In Search of Triggering Conditions for Spontaneous Visual Perspective Taking

Xuan Zhao (xuan_zhao@brown.edu)
Corey Cusimano (corey_cusimano@brown.edu)
Bertram F. Malle (bertram_malle@brown.edu)

Department of Cognitive, Linguistic, and Psychological Sciences, Brown University,
190 Thayer Street, Providence, RI 02912 USA

Abstract

Visual perspective taking (VPT) – people's ability to represent the physical world from another person's viewpoint – plays a fundamental role in social cognition. However, little is known about whether and when VPT can be triggered spontaneously without any explicit verbal prompting. In six studies, we measured spontaneous VPT as the tendency to read an ambiguous number from another agent's imagined perspective ("6") rather than from one's own default visual perspective ("9"). We found that the likelihood of spontaneous VPT varied systematically with the target agent's behavior. The strongest trigger for spontaneous VPT was the agent's goal-directed reaching, followed by object-directed gaze, and lastly the agent's mere presence in the scene. Furthermore, observing an agent's reaching or gaze toward an object triggered viewers' spontaneous VPT even for objects with which the agent was currently not engaged.

Keywords: visual perspective taking; nonverbal behaviors; social cognition; self; egocentric; theory of mind.

Introduction

Physical space sets minds apart. What is visible to one person might be occluded from another person's view, and what one person sees as a figure "6" might appear to be a "9" to the viewer from the opposite. To overcome such differences in point of view, humans have evolved the capacity for visual perspective taking (hereafter, VPT). With this capacity, people determine the visibility of an object to another person ("Level-1 VPT") or its visual aspects relative to that viewpoint ("Level-2 VPT") (Flavell, Everett, Croft & Flavell, 1981). Through VPT, people identify shared knowledge (Clark, 1992), establish common ground (Clark & Brennan, 1991), and resolve referential ambiguity in communication (Duran, Dale & Kreuz, 2011).

A considerable amount of research on VPT centers on the question to what degree and at what levels of accuracy people demonstrate VPT. While some evidence suggests that VPT is rare, effortful, and error-prone (e.g., Keysar, Barr, Balin & Brauner, 2000), other studies indicate that even young children readily see the world from another person's viewpoint (e.g. Moll & Meltzoff, 2011). However, previous research rarely studied the conditions under which alternative visual perspectives became salient. Instead, in many cases, explicit experimenter instructions simply *required* participants to take another person's viewpoint (e.g., Michelon & Zacks, 2007), a paradigm characterized as "instructed perspective taking" (Zwickel & Müller, 2013) . However, only when experimental settings allow participants to freely decide what perspective to take can we identify the favorable triggering conditions for spontaneous VPT – especially the more advanced Level-2 VPT, which people often have difficulty engaging in.

The current project therefore employs a free-response approach to investigate whether small differences in a target agent's nonverbal behaviors influence people's readiness to take that person's visual perspective. More specifically, we focus on gaze and reaching as potential triggers of spontaneous VPT, because both are taken as minimal signs of another person's mental agency: Another's gaze invites an inference of knowledge; another's reaching invites an inference of preference or desire (Woodward, 1998). Neither of the two, however, requires a shift in visual perspective; in fact, gaze has been shown to be powerful in guiding the observer's own attention toward the gazed-at object (Driver, Davis, Ricciardelli, et al., 1999), and reaching, according to a prominent view, triggers the observer's own action program of reaching (Rizzolatti, Fogassi & Gallese, 2001). So it would be by no means trivial if these signs of agency were able to trigger VPT – as if by merely recognizing other minds, human perceivers were ready to adopt their point of view. Beyond the general power of these triggering conditions, we further hypothesized that goal-directed reaching would be a more effective trigger than gaze because it conveys a stronger and clearer intention to causally alter the physical environment. Research on the spontaneous activation of spatial perspective taking also lends support to this hypothesis, as people tend to describe an object's physical location from another person's viewpoint upon seeing that person's goal-directed reaching (Tversky & Hard, 2009).

