Cognition 260 (2025) 106122

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Full length article

Kinematics in context: Predicting other's action intentions entails the perception of affordances

Ayeh Alhasan [a,*], Eyal Karin [a], Nathan Caruana [b], Emily Cross [c], David Kaplan [a], Michael J. Richardson [a]

[a] School of Psychological Sciences, Macquarie University, Sydney, 2109, NSW, Australia
[b] College of Education, Psychology and Social Work, Flinders University, Adelaide, 5001, South Australia, Australia
[c] Department of Humanities, Social and Political Sciences, ETH Zürich, Zürich, 8092, Switzerland

ARTICLE INFO

Dataset link: https://osf.io/4dj9r/

Keywords:
Intention
Intention prediction
Action prediction
Action observation
Action understanding
Kinematics
Environmental-context
Affordances

ABSTRACT

Intention prediction is essential for successful social interaction, but traditional research focusing solely on movement kinematics often overlooks the array of action possibilities in natural settings. This study employs a mixed-methods approach to explore intention prediction, analysing free-text responses from participants who watched videos of an actor reaching for a cup, bottle, or spoon, each with a distinct intention. Each video included varied environmental contexts to suggest specific intentions (e.g., full cups for drinking, empty cups for clearing) or presented ambiguous contexts (e.g., half-full cups). We found that participants' intention predictions depended on the variety of action possibilities presented by both kinematics and context. Participants tended to identify the primary action possibility of the grasped item as the intended action when both kinematics and context supported its feasibility. Predictions diversified when kinematics or context suggested that the object's primary action was less likely. Our findings suggest that while intention predictions can sometimes be inaccurate, they align with the (most functional) action possibilities (i.e., affordances) indicated by the actor's movements within a given context.

1. Introduction

Understanding how people discern, predict, and act on the intentions of others has been a long-standing area of research in the neurocognitive and psychological sciences. The importance of intention recognition for effective social interaction and teamwork (Helm et al., 2020; Knoblich & Sebanz, 2008; Williams, 2009; Williams & Jackson, 2019), as well as its relevance in the development of robust human–machine systems (Kim et al., 2017; Lin & Lukodono, 2021; Tahboub, 2006; Zhang et al., 2019), has continued to drive research in this domain over the past several decades. Although a significant body of this research has focused on neurocognitive processes that support action observation and intention prediction (Caligiore et al., 2013; Kilner, 2011; Rizzolatti et al., 2014), identifying what information most accurately reveals the intentions of others has been of equal importance (Becchio et al., 2018; Johansson, 1973; McEllin et al., 2018; Runeson & Frykholm, 1983; Stapel et al., 2012).

Of particular relevance here is the growing body of research showing that the way people or objects move in space (i.e., their kinematics) offers valuable information about the motor intention or action goal of an individual – e.g., reaching to grasp and drink from a cup on the table – (Becchio et al., 2018; Cavallo et al., 2016; Manera et al., 2010; Runeson & Frykholm, 1983). Observers can often detect this information and in certain situations use it to predict actions before they are completed, including discerning deceptive intentions (Elsner et al., 2012; Manera et al., 2010; Runeson & Frykholm, 1983; Sebanz & Shiffrar, 2009). Such studies frequently employ simplified stimuli like point-light displays, which record movements using motion-tracking markers, small lights, or reflective stickers placed on crucial points of an actor's body and limbs, so that observers can focus solely on bodily motion without further contextual details (Johansson, 1973; Runeson & Frykholm, 1983).

Although research on how kinematics reveal others' intentions has emphasized the importance of movement kinematics for predicting intentions, numerous studies have also shown that individuals often struggle to accurately predict the intentions of others based solely on kinematic information (Naish et al., 2013; Rutkowska et al., 2021). This is due to the fact that movement kinematics and intentions typically exhibit a many-to-many relationship (Bach et al., 2005; Baird, 1999; Baldwin & Baird, 1999; Clark, 2016; Csibra, 2008; Jacob & Jeannerod,

* Corresponding author.
E-mail address: ayeh.alhasan@mq.edu.au (A. Alhasan).

https://doi.org/10.1016/j.cognition.2025.106122
Received 2 August 2024; Received in revised form 9 March 2025; Accepted 12 March 2025
Available online 27 March 2025

2005; Kilner et al., 2007; Malle et al., 2001; Searle, 1984; Thompson et al., 2019; Uithol et al., 2011). For instance, the same kinematics, such as reaching out to grasp a cup, can reflect multiple intentions, including drinking, pouring, moving, cleaning, passing, or even throwing. Conversely, the same intention (e.g., drinking water) can be realized in various ways, such as by grasping and lifting a cup or a bottle, from a tap, or water cooler, or by opening one's mouth in the rain. For kinematics alone to unambiguously specify an action intention, there would need to be a one-to-one mapping between each distinct movement pattern and specific action intention.

