Center for Brains, Minds & Machines

CBMM Memo No. **038** 10/26/2015

Predicting Actions Before They Occur

by

Maryam Vaziri-Pashkam, Sarah Cormiea, Ken Nakayama

Abstract:
Humans are experts at reading others' actions in social contexts. They efficiently process others' movements in real-time to predict intended goals. Here we designed a two-person reaching task to investigate real-time body reading in a naturalistic setting. Two Subjects faced each other separated by a plexiglass screen. One (Attacker) was instructed to tap one of two targets on the screen and the other (Blocker) was told to tap the same target as quickly as possible. Reaction times were fast, much faster than reaction times to a dot projected on the screen moving in the same manner. This suggests Blockers use subtle preparatory movements of Attackers to predict their goal. Next, using video recordings of an Attacker, we showed that removing the preparatory cues slows reaction times and changing them could trick the Blockers to choose the wrong target. We then occluded various body parts of the Attacker and showed that reaction times slow down only when most of the body of the Attacker is occluded. This suggests that preparatory cues are distributed over the body of the Attacker. We saw no evidence of learning during the experiment as reaction times remained constant over the duration of the session. Taken together, these results suggest that in social contexts humans are able to use their knowledge of the biomechanical constraints on the human body to efficiently process preparatory cues from the body of their interaction partner in order to predict their intentions well before movement begins.

This work was supported by the Center for Brains, Minds and Machines (CBMM), funded by NSF STC award CCF-1231216.

Introduction

To navigate the social environment we often need to predict the goals of other agents from their movements. Imagine greeting a friend with a handshake. You need to predict where they want to meet your hand. This prediction needs to be made early in the movement otherwise your friend's hand will be awkwardly waiting in the air until your hand gets there.

What makes these predictions possible? Human bodily movements follow specific patterns due to biomechanical constraints (Johansson, 1973). Moving a hand towards a target on a table for example may require lifting the elbow, abducting the arm and moving the trunk towards the target. Other more distributed adjustments are also necessary, to position the center of gravity of the body appropriately. Does the visual system have knowledge of these biomechanical constraints? Could they be used to predict the goals and future actions of others in a social context? Most studies of action anticipation have focused on competitive sports and have shown that athletes can process predictive bodily cues to anticipate the goals of their opponents. In a study of soccer penalty kicks, Diaz et al. (2012) showed that both local cues such as the angle of the non-kicking foot as well as distributed cues over multiple joints of a kicker are predictive of the kick direction well ahead of the foot to ball contact. Similar results have been found in tennis (Farraw & Abernethy, 2003), cricket (Muller, Abernethy & Farrow, 2006), badminton (Abernethy & Zawi, 2007), squash (Abernethy et al., 2001), baseball (Ranganathan & Carlton, 2007) and volleyball (Starkes et al., 1995).

However competitive sports is not the only situation in which action anticipation is necessary. Many everyday activities rely on the analysis of action goals and all of us are to some extent experts at reading common actions performed by other individuals. In these everyday activities movements are often more subtle and span smaller portion of the human body compared to larger movements in sports and as a results action anticipation might be more difficult in these activities. In addition, anticipation in competitive sports is often subject to explicit training. This is not present in common everyday interactions.

Despite its importance only a few studies in the literature have explored action anticipation in more common everyday movements. Louis-Dam et al. (1999) showed that in a two-step action in which actors reached for a target in order to move it to a new location, observers were able to predict whether the intended location was close or far based on the pattern of the initial reach towards the target. Similar predictive abilities were found in weight-lifting (Runeson & Frykholm, 1983), speech (Abry et al., 1994), writing (Kandel et al., 1993, Orliaguet et al., 1997) and sign language (Fennel et al., 1999). To further these findings in this article we explored action anticipation in a simple reaching task.

Previous studies of everyday action prediction have mostly relied on off-line reports of the subjects in response to videos of moving actors. In a typical experiment a video or a movement sequence is played and cut at various time points and subjects are asked to decide the future outcome based on partial information (Runeson & Frykholm, 1983). These psychophysical studies are limited because they allow the subjects time to reflect on what they have seen and are not representative of the interactions that occur in the everyday world. In everyday life reading of actions occur in the moment and subjects are continuously predicting the goals of others in order to decide about their own actions. Only by measuring the actions of observers responding to the movements of others in real-time can we hope to characterize such naturally occurring processes. In fact studies of competitive sports have shown that real-time responses in naturalistic settings might provide different results from off-line responses (Ranganathan & Carlton, 2007; Farraw & Abernethy, 2003).

In this study we introduce a real-time approach in studying action anticipation. Using a motion tracking device we measured the movements of one subject in response to another. This design allows for moment to moment analysis of the respondents' movements to determine if they anticipate their opponent's goals. The task was a competitive reaching task in which one subject (Attacker) had to choose a target and touch it with their finger and another (Blocker) had to block the same target by touching it soon after the attacker. We found that Subjects were surprisingly fast in responding to their opponent, much faster than two-choice reaction times with a similar design. Reaction times were fast from the beginning of the experiment with no need for training. In subsequent experiments we demonstrated that the Blocker could use predictive cues present well ahead of the finger movement of the Attacker to reduce their reaction time. We showed that removing the predictive cues slowed down the blockers and inaccurate cues tricked the blockers into reaching for the wrong target. In the next experiment we explored the location of the predictive cues and showed that they are distributed over various body parts of the Attacker. Together these results demonstrate that in a simple reaching movement humans are able to efficiently read out cues from multiple body parts of their opponent for movement anticipation and are able to readily use these cues to guide their own actions.

Results:

Experiment 1: Fast Reactions

Subjects sat facing each other separated by a plexiglass screen. One subject was assigned the role of Attacker and the other the role of Blocker. The Attacker was instructed to pick one of two targets and reach for it with their finger. The Blocker was instructed to reach for the same target and get there as fast as possible (Figure 1, also see Supplemental Video 1).

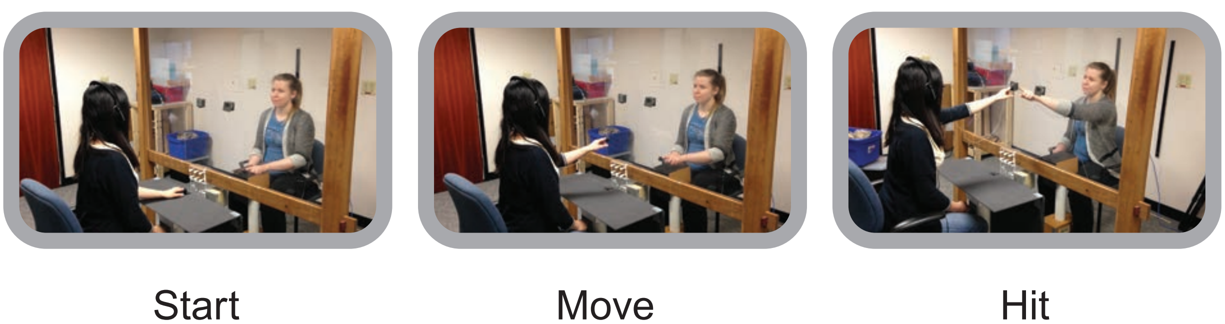


Figure 1. Set up of Experiment 1. Subjects completed a competitive reaching task while seated facing each other separated by a plexiglass screen. Two targets were affixed to the screen equidistant from both Subjects. Attacker is pictured on the left, Blocker on the right.