Study 1a: Spontaneously Seeing a "6"

Methods

Stimuli. To capture people's spontaneous VPT, we created a single-trial task in which naïve participants viewed one of four photographs depicting a young male sitting at a table with neutral facial expression (Figure 1). Placed on the table was a red wooden digit "9", which could also be read as a "6" from across the table. All photographs were taken with a 20° angle down upon the actor and the table, so that both the number and the actor's movements were clearly visible to the participants.

Design. There were four conditions in this between-subjects study, where the actor was either 1) looking away from the

object, thus being merely present in the scene (*Presence*), 2) gazing at the object (*Gaze*), or 3) reaching for while gazing at the object (*Reaching*). In a control condition, 4) neither the actor nor his chair was present in the scene (*No Actor*) (See Figure 1).

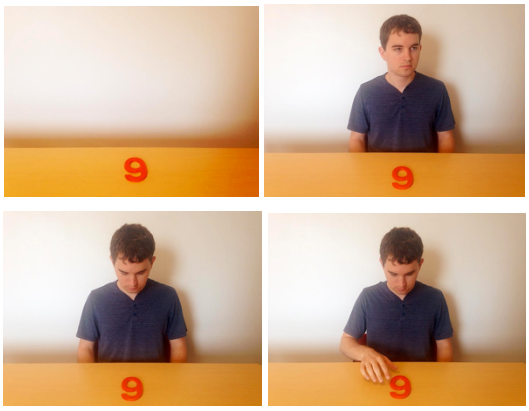

Figure 1. Four conditions in Study 1: No Actor (control), Presence, Gaze, and Reaching.

Procedures. All participants were recruited on Amazon Mechanical Turk and were randomly assigned to one of four conditions. After providing consent, each participant saw a photograph and a question below: "What number is on the table?" Participants typed their answers in a text box below the question and clicked "continue" to submit their answer. We recorded a total response time (TRT) between the participant's opening the photograph page and clicking on the "continue" button. On the next page they provided demographic information and received a payment code.

Results

Twelve participants who had TRTs three standard deviations beyond the mean of their respective conditions were removed from further data analysis (a criterion to exclude outliers in all studies in this paper). Of the remaining 236 participants (mean age = 29.7, 46% females, N = 56-64 per condition), all answered either "6" or "9". A response of "6" counted as spontaneous VPT, while "9" counted as seeing from a "self perspective."

A logit analysis with Helmert contrasts showed that, compared to the control condition where no actor was present, the three actor-present conditions elicited significantly higher VPT rates, $z = 3.4$, $p < .001$ (see Figure 2). Compared to the mere presence condition (12.5%), the average of gaze and reaching conditions elicited a significantly higher VPT rate, $z = 4.0$, $p < .001$, while the gaze condition (42.1%) and the reaching condition (45.8%) did not differ from one another.

A one-way ANOVA on TRTs for only those VPT trials (where people answered "6") revealed that people were significantly faster in taking the actor's perspective when he was reaching for the number (M = 14.6s) than when he was gazing at the number (M = 11.6s), p = .019 (See Figure 2).

TRT data also seemed to suggest that participants in the gaze and reaching conditions taken together spent less time on taking the actor's perspective than those in the presence condition (M = 15.8s), but there were only eight successful VPT trials in the presence condition, and the TRT difference was not significant, p = .119.

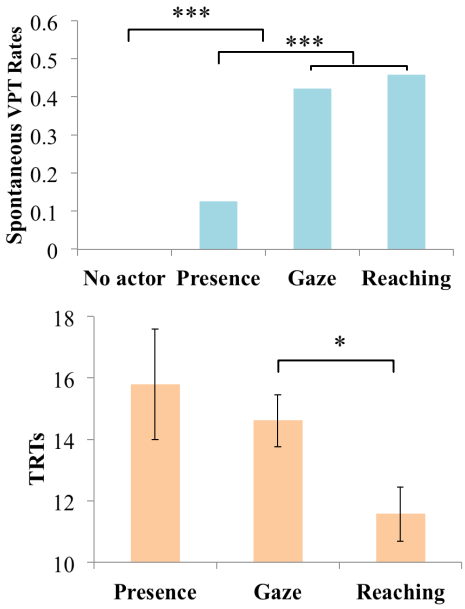

Figure 2. Spontaneous Visual Perspective Taking (VPT) rates (above) and mean Total Response Times (TRT) (below) in Study 1a.

