic cues also influence how we perceive actions in the world. It could also have implications for which perspective we use when we are looking at a visual scene and have language that highlights the potential action within.

Individuals establish linguistic frames of reference: intrinsic (object-centered coordinate system, where the coordinates are determined by the "inherent features"), absolute (fixed direction provided by gravity), and relative (viewer-centered, a viewpoint V, a figure, and a ground distinct from V, a triangulation of three points and utilizes the coordinates fixed on V to assign directions to figure and ground) and make distinctions about how to use these frames of reference with or without a deictic center (Irwin & Carlson-Radvansky, 1996). These frames of reference that individuals establish are not fixed however. Describers will frequently switch perspectives, indicating that although perspective is needed locally to define spatial relations, the same perspective is not needed throughout to insure coherence. This switching of perspectives hints that we are not necessarily taking a two-step approach to perspective taking but a continuous one where we are drawn to one over the other. Perhaps we settle into one type of perspective and use those descriptions and our intuitions regarding those descriptions are off base and so we settle into a different description to help the addressee better understand the situation. Selection of a perspective may depend in part on how an environment has been experienced but it also depends on characteristics of the environment, with single paths and landmarks about equivalent in size encouraging a route (landmarks are described relative to an observer moving through the environment in terms of the observer's front, back, left, and right, using an intrinsic frame of reference) rather than a survey (landmarks are described relative to one another as if from above, in terms of the canonical direction, using an extrinsic frame of reference) perspective (Taylor & Tversky, 1996).

These multiple perspectives offer an array of understanding for the addressee and the addresser. We see this multiple perspective approach in natural language and in literature. Adoption in language comprehension by showing a narrative style has an impact on whether comprehenders view a described scene from an internal or an external perspective. In some circumstances, third person language can be just as effective at transporting the comprehender into the described experience of a character as second person language can (Sato, Sakai, Wu, & Bergen, 2012). The perspective chosen has implications for our understanding and comprehension. As previously mentioned, language is not strictly understood through the words we use but through other means as well. Language isn't accessed through purely spatial information but conceptually through affordances even when the linguistic cues do not describe any action (Borghi, Glenberg & Kaschak, 2004). Affordances provide comprehenders with implied action and this too provides either directly or indirectly information the comprehender should regard. Language and affordances provide information about what information the comprehender should regard and possibly which perspective one should take. If those cues result in a switching of perspectives could those cues be overriding the two-step process? Or could it be that one settles into one perspective and when it is ineffective switched to the other perspective? When agent information is missing people will fixate on the interactive region of objects: whereas instead of when the agent is mentioned, people will fixate on the agent performing the action (Fausey, Matlock, & Richardson, 2007). These actions are often perceptually simulated or figurative they are not literally happening and that perceptual simulation can hinder linguistic cues. When the linguistic cue activates a perceptual representation that shares a few features with the physical stimulus, perception is delayed (Estes, Verges, & Barsalou, 2008). Although those authors might use this as evidence to predict a modal cognition, it seems more likely that this effect is produced as a continuity of the mind where highly connected parts of the brain are interacting and overlapping, not allowing for a strictly feed-forward process but one that results in a delayed perception due to settling time (Spivey, 2007).

Language is complex in its own right. Vision is equally complex and often teasing these two apart can be difficult. Although neurologically speaking the language system and visual system are considered separate systems they often recruit many of the same neurological pathways to process information (Cohen, Lehéricy, Chochon, Lemer, Rivaud, & Dehaene, 2002; Hickok & Poeppel, 2007). However, to break down the different systems and how they influence perspective taking let us now turn to vision.

Perspective Taking: Vision

There is a great deal we can deduce about which perspective an individual may take based on visual cues we pick up from the environment. Much has been studied about the human visual system and how one perceives the world based on that visual system and Gestalt principles. Given that our brains are constrained to one body and one visual system within that body it is puzzling that as humans we are able to take another person's perspective at all. However, we find ourselves in situations where we do this efficiently and we are able to communicate it effectively. Sometimes, there are missteps along the way but typically we are able to navigate these types of situations, and the environment that surrounds it, with relative ease.

The field has learned that there are multiple cues that play a part in which perspective a person takes. Depth cues lead people to take an intrinsic perspective (Bryant & Tversky, 1999), having a person present in the visual scene will lead people to take the perspective of that person (Tversky & Hard, 2009), and motor processes can at least in part guide mental images (Kosslyn, Thomposon, Ganis, 2006; Wexler, Kosslyn & Berthoz, 1998). We also know that perspective-taking predominately relies on self-initiated emulation of a body rotation unlike mentally rotating the objects which seems to involve a different cognitive operation (Kessler & Thompson, 2009). Although the visual system may recruit many of the same overlapping mental processes to take another perspective as it would if it were taking an egocentric perspective there seems to be other processes at work as well.

Taking another's perspective might include some basic understanding of the following environmental cues such as physics, occlusion, visual cliffs, color, and depth. These visual cues signal the brain and the visual system that something in the environment is worth your attention. Knowledge about many of these visual cues is formed as early as infancy or as a toddler (Kohler, 1929; Spelke, 1982; Wertheimer, 1923/1958), and the visual system uses that knowledge to help us navigate the world. In spite of the fact that we view the world from our own frame of reference, we see objects from multiple views in multiple states and that builds up our repertoire about how things appear, and over time we use that information to replicate how the world could look different from someone else's perspective (Ballard, 1991; Finke, 1986) Sometimes we use mental imagery to do this which recruits many of the same neurological underpinnings as our visual perceptual system in an effort to understand the world at hand or perceived world from another perspective (Kosslyn, Thompson, & Ganis, 2006).

Many of these visual cues provide opportunities for action. These opportunities often translate into motor movements. This embodied approach to cognition is what Gibson (1979) had in mind when he so compellingly pointed out there are affordances that inform our visual system about our environment. Because this way of understanding cognition does not include just the brain but also includes the environment (Chemero 2009; Clark, 2008; Gibbs, 2005; Gibson, 1979), it often falls under attack (Gallese & Goldman, 1998; Rupert, 2004; Stuart, 2002). When it comes to perspective taking, Ellis & Tucker (2000) take this idea of affordances one step further and report evidence for something they term microaffordances, where object properties that are irrelevant to the current behavioral goal nonetheless automatically facilitate components of action potentiation. There are other indications that shed light on the functional role of the motor system in encoding visually presented objects through experiments "…the sight of the affording features of an object selectively recruits the motor representations that enable an individual to act upon it, but also, and above all, that such recruitment is spatially constrained, as it depends on whether the object falls within the actual reaching space of an onlooker endowed with the suitable motor repertoire" (Cardellicchio, Sinigaglia, & Costantini, 2011, p. 1372). This understanding of the world can lead us to take a perspective that translates into something other than an egocentric meaning. What if affordances are available perceptually and are also being constrained by an agent, cultural norms, language, and visual motor cues but those affordances belong to a non-egocentric perspective? It seems plausible that through competition albeit biased toward the egocentric perspective the visually stimulated individual has enough experiences, affordances, and neural recruitment to be pulled toward another perspective that wins in the end.

One way we determine functional roles in the brain is when we discover deficits. When something goes wrong with a normal human brain we will see shortfalls in perspective taking such that those who lack inhibition control also lack appropriate perspective interpretations (Brown-Schmidt, 2009). There is also evidence in rats that gives us hints that this is not strictly a human ability but instead that the egocentric frame of reference contrasts to the allocentric (or a god-like perspective) mapping and supports the idea that the spatially specific firing patterns of parietal cortex neurons maintain neither a purely egocentric frame of reference nor a purely allocentric frame of reference, but, instead, a frame of reference that might best be termed 'route-centered' (Nitz, 2009). Seeing that even in rats there are these "route-centered" frames of reference that exist yet can not exactly be categorized as egocentric or allocentric it could be the case that humans also have other frames of reference as well that aren't exactly egocentric or allocentric either but possibly "route-centered" or possibly something more akin to affordance-centric. Many perspective frames of reference other than egocentric have been suggested and as we move forward we will unpack a few of those seen throughout multiple disciplines.

Not only are there multiple environmental cues that influence our perspective there are also multiple theories for understanding perspective taking. Zacks and Michelon (2005) put forth a multiple systems framework for perspective taking, that spatial image transformations draw on a number of general-purpose spatial processing neural resources and a small number of transformation-specific neural resources. Therefore, one cannot equate tasks with either representations or transformations. While Kessler and Wang (2012) analyzed data that explain how people approach perspective taking differently either as *systemisers* (males/low-social-skills) or *embodiers* (females/high-social-skills) because simply put, systemisers view it strictly as a spatial task while embodiers view it as a expression of empathy. However, some argue that this kind of data may not be so much of a deliberate process but the context that is picking up on that mental state might trigger a non-egocentric perspective (Zwickel & Muller, 2010). Clearly, more work in this area needs to be conducted.

These other-centric cues individually may not persuade someone to consider a continuous account of perspective taking, in which both egocentric and other-centric perspectives are simultaneously active. However, when added up there are systematic ways that individuals take another perspective. If we use a graded spatial attraction explanation for perspective taking and the egocentric perspective is competing against another perspective there are clearly many factors that can attract one to a non-egocentric perspective.

Perspective Taking: Social Psychology

Social psychology in regards to perspective taking has focused a great deal on the social conditions under which people will take another perspective other than their own. It is clear that there are situations that induce individuals to take another perspective, as opposed to an egocentric perspective and here are some situations in which those cases arise.

When people observe a person in the visual scene and the actions of that person are highlighted people will describe the scene from that person's perspective (Tversky & Hard, 2009). Shelton & McNamara, 1997, claim that mental representations of large (i.e. navigable) spaces are viewpoint dependent, and that two views of a spatial layout produce two viewpoint-dependent representations in memory. There is also research that shows people use an egocentric perspective to anchor things and then adjust accordingly and that time constraints matter along with incentives (money in this case) (Epley, Keysar, Van Boven, & Gilovich, 2004). Information about social relationships can influence the way people conceptualize physical space. For example, when one says things like, "I'm close to my friend," one conceptualizes the physical space between oneself and one's friend, to actually be measured in real space as narrower than it is between other people and social groups (Matthews & Matlock, 2010, 2011). This perception of the social relationships we are in can also influence whose perspective you take when describing those scenes.

The constraints of our own bodies and the bodies of others come into play in multiple ways when dealing with perspective taking. Carlson-Radvansky and Logan (1997) believe that preferences for using particular reference frames are exhibited through weights assigned to the spatial templates, such that when they are combined, a composite of space surrounding the reference object reflects such biases. So not only do we take into account our own space but the space of the objects we are interacting with in the world.

In an experiment using a confederate, Hanna and Tanenhaus (2004) showed that the referential domain within which the addressee's uniqueness constraints were interpreted was immediately modulated by an awareness of the speaker's goals, as defined by the task-based constraints and signaled by a non-verbal cue. In this case the goal at hand will also influence whose perspective we use. Pragmatic constraints can determine whether or not perceptually salient entities are included in the domain used to compute the uniqueness conditions of definite reference.

Others report similar findings as that of Hanna and Tanenhaus (2004), but their interpretation is that addressees only use mutual knowledge as a tool for error correction, by reducing the probability of considering a non-shared object and it allows error correction when such referents are considered (Keysar, Barr, Balin, & Brauner, 2000). However, this doesn't seem to be the most likely account given that we know there is a cognitive cost for switching perspectives, but the cost is small and transient. Switching perspectives can entail changes of target object, terms of reference, and viewpoint (Tversky, Lee, & Mainwaring, 1999). Also, listeners will invest in an assumed other's perspective despite cognitive cost (Duran, Dale, & Kreuz, 2011). In addition to the perspective switching and investing in an assumed other's perspective, space, actions & our understanding of actions within that space are constrained within that space and even our mental spaces subserve thinking in many other domains (Tversky, 2005). Zwickel and Muller (2010) believe that some of these social contextual cues that Tversky and Hard (2009) describe are not just a detecting of another human that elicits the effects but the detection of a mental state and that sometimes action cues are a way of detecting a mental state. This account is probable and reinforces the non-egocentric perspective taking we see arise under many social circumstances that otherwise would produce egocentric effects.