Each trial started by a "beep" sound audible to both Subjects and ended with auditory feedback indicating the winner of the trial (see Supplemental Methods). Figure 2c shows 30 sample trajectories of one pair of subjects. The Blockers' final accuracy was high (92.7% ± 6.29).

The reaction time (RT) for each trial was calculated as:

$$RT = T_b - T_a$$

Where T_a stands for the first point in time at which the speed of the Attacker in the 3D space exceeded a threshold and T_b stands for the first point in time at which the horizontal speed of the Blocker's finger towards their final target exceeded that same threshold (see Supplemental Methods). Figure 2a shows a sample trial in which the Blocker directly moved towards their final target. Figure 2b shows another more infrequent trial in which the Blocker changed their movement direction half way through the trial. In both cases T_b determines the start of the movement towards the final goal.

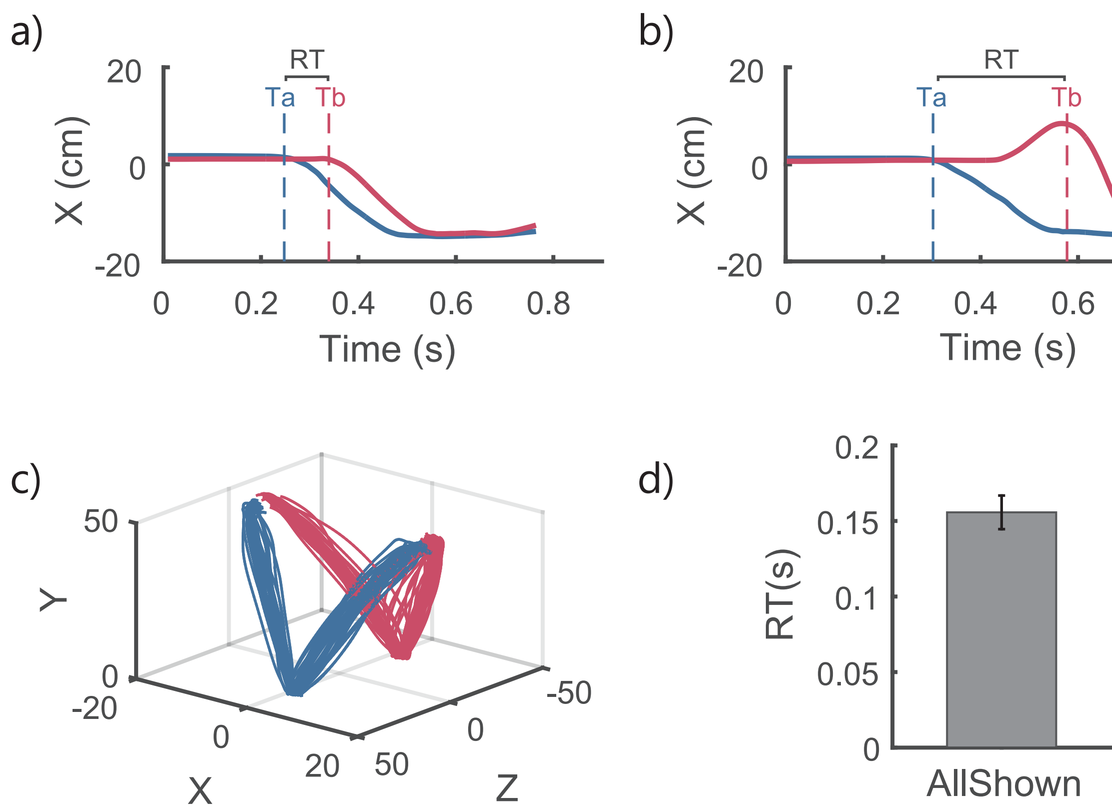


Figure 2. a) A typical trial, lateral position (X) of a subject plotted against time. The blue line depicts the Attacker's motion and the red line depicts the Blocker's motion. The dashed lines represent the start of the finger movements of the Attacker (T_a) and the Blocker (T_b). **b)** A less common trial, the Blocker changes direction mid-movement. **c)** 3D plot of all trials from a representative pair of subjects. **d)** Average RT in Experiment 1. The error bar shows ±SEM.

Results of this analysis revealed very fast RTs for the Blocker in response to the Attacker (155.7 ms, See Figure 2d), especially when compared to similar hand movement reaction times previously reported in the cognitive psychology literature (Song & Nakayama, 2006). To our knowledge this is the first time that such fast reactions have been reported in naïve Subjects in a competitive interaction. What is the source of these fast reaction times? We will address this question in the next two experiments.

Experiment 2: Human vs Moving Dot

The previous experiment demonstrated fast reaction times in response to the Attacker's finger movement. Are blockers only focused on the finger movement of the Attacker to achieve fast reactions or do they gather cues from other body parts of the Attackers? If the finger movement is the only source of information for the Blockers they should be equally fast reacting to any object that moves similar to the Attacker's finger. To investigate this possibility in Experiment 2 we measured the reaction times of the Blockers in response to a dot moving on a screen in an essentially identical manner to an Attacker's finger. We converted 3D Attacker movement paths in Experiment 1 to 2D paths and used it to position a moving dot on a display back-projected on the screen (see Supplemental Figure 1). We measured the RTs in response to this moving dot (Dot condition) and compared it to the RTs in response to the human Attackers (Human condition). The Blockers were instructed to try to beat the moving dot to its target. The task in the Human condition was exactly the same as the task in Experiment 1 with the exception that the Attackers were told which target to choose via headphones to assure that the left and right targets are chosen equally often (see Supplemental Methods). Subjects were highly accurate in both experimental conditions (Dot: 99.7%, Human: 97.6%) with a slightly higher performance in the Dot condition ($t(18) = 3.03$, $p < 0.01$). However, as Figure 3a shows, the Blockers' RTs were ~116 ms slower in the Dot condition compared to the Human condition (Dot: 301.2 ms compared to Human: 184.4 ms; $t(18) = 8.15$, $p < 0.0001$).