Study 1b & 1c: Instructed Self-perspective and Other-perspective (VPT) Judgments

To better interpret the spontaneous VPT rates in Study 1a, we conducted Studies 1b and 1c. We aimed to assess the lower and upper bounds of people's VPT rates when people were explicitly asked to take either their own or someone else's perspective.

In Study 1b, when participants were asked to report a number from their own visual perspective, those who still reported a number from the actor's viewpoint can be characterized as spontaneous VPT. In Study 1c, when participants were explicitly requested to report a number from the actor's perspective, those who followed the instruction and answered "6" fit the category of instructed VPT, and their performance represented the optimal VPT performance people could achieve in the current setup.

Methods

Studies 1b and 1c applied the same stimuli, design, and procedures as Study 1a, except for the free-response questions. Instead of asking the perspective-neutral question "What number is on the table?", Study 1b asked "What number can you see?", and Study 1c asked "What number can he see?" The actor-absent control condition was omitted in Study 1c because asking what another person could see was meaningless with no one being present.

Results

In Study 1b, 236 out of 249 participants (mean age = 31.0, 54% females) entered data analysis (N = 55-63 per condition). No participants in the presence or gaze condition provided other-perspective judgments, while one participant in the control condition mentioned both perspectives. In the reaching condition, by contrast, 17.2% of participants took the actor's perspective and answered "6", even though they were explicitly asked to adopt their own perspective. In a logit analysis, the spontaneous VPT rate in the reaching condition was significantly higher than that in the remaining conditions, z = -3.66, p < .001.

In Study 1c, 189 out of 204 participants (mean age = 30.1, 50% females) entered data analysis (N = 55-63 per condition). VPT rates of presence, gaze, and reaching conditions were 73.8%, 80.0%, and 69.0%, respectively. Neither VPT rates nor TRTs of the VPT trials in the three conditions significantly differed from each other.

Discussion

Three preliminary conclusions regarding spontaneous VPT can be drawn from Studies 1a, 1b & 1c.

First, Study 1a has shown that, compared to a person's mere presence, the person's gaze and reaching behaviors significantly increased the observer's tendency to take the actor's perspective. Second, as suggested by TRT differences between the gaze and reaching conditions in Study 1a and the high spontaneous VPT rate (17.2%) in the reaching condition in Study 1b, goal-directed reaching may be a more effective trigger than goal-directed gaze. Third, as indicated by Study 1c, when people were explicitly instructed to take another person's perspective (the typical "instructed VPT" paradigm), the three conditions no longer differed in their effectiveness in triggering VPT. This suggests that an instructed VPT paradigm could completely obscure the differences among VPT triggering conditions. Considering the important role VPT plays in action coordination and social interaction (and the fact that it is rarely verbally requested by the other interactant), future research should more closely heed the distinction between spontaneous and instructed perspective taking processes.

Although VPT can be induced spontaneously, Study 1c also provided evidence that VPT requires extra effort: An average VPT rate of 74.3% seems underwhelming when the VPT task is explicit and straightforward. In this light, the high *spontaneous* VPT rates in the gaze and reaching conditions in Study 1a are all the more impressive.

These studies confirm the hypothesis that people's propensity for spontaneous VPT varies as a function of an observed agent's specific behaviors, and triggering VPT becomes more effective from mere presence to goal-directed gaze to reaching. However, one question arises from the current paradigm: Do perspective-takers describe the number from the actor's visual perspective only when he is currently engaging with it? In other words, would people stop to represent the number from his viewpoint when his gaze or reaching behavior is directed towards an irrelevant object? If so, then the spontaneous VPT activated by an actor's goal-directed behavior is *object-specific*. However, if people still describe the number according to the actor's viewpoint even when his action is not directed towards that number, then VPT is activated *globally*, where people also take the actor's perspective to represent other objects of potential engagement.

