Affordance matching predictively shapes the perceptual representation of others' ongoing actions.

Katrina L. McDonough[a*], Marcello Costantini[b], Matthew Hudson[a1], Patric Bach[a].

[a] Plymouth University, School of Psychology, Plymouth, PL48AA, UK

[b] Department of Psychological, Health, and Territorial Sciences, "G. d'Annunzio" University of Chieti-Pescara, Italy and Institute for Advanced Biomedical Technologies - ITAB, University "G. d'Annunzio", Chieti, Italy

Running Head: prediction of ongoing actions.

*Corresponding author: Katrina L. McDonough

Address: Link 301a,
Plymouth University,
Drake Circus,
Plymouth,
Devon,
UK,
PL4 8AA

Email: katrina.mcdonough@plymouth.ac.uk

---

[1] Present Address: School of Business, National College of Ireland, Mayor Street, Dublin 1, Ireland.

Abstract

Recent predictive processing models argue that action understanding is a predictive process, in which goal inferences are constantly tested by comparing predictions of forthcoming behaviour against the actual perceptual input. In a recent series of studies, we showed that these predictions can be visible as a subtle shift in perceptual action judgments towards these inferred goals. Here we test whether this perceptual shift occurs even when goals are not explicitly given but have to be derived implicitly from the unfolding action kinematics. In two experiments, participants watched an actor reach towards a large object and a small object forming either a whole hand power grip or a precision grip. During its course, the hand suddenly disappeared, and participants made perceptual judgments about the last seen position on a touch screen. As predicted, judgments were consistently biased towards apparent action targets, such that power grips were perceived closer to large objects and precision grips closer to small objects, even if the actual kinematics were the same. Strikingly, perceptual shifts were independent of participants' explicit goal judgments, and were of equal size when action goals were explicitly judged in each trial (Experiment 1) or not judged (Experiment 2). Moreover, across trials and across participants, explicit goal judgments and perceptual shifts were uncorrelated. This provides evidence, for the first time, that people make on-line adjustments of predicted actions based on the match between hand grip and object goals, distorting the perceptual representation of the action. These distortions may not reflect high-level goal assumptions, but emerge from relatively low-level processing of kinematic features within the perceptual system.

Keywords: action understanding, action prediction, perception, social perception, predictive processing, representational momentum

Affordance matching predictively shapes the perceptual representation of others' ongoing actions.

The ability to understand – and predict – other people's behaviour is a cornerstone of human social cognition and makes people's sophisticated interactions with others possible. A parent constantly monitors their child's goals, and intervenes when it reaches for the hot cup of coffee instead of the toy right next to it. In sports, players foresee each other's behaviour, fluently passing a ball to a team mate's future position. In contrast, deficits in the ability to understand others' behaviour are a hallmark of several conditions that bring with them marked impairments in social interactions, such as autism (e.g., von der Lühe et al., 2016).

For over twenty-five years, the conventional view of action understanding has been of a simple bottom-up process, in which incoming visual information about the behaviour of others' is matched to one's higher-level motor – or conceptual – representations (Gallese & Goldman, 1998; Oosterhof, Wiggett, Diedrichsen, Tipper, & Downing, 2010; Rizzolatti, Fogassi, & Gallese, 2001). However, such a mechanism is only applicable when the action kinematics unambiguously discriminate between the possible intentions that drive the action (Jacob & Jeannerod, 2005; Uithol, van Rooij, Bekkering, & Haselager, 2011). Social behaviour is far less clear-cut; many actions that are superficially similar in their visual appearance can serve any number of different goals: a kicking action could be part of a football game, a fight, or a ballet routine. Such bottom-up processes are therefore unable to discriminate the intentions underpinning them (e.g. Bach & Schenke, 2017). Recently, theorists have started to conceptualise the capacity to understand others' actions as a predictive process (e.g., Bach, Nicholson, & Hudson, 2014; Bach & Schenke, 2017; Csibra, 2008; Donnarumma, Costantini, Ambrosini, Friston, & Pezzulo, 2017; Kilner, Friston, & Frith, 2007a, 2007b). Predictive models see social perception as an active process of hypothesis testing. Any assumption about others' goals and beliefs – derived from prior knowledge about the individual (e.g., Joyce, Schenke, Bayliss, & Bach, 2015; Schenke, Wyer, & Bach, 2016) as well as from contextual cues, such as nearby objects (e.g., Bach, Knoblich, Gunter, Friederici, & Prinz, 2005; Nicholson, Roser, & Bach, 2017), gaze and emotional expressions (Adams, Ambady, Macrae, & Kleck, 2006; Frischen & Tipper, 2006) or motor cues (e.g., Donnarumma et al., 2017) - is translated into predictions about which behaviour should be observed if these assumptions were correct. These predictions can then be used to test one's prior assumptions, triggering revisions if they fail to explain the perceptual input. Such top-down predictions not only aid perception, by filling in gaps in the input (e.g. in the case of occlusion, Prinz & Rapinett, 2008) or by compensating for the considerable noise during motion perception (Hammett, 1997), but also support fluent social interactions. The ability to foresee others' upcoming actions allows for "bridging the gap" between relatively slow perceptual analysis of others' behaviour and the need for fluent responses in social interactions, solving tasks such as the passing to a teammates future position or intercepting a child's actions mentioned above (Nijhawan, 1994, 2008). Indeed, we actively make our actions more predictable to facilitate the execution of joint actions with others (Vesper, van der Wel, Knoblich, & Sebanz, 2011).

We recently developed an experimental paradigm that can make these predictions visible (Hudson, Bach, & Nicholson, 2017; Hudson, McDonough, Edwards, & Bach, 2018; Hudson, Nicholson, Ellis, & Bach, 2016; Hudson, Nicholson, Simpson, Ellis, & Bach, 2016; McDonough, Hudson, & Bach, 2018, June 8). This paradigm rests on the assumption that, if predictions indeed act on perceptual representation (Bar, 2004; de Lange, Heilbron, & Kok, 2018; Ekman, Kok, & de Lange, 2017) then every prediction one makes about another person may subtly bias the perception of their forthcoming actions, especially in case of uncertainty, such as the visual blurring during motion perception. Thus, in the same way as prior expectations in the non-social world cause us to see a colour differently (Bloj, Kersten, & Hurlbert, 1999; see for an application to the blue/gold dress illusions, Chetverikov & Ivanchei, 2016) or shapes as either convex or concave based on the surrounding illumination (Adams, Graf, & Ernst, 2004), our prior knowledge of other people – their goals and intentions – may subtly shape the perceptual experience of their actions.