Another way we find these competing dynamics at play are found in the brain when abnormalities are discovered in individuals who have malfunctioning patterns of brain activity. In a study with schizophrenics and a normal control group, the investigators contrasted the subjects' visual perspective taking deficits in the following manner. Patients with schizophrenia, who show selective deficits of visual/or cognitive perspective taking, do so because of an impairment of allocentric simulation. Schizophrenic patients show clear evidence for the co-occurrence of selective mentalizing deficits and a tendency to make egocentric errors when asked to judge how a visual array will appear from another viewer position (Langdon, Coltheart, Ward, & Catts, 2001). This deficit is found in schizophrenics and schizophrenia is often characterized as a social dysfunction. Ironically, their visual perspective taking deficits impair their ability to take someone else's perspective suggesting that taking another person's perspective is a combination of a social aptitude and a form of visual imagination. Perspective taking of another person is an ability we have as social beings, and context matters in terms of how we interpret that fact.

One last area where we see perspective taking in regards to its social aspects is with technology. People will often anthropomorphize humanoid robots and respond to them emotionally as if they were humans (Breazeal, 2003; Picard, 1997). So, it only seems natural that we would interact with these robots and technological devices in similar ways that we do with humans and our perspective taking is influenced as well.

Again, any of these social aspects independently may not persuade someone to retreat from an egocentric default account of perspective taking. However, when you add up the different factors that result in someone taking another perspective it seems that a competition account could reasonably explain the instances mentioned while easily taking into account the egocentric bias at work.

As already discussed perspective taking can be influenced through our social relationships and that is also reflected in the language we use to describe it. Proximity influenced similarity ratings for pairs of words and pictures, the closer, the more similar. However, for conceptual judgments it was the opposite (Casasanto, 2008). Once again, metaphor creeps into our thinking and perspective taking is no exception. In English we say things like, “if you could walk a mile in someone else’s shoes,” or “if you could just see it from my point of view.” These phrases metaphorically convey meaning about trying to understand what someone else is thinking from another perspective. In the realm of metaphor and thought, time and space are closely linked (Gibbs, 2008; Lakoff & Johnson, 1980). This is an area of linguistics and perspective taking that has been neglected. It would be interesting to see how grammatical aspect influences perspective taking and our understanding of it.

Once again, we see more features that lead to a perspective-taking account other than the default egocentric one. When added up these multiple instances of taking another perspective point to a competition account of perspective taking. These studies show how an individual may have a bias toward an egocentric perspective but the other perspective wins out in the end due to multiple linguistic inputs. From visual perspective-taking, we have learned several things that encourage either another perspective or differences in the way we view the world around us.

Perspective Taking: Back and Forth

Another observation in this line of research is that studying perspective taking is in many ways one-sided. A great deal of participants are brought in to the lab individually to see which perspective that person takes. Tasks are insufficiently interactive, and do not use dyads to study perspective-taking. Although there is often an assumed other person available, little work has been done to uncover the perspective taking of the other person involved in the task and if that other person's perspective influences the person participating in the experiment. Remembering back to the previous coffee shop scenario with your friend, regardless of the perspective one person takes, the other person's interpretation of it is based on that perspective. In addition to that, one of the people in that scenario will have to take someone else's perspective to parse the situation at hand. This type of contextual thinking urges us to take a deeper look at the back and forth and the *others* in a visual scene to better understand the context in which people choose a perspective.

When a person is present in a visual scene we take that other person's perspective more often when describing the actions in that scene (Tversky & Hard, 2009). Speakers with a partner set spatial perspectives differently from solo speakers. The egocentric bias is not absolute. Solo speakers were almost uniform in being non-egocentric. And partners following the speakers seemed to follow a norm of reciprocity of fairness, matching their partner's perspective (in the opposite direction) (Schober, 1993). There is also a common ground we build up through language and its accumulation and this is foundational for all collective actions (Clark & Brennan, 1991). Figurative language can also have an immediate impact on the way we view the world specifically that when fictive motion is used our mental simulations will match real life physical patterns, i.e. fictive motion descriptions of difficult terrain was more difficult to comprehend versus a much simpler terrain was easier to process (Richardson & Matlock, 2007). From studies regarding vision, we know that optical stimulation is continuous and so is our processing of it. There is a gradual accumulation of information as we continuously process temporal dynamics in visual processing and motor processing. Interestingly, not only do we build up an accumulation of past state spaces but we also make anticipatory visual movements for where they are likely to go next (Spivey, 2007). This is important because if we can change that trajectory we can influence decisions and behaviors.

Taking another perspective can help us solve real world problems as demonstrated in the following example. Perspective taking facilitation effects are also beneficial at the bargaining table. Considering the world from another individual's viewpoint can have increased mediation in negotiations compared to empathy or connecting emotionally to another individual. Those individuals who took another's perspective increased the discovery of hidden agreements and the ability to both create and claim resources while negotiating (Galinsky, Maddux, Gilin, & White, 2008). It is also evident that when speakers have misjudgments about their listeners in conversation either through overestimating or underestimating their knowledge this can work against them because the listener assumes either the speaker is talking over their head or talking down to them (Nickerson, 1999). This inability to ascribe metal states to oneself and/or to others has been suggested as a basic characteristic of autism (Leslie & Frith, 1988) suggesting that we have deficits if we are not able to take another's perspective.

As is often the case with perspective-taking tasks people are asked to provide one perspective as an answer or decision. However, we know that people switch perspectives often in conversation (Taylor & Tversky, 1996; Tversky, Lee, & Mainwaring, 1999). In work studying phase transition, people take time to develop their high-level cognitive domains and their understanding of low-level dynamics to create a discrete-like mode of one phase over another (Spivey, Anderson, & Dale, 2009). Perhaps perspective taking is another type of phase transition in cognition. We use one perspective and phase transition into another perspective, and they have a discrete-like mode of one perspective over another at any given point in time. In Pfaff's (2007) book, he lays out four steps in being able to apply the golden rule, do unto you as I would have you do unto me. In his third and crucial step, the person who is contemplating taking a harmful act on another individual, "blurs the difference between the other person and herself. Instead of seeing the consequences of her act toward the other person… she loses the mental emotional difference between the other person and herself." This "blurring" seems very similar to a phase transition where there appear to be two discrete states but somehow those two states get blurred in transition. Similar effects of social impacts on the dynamical systems approach of interpersonal synchrony appear in the minimal group design when people synchronize more if they think their partner is on a different team (Miles, Lumsden, Richardson, & McRae, 2011), or that if participants think it is a person they are interacting with (rather than a computer) they do not align as much spatially. This suggests again that if you feel more similar to a partner, you may "put more onto" that person, rather than worry about bridging with them (Duran, Dale, & Kreuz, 2011). Either way we see how two perspectives are pitted against one another either through verbal description or through relationship pairings and depending on the circumstances we can attribute the choice of a perspective. Yet those choices may not be discrete initially. It is likely you are given two choices and over time you settle into one over another depending on those circumstances previously described. As evidence suggests is the case when people are tracking others' beliefs in a Theory of Mind of object-tracking task (Van der Wel, Sebanz, & Knoblich, 2014).

There is a tightly coupled relationship between language, vision, and action that we see in mousetracking experiments and eye-tracking experiments that tease apart the graded spatial attraction as the participant settles into one of the competing objects, words, and/or decisions (Spivey, Grosjean, & Knoblich, 2005). As it has been suggested by other authors, the formulation and testing of models that integrate different types of reference frames in effector-centered representations and how they interact with those that update object centered representations should be a long-term goal. These interactions have not been well characterized to date and accounting for how people can plan actions that involve moving in the world, while other objects are in motion, and interacting with those objects are very complex (Zacks & Michelon, 2005). Although this research does not address all of those components it does seek to address some of those concerns.

In the field of cognitive science there is a common emphasis on an egocentric bias in perspective-taking tasks. Although it seems in many ways there are pointers to a more graded spatial attraction toward competing perspectives that results in a more continuous account of how we choose a perspective. In examining perspective taking this journey for me started with a simple study (Tversky & Hard, 2009). When I first read this study I was struck by the simplicity of this experiment that found participants more often took the other perspective in the presence of a person in the scene (see Figures 2 and 3). However, in studying affordances, an opportunity for action in a given scene, it occurred to me that there were many affordances available in these scenes and wondered if this was in part driving their effect? It was important to me to replicate this study while minimizing the affordances available to the viewer. At the same time I also wondered if this agency only extends to humans or could it be achieved with other agents as well? To answer these questions is what led to this first experiment.

These experiments focus on furthering perspective taking research to determine this more all-inclusive approach for taking a perspective. The first experiment looks at past research results, replicates those results and builds a foundation for future experiments that work at looking at the graded spatial attraction of two perspectives over time. These experiments will show results that are consistent with a continuous approach to perspective taking and further the field of cognitive science in this direction.

Chapter 2: Experiment 1 - Agency & Affordances in a Perspective Taking Task

Tversky and Hard (2009) reported a simple but elegant study where they found that when another person was present in the visual scene, people were more likely to take the perspective of that other person when they described the scene (Figure 3a). The present experiment was intended to replicate that study with a few additional changes. In the former study (Tversky & Hard, 2009), they used images that contained many potential action affordances in the scene (see Figure 2). For instance, there are water bottles for drinking, books for reading, couches and chairs for sitting, candles for burning and printers for printing, to name a few. The experiment focuses on the person (or absence of), a water bottle and book, but it is possible that individuals are invoking any number of affordances in the scene while completing the task. The goal of this experiment was to eliminate as many affordances as possible to tease apart if the affordances or the agent drives the effect. The hypothesis remained the same: A person depicted in the scene prompts individuals to take the perspective of that person.

It is also the case that humans will anthropomorphize robotic agents (Breazeal, 2003; Picard, 1997), so it is speculated that other agents could also produce this effect, specifically robotic agents. In addition to the condition with a human present, a toy robot agent was introduced to see if a non-animate agent also could produce such an effect. In this case, the toy robot was unable to actually to perform the task, but we hypothesized that we would see similar results to a human agent in the scene if participants are anthropomorphizing the toy robot agent and using agency as a predictor for which perspective they would choose. A control photograph was also included that had no agent to rule out the possibility of a confound of affordances or inherent quality of the scene itself inducing this other-perspective taking.

Experiment 1

In addition to replicating the Tversky & Hard (2009) study, other conditions were added to examine the role of agency in this paradigm. First, we added conditions that turned the human agent ninety degrees to the left or the right (see Figures 4e and 4g) to see if it was just the presence of a human or if there was a particular affordance of the human agent that was producing the effect. Three other conditions were also added that included a picture of a toy robot (see Figures 4a, 4b, and 4c) instead of a human. In this first experiment, we examined participants' descriptions of seven different spatial scenes (see Figure 4). In the "Control" condition (see Figure 4a) with no agent in the visual scene, it was assumed that participants would take an egocentric perspective. The control was used as a way of comparing the other measures of Figures 4b-g to determine a difference between a "default" egocentric perspective and what could be a potential other-centric perspective.

Method

Participants

A total of 352 UC Merced undergraduates from lower division Cognitive Science or Psychology classes participated in this experiment for partial course credit.

Materials

The task was given to participants in a packet of several pages of unrelated materials. The other materials involved other pen and paper tasks ordered in such a way that there were no other perspective-taking tasks that surrounded this particular task. There were also no other spatial tasks involved near this task. They were given instructions to complete the packet of materials in the order they were received in the packet and in one sitting. The packet took between thirty to sixty minutes to complete.

This specific task consisted of one page with the following phrase at the top of the page, “In the picture below, you see the last frame of a video clip.” This dialogue was used to engage the viewer in mental imagery that some action had taken place. Below those words appeared one of the photographs from Figure 4.