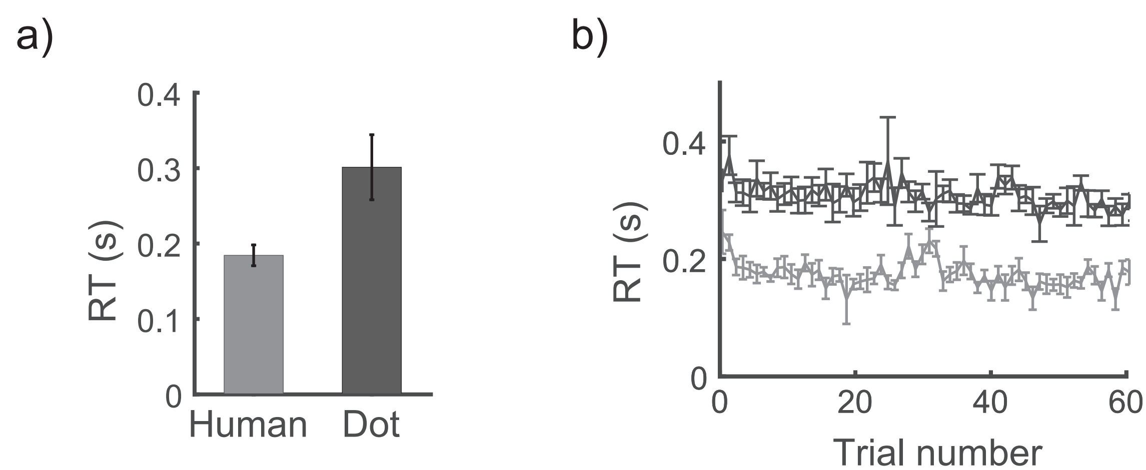


Figure 3. Results of Experiment 2. **a)** Average RTs in the Human and Dot conditions. Error bars show ±SEM. **b)** Average RTs over the course of the experiment. The x-axis represents RTs and the y-axis represents trial number. Subjects showed no sign of learning over the course of the experiment. Error bars show ±SEM.

These results replicate those of the first experiment demonstrating fast reaction times in response to human Attackers. Furthermore they show that the Blockers cannot be focused solely on the finger position of the Attacker, the Dot condition in this experiment preserves the finger movement information and the Blockers are slower in response to the dot.

Are subjects' reaction times fast from the beginning of the experiment or do they learn through the course of the experiment? To answer this question we explored the change in reaction times over the course of the experiment (Figure 3b). A linear regression analysis (see Supplemental Methods) showed no significant change of RT over the course of the experiment for the Human ($t(9) = 1.17$, $p = 0.27$) condition. The Dot condition however showed a marginally significant decrease in RT over time ($t(9) = 2.11$, $p = 0.064$). Subjects were fast in the Human condition from the beginning with no evidence of learning. Overall these results suggest that naïve Blockers are able to use bodily cues other than the Attackers' finger movement to predict the intended target. These cues are possibly available prior to the finger movement and the Blockers can use them without any training.

Experiment 3: Removing the Predictive Cues in Videos

If the fast reactions are driven by body movements present before the Attackers start to move their finger, removing this information should slow down the responses. To test this we videotaped an Attacker and modified the videos to remove the predictive cues and measured the reaction times of the Blockers in response to the modified videos. The experiment had three conditions: In the Real condition Blockers played against a real confederate Attacker. In the FullVid condition they played against video clips of the same Attacker projected at life size on the plexiglass screen. This was to ensure that playing against the videos is comparable to playing against a real subject. In the CutVid condition we replaced all frames before the start of the Attacker's finger movement (T_a) with the first frame of the clip in which no movement has yet occurred (Figure 4a, see also Supplemental Methods). This eliminates all possible predictive cues before the Attacker's finger starts to move. Subjects were highly accurate in all three conditions (Real: 99.5%, FullVid: 98.5%, CutVid: 100%) with no significant difference between the conditions ($ps > 0.05$). The reaction times in the Real and FullVid conditions were both fast and were not significantly different from one another ($t(9) = 1.072$, $p = 0.31$). This replicates the results of the previous experiments and in addition shows that all the cues responsible for fast reaction times are preserved in the video clips. The reaction times in the CutVid condition however were significantly slower than both the Real (RT difference 116 ms, $t(9) = 15.02$, $p < 0.0001$) and the FullVid conditions (RT difference 105 ms, $t(9) = 8.12$, $p < 0.0001$). Removing the video frames with preparatory information slows down the Blockers. Note that the difference between the full and cut videos was subtle. In fact when the two videos are compared with one another it is hard to tell the two apart (see Supplemental Video 2). We debriefed the subjects after the experiment and 7 out of 10 subjects did not notice that the videos were cut in the CutVid blocks. Surprisingly this subtle difference could provide the subjects with sufficient information for a speeded accurate response. This suggests that Blockers indeed use predictive information before the finger movement to determine the goal of the Attacker and to reduce their reaction times.

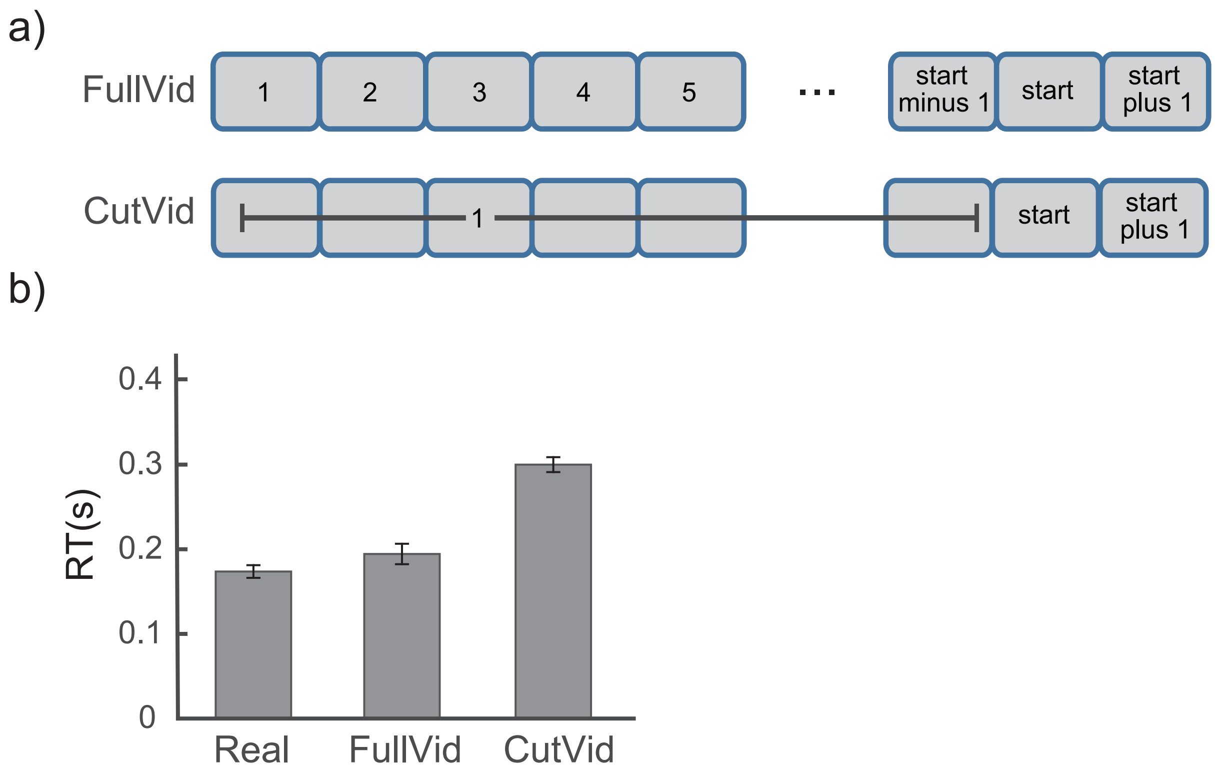


Figure 4. a) Schematic of the two experimental conditions in Experiment 3. In the CutVid condition all the frames before the start of the Attacker were removed and replaced with the first frame of the trial clip. **b)** Results of Experiment 3, Average reaction times for the three experimental conditions Error bars show ±SEM.