This is indeed what we observed. Participants heard an actor make a statement about their goal – “I’ll take it” or “I’ll leave it” – before they saw him reach for or withdraw from an object. The action disappeared mid-motion and participants indicated the perceived vanishing point, either by comparing it to a probe stimulus shortly after stimulus offset (Hudson, Nicholson, Ellis, et al., 2016; Hudson, Nicholson, Simpson, et al., 2016) or by indicating its disappearance point on a touch screen (Hudson et al., 2017; Hudson et al., 2018). The results reliably revealed predictive biases on these perceptual judgments. First, hands were generally reported to have disappeared further along the trajectory than what was actually seen, capturing lower-level predictions based on the action’s prior course (e.g. representational momentum, Freyd & Finke, 1984; Hubbard, 2005). Second, and more importantly, they revealed an influence of goals attributed to the actor: hands were reported to have disappeared further towards the object when the actor said they would take it and further away when the actor said they wanted to leave it.

Other studies replicated and extended these findings, showing that similar distortions can be induced when the participant instructs the observed actor (Hudson, Nicholson, Ellis, et al., 2016) and that the observed actor’s long-term reliability to do as they said modulates the strength of the prediction effects (Hudson et al., 2017). Most recently, we showed that similar effects can be elicited by the prior object context, such that hands reaching straight for an object are perceptually judged to veer slightly upwards if they would need to reach over an obstacle, and slightly downwards when reaching unnecessarily high (Hudson et al., 2018).

These data show that the goals inferred from the prior information are indeed translated into predictions about the upcoming action, which then bias perceptual judgments towards these expectations. Yet, in all these studies the goals or environmental constraints were explicitly given prior to action onset. In the real world, people typically do not always say what they are about to do. Instead, an action's goal often has to be derived dynamically once the action is underway and its kinematics become apparent (Ambrosini, Costantini, & Sinigaglia, 2011; Ambrosini et al., 2013; Bach, Bayliss, & Tipper, 2011; Sartori, Becchio, & Castiello, 2011). We have argued (Bach et al., 2014) that the affordances of the goal objects could play a major role in deriving a person's goals once an action is underway. Viewing an action (e.g. a hammering motion) that matches the affordances of an available goal object (e.g. a hammer) would immediately signal to an observer what the goal of the action would be (e.g., Bach et al., 2005; van Elk, van Schie, & Bekkering, 2014). And indeed, there is now ample evidence that people spontaneously derive the target of a reach, by matching the hand's grip configuration – i.e. either a small "precision" grip or a large "power" grip – to the available large or small objects in the environment (e.g., Ambrosini et al., 2011; Ambrosini et al., 2013; for a review, see Bach et al., 2014). For example, eye movements reveal that people anticipate the target of an ongoing reach by matching the unfolding grip shape (large or small grip) to the surrounding objects (e.g., Ambrosini et al., 2011; Ambrosini et al., 2013), automatic imitation effects are larger for actions that fit a goal object (e.g., Bach et al., 2011) and larger motor evoked potential are elicited by the same kinematics if they fit an available goal object (Southgate, Johnson, Karoui, & Csibra, 2010).

Here we test, for the first time, whether people use such grip-object matching not only to derive the action’s goal or target (e.g., which goal object is selected), but whether they also use these goals, as assumed by perceptual prediction models (e.g., Kilner et al., 2007a, 2007b), to predict the forthcoming action, even if it is already well underway. If this is the case, we should find that the match of an unfolding hand grip to one of two objects in the environment should again induce such perceptual biases, and they should be measureable – as in our prior work – in subtle distortions in perceptual judgment about these actions. Demonstrating such distortions is crucial to show that, during action observation, people go beyond simple goal inference (e.g. identifying the target of a reach), but that they use this information to then predict which future course the action will take, predicting the subtle kinematic change towards the identified goal object.

In two experiments, we presented participants with videos of an actor’s hand starting at rest and then reaching towards the centre point between two adjacent potential target objects – one small, one large –, with the hand forming either a whole-hand power grip or a precision grip. The hand disappeared mid-motion, at an equal distance away from either object, and participants were required to indicate the final location of the hand’s index finger on a touch screen monitor. If observers identify the goals of the action by matching the observed grip to the two objects’ affordances and then form a perceptual prediction about its future course, then perceptual judgments should show specific biases: the located disappearance points should be reported closer to the corresponding object than they actually were, and away from the alternative (mismatching) target object. Therefore, although the hand actually reached between the two objects, reaches with a precision grip should be reported closer to the smaller object and reaches with a power grip should be perceptually biased towards the larger object.

A crucial question is whether any such effects emerge from a general top-down mechanism, such that high-level attributions of others' goals penetrate lower-level perceptual representations, or whether any perceptual biases emerge from "encapsulated" interactions in the perceptual system itself, which has already been shown to be sensitive to such matching hand-object interactions (e.g., Bracci & Peelen, 2013). We therefore manipulated, across the two experiments, whether participants had to explicitly derive the action's goals. In Experiment 1, we had participants say into the microphone, after each trial, which object they thought the hand was reaching for, requiring explicit goal attribution in each trial. In Experiment 2, no such verbal responses were given while participants still reported the perceived disappearance points. These localisation judgments therefore measure spontaneous and implicit goal inferences and the resulting predictions. The difference between the experiments will reveal the extent to which perceptual biases emerge from a more or less encapsulated, automatic visual predictions system that relies on perceptually available "local" stimulus features (e.g., hand pre-shape, available objects, for a review, see Scholl & Gao, 2013), or whether these processes can be penetrated by higher-level information, such as the explicit attribution of goals to another. Moreover, the combination of explicit verbal goal judgments and implicit perceptual judgments in Experiment 1 will also allow us to test, across participants and across trials, the relationship between these measures.

Method

Participants

Sixty-two participants took part in Experiment 1 (mean age = 20 years, SD = 3.4, 52 females) and 63 participants took part in Experiment 2 (mean age = 21 years, SD = 5.5, 50 females). Eleven additional participants were excluded due to performance assessed against several a priori criteria (see Results). All were right handed and had normal/corrected-to-normal vision, and were recruited from Plymouth University for course credit. The study was approved by the University of Plymouth Ethics Committee, in accordance with the declaration of Helsinki. A power analysis revealed that a sample size of 62 provides .80 power to detect effects in the predicted direction with Cohen's *d* = .31, and effects in either direction with Cohen's d = .36. Our prior studies investigating similar effects with the same method (Hudson et al., 2018; McDonough et al., 2018, June 8) revealed that effect sizes are typically larger (*d* = .44 to *d* = .60).

Apparatus

Presentation (NeuroBS) software was used to present the experiment via a HP EliteDisplay S230tm 23-inch widescreen (1920 x 1080) Touch Monitor. Verbal responses for Experiment 1 were detected using Presentation's Sphinx speech recognition engine via a Microsoft LifeChat LX-3000 Headset.