There were seven conditions for this experiment: the “Control,” the three views of the “Human,” and the three views of the “Robot.” The “Control” condition contained a photograph of a table, a water bottle, and a book shown in Figure 4a. Below the photograph was the question, “In relation to the cup, where is the book placed?” Participants were tested on which perspective they would choose when there was nothing in the picture that would necessitate taking anything other than an egocentric perspective (see Figure 4a). This was compared to either a person present and facing the viewer (see Figure 4f), or a human present facing ninety degrees to the “right” or “left” angles (see Figures 4g, and 4e, respectively). Alternatively, this was compared to a miniature robot facing the viewer (see Figure 4b), or the miniature robot seen at ninety degree angles to the “right” or “left” (see Figures 4d and 4c, respectively). Below the test question, six additional questions were included on the page: A distracter question about their favorite band, as well as questions about their handedness, gender and age. These questions were all open-ended, free response.

Procedure

Participants were randomly assigned to one of the seven conditions. Participants read the instructions at the top of the page, viewed the image, and then answered the questions at the bottom of the page. The test and following questions were open-ended, free response.

The author coded the data. Participants that responded with their own perspective while viewing the photograph were coded as a response of *egocentric* such as, “right,” “to the right,” and “a foot or two to the right of the bottle.” Responses were coded as *other-centric* such as, “left,” “it is on the side of it to the left,” and “underneath bottom left.” Responses were coded as *neutral* if the response was void of spatial information in reference to the participant, for example, “on the other side,” “across the table,” and “the opposite side.” In the *neutral* cases it could still be the case that they are taking either an egocentric or an *other-centric* perspective, but their response did not convey that perspective linguistically. It could also be the case that they are taking a different perspective, such as the bird's eye or omniscient perspectives (cf. Levinson, 1996; Sato, Sakai, Wu, & Bergen, 2009), but again, it was not possible to determine that from their response. However, in order to specifically compare egocentric responses to other-centric responses, data from the 145 participants who gave *neutral* responses were not included in the analysis.

Results and Discussion

Results show that there was a reliable effect (χ^2 (1, N=75) = 13.63, p = <.001) for the finding that if there was a human in the visual scene facing the viewer, the viewers were more likely to take the "other" perspective compared to the "Control." This replicates previous findings by Tversky and Hard (2009).

However, there were other interests of agency in this experiment and it is important to tease apart some of their subtleties. Initially, the robot facing the viewer (see Figure 4b) compared to the human facing the viewer (see Figure 4f) compared to the control (see Figure 4a) were analyzed to see if there were differences between their responses for egocentric versus other-centric. There were reliable differences across the seven conditions (χ^2 (2, N=111) = 15.79, p = <.001). Although significant, it would help to know which condition was really driving this effect. We know from the replication effects reported previously that a human agent facing the viewer results in more individuals taking the other perspective. However when the robot is facing the viewer we do not find the same results (χ^2 (1, N=76) = 1.79, p = .181). It seems to be the case that participants are more willing to take the perspective of a human agent than a robot agent in this case.

Another part of this experiment investigated whether there were differences between the ninety degree turns (left or right) of the human and robot agents compared to full facial views (agents facing the viewer). As it turns out, when the human was facing the participant and the action in the scene was highlighted (i.e. "In the picture below, you see the last frame of a video clip.") individuals used the other person's perspective to describe the scene more often (40% of the time) than when the person in the scene was turned ninety degrees to the left or right (21% of the time; overall: χ^2 (1, N=96) = 3.84, p = .05) whereas, with the toy robot we see the following (χ^2 (1, N=95) = 1.32, p = .25). Again, it seems to be the case that participants are more willing to take the perspective of a human agent than a robot agent in this case, particularly when the human is facing the viewer.

These results are consistent with the Tversky and Hard (2009) replication study and the attempts to have individuals take the other perspective with the toy robot were not significant. It appears from these results that humans alone are driving this effect and agency alone is not the main contributor to adopt this other-centric perspective.

Although there is an increase in "other" perspective taking in the 90 degree turn cases compared to the control, there is an even greater increase in the condition of the human agent facing you. It could be the case that when humans are facing an individual the person being faced associates that other individual with a greater amount of engagement for affordances. Is it the case that people are picking up subtle cues from eye contact from the person in the scene or body language that suggests interaction with the participant? And if so, why do we not see that same effect from the robot? If people anthropomorphize the robot why do individuals not pick up on that interaction with the robot? More work in this area will need to be done in order to tease apart the cues individuals are soliciting and how they use them to determine the perspective they choose. Regardless, it is clear that agency of the human nature do produce results of other-centric perspective taking. However, if individuals are picking up on subtle cues possibly ones such as eye contact or body language it might be the case that in “Robot” condition those anthropomorphic cues were not available in this scene. As previously mentioned, the toy robot would not actually be able to perform the action suggested in this experiment. Although the cup is small, it is about half the robot’s size. Is it possible that when participants viewed the image they discounted the potential action of the robot due to the size of the cup? Further investigation would be done with the toy robot to see if is the case.

One area that was not explored in this experiment was the role of language. We know from research mentioned earlier (i.e.,Clark & Krych, 2004; Lakoff & Johnson, 1980; Talmy, 2000; Tversky & Hard, 2009) that language can have an effect on which perspective people choose, but does it affect this specific visual perspective-taking task? And if so is it the visual scene alone or the language that produces this change of which perspective an individual might take? Due to effects such as fictive motion (Matlock, 2004, 2006) and embodied cognition (Tversky & Hard, 2009), it is reasonable to imagine that there would be a difference with respect to how the linguistic query places emphasis on different properties of the scene, specifically due to verbs that highlight the action and point toward the agent taking that action in the visual scene. Language was an area that Tversky and Hard (2009) also researched through mentioning the actor (or not) involved in the photograph. The next experiment explores the role of language more thoroughly in this perspective-taking paradigm.

Chapter 3: Experiments 2-4 – Language & Agency in a Perspective Taking Task

Language is a powerful tool. The ways in which we wield it are as varied as the people who use it. Although not always exacting, merely approximate, humans use conventions, rules, and rules of thumb for expressing ideas that exist in our mind. In the following experiment I tried to be as exacting as possible when manipulating the language stimuli for these conditions while still maintaining the conventions and rules regarding grammar and syntax that govern the English language in a scientific and systematic way. Also, certain aspects of the test questions were manipulated based on our prior knowledge of those grammatical and syntactic aspects and how subjects would engage with a photograph. We assumed participants would view the photograph and access mental imagery or mental animation due to the manipulation of the verbs and the agent performing the verb if such agent was referenced. This was in part to replicate previous findings (Tversky & Hard, 2009) and tease out the language and agent aspects of this task. Another focus of this experiment was to use the toy robot agent (non-human) to see if individuals will only take the other perspective when the human was present or if other agents could produce this result. This led to the following hypothesis: If figurative language is used with action while simultaneously highlighting an agent figuratively taking that action, individuals would take the other-centric perspective more often than if the agent was missing from the photograph and the action was not highlighted.

Experiment 2

In Experiment 2 I used a similar approach to Experiment 1 to explore the roles action verbs and agency might play in this visual perspective-taking task. While this experiment focuses on the language used in the task, the visual stimuli are also being manipulated in a systematic way to achieve these comparisons. We replaced the human with a toy robot as we had done so previously but also swapped out the cup with a water bottle that was smaller to adjust for a more realistic affordance for the robot.

Method

Participants

A total of 317 UC Merced undergraduates from lower division Cognitive Science or Psychology classes participated in this experiment for partial course credit.

Materials

The task was given to participants in a packet of several pages of unrelated materials. This specific task consisted of one page with the following phrase at the top of the page, “In the picture below, you see the last frame of a video clip.” Below those words appeared either the photograph from Figure 5a or 5b.

There were two photographs for this experiment: Control and Robot. The “Control” conditions photograph contained a table, water bottle, and book shown in Figure 5a. The “Robot” conditions photograph contained a table, water bottle, book and a miniature robot shown in Figure 5b. Below the photograph, was one of the following four questions: “In relation to the bottle, where is the book?” (n=59); “In relation to the bottle, where is the book placed?” (n=59); “In relation to the bottle, where is the robot’s book?” (n=59); “In relation to the bottle, where does the robot place the book?” (n=59). Individuals were tested on which perspective people would choose when there was nothing in the picture that would necessitate taking anything other than an egocentric perspective versus when there was a miniature robot whose presence hinted at the kinds of motor affordances that might warrant taking the other-centric perspective. The question participants were asked was manipulated by either highlighting the action in the visual scene (i.e., “In relation to the bottle, where is the book placed?”) and participants saw either the Control photograph or the Robot photograph. Contrast that to the non-highlighted action (i.e., “In relation to the bottle where is the book?”) when viewing either the Control photograph or the Robot version. Another condition was conducted to highlight the possible agent taking the action in the visual scene (i.e., “In relation to the bottle, where is the robot’s book?”) but this condition does not apply to the Control condition. The last conditions combined highlighting the agent and the potential action of the agent (i.e., “In relation to the bottle, where does the robot place the book?”) compared to the control version (i.e., “In relation to the bottle, where was the book placed?”). Below the test question, were three additional questions included on the page, “What is your major at UC Merced?” (a distractor) and “Are you right-handed?,” “Or left-handed?,” all open-ended, free response questions.

Procedure

Participants were randomly assigned to one of the seven conditions. In a strict comparison I would have employed a two by four factorial design with both independent variables manipulated between subjects for a total of eight conditions. However, there are pairings for all the conditions except in the control condition when the agent is highlighted (“In relation to the bottle, where does the robot place the book?”). Participants read the instructions at the top of the page, viewed the image, and then answered the questions at the bottom of the page. All questions were open-ended, free response.

The data were coded exactly as they were in Experiment 1. Participants that responded with their own perspective were coded as a response of *egocentric*. This included responses like, “right,” “to the right,” and “a foot or two to the right of the bottle.” Responses were coded as *other-centric* if they contained a perspective other than the egocentric response such as, “left,” “it is on the side of it to the left,” and “underneath bottom left.” Responses were coded as *neutral* if they were void of spatial information in reference to the participant, for example, “on the other side,” “across the table,” and “the opposite side.” Responses from six participants were removed because they either contained multiple perspectives or they were unanalyzable (e.g., “in the square,” “I don’t see any book”). In order to specifically compare egocentric responses to other-centric responses, data from the 117 participants who gave *neutral* responses were not included in the analysis.

Results and Discussion

Results show that participants were twice as likely to take the other-centric perspective when viewing the “Robot” photograph (31.4%) compared to the “Control” photograph (15.7%); χ^2 (1, N = 236) = 8.08, p = .004 (see Figure 6). As predicted, the miniature robot activates a possible mental imagery of motor affordances enough to persuade some individuals to take the other-centric perspective. This finding complies with robotics work (e.g., Breazeal, 2003; Picard, 1997) showing that simple visual and auditory properties (such as synthetic facial expressions and vocalizations) can humanize a human-robot interaction situation in ways that lure people into accommodating the robot’s needs and even (unwittingly) doing some of the intelligent behavior *for the robot.* However, that mental accommodation may only support the other perspective when the object is motorically affordable.