Since no such one-to-one mapping exists between movement kinematics and intentions, it becomes essential to consider the environmental context in which these movements occur. The degree to which movement kinematics specify the intentions of another agent in real-life scenarios depends on the number of action possibilities that a given environmental context affords them (i.e., the number of affordances) (Gibson, 1979; Mark et al., 2015; Marsh et al., 2006; Richardson et al., 2010, 2008; Turvey, 1992). In other words, the fewer action possibilities a given context offers, the greater the specificity of movement kinematics for predicting action intentions. By contrast, the greater the number of action possibilities within a given context, the less specific the movement kinematics will be for predicting intentions. Thus, while point-light displays and related paradigms are suitable for testing hypotheses about the isolated role of kinematic information in intention prediction, these paradigms cannot address how both kinematic and contextual information might collectively determine our ability to predict others' action intentions.

Not surprisingly, the critical role of environmental or contextual information for predicting the intentions of others has been demonstrated in numerous studies (e.g., Azaad and Sebanz 2023, Iacoboni et al. 2005, Stapel et al. 2012). Many of these studies, however, have relied on still images to present participants with either context-only information (e.g. image of objects in a scene) or context combined with the end-state grasp configuration (e.g. hand grasping a mug by the handle) (e.g., Boria et al. 2009, Ortigue et al. 2010, Uithol et al. 2023), thereby neglecting the influence of kinematic information on intention prediction. Although, more recent studies have attempted to investigate the role of kinematic information in context using reach-to-grasp videos (for e.g., Amoruso et al. 2018, Amoruso and Urgesi 2016, Betti et al. 2022, Finisguerra et al. 2020), these studies often truncate the videos just before contact with the object (i.e., excluding the end-state grasp configuration information), focusing exclusively on pre-contact kinematics (for an exception, though, see Naish et al. 2013). The latter approach is predicated on the assumption that the end-state grasp configuration specifies the intention (e.g., Boria et al. 2009, Ortigue et al. 2010, Spunt and Adolphs 2014, Uithol et al. 2023), while kinematics primarily offer prospective information about this end-state. To date, however, no research has directly assessed this assumption, and it remains unclear whether observing an actor's movements (i.e. kinematics) as well as their end-state limb or end-effector (e.g., grasp) configuration provides additional information about intentions beyond what is available from the end-state grasp alone.

Finally, previous work investigating the ability of observers to predict the action intentions of others has almost exclusively relied on a two-alternative forced choice (2AFC) design (e.g., Amoruso et al. 2016, 2018, Amoruso and Urgesi 2016, Cavallo et al. 2016, Koul et al. 2019). This approach, which provides participants with a binary choice between two possible action outcomes, might oversimplify the task of predicting intentions (e.g., Järvilehto 1998, Yeshurun et al. 2008). The use of a 2AFC paradigm narrows the range of possible actions (affordances) available for the participants to choose from, and does not accurately reflect the many-to-many relationships between movement kinematics, contextual information, and action intentions. Provided with two choices and two distinct kinematic profiles, participants might simply associate each option with the most sensible kinematic presentation. Consequently, this paradigm might not provide a comprehensive view of the full range of determinants of action intention predictions. What is needed is an exploration of how kinematics and context (kinematics-in-context) information operate together to specify the action intentions of others.

1.1. Current study

The primary objectives of the current study were twofold. First, this work examined the widely held but empirically untested assumption that the end-state grasp configuration of a reaching action was the most informative for predicting intentions. We did this by investigating how participants' responses varied as a function of the type of stimuli used (still images vs. videos). The second and more important aim, was to explore the diversity of responses and types of intentions predicted when participants were allowed to respond freely, rather than being constrained by a forced-choice 2AFC design.

To address these objectives, the current study utilized a reach-to-grasp paradigm (Amoruso et al., 2016, 2018; Amoruso & Urgesi, 2016; Cavallo et al., 2016; Koul et al., 2019). Importantly, both the action intentions of an actor and the environmental context were systematically manipulated. Participants observed a hand reaching to grasp a cup, bottle, or spoon (see Fig. 1 for examples of the cup and bottle stimuli). The actor's intention was to drink, pour, eat, or clear the table with only three possible intentions assigned to each object. Videos showed the reaching kinematics including grasp. An ambiguous kinematics condition was also included, in which participants viewed a hand being placed on the table without a clear object-directed intention. The environmental context was manipulated to produce three different contexts for each item. Two contexts indicated one intention over the other (e.g., full cups indicating the intention to drink), while a third context was ambiguous by allowing either intention to unfold (e.g. half-full cups indicating the intention to drink or clear). The combination of the three intentions and three contexts resulted in nine unique conditions per object.