Stimuli

Example stimuli can be seen in Figure 1A. Stimuli were derived, using photo manipulation, from a prior stimulus set of video stimuli from one of the authors (Costantini, Ambrosini, & Sinigaglia, 2012). The experimental videos (950x540) depicted an actor's arm, from the side view, reaching towards a location in-between a small target (a strawberry) and a large target (an apple). Both objects were located at an equal distance away from the hand, with one closer to the foreground and lower down on the screen and one closer to the background and higher up on the screen (object positions counterbalanced across trials). The actor's hand started at rest in a neutral closed fist posture, and then began to reach, forming either a whole-hand grip or a precision pre-shape. Four reach videos were used for each pre-shape condition, which together with two target layouts (small object to the front, large at the back, or vice versa), created a total of 16 different videos. Each video was converted into 9 frames, where frame 9 showed the actor's hand at maximal pre-shape, halfway between the starting position and the objects, and halfway between the two objects, so that only pre-shape information would predict which object would be reached for and not hand position. The shadow of the hand was digitally removed so this information could not aid localisation.

Response images for both experiments were created by digitally removing the actor's arm from the scene, so that only the target objects and the background remained. Presenting this frame immediately after the action sequence gave the impression of the hand disappearing from the scene. A second response image for Experiment 1 was identical to these images, with the addition of four question marks positioned at each corner of the screen. These served as cues for participants to make their verbal responses about which object they believed was the target. All editing was completed using Adobe CC Photoshop.

Procedure

An example trial sequence can be seen in Figure 1B. Participants completed a total of 192 trials, consisting of four blocks of 48 trials (each representing all 16 different trials three times), with breaks in between. At the start of each trial, participants saw an instruction to "Hold the spacebar", to which they pressed the spacebar with their right hand and kept it depressed until the end of the action sequence. This ensured that they did not track the observed action with their finger, and could only initiate their response once the action sequence had disappeared. They then saw the first (neutral) frame of the action sequence for

1000ms, followed by successive frames at 80ms intervals. The final frame was randomly chosen in each trial as either frame 8 or frame 9, to increase variability of the hand's final position. This final frame was then immediately replaced with the response image. Participants released the spacebar and, with their right hand, touched the screen where they thought the final position of the tip of the observed index finger was. For Experiment 1, the touch response was immediately followed by the second response frame where participants were required to say into the microphone which target object they thought the actor was reaching towards (either "apple" or "strawberry"). Once the verbal response was registered, the next trial began. For Experiment 2, the next trial began as soon as the touch response was registered.

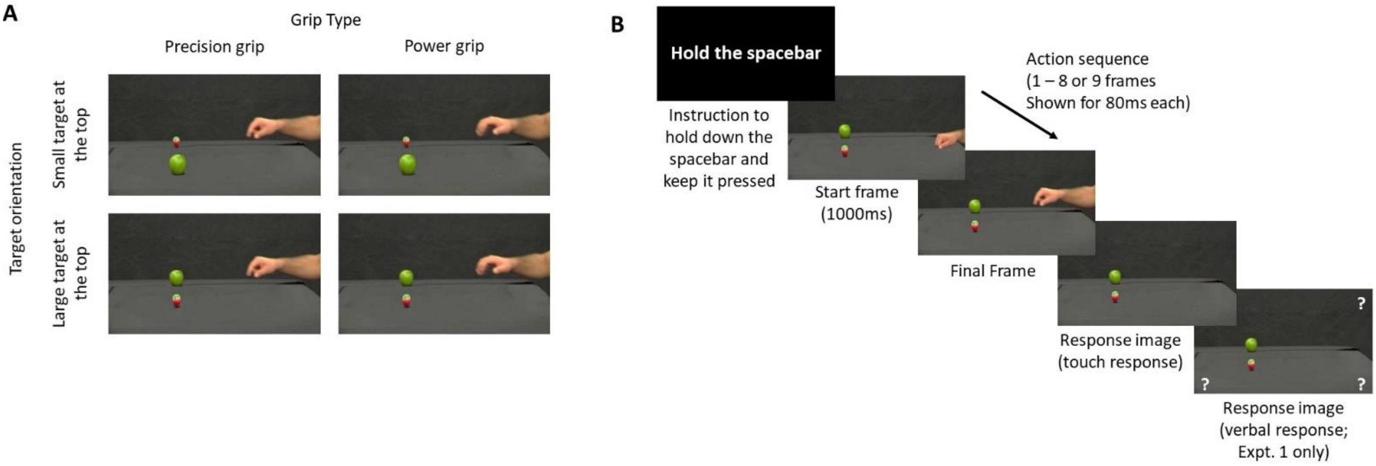

Figure 1. Panel A. Experimental conditions and trial sequence. The objects were arranged with either the small object (strawberry) on top and the large object (apple) on the bottom (top row), or with the large object on top and the small object on the bottom (bottom row). The actor's hand reached with either a precision (small) grip (left column), or with a power (large) grip (right column). Panel B. Example of a trial sequence, showing a "large object on top" configuration with a small grip.

Results

Exclusion criteria

Data filtering was identical to Hudson et al. (2018). In both experiments, individual trials were excluded if the correct response procedure was not followed (e.g. lifting the spacebar before the response image as presented; 2.8% of total trials), or if response initiation or execution times were less than 200ms or more than 3SDs above the sample mean (2.4%, Initiation: mean =355.5ms, *SD*=143.7; Execution: mean =646.0ms, *SD*=240.0). Participants were excluded when too few trials remained after trial exclusions (<50% trials, 5 participants), if their average distance between the real and selected positions was more than 3SDs away from the sample mean (mean =39.0 pixels, *SD*=17.0, 2 participants excluded), or if the correlation between the real and selected positions was more than 3SD below the median r value (X axis: median *r* =.762, *SD* = .113; Y axis: median *r* =.860, *SD* = .098, 2 participants excluded). Two further participants were excluded from Experiment 1, one because they selected the top object as the most likely target object in *all* trials, and one for showing an abnormally large effect in the predicted direction (e.g. 15 times larger than the sample mean) so that we suspected a misunderstanding of the task (e.g. touching the likely target object instead of the hand disappearance point). Removal of these two participants does not affect the results. This left a total of 62 participants in experiment 1 and 63 participants in experiment 2.

Data analysis

Analysis was conducted on the predictive perceptual bias by subtracting the real final coordinates of the tip of the index finger from the participant’s selected coordinates on each trial (see Figure 2). This resulted in separate difference scores along the X and Y axis where positive X and Y scores represented a rightward and upward displacement respectively, and negative X and Y scores represented a leftward and downward displacement respectively. A score of 0 on both axes indicated that the participant selected the real final position exactly. These difference scores were entered into a 2x2x2 mixed ANOVA for the X and Y axis separately, with Grip type (power vs precision) and Object location (large target on top vs small target on top) as repeated measures factors and Experiment (1: explicit prediction vs 2: implicit prediction) as a between-subjects factor. Note that our hypotheses predict perceptual distortions on the Y axis, but not the X axis.