The results showed no significant difference in the responses when comparing the Control (“In relation to the bottle, where is the book placed?”) to the Robot (“In relation to the bottle, where is the book placed?”) conditions (23.8% and 23.7%); χ^2 (2, N = 236) =.001, p = .99. Linguistically, these two conditions are exactly the same, it is only the visual stimuli that changed. However, when the linguistic query was manipulated within the Robot condition, participants were more likely to give other-centric responses (about 83% of the time) with the action-based query (“In relation to the bottle, where does the robot place the book?”) compared to the possessive-based query (“In relation to the bottle, where is the robot’s book?”) (39.4% of the time); χ^2 (1, N = 62) = 12.06, p =.001 (see Figure 7). As predicted, here linguistic properties influenced which perspective people choose. In these conditions, the visual stimulus that participants received was the same, and both sentences referenced the robot in the photograph. The only difference was that one sentence asked, “Where is the book?” and the other asked, “Where does the robot place the book?” The “does place” sentence, which puts focus on the event carried out by the agent (robot) across the table from the viewer, induced a significant increase in the number of other-centric descriptions of the scene. It appears from these results that toy miniature robot agents and not just humans can illicit the other-centric perspective and if language is used to facilitate taking the other perspective by highlighting that agent’s actions more individuals will choose the other perspective. Individuals might be picking up on the opportunity for motor affordances in the visual scene due to the intrinsic value of the affordance itself or through language that facilitates such affordances. But how far does this limit go? There is clearly an agent, a toy robot in the visual scene that could be doing the action in this experiment. Although it is unrealistic that this toy robot could actually perform the proposed action in the scene people still use those visual cues, and combined with the linguistic input of highlighting the action participants will use the other-centric perspective to describe the scene. We know this effect is also robust when human agents are used in the scene. Testing this “limit” of affordances seemed like an important next step in this line of research. Is it agency itself or is it more generally affordances that are driving this effect? We know from these data that language can play a part in this paradigm but I wanted to begin to understand the extent to which affordances can also account for this other-centric perspective taking.

Experiment 3

Experiment 3 used a similar design to Experiment 1 to explore the role affordances might play in this visual perspective-taking task. In this next experiment affordances were examined in a perspective-taking task to see if participants would still be willing to take the other-centric perspective when there were very few affordances available to them. Individuals were tested with stimuli that had even fewer visual immediately available affordances than a toy robot. Would respondents choose the other-centric perspective even when there was no agent present in order to facilitate the “other” perspective? A chair was used to hint at and allow for agency in a new set of photographs, the scenes however did not include any actual agents (see Figure 8). Is the mere presence of the motor affordances available to the viewer, in the form of an empty chair, sufficient to get people to describe a scene from a “chair’s” perspective? This led to the following hypothesis: A hint of agency, in the form of an empty chair facing the viewer would allow for a greater other-centric perspective-taking when compared to a chair that faces the same direction that the viewer is facing.

Method

Participants

A total of 278 UC Merced undergraduates from lower division Cognitive Science or Psychology classes participated in this experiment for partial course credit.

Materials

The task was given to participants in a packet of several pages of unrelated materials. This specific task consisted of one page with the following phrase at the top of the page, “In the picture below, you see the last frame of a video clip.” Below those words appeared a photograph. Participants saw one of three photographs. Figures 8a, 8b, and 8c correspond to the “Viewer-centric” condition (chair in foreground), the “Control” condition (no chair), and the “Viewer-facing” condition (chair in background). Below the photograph were three questions: “In relation to the bottle, where was the book placed?”, “What is your major at UC Merced?” (a distractor question), and finally “Are you right-handed?” “Or left-handed?” questions to determine handedness.

Procedure

Participants were randomly assigned to one of three conditions. Participants read the instructions at the top of the page, viewed the image, and then answered the questions at the bottom of the page. The questions were open-ended, free response.

The data were coded exactly as they were in Experiment 1. In order to specifically compare egocentric responses to other-centric responses, data from the 68 participants who gave *neutral* responses were not included in the analysis.

Results and Discussion

Participants produced other-centric descriptions of the scene more often in the “Viewer-facing” condition (30.9% of the time) (see Figure 8c for stimuli), where the chair faces the viewer of the photograph, than they did in the “Viewer-centric” (7.1%) (see Figure 8b for stimuli) or “Control” (12.5% of the time) (see Figure 8a for stimuli) conditions; χ^2 (2, N = 210) = 15.37, p =.001 (see Figure 9). Pairwise comparisons between the three conditions were conducted to provide a more in depth look at the conditions: “Control” versus “Viewer-facing” χ^2 (1, N = 140) = 7.02, p =.008; “Control” versus “Viewer-centric” χ^2 (1, N = 142) = 1.15, p = .28; and “Viewer-facing” versus “Viewer-centric” χ^2 (1, N = 138) = 12.71, p < .001, (see Figure 9).It is worth noting that the “Viewer-centric” condition had numerically fewer other-centric responses than the “Control” condition, but the difference was not statistically significant. It is possible that this is due to a ceiling effect where both conditions were too saturated with egocentric responses for an effect of a chair encouraging even more egocentric responses to be elicited. The key finding here is that the affordances made available by an empty chair can apparently activate enough of a perceptual simulation of sitting, or perhaps enough of a social language internal script to afford sitting, to induce a sizeable increase in the frequency of other-centric descriptions of the scene (taken from the perspective of the empty chair). It could also be the case that individuals could be sensitive to the affordances associated with a chair and not of an agent *specifically*. It is possible that mental imagery could be facilitating such a potential agent that results in the viewer taking the non-existent but imagined agent's perspective become apparent. Conversely, it could be the case that the affordances of a chair are so strongly associated with an individual *sitting* in the chair that participants take the chair's perspective due to the affordances of the chair because it *hints* at agency. Future work will be aimed at teasing apart these possible underlying explanations.

This finding of affordances was notable because it suggests that people clearly pick up on subtle cues available in their environment. However, in this particular study participants were given little context for the visual scene they observed. Previous research in perspective taking has suggested many social factors that affect an individual's interpretation of a visual scene and the perspective they ultimately adopt or express. Therefore, I wanted to see if a social context would influence perspective taking and induce what I am calling a social affordance. Schilbach, Timmermans, Reddy, Costall, Bente, Schlicht, and Vogeley (2013) define such affordances as the possibilities for interaction provided by others, in terms of an activation of motor programs that would allow for interpersonal coordination of behavior. Kiverstein (2015) distinguishes their definition of social affordances from Schilbach, et al.'s in that theirs are not just a special category of affordances that exist and that take the form of possibilities for coordinated behavior, but that human environment offers count as "social" because the affordances humans pick up on belong to an intersubjective shared reality and are bound up with normatively constrained social practices. Given that the previous experiments relied on the environment (e.g. an empty chair) as an affordance but one that would be shared if you perceived that chair to be part of a social practice that would invoke another person, the evidence leans toward Kiverstein's definition. This next experiment was conducted to show an example of this within the context of a social setting explicitly to forward the idea of social affordances.

Experiment 4

A considerable amount of psycholinguistic work has treated the egocentric perspective in language processing as the default mode that is enforced by the cognitive architecture of the language processing system (Epley, Keysar, Van Boven, & Gilovich, 2004; Horton & Keysar, 1996; Keysar et al., 2000). In that account, factors that might encourage the accommodation of another's perspective (in either comprehension or production) come into play during a second stage of processing, after the initial egocentric anchoring-and-adjusting point has been assumed. An alternative approach treat's both egocentric and other-centric biases as competing against one another simultaneously and on equal footing in a parallel competitive constraint-based model (Hanna & Tanenhaus, 2004; Hanna, Tanenhaus, & Trueswell, 2003; Nadig & Sedivy, 2002). The present work does not distinguish between these two different real-time processing accounts, but instead attempts to introduce a new type of situational factor (or hint at social affordances) that can act as an attractor toward accommodating an other-centric perspective when describing scenes. Both the egocentric anchoring-and-adjustment approach and the parallel competitive constraint-based approach will need to accommodate this rather surprising kind of situational factor: environmental social affordances that merely hint at an alternative perspective.

In this experiment a social scenario was set up where the participant would be required to give a description of the perspective they were using based on the characters in the scenario. The variable manipulated was the “friendliness” of one of the characters. This led to the following hypothesis: The more friendly you were with “John” the more likely you would be to take his perspective.

Method

Participants

A total of 392 UC Merced undergraduates from lower division Cognitive Science or Psychology classes participated in this experiment for partial course credit.

Materials

The task was given to participants in a packet of several pages of unrelated materials. This specific task consisted of one page with the following narrative at the top of the page:

> Imagine that UC Merced has a committee of students that helps high-schoolers. The committee travels to high schools in the area and discusses how to apply to college. You and two other students have been selected to be on this committee. You are sitting in the seat labeled 'You.' John is sitting in the seat labeled 'John.'

Then they were given one of three pieces of social information: 1) "You do not know John, and have never talked to him." 2) "You have seen John, but have not talked to him." 3) "You are really good friends with John, and talk to him often."

A photograph (see Figure 10) was placed below the narrative and was the same for every condition. Below the photo were three open-ended response questions: 1) What would be the most important thing <u>you</u> would tell the high school students? 2) Your committee has been asked to give the panel a name, what would <u>you</u> name it? 3) Describe where Mary is sitting. The first two questions were distracters and the third helped determine which perspective they used when viewing the visual scene. After the third question several check boxes were presented for the participant to indicate their handedness, gender, and age.

Procedure

Participants were randomly assigned to one of the three social informational conditions. Participants read the narrative at the top of the page, viewed the image (see Figure 10), and then answered the questions at the bottom of the page. The questions were open-ended, free response, and closed-ended check boxes.

The data were coded similarly to Experiments 1, however, in this experiment the responses that took both perspectives into account were included as part of the analyses. Responses that included both perspectives were actually a redundancy of the perspective they were taking or a reiteration of the perspective used.

Results and Discussion

Participants produced descriptions of the scene taking another's perspective (John's in this case) more often in the "strangers" condition (16% of the time) compared to the "friends" (10% of the time) or "acquaintances" conditions (12% of the time) (see figure 12); Kendall's tau-b ($N = 392$) = -.105, $p = .026$ (see figure 11). This was surprising because it was hypothesized that people would take John's perspective more often when they were under the impression they were friends with John. One might assume this to be the case given that social conventions dictate that being "on your friend's side" might mean that you would also take their perspective in matters. However, after further reflection it is also possible that the social dynamics of a committee are such that you hope to come to a consensus. And so, if one is friends with John it is possible that one would think they already know his perspective and don't need to try to understand his perspective. Rather, only when you do not know an individual well, you feel the social pressure to try to understand their perspective.

Accounting for the dynamics of perspective taking is not often found in the theories relating to this phenomenon. Social affordances could be a powerful cue in the environment that people use as a means to choose a perspective when describing a visual scene. However, as previously referenced literature in synchronization has acknowledged it could be the case that people tend to align more with strangers or non-friends as a compensatory function or accommodation of perceived separation (i.e., Miles, Lumsden, Richardson, & McRae, 2011; Duran, Dale, & Kreuz, 2011). This pattern was observed in the data presented here where the scenario described an interaction between the participant and another person (in this case “John” as a stranger, not “John” as a friend). Joint task research has shown examples of participants aligning more if they think their partner cannot see the same stimuli (Richardson, Dale & Tomlinson, 2009). In this research it could be the case that participants align more with the partner that is perceived to not see the same stimuli (in this case, the partner being a fellow committee member and the unseen stimuli are thoughts, trust, ideas, etc.).

Although Experiments 2, 3, and 4 have been crafted in an effort to compare egocentric versus other-centric perspectives, it is difficult to understand how the graded attraction of the two perspectives are settling over time. It may be the case that when performing these tasks, people are actively torn between two perspectives but settle into one and use it to describe the visual scene they are experiencing. On the other hand, it could also be the case that people initially chose a certain perspective then occasionally change perspectives and use the second perspective for their description. Experiment 5 attempts to examine this possible temporal shift in more depth to see if one can get a better understanding of the competition that individuals encounter during active perspective taking.