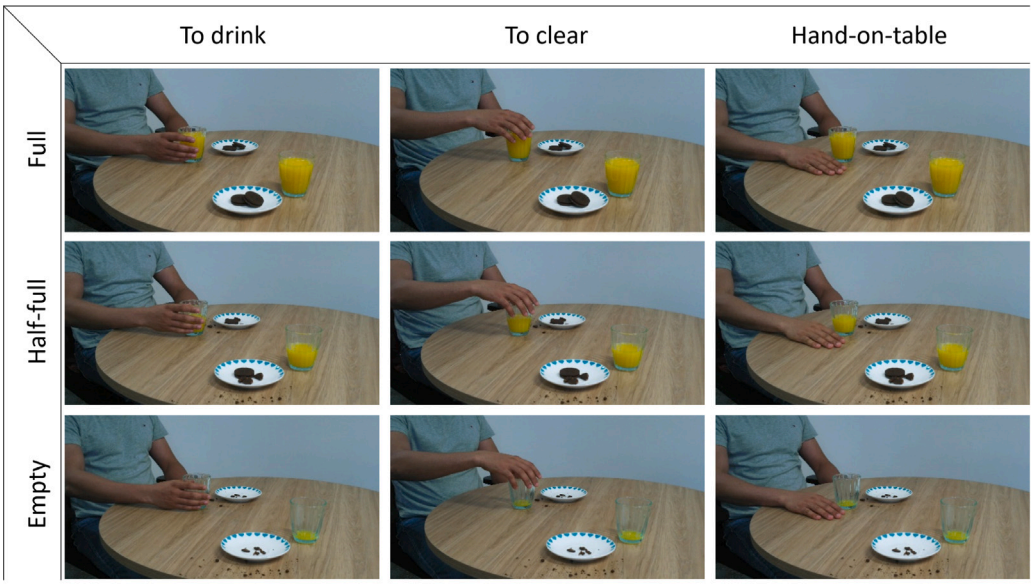
To explore the role of kinematics in intention prediction, three types of stimuli were presented to participants: (1) images of contextual scenes before action (i.e., context-only; pre-image), (2) images of the end-state configuration of the hand and context at grasp (i.e., context and end-state grasp; post-image), and (3) video clips of the reaching kinematics up to and including the grasp (i.e., context, end-state grasp and kinematics; video). By manipulating the stimulus types, we aimed to understand whether kinematics provide additional intention-specifying information beyond that offered by the end-state of a grasping action.

The current study adopted an exploratory approach, with the goal of examining whether there were patterns in participants' responses and whether these patterns generate insights into the degree to which contextual and kinematic information potentially moderate intention predictions. Although no hypotheses were defined a priori, we did anticipate that the richness and diversity of responses will provide a more comprehensive understanding of how individuals predict intentions from observed actions across varying conditions. And, in turn, highlight the methodological challenges associated with relying on 2AFC responses and kinematic or context free stimuli.