Y axis

We predicted that reports of perceived disappearance points would be distorted towards the apparent target object of the reach, such that reaches would appear to have terminated slightly higher if matching a target object at the top and lower for a target object at the bottom. Overall, the analysis of displacements along the Y axis revealed a main effect of grip type, $F(1,123) = 1130$, $p < .001$, $\eta_p^2 = .902$, showing perceived disappearance points of power grips were displaced further downward than precision grips. This was expected since the power grip is larger and therefore has a lower centre of gravity (Coren & Hoenig, 1972; see also Hudson et al., 2017; Hudson et al., 2018). Importantly, and as predicted, there was an interaction between grip type and object location, $F(1,123) = 16.8$, $p < .001$, $\eta_p^2 = .120$. The disappearance point of power grips was perceptually reported higher when the large target object was placed at the top (-15.1px) than when the large target object was placed at the bottom (-16.0px, $t(124)=2.75$, $p=.007$, $d=.25$). Conversely, the disappearance point for precision grips was reported to be higher when the small target object was at the top (-1.5px) compared to when the small target object was at the bottom (-2.8px, $t(124)=3.64$, $p<.001$, $d=.33$). There was no three-way interaction between grip type, object location and experiment ($F(1,123) = .666$, $p = .416$, $\eta_p^2 = .005$). Indeed, the relevant interaction of grip and object location was present in both experiments, irrespective of whether participants explicitly reported the action's goals after the perceptual judgments (Experiment 1: $F(1,61) = 9.56$, $p = .003$, $\eta_p^2 = .135$, Experiment 2: $F(1,62) = 7.20$, $p = .009$, $\eta_p^2 = .104$, see Figure 2 . Two one-sided tests (TOST) procedure indicated that the observed effect size (d=.16) was significantly within the equivalence bounds of $\Delta L = -.51$ and $\Delta U = .51$, $t(113.89) = -1.95$, $p = .027$. There were no further main effects or interactions (all $F < 1.62$, all $p > .205$).

X axis

We did not have any prediction for the X axis, and all effects are therefore subject to alpha inflation due to multiple comparisons in an ANOVA (Cramer et al., 2015) and should be evaluated against a Bonferroni-adjusted alpha of p < .017. Only the main effect of grip type, $F(1,123) = 503$, $p < .001$, $\eta_p^2 = .804$, passed this adjusted threshold, with the perceived disappearance point of power grips more leftward than precision grips, which again reflects leftward centre of gravity for power grips. There were no further main effects or interactions (all $F < 4.07$, all $p > .046$).

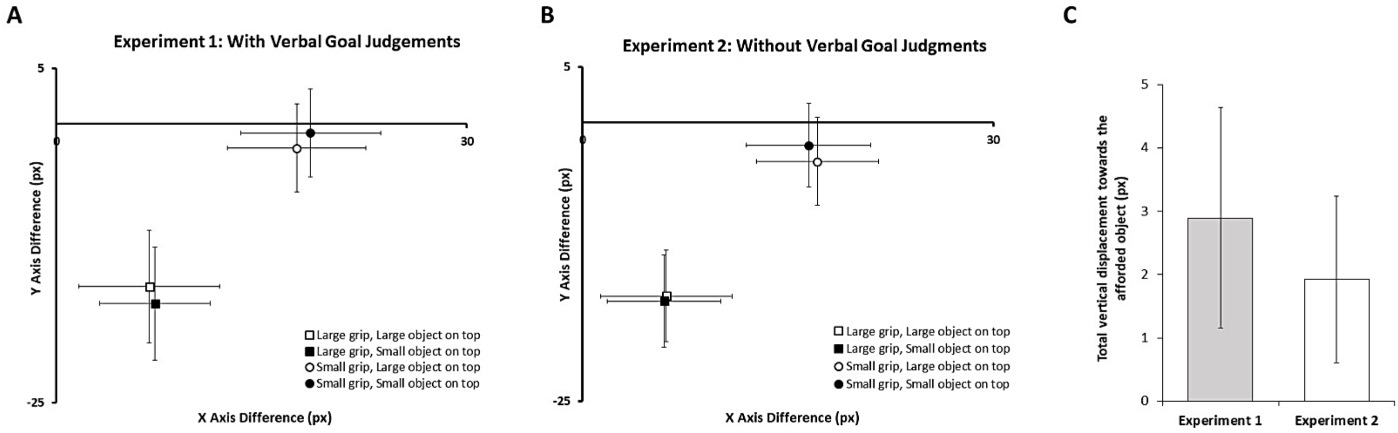

Figure 2. Panel A. Grip type x object interaction for experiment 1. The difference scores between the real final position and the selected final position is plotted for the X axis and Y axis. Panel B. Grip type x object interaction for experiment 2. Panel C. A comparison of the

size of the Y axis interaction in pixels, equivalent to the total amount by which each grip type was distorted towards the congruent object. Error bars depict 95% confidence intervals.

Relationship between perceptual shifts and explicit goal judgments

After each observed action in Experiment 1, participants reported which goal action they thought the action was directed to (i.e. the apple or the strawberry). To test whether participants relied on grip information in these explicit goal judgments, we coded the actual match (i.e. in the stimuli) between a hand grip and an object as 0 and 1 for grips that matched the top or bottom object respectively (i.e. large grip when the large object was at the bottom and small grip when the small object was at the bottom vs. large grip when the large object was at the top and small grip when the small object was at the top), respectively. Verbal goal judgments for each trial were similarly coded as 0 and 1 for goal objects identified at the bottom and goal objects identified at the top, respectively. We then simply derived, for each participant, the proportion of verbal goal judgments that corresponded to the actual match with the goal object. A simple t-test against chance probability (50%) revealed that explicit judgments corresponded well with the actual hand-object match (*M* = 66.7%, *SD* = 16.7%; *t* =31.5, *p* <.001). This confirms that the grip-object-match did not only inform perceptual displacements (see main analysis), but also participants' explicit goal object judgments.

The crucial question was whether this explicit use of grip match information was related to its implicit use in perceptual judgments. If perceptual predictions are shaped by high-level goal attributions or vice versa, then perceptual displacement and explicit judgments in Experiment 1 should be correlated, both across participants and across trials within participants.