Experiment 4: Incongruent Predictive Cues in Videos

If the preparatory information is predictive of Attackers' goals then incongruent preparatory cues should cause Blockers to react slower or choose the incorrect target. To see if this is the case, we divided the videos into two epochs, frames before the start of the Attacker (epoch 1) and those after (epoch 2). We then pasted epoch 1 from one trial clip to epoch 2 of another with either the same Attacker choice (Congruent condition) or a different Attacker choice (Incongruent condition; for further details see Supplemental Methods). If the preparatory information is predictive of the Attacker's choice, the Blockers should choose the incorrect target in the Incongruent condition. We first measured the average accuracies and reaction times of the subjects in the Congruent and Incongruent conditions. At first glance the results showed no difference between the accuracies ($t(18) = 0.82$, $p = 0.41$) and the reaction times ($t(18) = 1.6$, $p = 0.12$) of the two conditions. However a closer look at the data from individual subjects revealed that in the Incongruent condition subjects chose different strategies. Some had slow reaction times and were highly accurate, while others responded early and were inaccurate (see Supplemental Figure 2). The average accuracy and reaction time thus does not fully capture the intricacies of movement in the Incongruent condition. To get around this problem and to have a closer look at the Blockers' behavior during the trial, we ran a separate analysis. We used the instantaneous lateral direction of the finger to define the average accuracy across all Blockers at each time-point during the trial (see Supplemental

Methods). Accuracy rose smoothly in the Congruent condition. In the Incongruent condition however, the initial accuracy was lower than chance level indicating that on average subjects chose the incorrect target initially and then corrected their choice later in the trial (Figure 5). These results demonstrate that through systematic manipulation of the preparatory information it is possible to trick the Blockers into choosing the incorrect target.

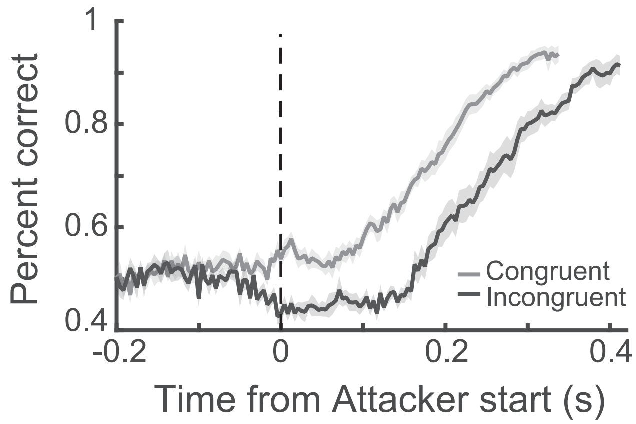


Figure 5. Results of Experiment 4. The average accuracy for each time point determined as the proportion of trials in which the Blocker was moving towards the correct target. The light gray and the dark gray lines depict the results for the Human and Dot conditions respectively. The shaded regions represent ±SEM.

Experiment 5: Where in the body?

Where are these predictive cues? Are they focused on a single body part? If so covering that part should slow down reaction times. To test this we covered the screen with occluders to prevent the Blockers from seeing various body parts of the Attackers. In a task similar to Experiment 2 we first measured the reaction times of the Blockers in three conditions (Supplemental Figure 2): In the All condition Blockers were able to see the Attackers from the waist up. In the Top condition Blockers were able to see the shoulders and the head of the Attackers. In the Torso condition the head and shoulders were occluded and the Blockers were able to see the torso of the Attackers. For comparison, in a separate set of subjects we measured reaction times in response to a dot moving on the screen (similar to the Dot condition in Experiment 2) with a minor modification to the inter-trial interval to make it less predictable (see Supplemental Methods).

We replicated the results of Experiment 2 in this experiment. In the All condition the reaction times were 105 ms faster than those in the Dot condition. Note that the reaction times in this experiment were overall slower than those in Experiment 2 as the inter-trial interval was unpredictable (Klemmer, 1956). Nevertheless the difference between the two conditions was preserved in this Experiment ($t(18) = 4.62$, $p < 0.001$). These results once again show that in the All condition Blockers use preparatory cues before the start of the Attacker for speeded reactions.

The reaction times in the Torso condition were not significantly different from the All condition ($t(9) = 1.21$, $p = 0.26$) and were ~94 ms faster than the Dot condition ($t(18) = 4.47$, $p < 0.001$). In other words preparatory cues in the torso and arms were enough to help the Blockers achieve fast reaction times. The reaction times in the Top condition were slightly slower (~35 ms) than the All condition ($t(9) = 3.72$, $p < 0.01$) and faster (70 ms) than the Dot condition (t(18) > 3.06, $p < 0.01$) indicating that there are preparatory cues in the head and shoulders but they are available slightly after the torso and arms. Note that in the Torso and Top conditions Blockers view completely non-overlapping regions of the Attacker's body. Thus the fact that both of these conditions have faster reaction times than the Dot condition demonstrates that the preparatory cues are distributed over large swaths of the body.