Chapter 4: Experiment 5 – Mousetracking in a Perspective Taking Task

So far, the research that has been conducted in this dissertation has been experiments performed with pen and paper, pitting an egocentric perspective against another. It should be evident that there are multiple perspectives that can be used to describe a visual scene. There are several factors that can elicit an other-centric perspective compared to an egocentric perspective, such as agency, language, and social contexts. There appears to be a systematic strategy for when people describe a scene from the other perspective and not rely solely on their traditionally default egocentric frame of reference. However, one question still remains: is taking the other perspective a two-step process, where one initially takes the egocentric perspective and then switches to another perspective (Barr, 2008; Keysar, Barr, & Brauner, 2000), or are there two simultaneously-active perspectives competing over time before settling in on one (Hanna & Tanenhaus, 2003, 2004)? The next experiment was conducted in an effort to answer this lingering question. Using what was learned from the other experiments, I attempted to maximize the action potential of participants through agency, affordances and language in this experiment to measure the differences between mouse movements of individuals in two different conditions.

More commonly, researchers accept the former two-step processing account of perspective taking (Epley, Keysar, Van Boven, & Gilovich, 2004).

However, I hypothesize here that disparate perspectives compete over time before settling into a chosen perspective.

Experiment 5

The purpose of this experiment was to test the interaction between two perspectives (egocentric and other-centric), over time and observe how it manifests through the motor movements of those individuals' via computer mouse movements. The mouse movements people make are indicative of their thoughts (e.g., Freeman, Dale, & Farmer, 2011; Freeman & Ambady, 2010; Spivey, Grosjean, & Knoblich, 2005), thus measuring those movements at a rapid interval gives us a glimpse of what people are thinking over that period of time. This measurement can show the patterns of activation for two different visual scenes and then make various comparisons. We have seen different patterns of responses for the other experiments reported. Traditionally, the expectation in this experiment would be to see a pattern of responses that would mimic those results. However, I expect to see a measured spatial attraction toward the other-centric perspective in the person condition even when the egocentric perspective is chosen.

> This was determined by the trajectories measures to determine how much deviation or curvature exists in the trajectory when comparing the person condition to the no-person condition. Prior studies have used two measures, which are fully implemented in this Mouse Tracker experiment: maximum deviation (MD) and area-under-the-curve (AUC). For both of

these measures, MouseTracker first computes an idealized response trajectory (a straight line between each trajectory's start and endpoints). The MD of a trajectory is then calculated as the largest perpendicular deviation between the actual trajectory and its idealized trajectory out of all time-steps. Thus, the higher the MD, the more the trajectory deviated toward the unselected alternative. The AUC of a trajectory is calculated as the geometric area between the actual trajectory and the idealized trajectory (straight line). Area on the opposite side (i.e., in the direction away from the unselected response) of the straight line is calculated as negative area (Freeman & Ambady, 2010).

Method

Participants

Fifty-seven individuals (41 females, 15 males, 1 declined to answer) between the ages of 18 and 32 (M = 20.0) from the University of California, Merced participated in this experiment. They were given partial course credit for their participation.

Materials

A 24" Apple iMac (Apple, Inc., 1997-2014) running Microsoft Windows 7 (Microsoft Corporation, 2001-2014) and MouseTracker (Freeman & Ambady, 2010) software was used in this experiment. All trajectories are rescaled into a standard MouseTracker coordinate space. The top-left corner of the screen corresponds to "-1, 1.5" and the bottom-right corner corresponds to "1, 0". In this standard 2-choice design, this leaves the start location of the mouse (the bottom-center) with coordinates “0, 0”. This standard space thus represents a 2 × 1.5 *X-Y-* coordinate rectangle, which retains the aspect ratio of most computer screens. Photographs of the objects and the backgrounds used for the experiment were taken with a Canon Power Shot SD 1000 Digital Elph (Canon, U.S.A., 2015) digital camera. The female voice was used for each of the trials and recorded using Praat software (Boersma & Weenink, 2014). Mouse movements were recorded at 70 Hz using a wired Apple mouse.

Procedure

Participants were brought into the laboratory to perform the task individually at a computer terminal. After consenting to participate, research assistants explained the task and sat them at the computer station. Instructions on how to proceed were displayed on the computer. Each participant was randomly placed in one of two conditions: the no-person condition contained a table, on which a basket was placed (Figure 14); the person condition was similar to the no-person condition except that the image included a person sitting directly across the table facing the participant (see Figure 13). Participants were informed of a square at the bottom-center of the screen labeled “start”. They would click on the “start” square to initiate each trial. Once the participants moved the cursor outside the small start box their initiation time would end. Participants were given several practice trials and then continued on to the experimental trials. Once they clicked on the start square, their cursor would become an object that they would need to place somewhere. The instructions were delivered verbally by the same recorded female voice throughout the experiment. A female voice was chosen in an effort to counterbalance the male photograph that participants viewed in the person condition (see Figure 13). It was critical that participants did not associate the delivery of the instructions with the male subject in the photograph. Participants were verbally instructed to “Place the cursor which had turned into the object (see Table 1) to (either the right, left, or inside) the basket.” Twenty different objects were used (Table1), each object was presented three different times, once each for “right”, “left”, and “inside”. The “inside” trials were fillers, included to help prevent participants from figuring out the experimental hypothesis. The order in which each object appeared, and the direction of the verbal instructions were randomized for each participant. The participants ended the trial by dropping the “object”/cursor on either side of the basket, designated by either the right or left half of the screen.

Due to the ambiguous nature of the questions and my interest in their interpretation, no feedback was given regarding accuracy of their response. After 5000 milliseconds without a response, that trial would end and the next trial would automatically begin.

The data was processed using MouseTracker Analysis Tool (Freeman & Ambady, 2010, 2011) and analyzed with R Statistical package (R Core Team, 2013). Due to the ambiguous instructions of the task there were no right or wrong answers in this task. For consistency, the responses are described from a participant's egocentric perspective. As in the previous experiments, the egocentric perspective assumes that the right would be to the right of the screen as the participant is facing the screen. We collected 2280 critical trials from the 53 participants (only right or left trials were used, "inside" trials were used as a filler), 63 trials were discarded due to the incompletion of the trial as a result of the time requirements.

The hypothesis was that those individuals in the person condition (see Figure 13) would take an other-centric perspective more often than those in the no person condition (see Figure 14). Moreover, I hypothesized that there would be a greater difference in mouse curvature towards the competing target for the person condition than in the no person condition.

Results and Discussion

Results show that indeed people take the other-centric perspective more often in the person condition when compared to the no person condition, a Welch's *t*-test, $t\,(1636) = 11.89$, $p < .001$. The 95% confidence interval for the effect of perspective on condition is between 0.129 and 0.180. (see Figure 14). To rule out the possibility that a particular object could potentially drive a specific perspective, the stimulus items were analyzed individually. This was not the case; all items appeared equally in both perspectives and were each used on average about 14% of the time.

Of primary interest was the examination of the mouse movement trajectories and response times. To permit averaging and comparison across multiple trials with different numbers of coordinate pairs, the *x, y* coordinates of each trajectory were time-normalized into a 101 time-steps using linear interpolation (Freeman & Ambady, 2010). From the normalized-time trajectories (AUC), (MD), and reaction time were calculated. In this experiment the AUC measure reflects the extent in which the other-centric response location influenced the movements towards the egocentric response location. Larger values of AUC represent a stronger influence of the other-centric location while MD pinpoints where in the trajectory that the reflection is the greatest. Reaction time (RT) is measured from the end of the time initiated duration until the trial ends when the object is placed. Initiation time (IT) is measured from the start of the trial until the cursor is moved outside the start box. Time of maximum deviation (MDT) is measured by time in milliseconds at which point the maximum deviation occurs.

For the reported analyses, each participants' mean values were entered into a 2 (Condition: person or no-person) X 2 (Response: egocentric or other-centric) repeated-measures ANOVA with condition (person versus no-person) entered as a between-subject variable (See Barr (2008) for a discussion regarding time-constrained psycholinguistic methods such as eye tracking or mouse tracking). This type of multilevel logistic regression allows us to assess the effects of a continuous variable (time) on a categorical variable (discrete mouse location) while still using a repeated measure (each subject performing 40 critical trials). After exploring many measures such as MD, AUC, and RT to determine if there were differences between participants who were in the person condition compared to the no-person condition, only two were found. The first measure was IT, the time from starting the trial until moving outside the start box. Participants in the person condition were faster in their IT than those in the no-person condition ($F(1, 2217) = 4.039$, $p < .04$) (Figure 15).The second measure is MDT, the time in milliseconds at which point a trajectory maximally deviated toward another response alternative (the other-centric perspective). Participants in the person condition were quicker to deviate maximally than those in the no-person condition ($F(1, 2217) = 7.090$, $p < .007$) (see Figure 16). Figure 17 shows the trajectories of the two conditions compared to one another for IT and MDT.

A distributional analysis was conducted to exam if the variables that were significant were bimodal and did not adhere to the normal distribution for a linear model. Table 2 provides the results. There is evidence for bimodality for IT in the person condition (see Freeman & Ambady, 2010; Freeman & Dale, 2013 for a recent discussion on determining bimodality). In the mixed effects model used to analyze these variables the bimodality is still accounted for and therefore significant. However, it is interesting that only in the case of the person-present condition, and when a participant takes the other perspective, do we see bimodality arise, and only with IT not MDT. It could be the case that participants either have a sharp change of mind before responding or we are observing two different types of responses, one that is more direct and one that requires more time before responding. It is also possible that there are two types of subjects one that is slow but deliberate while the other is fast but impetuous. Both conditions are ambiguous in their instructions. It never explicitly states which perspective one should choose when performing this task. However, in the no-person condition it is almost as if it is even more ambiguous than the person condition and the person provides an anchor of sorts irrespective of which perspective you choose (egocentric or other-centric). However, the data points to the more likely case of having two types of subjects; one that is slow and deliberate while the other is fast and impetuous. Especially given that we basically see no differences in overall RT, MD, and AUC. It is only in the case when both a person is present and the other-centric perspective is taken do we find this unique case of bimodality.

As can be seen in Figure 17, trajectories for IT & MDT in the person condition are a bit “drawn away” from the trajectories for the no-person condition. To analyze the mouse trajectories, I computed an average trajectory across all participants for each of the two spatially equivalent conditions. Before averaging, all trajectories were lined up to a common x, y starting position (0, 0). They were then individually normalized by resampling the time vector at 101 equally time spaced values and computing, by means of linear interpolation, the corresponding mouse-coordinate values (separately for the x and y coordinate vectors) (Spivey, Grosjean, Knoblich, 2005). MouseTracker (Freeman & Ambady, 2010) computes the 101 timesteps automatically and through further analysis it was found that time steps 85 – 101 are significantly different when comparing trajectories of person versus no person conditions. This gives us additional information that when a person is present participants are drawn toward the other-centric response even when they choose the egocentric perspective.

Initially, this discussion began with the assumption that there would be differences in curvature due to opposing perspectives to support the simultaneous competition approach of perspective processing. As it turns out we see very little differences in the overall curvature (MD, AUC). However, we do see differences in the IT and the MDT. According to the two-step account this might actually be support for their case, especially since the only differences we see happen so early in the process (i.e., 1500 – 2000 ms for IT). If we take a closer look though, a case for the competition account is more likely. In this experiment, it is generally the case that once a person started toward an egocentric perspective, they almost always ended up at an egocentric perspective; and vice versa. Their RT, MD, and AUC are nearly the same regardless of which perspective they used. However, when choosing the other perspective when a person was present, participants were faster in their IT and their MDT. If individuals were taking their own perspective and then taking the other perspective, one would assume they would be slower at initiating their response. One might also assume their deviation toward the opposing perspective would happen later, as if they were still considering their own perspective longer. This is not the case. Instead we see people taking longer to initiate an egocentric perspective when a person is present. We also see other-centric mouse movement responses in the person condition even when they chose the egocentric response.

A considerable amount of psycholinguistic literature has accepted the egocentric perspective in language processing as a default mode that is enforced by the cognitive architecture of the language processing system (Epley, Keysar, Van Boven, & Gilovich, 2004; Horton & Keysar, 1996; Keysar,Barr, Balin, & Brauner, 2000). In that account, factors that might encourage an accommodation of another's perspective come into play during a second stage of processing after an initial egocentric anchoring point has been assumed. An alternative approach has been to treat egocentric biases and "other" centric biases as competing against one another simultaneously and on equal footing (Hanna & Tanenhaus, 2004; Hanna, Tanenhaus, & Trueswell, 2003; Nadig & Sedivy, 2002). These results serve to reinforce the simultaneous competition account.