Across-participant relationships

We first tested whether any relationships were apparent on the subject-by-subject level. We therefore correlated each participant's proportion of verbal goal judgments that matched actual grip-object links with the crucial interaction contrast value that marks the predictive perceptual shift due to matching grips to object affordances. Surprisingly, the two types of judgments were almost perfectly uncorrelated, *r*(59) = .08, *p* = .518, *N*=62. TOST procedure indicated that the observed effect size (*r*=.08) was significantly within the equivalence bounds of ΔL = –.36 and ΔU =.36, *p* = .011.

Across-trial relationships

Potential relationships were further investigated on a trial-by-trial basis. If explicit and perceptual judgments depend on one another, then trials judged explicitly to be directed towards the top object should also show a perceptual mis-location towards the top, and vice versa for reaches judged to be directed to the bottom object.

We first checked if trial-by-trial perceptual shifts reflected the actual target object location. We therefore correlated, for each participant separately, the actual target object location (coded as 0 or 1) for each trial with the size of the perceptual judgment displacement on the Y axis across all trials of the participant. Testing the resulting fisher-transformed correlation coefficients against zero with a simple t-test, revealed a positive mean correlation between perceptual shifts and target object location across participants (mean *r* = .03, *t* = 3.10, *p* = .003; *d* = .40), see Figure 3A. This trial-by-trial correlation between perceptual judgments on the y-axis and target object location was replicated in Experiment 2, again revealing a positive correlation (mean *r* = .02, *t* = 2.77; *p* = .007; *d* = .35), which did not differ from Experiment 1 (*t*(123)=.512, *p*=.609, *d*=.14). TOST procedure indicated that the observed effect size (*d*=.14) was significantly within the equivalence bounds of ΔL = –.51 and ΔU

=.51, *t*(121.5) = -2.06, *p* = .021. Replicating the results of the main analyses with an across-trials correlational measure, this analysis therefore confirms that actions in which the hand grip matched the object on the top, compared to a grip match to the bottom object, induced larger shifts upwards.

An identical across-trials correlation analysis was also conducted for the relationship between actual target object location and verbal goal judgment. This again revealed a positive correlation between verbal goal judgments and grip information (mean *r* = .44, *t* = 7.20, *p* <.001; *d* = .91), see Figure 3B. Across trials, actions with a grip match to the top were more likely to be judged to be reaching to the top object, compared to the bottom object.

As before, the crucial question was whether explicit verbal goal judgments about an action and the perceptual shifts showed a similar positive relationship. Strikingly, as in the across-participants analysis, there was no correlation between perceptual displacements and the explicit verbal goal judgments (mean r = .01, *t* = .694, *p* = .500, *d* = .09), see Figure 3C. TOST procedure indicated that the observed effect size (d=.09) was significantly within the equivalence bounds of ΔL = –.36 and ΔU =.36, *t*(61) = -2.15, *p* = .018. Thus, while the actual target location informed perceptual judgments and verbal goal judgments, the two types of judgments were not related to each other.

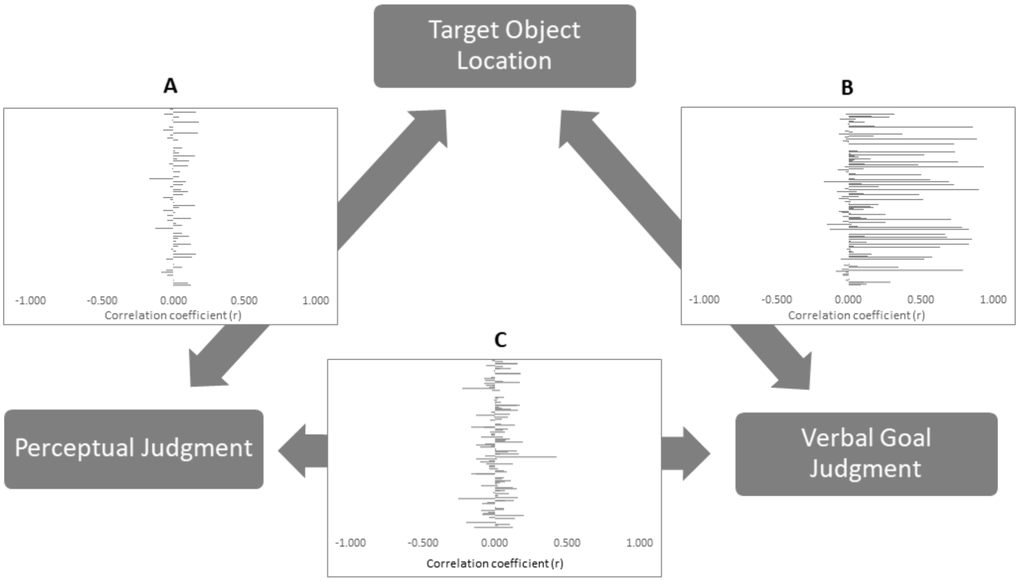

Figure 3. Each participant's correlation coefficient (r) for the correlation between perceptual judgments and the target object location (top object or bottom object, based on the grip-object match, Panel A), the correlation between verbal goal judgments and the target object location (Panel B), and the correlation between perceptual judgments and verbal goal judgments (Panel C).

Discussion

Prior work has shown that people integrate object and action kinematic information to derive the likely goal of observed actions, even while the action is still ongoing (Ambrosini et al., 2011; Bach et al., 2011; Eshuis, Coventry, & Vulchanova, 2009; for a review, see Bach et al., 2014). Here, we tested the hypothesis of predictive processing models (Bach & Schenke, 2017; Hudson, Nicholson, Ellis, et al., 2016; Hudson, Nicholson, Simpson, et al., 2016), that such goal inferences are immediately translated into perceptual predictions about the actions future path towards the inferred goal, and bias perceptual judgments towards these expected trajectories.

The data from two experiments supported this proposal. In each trial, participants observed the initial stages of a reach towards two potential target objects that differed in size, with either a whole-hand power grip or a precision grip, and were asked to perceptually report the hand’s last seen position after its sudden offset. The results revealed consistent biases in perceptual judgments towards action expectations derived from the compatibility between the emerging grip type and object size. While reaches with a power grip were reported to be closer to large objects, reaches with a precision grip were perceived to be closer to small objects, even when actions with the same kinematics were observed and only the location of the relevant target object changed. These perceptual mis-locations were present both when participants were explicitly asked to identify the goal objects in a secondary task (Experiment 1) and when the reach targets were completely task irrelevant and participants were only asked to accurately report the hand’s disappearance point (Experiment 2). Moreover, they were observed even though the actor’s hand started at rest before pre-shaping. The target object was therefore ambiguous before action onset and only became apparent once the action commenced and a specific grip type began to form.