To test whether goal-directed gaze and reaching activate spontaneous VPT in a global or object-specific fashion, we adjusted our paradigm for Study 2. In particular, because the reaching condition in Study 1 also encompassed gaze towards the same object, we disassociated them in Study 2 to investigate their individual triggering effects.

Study 2a & 2b: Spontaneous VPT Activated by Gaze and Reaching: Global or Object-Specific?

When people know where an object is located, they often reach for it while looking at another object. Inspired by such natural movements, we made gaze and reaching behaviors entirely independent. To this end, a horizontally symmetric diamond was placed on the table beside the familiar 6/9 number (see Figure 3). The actor could look at either the diamond or the number while simultaneously reaching for either object, resulting in four gaze-reaching combinations. In addition, he could merely look at either object without reaching, resulting in two gaze-only conditions.

If gaze activates spontaneous VPT globally, then people should show similar VPT rates when the actor is looking at either the diamond or the number; similarly, if reaching triggers VPT globally, then people should show similar VPT rates when the actor reaches for either object. Object-specific activation, on the other hand, predicts that VPT rates should drop significantly when the goal-directed behavior (either gaze or reaching) is directed at the diamond rather than the number.

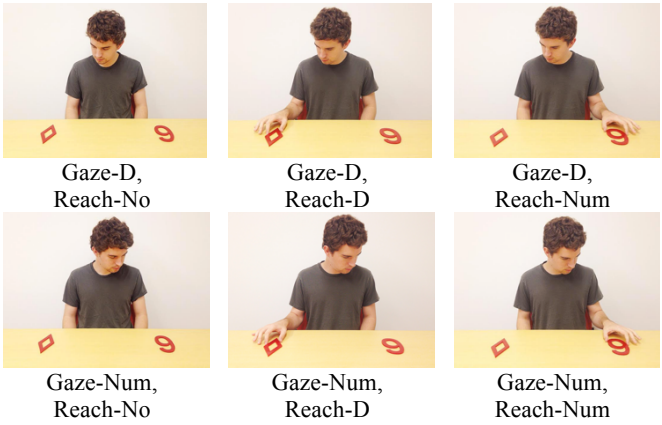

Figure 3. Six experiment conditions in Study 2a and 2b. "D" = Diamond. "Num" = the Number "6"/"9". The mere presence condition (baseline) is not shown above.

Methods

Stimuli. Photographs in Study 2a and 2b were very similar to those in Study 1 except that they displayed both a digit "9" and an equally sized, horizontally symmetric diamond shape on the table. These two symbols were placed on two sides of the table in front of the same actor (Figure 3).

Design. There were seven conditions in both studies. In the baseline condition, the actor was looking aimlessly to his left without goal-directed gaze, which was the same as the presence condition in Study 1. Other conditions were the six combinations of the actor's gaze direction (diamond vs. number) and his reaching behavior (no reaching, reaching for the diamond, reaching for the number). We tested the impact of the reaching manipulation by way of two Helmert contrasts: no reaching vs. some reaching, and reaching for diamond vs. reaching for number.

Procedures. Focusing on the activation conditions of spontaneous VPT, we again asked participants to provide free responses to the question "What number is on the table?" (Study 2a) and to the question "What number can you see?" (Study 2b). The remaining procedures were the same as those in Study 1.

Results

In Study 2a, 660 out of 688 participants (mean age = 31.3, 57.6% females) entered data analysis (N = 88-98 per condition). All participants answered either "6" or "9", except one who answered from both perspectives. VPT rates are shown in Panel A of Figure 4. A 2 × 3 logit analysis found no main effect of gaze direction (toward diamond or number) but a main effect of reaching behavior. The first Helmert contrast showed that people were more likely to take the actor's perspective when he was reaching for something than not reaching at all (23.3%), z = 4.47, p < .001; the second Helmert contrast showed that people were more likely to take the actor's perspective when he was reaching for the number (50.5%) rather than the diamond (35.7%), z = 2.90, p = .004. No interaction between gaze direction and reaching behavior was found.

