Understanding Action Prediction with Machine Learning and Psychophysics

Emalie McMahon (emalie.mcmahon@nih.gov)
Laboratory of Brain and Cognition, National Institute of Mental Health,
Bldg 10, Room 4C214, 10 Center DR
Bethesda, Maryland 20892, USA

Ray Gonzalez (raygon@mit.edu)
Vision Sciences Laboratory, Department of Psychology, Harvard University,
33 Kirkland Street #770
Cambridge, MA 02138, United States

Ken Nakayama (nakayama@g.harvard.edu)
Vision Sciences Laboratory, Department of Psychology, Harvard University,
33 Kirkland Street #770
Cambridge, MA 02138, United States

Leslie G. Ungerleider (ungerlel@mail.nih.gov)
Laboratory of Brain and Cognition, National Institute of Mental Health,
Bldg 10, Room 4C104, 10 Center DR
Bethesda, Maryland 20892, USA

Maryam Vaziri-Pashkam (maryam.vaziri-pashkam@nih.gov)
Laboratory of Brain and Cognition, National Institute of Mental Health,
Bldg 10, Room 4C205, 10 Center DR
Bethesda, Maryland 20892, USA

Abstract:

Predicting the actions of others is relevant in many social situations from extending a handshake to elaborate waltzes. To study the preparatory information in movements and how people are able to interpret these preparatory cues, we designed a partnered reaching task. In the competitive condition, one partner (the Blocker) had to beat the other (Attacker) to the target (see Vaziri-Pashkam et al., 2017), and in the cooperative condition, both participants were asked to tap the same target at the same time. In a psychophysical paradigm, different subjects viewed short clips of the Attacker's movements and were asked to predict whether the Attacker was going to point to the left or right target. Subjects were able to predict the direction of movement with between 80% and 90% accuracy before finger lift off. A follow-up searchlight analysis revealed that all body parts contained informative predictive cues with the head showing predictive information earlier in the movement for both conditions, but especially for the cooperative condition. These results reveal that subjects can use preparatory cues in the movements of others to predict action goals before the start of the movement and that these cues are exaggerated in the cooperative context to communicate the goal of actions.

Keywords: Action, Pattern Classification, Social Goal

Introduction

Walking down a busy street, remarkably, does not result in many people bumping into each other. When someone is walking directly toward you, you seamlessly predict the result of their action—a collision—and effortlessly adjust your path to pass by easily. Similar predictions of actions are made when the barista hands you your coffee. This prediction ability is essential to social interaction (Frith & Frith, 2006; Sebanz, Bekkering, & Knoblich, 2006), occurs spontaneously, without training (Vaziri-Pashkam, Cormiea, & Nakayama, 2017), and may arise from knowledge of biomechanical constraints of human action (Johansson, 1973).

We have the ability to predict the future course of an action sequence from the movements of others. Action prediction ability has been studied extensively in sports (Abernethy et al, 2001, Abernethy & Zawi, 2007, Aglioti, Cesari, Romani, & Urgesi, 2008, Knoblich & Flach, 2001, Muller, Abernathy & Farrow, 2006, Knoblich & Flach, 2001, Ranganathan & Carlton, 2007, Diaz, Fajen & Phillips, 2012). For instance, Aglioti et al (2008) demonstrated that elite basketball players we are more accurate at predicting whether a free-throw shot would go into the basket when only viewing part of the movement than novice basketball players. Predictive ability has even been shown for simple reaching actions (Louis-Dam, Orliaguet, and Coello, 1999; Martel, Bidet-Ildei, & Coello, 2011; Pesquita, Chapman, & Enns, 2016), and the removal of predictive information has been shown to increase reaction times (Vaziri-Pashkam et al. 2017).

These studies demonstrate the existence of goal predictive information at the beginning of a goal-directed movement. What remains unknown is the spatio-temporal profile of the predictive cues. Is the predictive information focused on a single body part? Or is it distributed over the body? Do informative body parts vary across time? Does the social goal of the actors affect the bodily profile of goal predictive information? The current study aimed to investigate when preparatory information becomes available to human observers, where the information is located in the body, and how the location and timing of information is affected by social context using a combination of psychophysics and machine learning.

Methods

Psychophysics

Twenty attackers were recorded in one of two conditions: cooperative or competitive, which differed only in the instructions. In the competitive condition, one partner (the Blocker) had to beat the other (Attacker) to the target (see Vaziri-Pashkam et al., 2017), while in the cooperative condition, both participants were asked to tap the target at the same time (Figure 1a). Attackers wore magnetic motion sensors that tracked the position of their finger and the session was recorded. To determine when enough information was available in the videos for humans to predict the goal of movement, the full videos were separated into individual clips containing single trials (Figure 1b). Then, in each video clip, the frame in which the Attacker lifted his/her finger off the table was determined based on the kinematic information from the finger. The videos of the Attackers were shown to twenty new participants (10 for each condition), and they were asked to determine the direction of movement after viewing a short video clip. The clips were presented from ~500 ms (~30 frames) prior to the finger lift off up to specific cut-frames. The cut-frames varied between -133 ms (-8 frames) and +33 ms (+2 frames) relative to the Attacker start point. The average accuracy of the participants was determined at each cut-point to construct a psychometric function across video clips for each of the 20 Attackers. The psychometric curve was fit with a logistic function to determine the time of 75% accuracy (T_{75}).