These perceptual displacements towards the expected kinematics support predictive processing models of social perception, which argue that any inferences about an observed action’s goal will (1) give rise to predictions about the action’s further kinematics, which can then (2) bias action perception towards these expectations (Bach & Schenke, 2017; Hudson et al., 2018; Hudson, Nicholson, Ellis, et al., 2016; Hudson, Nicholson, Simpson, et al., 2016; Kilner et al., 2007a, 2007b). They go beyond previous findings in which action expectations were explicitly induced prior to action onset, for example, by asking participants to instruct the (virtual) actor (Hudson, Nicholson, Ellis, et al., 2016), by hearing goal statements of the actor (“I’ll take it!”, Hudson et al., 2017; Hudson, Nicholson, Simpson, et al., 2016), or by presenting a static image of the goal object with or without obstructing objects in the way (Hudson et al., 2018; McDonough et al., 2018, June 8). Here, no such prior information was available. The actions started from a neutral position and the goals only became apparent once it was underway, from the subtle pre-shaping of the hands for the affordances of the goal object (i.e. precision grip when directed towards the small object, power grip when directed towards the large object).

Our results therefore show, first, that predictions are not just made before action onset, but are dynamically adjusted “on-line” as more information becomes available from the unfolding kinematics and are then integrated with the action’s perceptual representation (Donnarumma et al., 2017). Second, they reveal that matching of actions to potential goal objects in the environment plays a major role in this process, as previously hypothesized (see Bach et al., 2014 for a theoretical proposal and review). Third, they go beyond prior work that has already shown that people use such affordance matching to identify the goal of another’s action, guiding eye movements towards it, for example, in an anticipatory manner (e.g., Ambrosini et al., 2011; Ambrosini et al., 2013; Bach et al., 2011). Instead, they reveal that these predictions represent concrete expectations about the next step of a hand’s path through the scene, which interact with the perceptual representation of the kinematics that were actually observed.

A surprising finding was that this matching of actions to goal objects and the resulting perceptual biases appeared to be highly automatic and independent from explicit judgments. We had hypothesized that if perceptual mis-locations and explicit judgments inform each other, then those trials that were explicitly judged to be directed towards the top object should also show upwards mis-locations, and vice versa for judgments towards bottom objects. However, several findings argue against this interpretation. First, in Experiment 1 participants indicated verbally, after each action, which object they believed was the target, but no such response was required in Experiment 2, such that any perceptual bias therefore indexes only spontaneous, implicit goal inferences and prediction. Nevertheless, the perceptual bias towards the grip-matching target was evident – with virtually identical effect sizes – in both experiments. Second, in Experiment 1, correlational analyses across participants showed that the perceptual biases were independent of whether participants' verbal goal judgments revealed a reliance on grip information or not. Third and finally, correlational analyses across trials that directly relate perceptual displacements in a given to trial to verbal goal judgments in the same trial (Experiment 1) confirmed this lack of a top-down influence. Even though the actual goal object predicted both the direction of the perceptual bias and which object was explicitly reported as a target, the perceptual biases and verbal goal judgments remained uncorrelated. In other words, while the mechanisms for action prediction and the goal identification both rely on grip-object matching, the two mechanisms do not strongly inform each other: explicit goal judgments do not induce perceptual biases, nor do perceptual biases induce explicit goal judgments.

This apparent dissociation between explicit and implicit perceptual biases may appear surprising from a viewpoint of predictive coding models, according to which predictions and prediction errors ensure that top-level and lower-level judgments remain aligned (Clark, 2013; Friston & Kiebel, 2009; Hudson et al., 2017). Thus, any inferences on a higher level – for example, what the goal of the action is – would propagate downwards to lower levels and inform perceptual judgments. Conversely, any change in perceptual estimation – whether it is perceived to travel upwards and downwards – would, via prediction errors, inform resulting high-level judgments of action goals. This apparent conflict can be resolved, however, if one assumes that predictions can also emerge locally, from top-down interactions *within* the human perceptual system for the perception of biological motion (for a review, see Scholl & Gao, 2013), without drawing on information external to these networks such as high-level explicit action goal judgments (e.g., Firestone & Scholl, 2016). Indeed, several lines of evidence suggest that the perceptual system itself can detect many aspects of intentional behaviour, without the need for higher-level evaluation, such as whether one actor chases another (Gao, McCarthy & Scholl, 2010), whether an actor pays attention to their reach or whether an actor moves certain limbs in a particular direction, dependent on this attention orientation (e.g. moving an arm to the left when the actor is attending to this target position, Jellema, Baker, Wicker, & Perrett, 2000). The match between hand and goal object may therefore provide another feature from which such lower-level teleological interpretations of observed motion can be derived. This would then suggest that the relevant higher-order perceptual regions are not only sensitive to the goals implied by such matches, as found before (e.g., STS, Gao, Scholl & McCarthy, 2012; Saxe, Xiao, Kovacs, Perrett & Kanwisher, 2004; posterior temporal lobe, Bracci & Peelen, 2013), but that they also use them to predict the action's further path and bias the perceptual representations towards it, independently of the goals explicitly attributed to the other person.

Future studies now need to investigate from what kind of mechanism the perceptual biases emerge. Our prior studies point towards lower-level perceptual processes that determine participants' conscious perceptual experience of the actions, which then drives their explicit judgments. First, in our original studies (Hudson et al., 2017; Hudson, Nicholson, Ellis, et al., 2016; Hudson, Nicholson, Simpson, et al., 2016), these effects were measured not with touch screen judgments, but probe judgments. Participants compared the hand disappearance point with probes in the same, forward, or prior position. Even when these probes were presented only 250 ms after hand disappearance, the perceptual distortions were apparent, suggesting, at the very least, an effect in iconic memory. Second, in our most recent work testing perceived changes to action kinematics in the presence of obstacles, all such effects were eliminated when dynamic visual noise masks were presented briefly (560 ms) after action offset (Hudson et al., 2018), which are known to interfere with re-entrant top-down projections to early visual cortex (Fahrenfort, Scholte, & Lamme, 2007; Lamme, Zipser, &

Spekreijse, 2002). Together, therefore, these findings support a low-level locus of the effects that either reflects the top-down sharpening of the uncertainty during motion perception (i.e. motion blurring, Hammett, 1997), or the filling-in of the expected path after the unexpected sudden offset (Ekman et al., 2017). Neuroimaging studies would be useful to disentangle to what extent the perceptual changes we have measured here reflect changes to early perceptual systems, similar to that seen in various visual illusions and sometimes in motion illusions (e.g. apparent motions, Muckli, Kohler, Kriegeskorte, & Singer, 2005; predicted motion pre-play, Ekman et al., 2017).