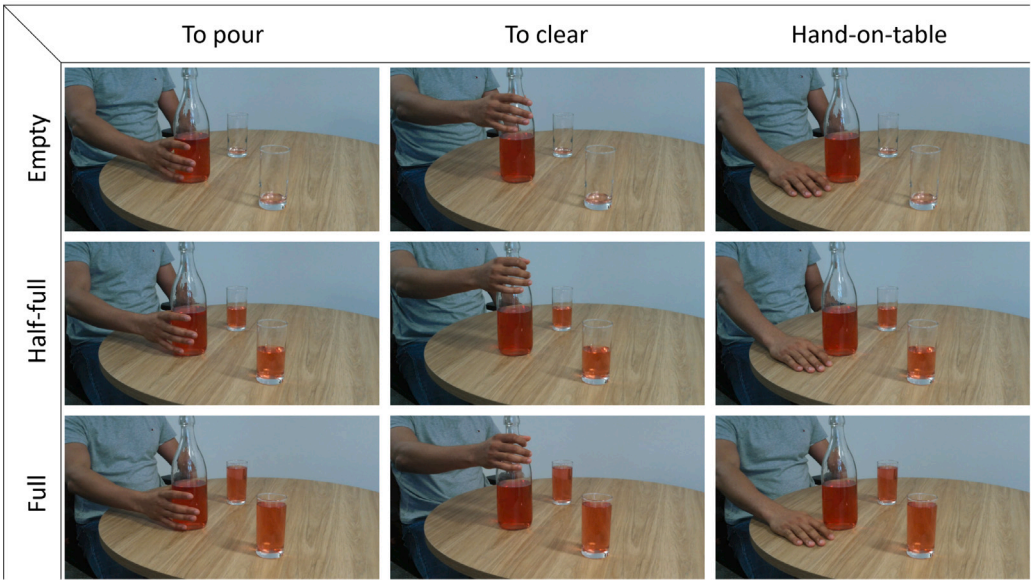
2. Method

2.1. Ethics

All procedures implemented in this study were approved by the Macquarie University Human Research Ethics Committee (ID: 2970). Participants gave their informed consent to participate in this study prior to participation.



(a) Cup



(b) Bottle

Fig. 1. End-state configurations of the nine different stimuli conditions for (a) the Cup and (b) the Bottle stimuli. Images from left to right depict different action intentions: the actor drinking from the cup, pouring from the bottle, clearing the table, or placing his hand on the table. Images from top to bottom illustrate the manipulation of environmental context, with cups being almost full, half full, or almost empty. These images reflect the exact visual stimuli presented to participants.

2.2. Participants

Participants were recruited using an online participant recruitment service, Prolific. A total of 120 participants were initially recruited for this study. Participation was open to all countries available on Prolific, but restricted to those individuals fluent in English. To ensure appropriate stimulus presentation, only participants who met the following criteria were recruited: (1) using laptops or desktops to complete the study, (2) had a stable internet connection with a PING of less than 50 ms and more than 15 mbps download speed. Additionally, we aimed to ensure minimal lag in video presentation. To do so, trials that exceeded the expected clip duration by more than 800 ms were flagged. Participants with more than 10 flagged trials were excluded. This resulted in the exclusion of 18 participants. Two more participants were excluded for providing responses in languages other than English. This resulted in a total of 100 participants, divided equally into two condition groups, with 20 females and 30 males in the pre-image/video stimulus condition and 23 females and 27 males in the pre-image/post-image stimulus condition (see below for details about the stimulus conditions).

2.3. Stimuli

2.3.1. Intention

The intended action of the actor used to create the stimuli for the current study was manipulated by asking the actor to perform three different actions. The first two actions were object-directed: one (intention A) involved the actor grasping the cup, bottle, or spoon to drink, pour, or eat, respectively, and the other (Intention B) involved removing the object from the table. The actor was instructed to grasp the item naturally, in the manner they deemed appropriate for achieving the intended goal. The third action (Intention C) was a non-object-directed action, with the actor being asked to place their hand on the table next to the object (i.e., cup, bottle, spoon). This manipulation resulted in

kinematic and end-state profiles that were either more predictive of a specific, object-directed action intention (intentions A and B) or not predictive of any specific object-directed action intention.

2.3.2. Context

For each intended action, we varied the contextual information by changing the action-specificity of the environmental scene. With each item, the environment was designed to specify Intention A, Intention B or not specify either intentions by equally affording both. For the cup, we varied the context between (1) cups full of juice and plates with cookies, (2) half-full cups and half-eaten cookies with cookie crumbs, and (3) almost empty cups and plates with cookie crumbs (see Fig. 1(a), top). For the bottle, we had a scene with a bottle and cups and a more subtle context manipulation where we manipulated the cups, rather than the bottle, to have (1) almost empty cups, (2) half-full cups and (3) almost full cups (see Fig. 1(b), bottom). For the spoon, the context varied between (1) a full bowl and fresh lemon, (2) a half-full bowl and squeezed lemon, and (3) an almost empty bowl and squeezed lemon (see Supplementary Material (Figures) for examples of the spoon and bowl stimuli).

It should be noted that the contexts were designed to always afford and render plausible the actor's intended action. For instance, in contexts specifying an intention to clear, cups were depicted with a small amount of juice remaining, making the intention to drink still possible. This enabled an investigation of how participants predict someone's intentions at the point of grasp, given contextual information that made their intended action seem improbable but not impossible.

2.3.3. Stimulus type: Video clips and pre- and post-images

Video clips were recorded using a Zoom Q3HD video recorder and edited using a custom MATLAB code to begin 1 s before movement onset and end at the point of grasp completion. Each clip ranged in duration between 2–3 s. The actor who was recorded was a 36-year-old male, who was naive to the purpose of the study.