Machine Learning

Using the Gunner-Farneback Optical Flow algorithm, the motion energy of each pixel in the video was calculated. A bagged SVM was trained on 50% of the optical flow data and tested on a left-out 50% frame-by-frame. This allowed for construction of a sigmoid function that could be compared to the behavioral psychometric results. The sigmoid curve was fit with a logistic function to determine the T_{75}. Next, using a searchlight algorithm (Pereira & Botvinick, 2011), accuracy of the classifier was determined for different spatial positions in each frame of the video to determine the spatio-temporal profile of the informative features in the video.

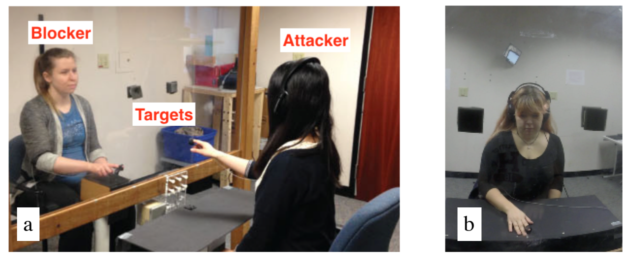

Figure 1: a) Set-up of the Attacker and Blocker. The attacker is told through the headphone which target (left or right) to point to and the blocker needs to beat her to the target (competitive) or arrive at the same time at the target (cooperative). b) Example video frame shown to participants in the psychometric experiment.

Results

Psychophysics

The average subjects' performance was greater for the cooperative condition than the competitive condition at some, but not all, time points (Figure 2). These data illustrate two critical points. The average cooperative curve is shifted to the left compared to the competitive curve, suggesting that the cooperative Attackers revealed more information prior to the movement start than did the competitive Attackers. The leftward shift of the cooperative curve is further illustrated by the T_{75} of the two curves. The average T_{75} was earlier in time for the cooperative (M = -47.07 ms) than the competitive condition (M = -23.45 ms, $t(28) = 2.44$, $p < 0.01$).

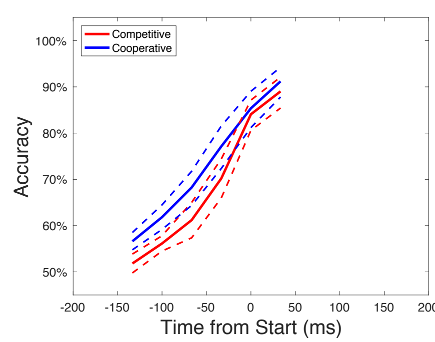

Figure 2: Average psychometric curves from the behavioral data for the cooperative and competitive conditions. Time zero is when the finger lifted off the table. Dotted lines represent 95% confidence intervals from bootstrapping.

Machine Learning

The SVM performance was greater for the cooperative condition than the competitive condition at some, but not all, time points (Figure 3). The average T_{75} for the SVM was more negative in the cooperative (M = -72.10 ms) than the competitive condition (M = -36.80 ms, $t(198) = 87.68$, $p < 0.001$). The average T_{75} for the SVM was also more negative for both conditions than the behavioral data. However, the T_{75} of the SVM was strongly correlated with the behaviorally determined T_{75} for both the competitive ($r(8) = 0.68$, $p < 0.05$) and cooperative ($r(8) = 0.72$, $p < 0.05$) conditions (Figure 4). In other words, if it was difficult to interpret the direction of movement of the Attacker for a human, it was also difficult for the SVM.

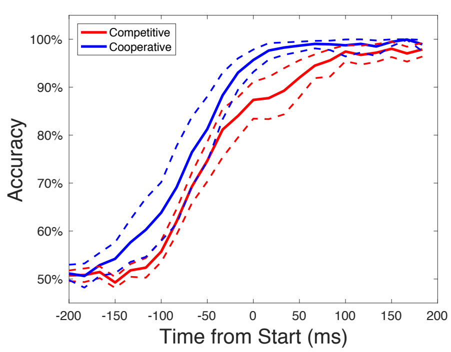

Figure 3: Average sigmoidal curves from the SVM for the cooperative and competitive conditions. Time zero is when the finger lifted off the table. Dotted lines represent 95% confidence intervals from bootstrapping.

The results of the searchlight analysis are shown in Figure 5. These results demonstrate the spatio-temporal profile of predictive information and how it varies across conditions. In both conditions the head is informative early and the information then moves to the arms and the rest of the body. Also, the head appears to be more informative in the cooperative than the competitive condition in the early frames, while the shoulders seem to carry a greater amount of early information in the competitive condition.