Conclusions

The present results reveal that the perceptual experience of others' actions is predictively shaped by the integration of the unfolding action kinematics with the affordances of available goal objects, as proposed by recent predictive models of social perception, (Csibra, 2008; Kilner et al., 2007[ab]; Bach, Nicholson & Hudson, 2014; Schenke & Bach, 2017). These integrations likely emerge at a relatively low-level, from processes within systems for the perception of biological motion, without influences from top-down evaluations of others' goals and intentions. Future studies must now resolve precisely via which mechanism predictions act on perceptual representations, how they help guide own actions towards future states in social interactions, and how prior knowledge is updated and revised if it consistently fails to explain the perceptual input.

Data Availability

Data, code and materials are available at xxxx

Authors' contributions

All authors devised the experiment. Stimuli were created by MC and edited by KLM. The program was developed by MH and KLM. Data was collected by KLM. Data were analysed by KLM and PB. The manuscript was written by KLM, PB and MH. All authors gave final approval for publication.

Funding

This work was supported by the Economic and Social Research Council [grant number ES/J019178/1] awarded to Patric Bach, and a University of Plymouth doctoral student grant awarded to Katrina L. McDonough.

Competing interests

The authors declare no competing interests.

References

Adams, R. B., Ambady, N., Macrae, C. N., & Kleck, R. E. (2006). Emotional expressions forecast approach-avoidance behavior. *Motivation and Emotion, 30*(2), 177-186. doi:10.1007/s11031-006-9020-2

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the 'light-from-above' prior. *Nature Neuroscience, 7*(10), 1057-1058. doi:10.1038/nn1312

Ambrosini, E., Costantini, M., & Sinigaglia, C. (2011). Grasping with the eyes. *Journal of Neurophysiology, 106*(3), 1437-1442. doi:10.1152/jn.00118.2011

Ambrosini, E., Reddy, V., de Looper, A., Costantini, M., Lopez, B., & Sinigaglia, C. (2013). Looking Ahead: Anticipatory Gaze and Motor Ability in Infancy. *PLOS ONE, 8*(7), e67916. doi:10.1371/journal.pone.0067916

Bach, P., Bayliss, A. P., & Tipper, S. P. (2011). The predictive mirror: interactions of mirror and affordance processes during action observation. *Psychonomic Bulletin & Review, 18*(1), 171-176. doi:10.3758/s13423-010-0029-x

Bach, P., Knoblich, G., Gunter, T. C., Friederici, A. D., & Prinz, W. (2005). Action Comprehension: Deriving Spatial and Functional Relations. *Journal of Experimental Psychology: Human Perception and Performance, 31*(3), 465-479. doi:10.1037/0096-1523.31.3.465

Bach, P., Nicholson, T., & Hudson, M. (2014). The affordance-matching hypothesis: how objects guide action understanding and prediction. *Frontiers in Human Neuroscience, 8*. doi:10.3389/fnhum.2014.00254

Bach, P., & Schenke, K. C. (2017). Predictive social perception: Towards a unifying framework from action observation to person knowledge. *Social and Personality Psychology Compass, 11*(7), e12312. doi:10.1111/spc3.12312

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5*(8), 617-629. doi:10.1038/nrn1476

Bloj, M. G., Kersten, D., & Hurlbert, A. C. (1999). Perception of three-dimensional shape influences colour perception through mutual illumination. *Nature, 402*(6764), 877-879. doi:10.1038/47245

Bracci, S., & Peelen, M. V. (2013). Body and Object Effectors: The Organization of Object Representations in High-Level Visual Cortex Reflects Body-Object Interactions. *Journal of Neuroscience, 33*(46), 18247-18258. doi:10.1523/jneurosci.1322-13.2013

Chetverikov, A., & Ivanchei, I. (2016). Seeing “the Dress” in the Right Light: Perceived Colors and Inferred Light Sources. *Perception, 45*(8), 910-930. doi:10.1177/0301006616643664

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(03), 181-204. doi:10.1017/s0140525x12000477

Coren, S., & Hoenig, P. (1972). Effect of Non-Target Stimuli upon Length of Voluntary Saccades. *Perceptual and Motor Skills, 34*(2), 499-508. doi:10.2466/pms.1972.34.2.499

Costantini, M., Ambrosini, E., & Sinigaglia, C. (2012). Out of your hand's reach, out of my eyes' reach. *Quarterly Journal of Experimental Psychology, 65*(5), 848-855. doi:10.1080/17470218.2012.679945

Cramer, A. O. J., van Ravenzwaaij, D., Matzke, D., Steingroever, H., Wetzels, R., Grasman, R. P. P. P., . . . Wagenmakers, E.-J. (2015). Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies. *Psychonomic Bulletin & Review, 23*(2), 640-647. doi:10.3758/s13423-015-0913-5

Csibra, G. (2008). Action mirroring and action understanding: An alternative account. In P. Haggard, Y. Rosetti, & M. Kawato (Eds.), *Sensorymotor foundations of higher cognition. Attention and performance XXII* (pp. 435-459). Oxford: Oxford University Press.

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? *Trends in Cognitive Sciences, 22*(9), 764-779. doi:10.1016/j.tics.2018.06.002

Donnarumma, F., Costantini, M., Ambrosini, E., Friston, K., & Pezzulo, G. (2017). Action perception as hypothesis testing. *Cortex, 89,* 45-60. doi:10.1016/j.cortex.2017.01.016

Ekman, M., Kok, P., & de Lange, F. P. (2017). Time-compressed preplay of anticipated events in human primary visual cortex. *Nature Communications, 8,* 15276. doi:10.1038/ncomms15276

Eshuis, R., Coventry, K. R., & Vulchanova, M. (2009). Predictive Eye Movements Are Driven by Goals, Not by the Mirror Neuron System. *Psychological Science, 20*(4), 438-440. doi:10.1111/j.1467-9280.2009.02317.x

Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2007). Masking Disrupts Reentrant Processing in Human Visual Cortex. *Journal of Cognitive Neuroscience, 19*(9), 1488-1497. doi:10.1162/jocn.2007.19.9.1488

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences, 39*. doi:10.1017/s0140525x15000965

Freyd, J. J., & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(1), 126-132. doi:10.1037//0278-7393.10.1.126

Frischen, A., & Tipper, S. P. (2006). Long-term gaze cueing effects: Evidence for retrieval of prior states of attention from memory. *Visual Cognition, 14*(3), 351-364. doi:10.1080/13506280544000192

Friston, K., & Kiebel, S. (2009). Cortical circuits for perceptual inference. *Neural Networks, 22*(8), 1093-1104. doi:10.1016/j.neunet.2009.07.023

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 2*(12), 493-501.