A two-way ANOVA on TRTs of VPT trials found that people were marginally faster to generate an other-perspective answer when the actor exhibited some reaching behavior than when he was not reaching at all, p = 0.085. No other effects were significant.

In Study 2b, 384 out of 420 participants (mean age = 30.2, 49.5% females) entered data analysis (N = 51-61 per condition); all but three participants answered either "6" or "9". VPT rates are shown in Panel B of Figure 4. According to a 2 × 3 logit analysis, people were overall more likely to take the actor's perspective when his gaze was directed at the number (14.8%) rather than the diamond (11.6%), z = 2.12, p = .034. VPT rates also showed a significant main effect of reaching behavior: according to the first Helmert contrast, people were more likely to take the actor's perspective when he was reaching for something than not reaching at all (5.2%), z = 2.63, p = .008; and the second Helmert contrast found that people were significantly more likely to perform VPT when he reached for the number (27.0%) rather than for the diamond (7.4%), z = 3.56, p < .001. In addition, no interaction effect was found between the diamond-reaching vs. number-reaching contrast and gaze direction.

Two-way ANOVA on TRTs in Study 2b revealed only one significant effect: people were faster in VPT trials when the actor was reaching for the number than for the diamond, p < 0.01. Neither main effect of gaze direction nor any interaction effects were significant.

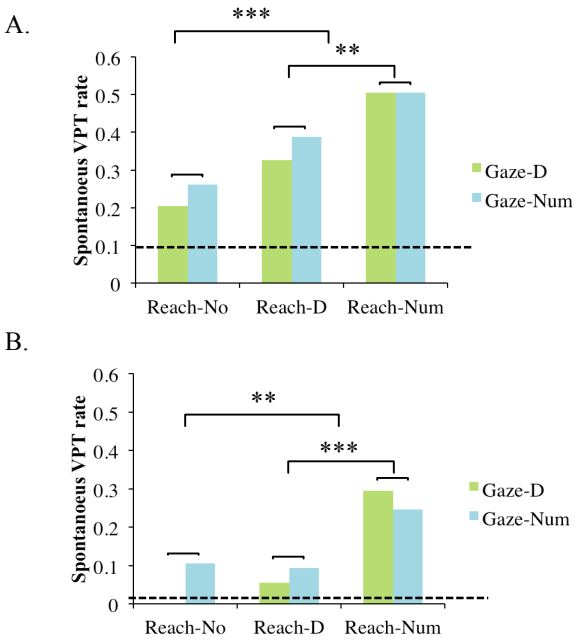

Figure 4. Panel A: Spontaneous VPT rates in Study 2a. Panel B: Spontaneous VPT rates in Study 2b. Dashed lines indicate VPT rates in baseline conditions where the actor was merely present. The six bars represent the 2 × 3 design of two gaze directions (diamond vs. number) and three reaching behaviors (no reaching, reaching for the diamond, reaching for the number).

Discussion

Both studies confirmed the conclusions of Studies 1a & 1b: Reaching was more effective than gaze at triggering spontaneous VPT. More importantly, these studies disassociated the triggering effects of gaze and reaching to reveal whether they activated spontaneous VPT in a global or object-specific fashion.

The triggering effect of reaching was always the strongest when it was directed toward the number and weakest when it was not displayed at all. However, when reaching was directed toward an irrelevant, adjacent object on the same table (the diamond), people's VPT rates were neither as high as those in the number-reaching condition, nor were they as low as when reaching was absent. This suggests that the activation of spontaneous VPT was not entirely object-specific, because people were still inclined to take the

actor's perspective to describe other things he might potentially reach for; and it was also not entirely global, because VPT rates were still highest for the object he was directly engaging with.