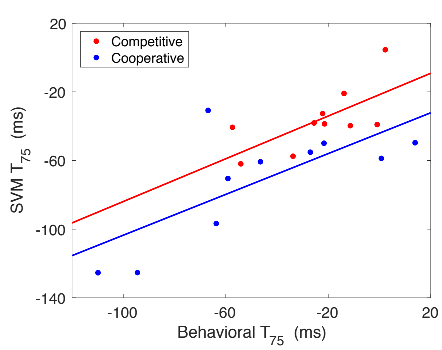

Figure 4: Correlation of the time of 75% accuracy (T_{75}) of the SVM compared to human T_{75}.

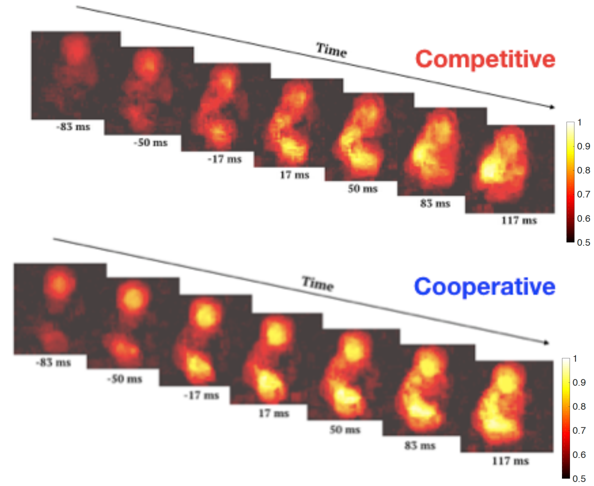

Figure 5: Informative features from the searchlight analysis of the Optical Flow information through time. Time zero is when the finger lifted off the table.

Discussion

The current study aimed to investigate when preparatory information is sufficient to predict an action goal. Attackers were recorded playing a competitive or a cooperative reaching task against an opponent. The videos were cut and, using a psychophysical paradigm, the accuracies of observers in predicting the goal of action from cut-videos were measured. We found that subjects were able to predict the direction of movement well before the movement began in both conditions, although there was more information available earlier in the cooperative condition than the competitive condition. Further, a support vector machine (SVM) was trained to decode the direction of movement and was found to accurately predict the direction of movement significantly earlier than human viewers. Despite outperforming human perception, the accuracy of the SVM was correlated with the behavioral accuracy.

Considering the SVM as an ideal observer, we borrowed a searchlight method from the field of neuroimaging (Kriegeskorte, Goebel, & Bandettini, 2006) to investigate which pixels in the videos of the movement were most informative in decoding the direction of movement. The searchlight analysis revealed early information in the head of the Attacker that becomes more distributed over the body as the movement progressed. The searchlight revealed that the distribution of information throughout the body differs in cooperation and competition. The presence of more information in the head and body parts explicitly involved in movement execution suggest that people may have some knowledge of preparatory cues and can leverage them to their benefit in cooperation.

Acknowledgments

We thank Francisco Pereira of the NIMH Machine Learning Core for assistance with the Searchlight Analysis.
This work was funded in part by the NIMH DIRP and in part 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, 30*(2), 233-52.

Abernethy, B. & Zawi, K. (2007). Pickup of essential kinematics underpins expert perceptions of movement patterns. *Journal of Motor Behavior, 39*(5), 353-67.

Aglioti, S.M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience, 11*(9), 1109-16.

Frith & Frith (2006). How we predict what other people are going to do. *Brain Research, 1079*(1), 36-46.

Diaz, G.J., Fajen, B.R., & Phillips, F. (2012). Anticipation from biological motion: the goalkeeper problem. *Journal of Experimental Psychology: Human Perception & Performance, 38*(4), 848-64.

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

Knoblich, G. & Flach, R. (2001). Predicting the effects of actions: interactions of perception and action. *Psychological Science, 12*(6) 467-72.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Inforamtion-based functional brain mapping. *Proceedings of the National Academy of Sciences*, *103*(10), 3863-8.

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.

Martel, L., Bidet-Ildei, C., & Coello, Y. (2010). Anticipating the terminal position of an observed action: Effect of kinematic, structural, and identity information. *Visual Cognition, 19*(6), 785-98.

Pereira, F. & Botvinick, M. (2011). Information mapping with pattern classifiers: a comparative study. *Neuroimage, 56*(2), 476-96.

Pesquita, A., Chapman, C. S., & Enns, J. T. (2016). Humans are sensitive to attention control when predicting others' actions. *Proceedings of the National Academy of Sciences*, *113*(31), 8669–8674.

Ranganathan, R. & Carlton, L.G. (2007). Perception-action coupling and anticipatory performance in baseball batting. *Journal of Motor Behavior, 39*(5), 369-80.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences, 10*(2), 70–76.

Vaziri-Pashkam, M., Cormiea, S. & Nakayama, K. (2017). Predicting actions from subtle preparatory movements. *Cognition, 168*, 65-75.