Chapter 5: General Discussion

These five experiments, 1) a replication of previous studies and a robot manipulation, 2) a robot and language manipulation, 3) three empty chairs, 4) committee with “John,” “Mary,” and “you,” and 5) mouse tracking with either a person or no-person, all work together to explain a different account of perspective-taking other than the egocentric two-step default account (e.g., Barr, 2008; Keysar, Barr, Balin, & Brauner, 2000). While we find a bias toward the egocentric perspective in all of the experiments conducted, we also see how there are many factors that persuade individuals to lean towards an other-centric perspective instead. These experiments show there is more competition than previously theorized (e.g., Hanna & Tanenhaus, 2004; Hanna, Tanenhaus, & Trueswell, 2003; Spivey, 2007; Spivey, Grosjean, Knoblich, 2005) that takes place when deciding which perspective an individual uses. During that competition, people often use their own perspective to describe a visual scene. However, when there are affordances, agency, and social norms that provide for an other-centric explanation, the other-centric description will more often be used than when they are not present.

In the first experiment (replication and robot manipulation), general findings were replicated to show that when a person is present in the scene and facing the viewer, people more often take the perspective of that person (Tversky & Hard, 2009). That research was expanded to also look at 90 degree turns (left or right) of the individual so that if the person in the scene was not facing you but still present, which explored whether the participant would take the other-centric perspective. Participants did take the other-centric perspective even when the person was turned 90 degrees, but not as often as when the person was facing the viewer.

This same experiment was conducted with a miniature toy robot to see if people would anthropomorphize the robotic agent and produce similar other-centric taking perspectives. Participants did not respond in the same way when the robot was present as they did when the human was present. The robot did not produce other-centric responses. One might initially take this as evidence that only human agents produce such other-centric points of view. In this particular experiment the affordances associated with the toy robot did not seem plausible even if individuals were anthropomorphizing the robot and mentally doing some of the work "for the robot," resulting in participants not taking the "other" perspective.

In the second experiment (robot and language manipulation) the same miniature toy robot that was used in Experiment 1 was used again and yet we see significant differences in the way individuals respond. The only difference between the robot photograph in the first experiment and the second experiment was that the cup in the first experiment was about half the size of the robot. In the second experiment there was a small water bottle that may have allowed for a more believable affordance for the robot. This experiment also explored the role that language played more thoroughly when deciding which perspective an individual used. When the focus was on the agent in the scene (toy robot) and the event being carried out was highlighted by using the verbs associated with that activity, participants chose the other-centric perspective more often than when that language was controlled for by not using agency or highlighted action.

One other language possibility we considered had to do with culture. While the population we assessed was predominately Latino, English was their primary language and they were also mostly English language learners. Although we hypothesized about Asian cultures specifically and if we might see differences in the way they would respond, we think we would see differences in these types of tasks in the opposite direction (Wu & Keysar, 2007).

There are a few phenomena that are persuading individuals to take the other-centric perspective. The first phenomenon is agency; any entity in the environment believably able to perform the task (even unwittingly) and that mentally the individual is able to cognitively work “for the robot” (that is actually unable to perform the task). Second, language associated with the other agent and that agent’s performance being highlighted. However, I consider the third phenomenon includes affordances (Gibbs, 2005, 2008; Gibson, 1979), which are more pervasive in this set of experiments than originally considered. If affordances of the agent have as much credibility to the task as the agent itself, then affordances may have more weight than are often associated with such tasks, particularly as something as simple as which perspective is used in a given scene.

The third experiment (three empty chairs) explored the possibility of affordances even more by use of a photograph that contained an empty chair either facing the viewer, facing the same direction as the viewer, or with the chair missing. Participants took the other-centric perspective more often when an empty chair was facing the viewer. This empty chair perspective could be interpreted a couple of ways. First, the chair either hints at agency or it hints at affordances associated with implied agents. Either way, if the default perspective is egocentric and there is no actual agent present to adopt the other perspective, then why would individuals choose to describe the scene in that manner? However, if there is a continuum of perspectives and a person is drawn towards one over the other (granted, in a very short amount of time), it seems individuals might be more likely to describe the scene in the perspective they were drawn toward. With the three versions of the stimuli (ego-centric chair, no chair, other-centric chair) there follows a gradation of other-centric responses. This points toward a graded competitive account of perspective taking and those affordances consistent with those two competing perspectives.

In the fourth experiment (committee with “John,” “Mary,” and “you”), the other-centric perspective was explored through a social setting of a theoretical committee comprised of “Mary,” “John,” and the participant (“you”). The level of friendliness with the participant and John varied across conditions. The surprising aspect of this finding is that people take the other-centric perspective more often when the scenario describes “John” as being a stranger compared to being a close friend (or an acquaintance). This result is agnostic in terms of the two-step egocentric processing (e.g., Barr, 2008; Keysar, Barr, Balin, & Brauner, 2000) compared to the continuous approach of settling into one perspective over another (e.g., Hanna & Tanenhaus, 2004; Hanna, Tanenhaus, & Trueswell, 2003; Spivey, 2007; Spivey, Grosjean, Knoblich, 2005). However, either camp would need to account for this rather interesting aspect of determining perspective-taking descriptions.

In this experiment, I introduced a term called social affordances, because the "human environment offers count as 'social' because the affordances humans pick up on belong to an intersubjective shared reality and are bound up with normatively constrained social practices" (e.g., Kiverstein, 2015; Schilbach,Timmermans, Reddy, Costall, Bente, Schlicht, & Vogeley, 2013; Van der Wel, Sebanz, & Knoblich, 2014). It is this type of social affordance we see being enacted in this committee scenario. Once again, affordances, albeit "social affordances," are in play when determining the perspective an individual chooses when describing a visual scene.

In the fifth experiment (mouse tracking with either a person or no-person present) I attempted to tease apart this subtle but graded competition account of perspective taking. Using mouse tracking (e.g., Freeman & Ambady, 2010; Freeman, Dale, & Farmer, 2011) as a way to watch that competition unfold over a short time period, one might see how individuals are drawn toward an other-centric perspective even when ultimately choosing an egocentric perspective.

Once again, individuals selected the other perspective more often when the person is present than when there is no person. Another interesting aspect of this experiment involved the initiation times of participants who took the other-centric perspective more quickly when there was a person in the scene compared to when there was no person in the scene. A final aspect of this experiment worth detailing is the difference in mouse trajectories in that individuals are drawn toward the other-centric perspective in the person present condition even when they choose the egocentric perspective. This happens toward the end of the trial and not during the initiation time when they seem to be deciding which perspective to use. This is of notable interest because with a two-step processing account where the default response would be to take the egocentric perspective and then take a second perspective, one would expect to find the opposite. It is as if the other person in the photograph provides an anchor in the opposite direction, you are competing between two perspectives and you are drawn toward either the other person's perspective or your own. The other measures such as area under the curve, maximum deviation and reaction time are basically constant. However, the maximum deviation time should be mentioned, the time in milliseconds at which point a trajectory maximally deviates toward the other-centric response is also faster when a person is present compared to no person. Participants deviated later in time when there was no person in the scene, meaning they are drawn toward the other response later in the trials than those viewing a person in the scene. Again, this supports the continuous processing account of perspective taking.

This line of research has provided a detailed account of how perspective taking is a competition of two (or multiple) perspectives and not a two-step process. It is likely the case that there is an egocentric bias, especially if coming from an embodied approach (Gibson, 1979) to cognition. However, I have shown that agency, highlighted action language, and social context all provide evidence that points toward a graded spatial attraction of one perspective and that individuals settle into one of those competing perspectives when presented with a visual scene. This research was carefully constrained to provide an interpretation of the visual scene that had two competing perspectives, egocentric and other-centric. While we often see results that people take an egocentric perspective, when provided with agency, action language, and social contexts that are consistent with the other-centric perspective people will choose that other perspective more often. If there were an egocentric default it would not be likely that we would see differences in individuals' responses otherwise.

This research includes aspects that the two-step account camp would need to accommodate in order to verify their version of perspective taking. First, there are factors such as agency and affordances that invoke another perspective more often than when such factors are not present. Next, there are "social affordances" that provide different perspective taking descriptions when there is an intersubjective reality shared and are bound up with normatively constrained social practices. Last, there are quicker initiation times when a person is present compared to when there is no person in the scene, suggesting that the other perspective can actually provide an anchor linguistically, visually and cognitively to easily settle into.

Together the experiments support a theory for perspective taking that involves a continuous spatial attraction model where two (or more) perspectives are pitted against one another and individuals settle into one over time. This differs from the two-step egocentric default approach in perspective taking that has formerly dominated the literature. While it may be true that if taking an embodied approach, the egocentric perspective (or at least in individualistic cultures) is weighted more heavily. Data suggest there are tendencies to be drawn toward the egocentric bias, but it does not appear that the underlying mechanisms involve taking two steps for reaching a chosen perspective.

Figure 1: Perspective-taking example. Which coffee mug is yours?

Figure 2a

Figure 2b

Figure 2c

Figure 2: Visual Stimuli for Tversky & Hard, 2009

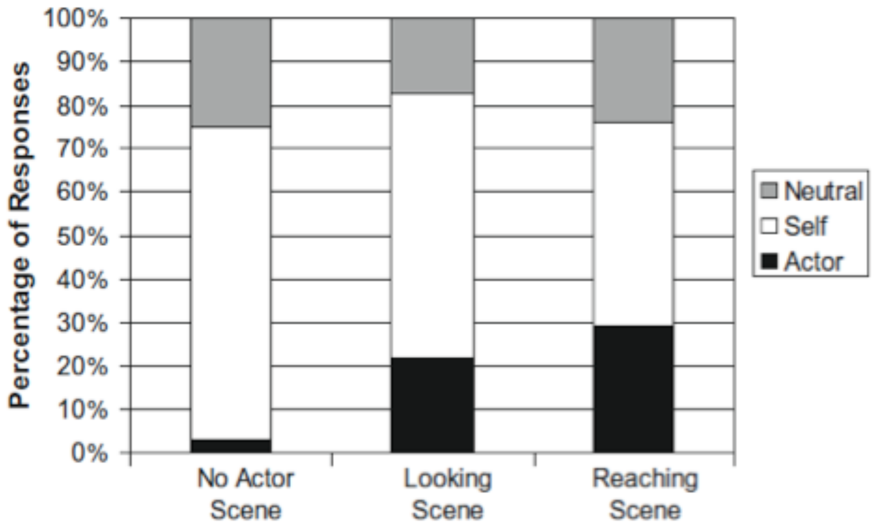

Figure 3a: Tversky & Hard, 2009; Mean response from an other, self, or neutral perspective as a function of scene.

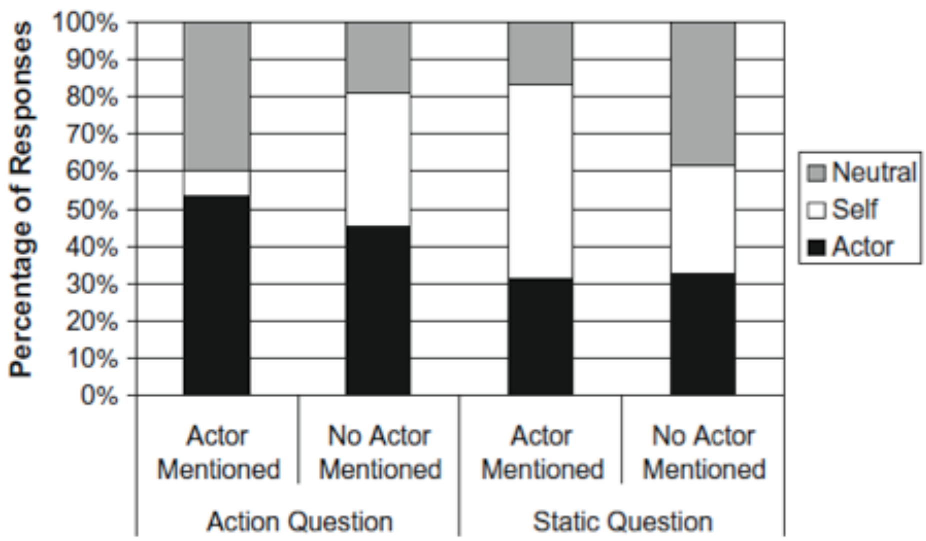

Figure 3b: Tversky & Hard, 2009; Mean response from an other, self, or neutral perspective as a function of question type.