The conclusion on the triggering effect of gaze contains more nuances. Study 2a found that people had similar VPT rates and TRTs regardless of the object the actor looked at, suggesting that gaze activated spontaneous VPT globally. Study 2b also found no differential effect of gaze direction on TRTs, yet there was a main effect of gaze direction. However, a close examination on VPT rates in Panel B of Figure 4 seems to suggest that the effect of gaze direction only survived when reaching was absent. In other words, with the presence of a stronger cue such as reaching, people were not sensitive to gaze direction; however, when gaze was the only trigger, people showed sensitivity to its specific target – just like reaching.

Study 3: Problem Solving by Taking Visual Perspective

Without being explicitly instructed, a considerable number of participants provided descriptions of the target number from another person's perspective, and we take such descriptions as evidence for the "spontaneity" of VPT. However, there is an alternative interpretation: Participants in all conditions may have recognized another person's perspective but deliberately selected from the two perspectives the "right" one, considering their interpretation of the scenes and their perception of the experimenter's expectations. According to this explanation, participants still spontaneously took the actor's visual perspective in the first place, but their differential responses to the various triggering conditions might have been more deliberate than spontaneous.

To eliminate the possibility that differences among triggering conditions were a mere reflection of participants' deliberate effort to infer a likely answer, we designed a task where there was one objectively correct answer that could be accessed only if the participant spontaneously and successfully represented another person's visual perspective. Once participants take the other's perspective, they would recognize that this perspective provides the right answer, and they would no longer consider an answer based on their self-perspective. Therefore, being able to provide this answer would serve as a reliable indicator of genuinely spontaneous VPT, and differential rates due to presence, gaze and reaching would indicate the inherent triggering effects of these conditions.

Methods

Stimuli. We created a scenario resembling that in previous studies, but instead of two objects on the table, the actor faced four numbers – 86, 87, 88, and 89 – with equal distance between them and the number "87" covered under a piece of white paper (Figure 5). Critically, the visible numbers were horizontally symmetric and therefore did not reveal their orientation when viewed from either direction; however, seeing a clear pattern in the displayed numbers requires participants to take the actor's visual perspective.

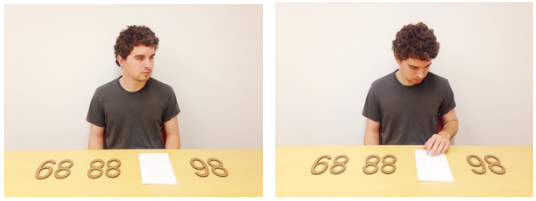

Figure 5. Presence and reaching conditions in Study 3. Gaze and control conditions were analogous to previous studies.

Design. As in Study 1a, participants were randomly assigned to one of four conditions. In the *presence* condition, the agent was looking aimlessly to his left. In the *gaze* condition, he was looking at the paper that covered "87". In the *reaching* condition, he was reaching for the paper. In a control condition, neither the person nor his chair was present (*no actor*).

Procedures. Amazon Mechanical Turk participants saw one of the four photographs and a question below, "What number is under the paper?" They typed their answers into a text box. After submitting their answers, participants were asked to indicate what computer device they used to complete the study, whether they turned their device upside down to view the photograph, and whether they had seen similar questions in the past. They also provided basic demographic information.

Results

432 participants completed this study. Those who had seen a similar puzzle before, who turned their device upside-down to view the photograph, and those whose TRTs were three standard deviations beyond the mean of their respective conditions were discarded, resulting in 377 participants in data analysis (mean age = 30.8, 59.2% females, N = 88-109 per condition). The percentages of participants who correctly answered "87" are shown in Figure 6. Notably, 66.3% of the participants gave the answer "78", which was the likely conclusion when one saw "68 88 ( ) 98" from one's own perspective.

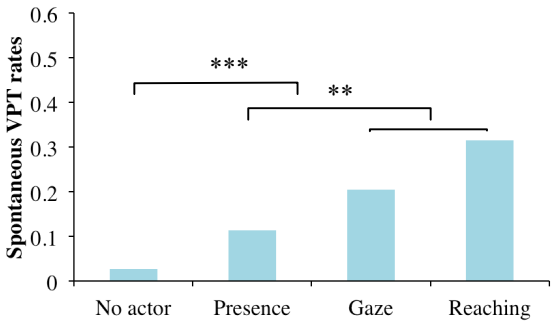

Figure 6. Spontaneous VPT rates in Study 3.