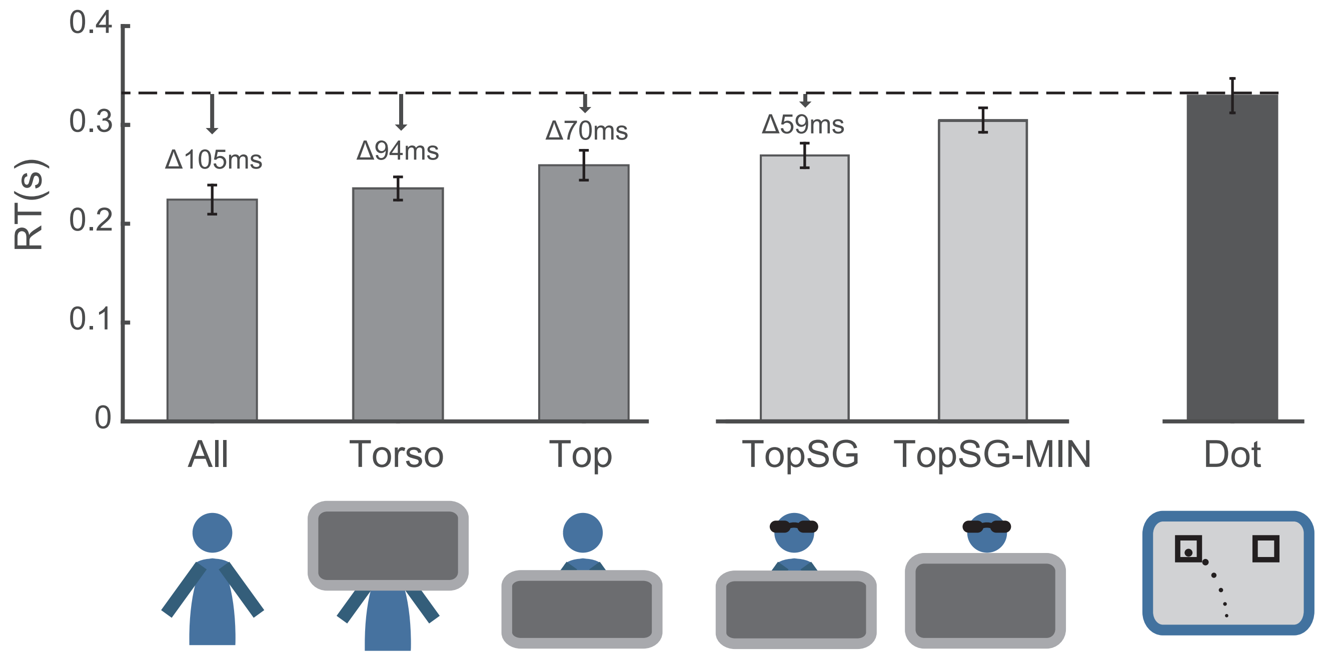


Figure 6. Results of Experiment 5. Average RTs in the 5 condition with the human Attacker compared to the RTs in the Dot condition.

To further restrict the Blocker's view in a new set of subjects we measured the reaction times in two extra conditions (Supplemental Figure 2): the TopSG condition which was similar to the Top condition but Attackers wore dark sunglasses to cover their eyes; and the TopSG-MIN condition in which the Attackers wore sunglasses and only their head from the chin up was visible. The reaction times in the TopSG condition were ~59 ms faster than the Dot condition (t(17) = 2.70, $p < 0.05$) and ~46 ms slower than the All condition (t(17) = 2.36, $p < 0.05$). In other words covering the eyes slightly slowed the reaction times but not as much as the Dot condition and subjects could use cues from the shoulders and the head to achieve fast reactions. The TopSG-MIN condition was the only condition in this experiment with slow reaction times comparable to the Dot condition (t(17) = 1.06, $p = 0.30$). This condition was ~82 ms slower than the All condition (t(17) = 4.21, $p < 0.001$). But even in this condition there was some predictive information. Note that in the TopSG-MIN condition subjects saw the finger only towards the end of the movement, while in the Dot condition movement of the dot starts from the bottom of the screen. Therefore the fact that the reaction times in the TopSG-MIN condition are not slower than the Dot condition indicates that even the head of the Attacker contains some predictive information. In fact to have a better control for the TopSG-MIN condition, we ran another block in which we only presented the Dot towards the end of the movement at the final target location (see Supplemental Methods). Reaction times in this condition were significantly slower than the TopSG-MIN condition ($t(17) = 2.95$, $p < 0.01$) supporting the presence of predictive cues in the head of the Attacker. Overall the results from all the conditions with the human Attacker determine that to perform this task subjects do not focus on a single body part, and that the preparatory cues are distributed over the body.

Further analysis of the reaction times over the course of the experiment in each condition (See Supplemental Methods) showed no signs of learning. A linear regression analysis showed no change in the reaction times throughout the block for all conditions with the human Attacker (ts < 1.64, ps > 0.14). Similar results were found in the Dot condition ($t(9) = 0.72$, $p = 0.49$). The Blockers were able to efficiently collect the predictive cues from any available body part of the Attacker to predict the movement goals without any need for training.

Discussion

In this study we measured the hand movements of subject pairs during a dynamic interaction. One subject (Attacker) was asked to reach for one of two targets and another (Blocker) was asked to reach for the same target as fast as possible. Reaction times of the Blocker were surprisingly short, approximately 100 ms faster than expected, suggesting that preparatory cues before the start of the Attacker's movement are predictive of their goals and Blockers are able to use those cues for such speeded responses. Experiments using video recordings of an Attacker confirmed this, showing that (1) with the putative preparatory cues removed, reaction times slowed correspondingly and (2) when these cues were reversed, blockers moved towards the incorrect target in the beginning of their trajectory. Furthermore we found that the preparatory cues were widely distributed over the body of the Attacker.

The results of the last experiment demonstrate that our ability to predict action goals is remarkably resistant to occlusion of body parts possibly due to redundant cues present in large swaths of the body. Reaction times slowed down only when most of the body of the Attacker was occluded. In the Top and Torso conditions for example, the visible body parts were fully non-overlapping. Yet the reaction times in both conditions were fast indicating that the informative cues were present in the head and shoulders as well as the torso and arms. Note that the informative cues in the Top condition cannot be only focused on the eyes as wearing sunglasses in the TopSG condition did not considerably change the speed of response. Even the TopSG-MIN condition was not fully stripped from predictive information. In sum distributed cues were present in the arms, shoulders, torso, and head that inform about the goal of movement. These results are in line with previous literature in sports showing that distributed cues over the body are informative about future actions (Huys et al., 2008). In an analysis of soccer kicks, Diaz et al. (2012) showed that other than local cues, correlated movements of multiple joints could be predictive of the direction of the kick. Our findings extend these results and show that these distributed cues can be used in a simple reaching task for fast responses. These results suggest that humans have implicit knowledge of the biomechanical constraints that govern bodily movements. In the human body, due to specific joint properties and skeletal connections, movement of one limb often requires moving other body parts. For example in a reaching movement the amount of torque in the elbow is directly affected by the movement in the shoulder joint (Hollerbach & Flash, 1982). Moreover moving a limb often requires postural adjustments in order to stabilize the body in the new form. The first muscle to contract in a reaching task are the torso and lower limb muscles in order to adjust the center of gravity (Belen'kiĭ et al., 1967). The knowledge of these movement principles is essential for predicting actions of others. For example in our task reaching for a target might engage the shoulders or might require leaning to one side. The shoulder and trunk movements in this case are integral parts of the reaching movement and in a sense cannot be isolated from the final finger movement. It is not inconceivable that subjects have a full body model of the left/right reach in their mind. They then gather cue for or against each model from the start of the trial and respond according to their prediction about the final target.

A prominent finding in our study was that subjects were fast from the start of the experiment and remained fast throughout the block. In other words no learning was required to achieve fast reactions. This stands in contrast with a seemingly similar study where significant improvements in performance in the course of the experiment was observed (Diaz et al., 2012). The critical difference between our study and that of Diaz et al., is that the latter study, lacked the essentials of a real interaction. Subject's viewed point light walkers on a computer screen. The videos were cut to show only the preparatory movements. Subjects in their own time, made a prediction about the goal of movement. However, in real life interactions humans do not explicitly report their predictions of others' actions. The process is automatic and implicit, part of the observer's sensory-motor loop (Gibson, 1979). Investigating the process of action reading outside a naturalistic context is problematic. For example subjects might have time to focus on specific body parts when they are not forced to respond immediately. Also viewing a few frames of a video might feel unnatural, force subjects to use strategies that are not generally used in a naturalistic setting. In our experiment using continuous recordings of hand movements we were able to show that not only subjects are able to read actions of others but also they can efficiently incorporate their predictions into their own actions for speeded responses, without any need for learning.