Figure 4a: Control

Figure 4b: Robot Egocentric

Figure 4c: Robot Left

Figure 4d: Robot Right

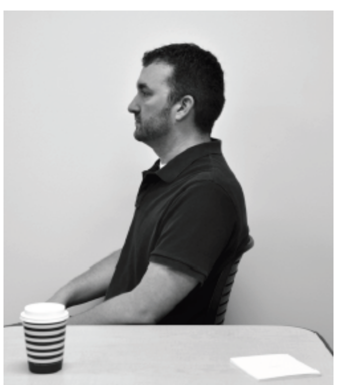

Figure 4e: Person Left

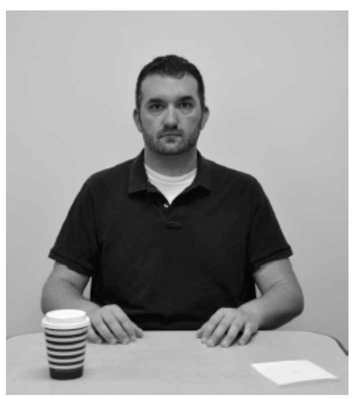

Figure 4f: Person Egocentric

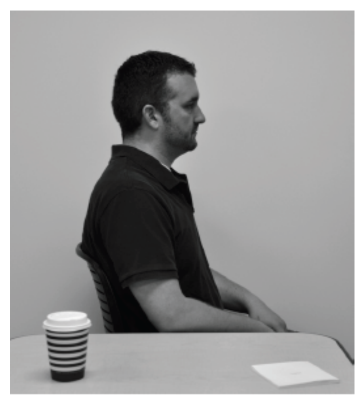

Figure 4g: Human Right

Figure 4:Visual stimuli for Experiment 1. a) control b) egocentric c) other-centric.

Task: In relation to the cup where is the book placed?

Figure 5a: Control

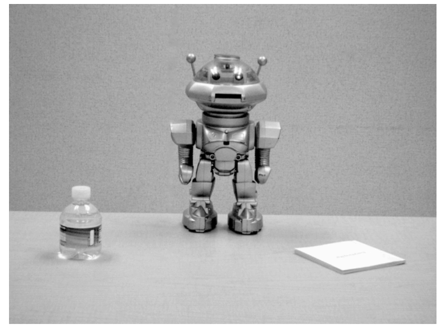

Figure 5b: Robot

Figure 5: Visual stimuli used in Experiment 2. Participants were given either photograph a) “Control” or b) “Robot.”

Task: Free response description of one of the following written stimuli:

In relation to the bottle, where is the book placed?

In relation to the bottle, where is the robot’s book?

In relation to the bottle, where does the robot place the book?

In relation to the bottle, where is the book?

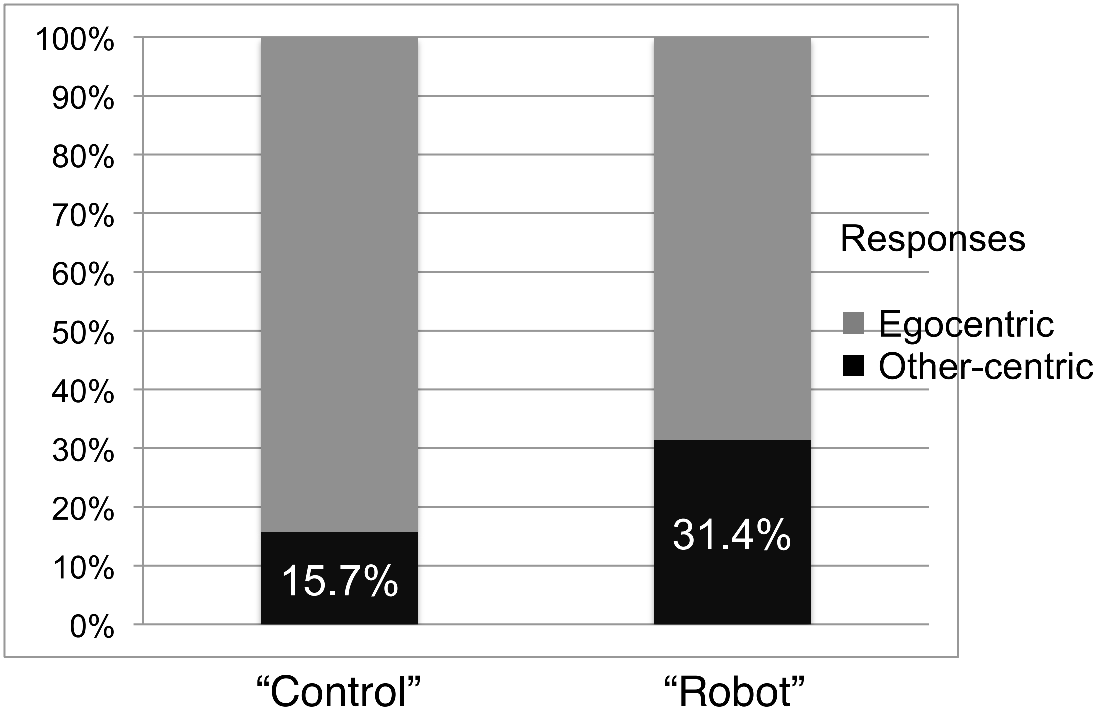

Figure 6: Experiment 2 results. Responses for two conditions, one using the "Control" photograph (see Figure 5a) and the "Robot" photograph (see Figure 5b).

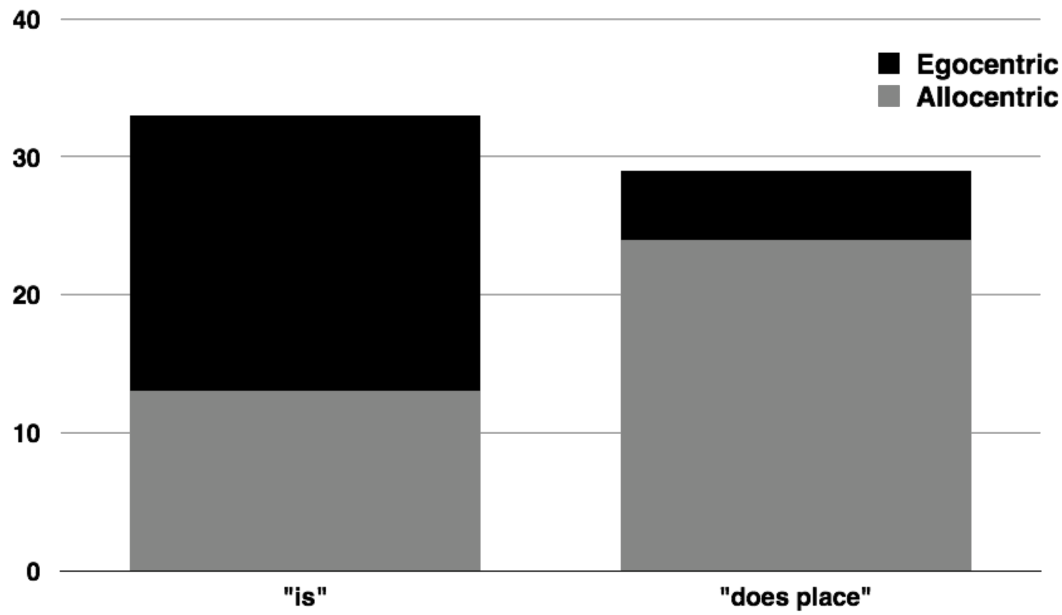

Figure 7: Experiment 2 results. Responses for either question "is" or question "does place": "In relation to the bottle, where does the robot place the book?" Versus "In relation to the bottle, where is the robot's book?

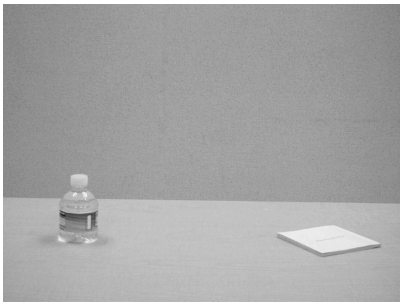

Figure 8a: Control

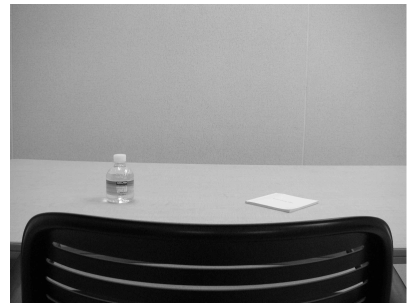

Figure 8b: Egocentric

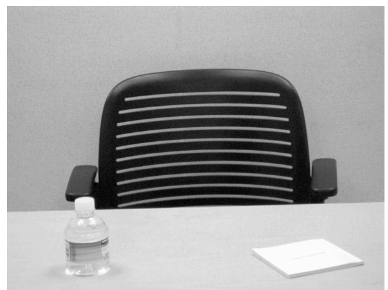

Figure 8c: Other-centric

Figure 8: a) Control b) Egocentric c) Other-centric. Task: Free response description of "In relation to the bottle where was the book placed?"

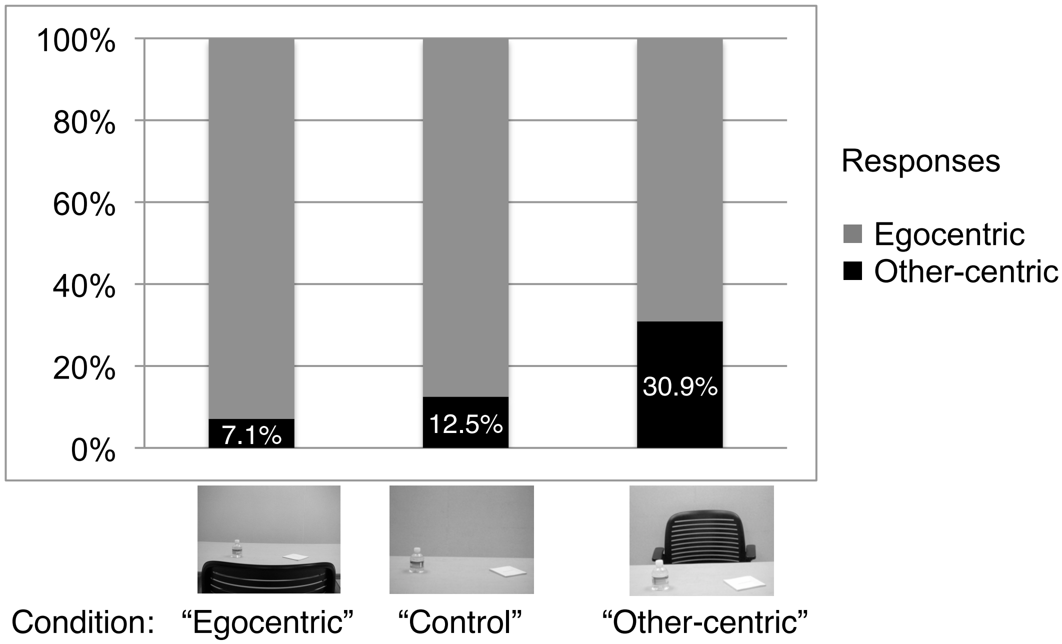

Figure 9: Experiment 3 results. Responses for conditions “Egocentric,” “Control,” and “Other-centric.”