Gao, T., McCarthy, G., & Scholl, B. J. (2010). The Wolfpack Effect: Perception of animacy irresistibly influences interactive behavior. *Psychological Science, 21*(12), 1845-1853. doi:10.1177/0956797610388814

Gao, T., Scholl, B. J., & McCarthy, G. (2012). Dissociating the Detection of Intentionality from Animacy in the Right Posterior Superior Temporal Sulcus. *Journal of Neuroscience, 32*(41), 14276-14280. doi:10.1523/jneurosci.0562-12.2012

Hammett, S. T. (1997). Motion blur and motion sharpening in the human visual system. *Vision Research, 37*(18), 2505-2510. doi:10.1016/s0042-6989(97)00059-x

Hubbard, T. L. (2005). Representational momentum and related displacements in spatial memory: A review of the findings. *Psychonomic Bulletin & Review, 12*(5), 822-851. doi:10.3758/bf03196775

Hudson, M., Bach, P., & Nicholson, T. (2017). You said you would! The predictability of other’s behavior from their intentions determines predictive biases in action perception. *Journal of Experimental Psychology: Human Perception and Performance, 44*(2), 320-335. doi:10.1037/xhp0000451

Hudson, M., McDonough, K. L., Edwards, R., & Bach, P. (2018). Perceptual Teleology: Expectations of Action Efficiency Bias Social Perception. *Proceedings of the Royal Society: B.*(285(1884)), 20180638. doi:10.1098/rspb.2018.0638

Hudson, M., Nicholson, T., Ellis, R., & Bach, P. (2016). I see what you say: Prior knowledge of other's goals automatically biases the perception of their actions. *Cognition, 146*, 245-250. doi:10.1016/j.cognition.2015.09.021

Hudson, M., Nicholson, T., Simpson, W. A., Ellis, R., & Bach, P. (2016). One step ahead: The perceived kinematics of others' actions are biased toward expected goals. *Journal of Experimental Psychology: General, 145*(1), 1-7. doi:10.1037/xge0000126

Jacob, P., & Jeannerod, M. (2005). The motor theory of social cognition: a critique. *Trends in Cognitive Sciences, 9*(1), 21-25.

Jellema, T., Baker, C. I., Wicker, B., & Perrett, D. I. (2000). Neural Representation for the Perception of the Intentionality of Actions. *Brain and Cognition, 44*(2), 280-302. doi:10.1006/brcg.2000.1231

Joyce, K., Schenke, K., Bayliss, A., & Bach, P. (2015). Looking ahead: Anticipatory cueing of attention to objects others will look at. *Cognitive Neuroscience, 7*(1-4), 74-81. doi:10.1080/17588928.2015.1053443

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007a). The mirror-neuron system: a Bayesian perspective. *NeuroReport, 18*(6), 619-623. doi:10.1097/wnr.0b013e3281139ed0

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007b). Predictive coding: an account of the mirror neuron system. *Cognitive Processing, 8*(3), 159-166. doi:10.1007/s10339-007-0170-2

Lamme, V. A. F., Zipser, K., & Spekreijse, H. (2002). Masking Interrupts Figure-Ground Signals in V1. *Journal of Cognitive Neuroscience, 14*(7), 1044-1053. doi:10.1162/089892902320474490

McDonough, K. L., Hudson, M., & Bach, P. (2018, June 8). Seeing the best way forward: Intentionality cues bias action perception toward the most efficient trajectory. doi:10.31234/osf.io/qwjtf

Muckli, L., Kohler, A., Kriegeskorte, N., & Singer, W. (2005). Primary Visual Cortex Activity along the Apparent-Motion Trace Reflects Illusory Perception. *PLoS Biology, 3*(8), e265. doi:10.1371/journal.pbio.0030265

Nicholson, T., Roser, M., & Bach, P. (2017). Understanding the Goals of Everyday Instrumental Actions Is Primarily Linked to Object, Not Motor-Kinematic, Information: Evidence from fMRI. *PLOS ONE, 12*(1), e0169700. doi:10.1371/journal.pone.0169700

Nijhawan, R. (1994). Motion extrapolation in catching. *Nature, 370*(6487), 256-257. doi:10.1038/370256b0

Nijhawan, R. (2008). Visual prediction: Psychophysics and neurophysiology of compensation for time delays. *Behavioral and Brain Sciences, 31*(02). doi:10.1017/s0140525x08003804

Oosterhof, N. N., Wiggett, A. J., Diedrichsen, J., Tipper, S. P., & Downing, P. E. (2010). Surface-based information mapping reveals crossmodal vision–action representations in human parietal and occipitotemporal cortex. *Journal of Neurophysiology, 104*(2), 1077-1089.

Prinz, W., & Rapinett, G. (2008). Filling the gap: Dynamic representation of occluded action. *Emerging Communication, 10*(223).

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

Sartori, L., Becchio, C., & Castiello, U. (2011). Cues to intention: The role of movement information. *Cognition, 119*(2), 242-252. doi:10.1016/j.cognition.2011.01.014

Saxe, R., Xiao, D. K., Kovacs, G., Perrett, D. I., & Kanwisher, N. (2004). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia, 42*(11), 1435-1446. doi:10.1016/j.neuropsychologia.2004.04.015

Schenke, K. C., Wyer, N. A., & Bach, P. (2016). The Things You Do: Internal Models of Others' Expected Behaviour Guide Action Observation. *PLOS ONE, 11*(7), e0158910. doi:10.1371/journal.pone.0158910

Scholl, B. J., & Gao, T. (2013). Perceiving Animacy and Intentionality: Visual processing or higher-level judgment. In M. D. Rutherford & V. A. Kuhlmeier (Eds.), *Social Perception: Detection and interpretation of animacy, agency, and intention,* (pp. 197-230): The MIT Press.

Southgate, V., Johnson, M. H., Karoui, I. E., & Csibra, G. (2010). Motor System Activation Reveals Infants' On-Line Prediction of Others' Goals. *Psychological Science, 21*(3), 355-359. doi:10.1177/0956797610362058

Uithol, S., van Rooij, I., Bekkering, H., & Haselager, P. (2011). What do mirror neurons mirror? *Philosophical Psychology, 24*(5), 607-623.

van Elk, M., van Schie, H., & Bekkering, H. (2014). Action semantics: A unifying conceptual framework for the selective use of multimodal and modality-specific object knowledge. *Physics of Life Reviews, 11*(2), 220-250. doi:10.1016/j.plrev.2013.11.005

Vesper, C., van der Wel, R. P. R. D., Knoblich, G., & Sebanz, N. (2011). Making oneself predictable: reduced temporal variability facilitates joint action coordination. *Experimental Brain Research, 211*(3-4), 517-530. doi:10.1007/s00221-011-2706-z

von der Lühe, T., Manera, V., Barisic, I., Becchio, C., Vogeley, K., & Schilbach, L. (2016). Interpersonal predictive coding, not action perception, is impaired in autism. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1693), 20150373. doi:10.1098/rstb.2015.0373