A logit analysis with Helmert contrasts showed that, compared to the minimal VPT rate in the baseline condition when no actor was present (2.8%), the three actor-present conditions induced significantly higher VPT rates, $z = 3.61$, $p < .001$. Compared with the mere presence condition (11.4%), the average of gaze and reaching conditions induced a significantly higher VPT rate, $z = 2.58$, $p = .01$. Finally, the gaze condition (20.4%) and the reaching condition (31.5%) differed marginally from each other, $z = 1.66$, $p = .10$.

One-way ANOVA on TRTs of trials in which participants correctly answered "87" did not reveal any significant effects. However, only a fraction of participants in each condition provided the correct answer of "87", resulting in only 3 to 29 data points per condition. A closer observation of the trend in TRTs across different conditions revealed that TRTs seemed to decrease from presence (21.6s) to gaze (19.0s) to reaching (18.4s).

Discussion

Study 3 deployed a problem-solving task that was less subject to participants' deliberate selection between two potential perspectives and more effective in capturing the activation of spontaneous VPT. The difficulty of the task lowered overall VPT rates, but it showed, as previous studies, that the proportion of people who took the other person's visual perspective increased from baseline to mere presence to gaze and then goal-directed reaching. Although only marginally significant, people's VPT rates tended to be even higher for reaching than for gaze.

General Discussion

In the present studies, we measured spontaneous VPT as participants' tendency to read an ambiguous number from another agent's perspective ("6") rather than from their own perspective ("9"). We found that the mere presence of the agent activated a low level of VPT; object-directed behaviors such as gaze and reaching markedly increased spontaneous VPT; and of those, reaching was more effective than gaze as a VPT trigger. In addition, observing an agent's goal-directed gaze or reaching toward one object triggered VPT even for objects with which the actor was currently not engaged.

A more general message our project aims to convey is that research on VPT needs to look beyond the debate on people's capacity of perspective taking and instead study social and contextual triggers that give rise to its activation. By taking a dynamic approach and viewing VPT as a cognitive tool that is more readily available under certain conditions, our project provides one initial step towards such exploration. However, we limited our search scope to the most fundamental "mental agency" behaviors in this project; in all likelihood, there are additional behavioral and social contexts that might evoke VPT in people's daily interaction.

For example, future research may examine how other nonverbal behaviors, such as eye contact and referential pointing, can influence people's VPT tendency, and whether specific relationships between interactants influence their spontaneous adoption of each other's viewpoint in dyadic interaction or joint action. Future research should also expand from visual perspective taking to other types of perspective taking, such as understanding and predicting other people's beliefs, desires, and emotions, and examine whether similar triggers are responsible for the different kinds of perspective taking, and how these different kinds relate to one another at the level of cognitive processing.

References

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition.* Washington, DC: American Psychological Association.

Clark, H.H. (1992). *Arenas of language use.* Chicago: University of Chicago Press.

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze Perception Triggers Reflexive Visuospatial Orienting. *Visual Cognition, 6*(5), 509–540.

Duran, N. D., Dale, R., & Kreuz, R. J. (2011). Listeners invest in an assumed other's perspective despite cognitive cost. *Cognition, 121*(1), 22–40.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual-perception – Further evidence for the level 1–level 2 distinction. *Developmental Psychology*, 17, 99–103.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science, 11*, 32–38.

Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. *Perception & Psychophysics, 68*(2), 327-37.

Moll, H., & Meltzoff, A. N. (2011). How does it look? Level 2 perspective-taking at 36 months. *Child Development, 82*(2), 661-673.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, *2*(9), 661–670.

Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: spatial perspective-taking. *Cognition, 110*(1), 124-9.

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

Zwickel, J., & Müller, H. J. (2013). On the relation between spontaneous perspective taking and other visuospatial processes. *Memory & Cognition, 41*(4), 558–70.