The subtlety of the preparatory cues used to predict Attacker's movements is particularly significant. This is evident in the videos of Experiment 3 (see Supplemental Video 2) where the difference between the full videos (that include preparatory cue) and the cut videos (with no preparatory cues) is barely noticeable. The ability to read these subtle cues further attests that humans are expert body readers. On reflection, this is perhaps not surprising knowing that humans are social species and the demands of a social environment often requires reading subtle cues from the body of other individuals to predict their intentions, emotions and other attributes.

One other possible source for fast reaction times could be the social and competitive nature of the task (Georgiou et al., 2007). The social demands of the task when subjects respond to real humans could introduce a sense of urgency in the subjects and cause the reaction times to be faster than when responding to a dot on the screen. However the results of experiment 3 suggest that this is not the case. In experiment 3 we used video recordings of an Attacker to be able to systematically manipulate the predictive cues. The reaction times in the cut videos in which the predictive information was removed was slower than those in response to full videos. Note that the social and competitive nature of the task is preserved to a large extent in the manipulated videos. The fact that subjects are slow in response to the videos with only a fraction of a second removed shows that the fast reaction times are not resulted from the urgency in a competitive social context and demonstrate the causal role of the preparatory cues in fast reactions.

Taken together these results show that humans rely on subtle preparatory cues, distributed over the body of others to anticipate their goals. More broadly, our study could open the door to future studies of human interaction in real-time settings, to have a deeper understanding of the human bodily movements and what humans are sensitive to during interactions. We believe that the study of biomechanics of human movements and their link to the perception could be instrumental in understanding the predictive aspects of social interaction. These studies could in turn be combined with machine learning and modeling techniques to improve human computer interface.

Materials and Methods:

Stimulus generation and data analysis were done on Windows computer with Matlab and Psychtoolbox software. Hand movements were tracked with a Fastrak electromagnetic position and orientation measuring system (Polhemus Liberty) with an update rate of 240 Hz. A small position tracking sensor (1.27 x 2.22 x 1.9 cm) was attached to the index fingertip of the right hand to record the 3D position of the fingertip.

For more information see supplemental methods.

Acknowledgements: This research was supported by NSF STC award CCF-1231216.

References:

Abernethy, B., Gill, D. P., Parks, S. L., & Packer, S. T. (2001). Expertise and the perception of kinematic and situational probability information. *PERCEPTION-LONDON-*, *30*(2), 233-252.

Abernethy, B., & Zawi, K. (2007). Pickup of essential kinematics underpins expert perception of movement patterns. *Journal of motor behavior*, *39*(5), 353-367.

Abry, C., Cathiard, M. A., ROBERTRIBES, J., & Schwartz, J. L. (1994). The coherence of speech in audiovisual integration. *CAHIERS DE PSYCHOLOGIE COGNITIVE-CURRENT PSYCHOLOGY OF COGNITION*, *13*(1), 52-59.

Belen'kii, V. E., Gurfinkel, V. S., & Pal'tsev, E. I. (1967). ON THE CONTROL ELEMENTS OF VOLUNTARY MOVEMENTS. *Biofizika*.

Diaz, G. J., Fajen, B. R., & Phillips, F. (2012). Anticipation from biological motion: the goalkeeper problem. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 848.

Farrow, D., & Abernethy, B. (2003). Do expertise and the degree of perception-action coupling affect natural anticipatory performance?. *PERCEPTION-LONDON-*, *32*(9), 1127-1140.

Fennel, I., Coello, Y., & Orliaguet, J. P. (1999, July). Visual Perception Of Motor Anticipation In Sign. In *Studies in perception and action V: Tenth International Conference on Perception and Action: Aug. 8-13, 1999, Edinburgh, Scotland* (p. 144). Lawrence Erlbaum.

Georgiou, I., Becchio, C., Glover, S., & Castiello, U. (2007). Different action patterns for cooperative and competitive behaviour. *Cognition*, *102*(3), 415-433.

Gibson, J. J. (1979). 1986. *The ecological approach to visual perception*.
Hollerbach, J. M., & Flash, T. (1982). Dynamic interactions between limb segments during planar arm movement. *Biological cybernetics*, *44*(1), 67-77.

Huys, R., Smeeton, N. J., Hodges, N. J., Beek, P. J., & Wiliams, A. M. (2008). On the dynamic information underlying visual anticipation skill. *Perception & Psychophysics*, *70*(7), 1217-1234.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Attention, Perception, & Psychophysics*, *14*(2), 201-211.

Kandel, S., Boë, L. J., & Orliaguet, J. P. (1993, September). Visual detection of coarticulatory anticipation or... guessing what has not yet been written. In *Virtual Reality Annual International Symposium, 1993., 1993 IEEE* (pp. 148-154). IEEE.

Klemmer, E. T. (1956). Time uncertainty in simple reaction time. *Journal of experimental psychology*, *51*(3), 179.

Louis-Dam, A., Orliaguet, J. P., & Coello, Y. (1999). Perceptual anticipation in grasping movement: When does it become possible. *Studies in perception and action V*, 135-139.

Müller, S., Abernethy, B., & Farrow, D. (2006). How do world-class cricket batsmen anticipate a bowler's intention?. *The quarterly journal of experimental psychology*, *59*(12), 2162-2186.

Orliaguet, J. P., Kandel, S., & Boë, L. J. (1997). Visual perception of motor anticipation in cursive handwriting: Influence of spatial and movement information on the prediction of forthcoming letters. *PERCEPTION-LONDON-*, *26*, 905-912.

Ranganathan, R., & Carlton, L. G. (2007). Perception-action coupling and anticipatory performance in baseball batting. *Journal of Motor Behavior*, *39*(5), 369-380.

Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: expectation, gender recognition, and deceptive intention. *Journal of Experimental Psychology: General*, *112*(4), 585.

Song, J. H., & Nakayama, K. (2006). Role of focal attention on latencies and trajectories of visually guided manual pointing. *Journal of Vision*, *6*(9), 11.

Starkes, J. L., Edwards, P., Dissanayake, P., & Dunn, T. (1995). A new technology and field test of advance cue usage in volleyball. *Research quarterly for exercise and sport*, *66*(2), 162-167.

Supplemental Methods:

Experiment 1: Fast reactions

Subjects: 11 pairs of subjects aged between 18 and 35 participated in this experiment.