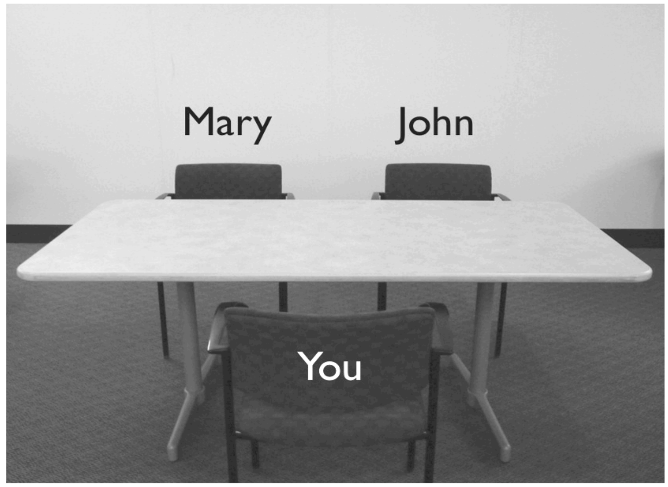

Figure 10: Experiment 4 Stimulus. Conditions: 1) You do not know John, and have never talked to him. 2) You have seen John, but have not talked to him. 3) You are really good friends with John, and talk to him often. Task: Describe where Mary is sitting.

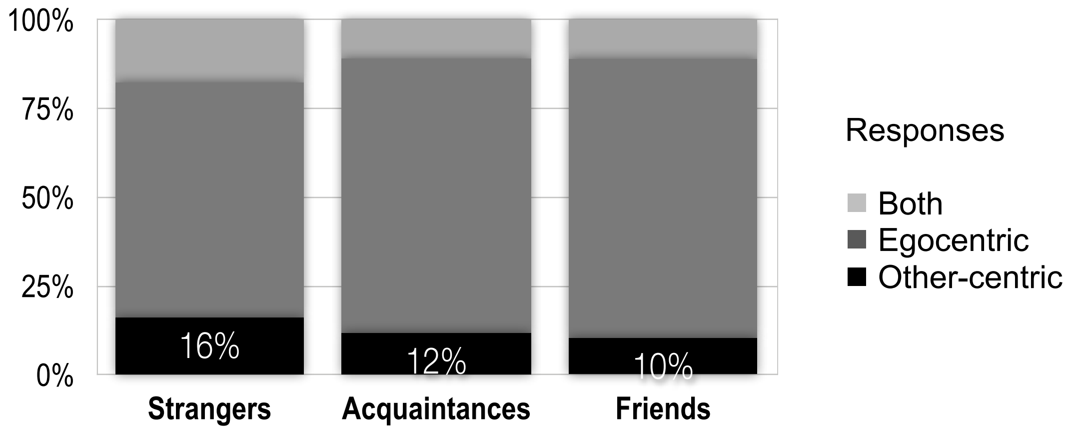

Figure 11: Experiment 4 results. People take "John's" perspective most often in the "Strangers" condition (see Figure 10 for visual stimuli).

Figure 12: Experiment 5 visual stimulus for the person condition.

Figure 13: Experiment 5 visual stimulus for no-person condition.

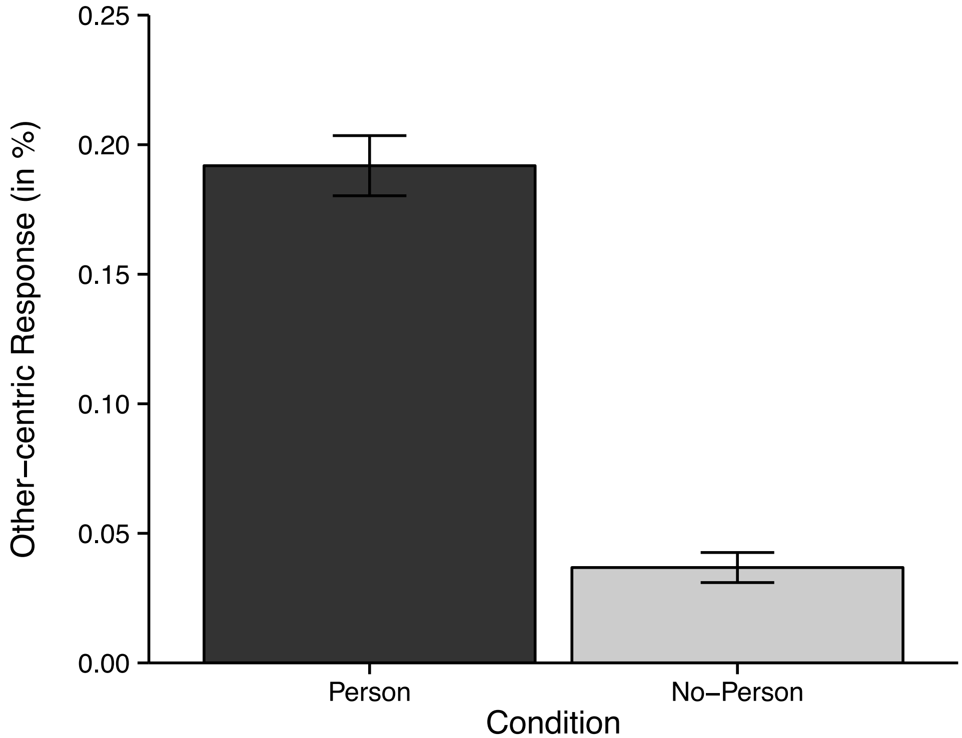

Figure 14: Results for Experiment 5, Response by condition. More people take the other-centric perspective in the person condition.

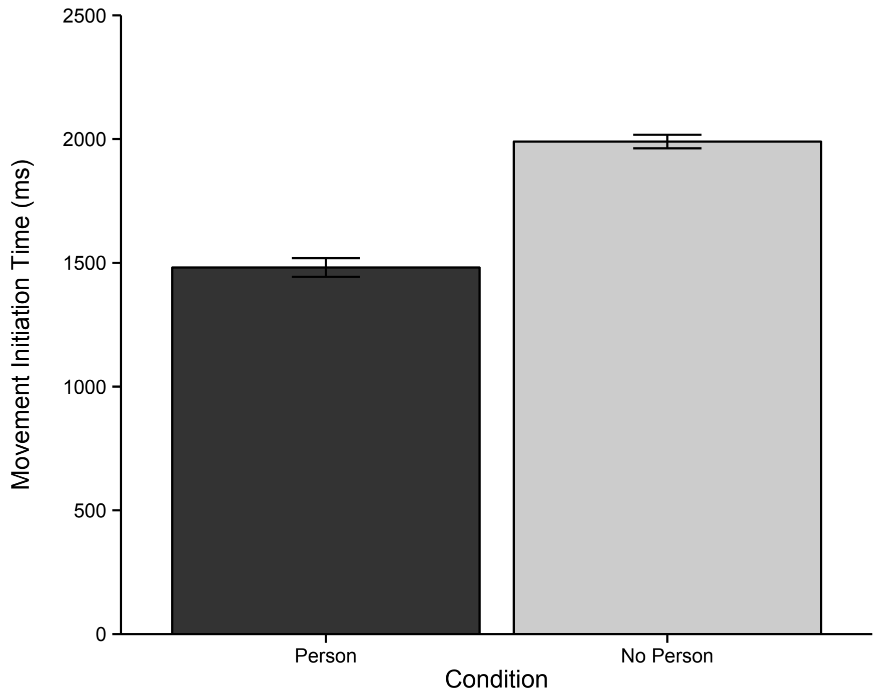

Figure 15: Results for Experiment 5, Initiation Time by condition.

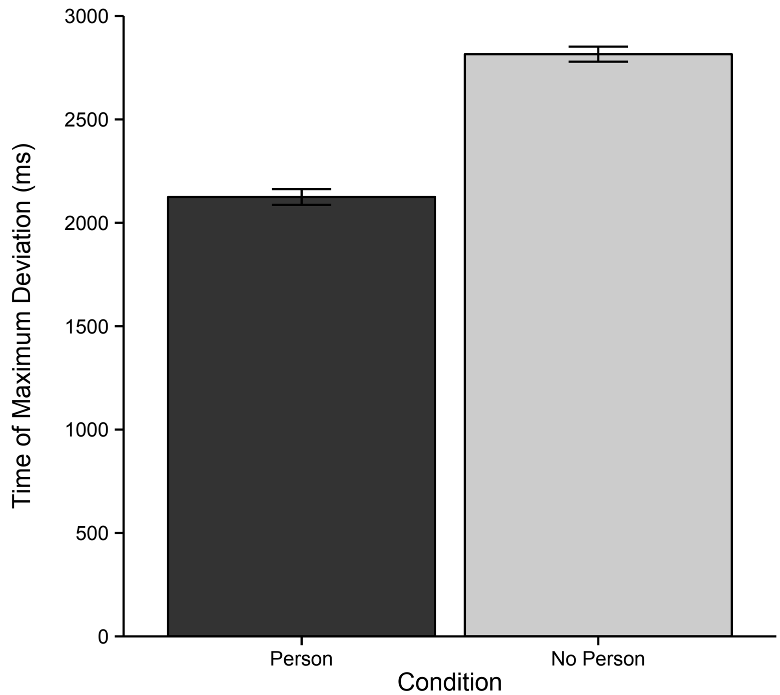

Figure 16: Results for Experiment 5, Time of Maximum Deviation by condition.

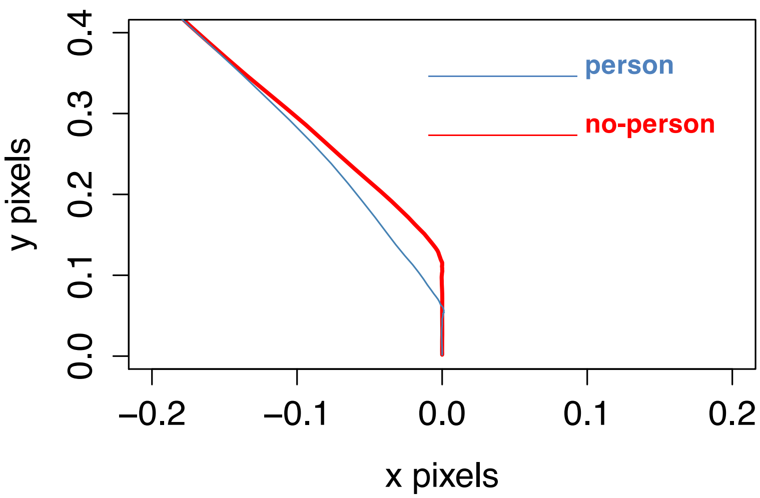

Figure 17: Experiment 5

Trajectories of computer mouse movements by condition.

Table 1: Table of Experiment 5 stimuli. Each object is an individual trial, repeated for "left", "right", and "inside," and randomized for each participant. They were given a recorded instruction to move these objects as a picture of that object appeared for them to drag and drop either to the left, right, or inside the basket.

Objects (alphabetically)	Objects (continued)
Apple	Matches
Berries	Mug
Bowl	Nuts
Candle	Orange
Candy	Pizza
Chips	Pliers
Clip	Roll
Cup	Salt
Gum	Scissors
Keys	Tape

Table 2: Table of distributional analysis of the mouse tracking Experiment 5 for the significant variables of initiation time (IT) and time of maximum deviation (MDT).

Statistics for the response distributions for each experimental group and condition.				
	Person Condition		**No Person Condition**	
	IT	*MDT*	*IT*	*MDT*
Trails	1131	1131	1086	1086
Skew	0.09	0.55	-0.37	-0.78
Kurtosis	-1.49	-0.77	-0.98	3.13
Bimodality	**0.664**	0.312	0.426	0.064