Stimuli and procedure: Two subjects sat across from each other (~1.2 m apart) separated by a large (1.2 m x 1.5 m) plexiglass screen (each subject was ~63 cm from the screen). Two small pieces of foam targets (5 cm x 5 cm) were affixed to the screen to serve as targets. Subjects were randomly assigned one of two roles: Attacker or Blocker. A beep, audible to both subjects, prompted the start of each trial, at which point the Attacker chose and reached for one of the two targets and the Blocker responded by reaching for the same target as fast as possible, attempting to beat the Attacker. The Blocker was announced winner if they hit the same target as the attacker within a certain time window relative to the Attacker. The size of the time window was adjusted for each pair so that the Blocker won in approximately half the trials. To do that in each trial (except for the first five trials), the time window was set to be equal to the median reaction in all the prior trials. If the Blocker failed, the Attacker was announced the winner of the trial. The Attacker was instructed to go directly to the target without any attempt to trick the Blocker. The Attacker sat behind a panel that covered their body from the waist down. Both the Attacker and Blocker started their movements from a flat resting spot placed 24 cm from the screen (Figure 1). Each pair of subjects completed 2 blocks of 30 trials in each experimental condition.

Analysis: In this experiment and all the subsequent experiments the reaction time (RT) for each trial was calculated as:

$$RT=T_b-T_a$$

Where T_a stands for the first point in time at which the speed of the Attacker in the 3D space exceeded a speed threshold (25 cm/sec for 15 samples in a row) and T_b stands for the first point in time at which the horizontal speed of the Blocker's finger towards their final target exceeded the same speed threshold. For the Blocker's start point we used the horizontal speed towards the target instead of the Euclidean speed in order to account for changes of mind. Figure 2 shows two sample trials one without (Figure 2a) and one with (Figure 2b) a change of mind. Note that calculating the reaction time based on the horizontal speed towards the target allowed us to accurately measure the true reaction time of the Blocker for their final choice in both types of trials.

To further ascertain that this choice of speed did not affect the results we recalculated the starting point of the Blocker in all experiments based on speed in 3D space. The reaction times were slightly reduced in this analysis, however the results in all experiments remained qualitatively similar.

Experiment 2: Human vs Dot

Subjects: 31 subjects aged between 18 and 35 participated in this experiment, 10 pairs in the Human condition and 11 single subjects in the dot condition. One subject from the Dot condition was removed from the analysis due to very low accuracy (67%). Including this subject did not qualitatively change the any of the results.

Stimuli and procedure: This experiment consisted of two conditions: The Human condition and the Dot condition. The procedure of the Human condition was similar to Experiment 1 except that Attackers wore headphones and were told, on a random basis, which target to choose at the start of each trial, thus eliminating any attacker direction bias. Auditory instructions were recordings of a human voice speaking the words “left” or “right” to indicate one of the two targets. The inter-trial interval in this experiment was set to 1 seconds. In the Dot condition stimuli were back-projected on a semi-transparent sheet affixed to the plexiglass screen using a ViewSonic projector (1024x768, 60 hz) at 70 x 50 cm. Twenty random motion paths were selected from Attacker data in Experiment 1 and used to create the dot stimulus for this experiment. The horizontal and vertical position of Attacker's finger were represented by the location of the dot on the screen. The trial started with a dot presented at the bottom of the screen equidistant from the two targets. Because the angular size of the finger increases when approaching, the dot's diameter started at 0.67 cm and reached 1.34 cm at the endpoint. The Blocker was instructed to tap on the same target as fast as possible. Each subject completed 2 blocks of 30 trials in this experiment for the human condition and 3 blocks of 40 trials for the Dot condition.

Analysis: The analysis in this experiment was similar to experiment one. Other than calculating the average RTs in each condition we also calculated the RTs over the course of the experiment. To do this for each subject we measured the RT in each trial for and each experimental condition. We then ran a linear regression analysis with the trial number as the independent variable and RT as the dependent variable. We then performed a one sample t-test on the regression slopes to determine if they are greater than zero or not. For this analysis, to have equal number of trials across the two conditions we only analyzed the first 60 trials of the Dot condition.

Experiment 3: Removing the predictive cues in videos

Subjects: 10 subjects aged between 18 and 35 participated in this experiment**.**

Stimuli and procedure: A lab member was video taped as an Attacker playing against 6 blockers. We positioned the camera lens (GoPro Hero3+ Black Edition) slightly above and in front of the Blocker's eyes to capture the visual screen from the point of view of the Blockers. In order to simulate the experience of playing against a real Attacker, we projected these videos at life size on the screen at a visual angle matching that of the real Attacker. A total of 747 videos of single trials (trial clips) were extracted from the videos and for each trial clip the frames that matched the start and end of the finger movement towards the target were identified. The experiment consisted of three conditions: Real, FullVid, CutVid. In the Real condition subjects played against the lab member shown in videos. In the FullVid condition subjects played against unedited trial clips. In the CutVid condition we removed all frames before Attacker start time (T_a) and replaced them with the first frame of the trial clip in which no movement has yet occurred (Figure 4a). A random set of trial clips were selected for each condition. No video was repeated for a given subject. Each subject completed 2 blocks of 30 trials for each experimental condition. The order of conditions was counterbalanced across subjects.

Experiment 4: Incongruent predictive cues in videos

Subjects: 20 subjects aged between 18 and 35 participated in this experiment, 10 in the Congruent condition and 10 in the Incongruent condition. One subject was removed from the incongruent condition due to very low accuracy (54%). Including this subject did not qualitatively change any of the results.

Stimuli and procedure: We used the videos from Experiment 3. We divided each video into two epochs. Epoch 1 consisted of frames before the start of the attacker and epoch 2 consisted of the frame in which the attacker started to move (T_a) and all frames after that. The experiment consisted of Congruent and Incongruent conditions. In the Congruent condition epoch 1 and epoch 2 from two trials with the same direction were combined. In the Incongruent condition epoch 1 and epoch 2 from two trials with different directions were combined. Due to slight variations in the videos from one trial to another, the transition point between the two epochs was not smooth. In order to ascertain that the two conditions were similar in this regard, we sorted the trials based on the pixelwise difference between the two frames around the transition point and matched the trial clips in the Congruent and Incongruent according to this difference. No video was repeated for a given subject. Each subject completed 2 blocks of 30 trials. Half of the subjects completed the Congruent condition and the other half the Incongruent condition.

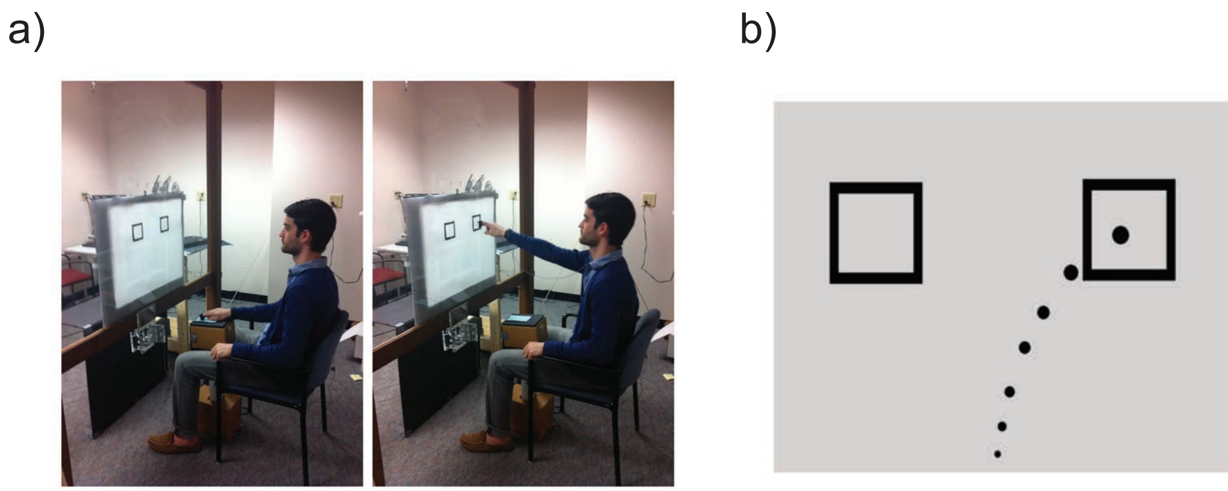
Analysis: In this experiment we measured the accuracy of subjects at each time point during the trial. To do this, in each trial we used the speed in the lateral direction to determine if the subjects are moving towards the correct target or not. We then averaged this instantaneous accuracy across trials and subjects to determine the time-course of the increase in accuracy during the trial.

Experiment 5: Where in the body?

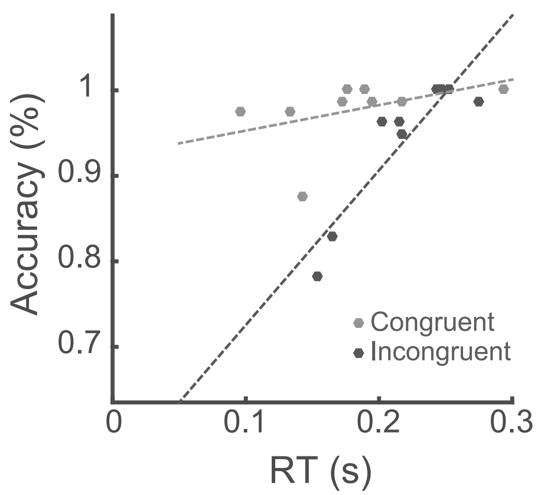
Subjects: 50 subjects aged between 18 and 35 participated in this experiment, 10 subjects participated in the Dot conditions, 10 pairs participated in the All, Top and Torso conditions and 10 pairs participated in the TopSG and TopSG-MIN conditions**.**

Stimuli and procedure: This experiment consisted of 5 conditions: All, Top, Torso, TopSG, TopSG-MIN and Dot. In each condition parts of the Attacker's body was obscured from the view of the Blocker using occluders on the screen and/or wearing sunglasses. The All condition and the Dot condition were similar to the Human and Dot conditions in Experiment 2 respectively with the exception that the inter-trial interval in this experiment varied randomly between 1 and 4 seconds in this experiment. This was to ensure that the predictability of the trial timing is not the source of the fast reaction times and that the results will generalize to a broader range of conditions with more unpredictability. In the Top condition only the Attacker's shoulders and head were visible. In the Torso condition only the waist to shoulders were visible. In the TopSG condition sunglasses were added to the Top condition. The TopSG-MIN condition was the same as the TopSG condition but the shoulders and neck were also obscured. To have a better control for the TopSG-MIN condition we ran an extra Dot condition in this experiment in which the dot appeared only in the final target position instead of moving from the start point to the target. Subjects completed two blocks of 30 trials for each condition with the Human Attacker and 3 blocks of 40 trials for each of the Dot conditions. In the cases that the subjects performed more than one condition the order of the blocks was counterbalanced across subjects.

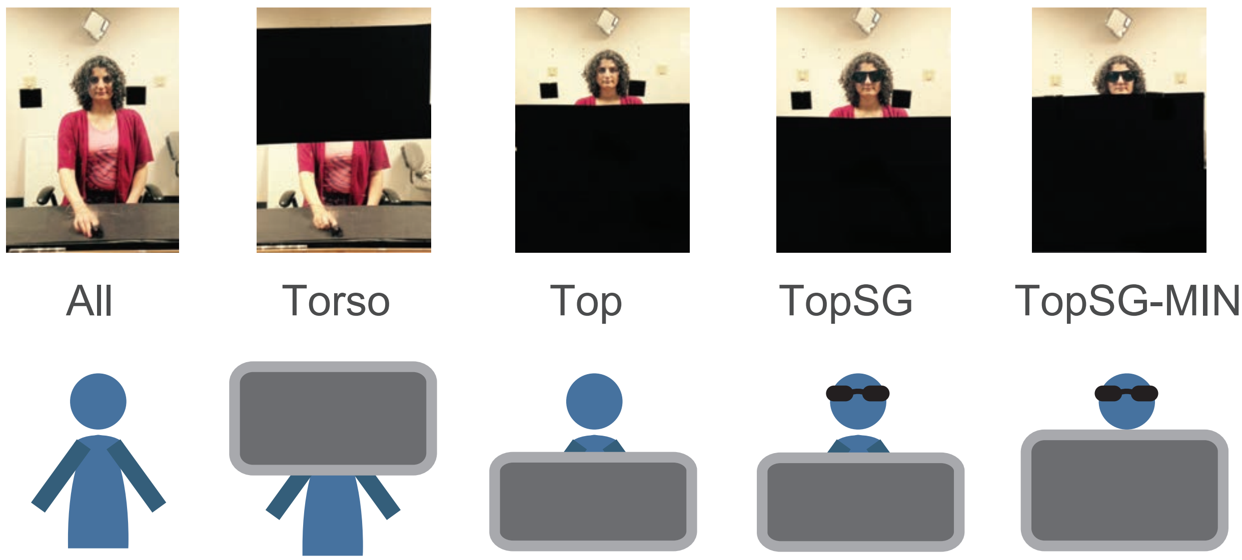
Analysis: In this experiment, other than calculating the average reaction times, we ran an analysis to determine the amount of learning by measuring the decrease in reaction time over the course of the experiment in each condition (see methods of Experiment 2).



Supplemental Figure 1. Set up of the Dot condition in Experiments 2 and 5. a) Stimuli were back projected on the plexiglass screen. b) Blockers responded to a dot moving from the bottom of the screen to one of the two targets.



Supplemental Figure 2. Individual Subjects' results in Experiment 4. Each dot represents one subject with the average Accuracy plotted in the y axis and the reaction time plotted in the x axis. In the Congruent condition subjects were accurate with a small positive relationship between the accuracy and reaction time. In the Incongruent condition however this relationship was much more strong. Subjects were either fast with low accuracy or had longer reaction times with higher accuracy.



Supplemental Figure 3. Set up of all the conditions in Experiment 5 with the human Attacker. In each condition parts of the body of the Attacker was occluded.

Supplemental Video 1. Sample video of a pair of subjects engaged in the two-person interaction.

Supplemental Video 2. Sample trial clips from Experiment 4. The left panel is a trial clip from the CutVid condition in which all frames before the start of the Attacker was removed and replaced with the first frame. The right panel is a trial clip from the unmodified FullVid condition. As it is apparent in the videos, the difference between the CutVid and FullVid conditions was very subtle.