From the video recordings, images of the first frame of each unique context scene were extracted, captured before the start of the reach-to-grasp action. In these pre-images, the actor's hand was in a resting position on the table, close to his torso. Each image was displayed for 2 s. The post-images corresponded to the final frame of each video, where the actor's hand was in its end-state, either grasping the items or placed on the table next to the item and away from his torso. Fig. 1 illustrates an example of the nine stimuli conditions from the post-images for the cup and bottle.

Note that the post-images and clips across three action intentions and three environmental contexts resulted in nine stimulus conditions per object, whereas the pre-images only corresponded to three stimuli conditions, one for each environmental context (i.e., action intention, movement, or end-state configuration were irrelevant for pre-images).

All stimuli used in the study are available on OSF (https://osf.io/4dj9r/) (Alhasan et al., 2024).

2.4. Procedure

The study used Qualtrics (https://www.qualtrics.com) to present stimuli and record the responses of the participants. As noted above, participants were pseudo-radomly assigned one of two stimulus condition groups. In each group, the participants viewed two blocks of trials. In the first block of trials, both groups were presented with the pre-image stimuli. After each image, participants were asked to answer the following question: 'What is this person's intent?' Responses were collected in free text format. For the second block of trials, one group viewed the post-image stimuli, while the other group viewed the video stimuli. After each image or video stimulus in the second block, participants were again asked: 'What is this person's intent?' The order of stimuli was randomized within each trial block and participants were presented with each unique stimulus only once. Note that a detailed kinematic analysis of the video stimuli was not performed, as pilot studies verified that the different intentions consistently produced three distinct observable kinematic patterns (e.g. grasping the cup from the side, grasping from the top, or placing the hand on the table).

2.5. Response coding and statistical analysis

In this study, a mixed-method approach (Creswell & Clark, 2017; Teddlie & Tashakkori, 2003) was employed to analyse participant response data. This approach effectively combines the detailed exploration of qualitative data collection with the systematic rigour of quantitative analysis, allowing for an in-depth examination of the various types of responses and clusters.

The analysis followed a structured sequence beginning with the initial coding of participant responses to extract the main verbs and verb statements. This was followed by a frequency analysis to identify unique verbs across different conditions, visualized using Sankey diagrams and word clouds. Generalized Linear Models (GLM) were used to quantify the effect of the stimulus type (pre-image, post-image or video) on the number of unique verbs, providing a measure of variance in participants' predictions.

Responses to the video stimuli were then classified into object-directed and non-object directed intentions, further subdivided into lower-order and higher-order intentions (Wagman et al., 2018) and verb clusters were identified (full cluster classification is detailed below). A secondary coding applied this classification schema to the identified verbs, and the frequency data were converted to proportions for detailed analysis. Finally, Generalized Logistic Mixed Random Effects Models (GLME) were used to quantify the effects of intention and context on response likelihood, providing a comprehensive understanding of participants' predictions of others' intentions.

2.5.1. Initial coding

The responses of participants were first manually coded by an experimenter (the first author of the study) and a paid research assistant, who was blind to the purpose of the study. Each coder extracted the main verbs and verb statements from the participants' responses. In cases where responses included more than one intention (e.g., 'to eat and drink'), each verb was coded as a separate intention. Missing data was kept to a minimum through survey design features that allowed participants to type freely and required minimal response to progress.

2.5.2. Frequency analysis

To compare the range of responses across the different stimulus types (pre-images, post-images, videos), a frequency analysis was first performed for each condition (action intention x environmental context) for each object (cup, bottle, spoon) to identify the unique verbs and the number of times each verb appeared in participant responses for each stimulus type. Sankey diagrams were used to identify the differences in response **type** and **variance** for the different conditions, items and types of stimuli (see Fig. 2). Word clouds were also created to illustrate the distribution of responses for each condition within the video stimuli (see Fig. 4 and Supplementary Material (Figures)).

2.5.3. Classification

Based on a preliminary analysis of Sankey diagrams and word clouds, the participant's responses were then classified into two main groups: *object-directed* and *non-object directed* intentions. Responses under object-directed intentions were split further into lower-order intentions, describing the immediate action with the object (e.g., grasping a cup, picking up, holding, moving a bottle) all grouped into a single 'Move Object' cluster, and higher-order intentions, which refer to the broader goal of the actor's action (e.g., to drink, eat, pour), classified into four clusters: 'Drink', 'Fill/Pour', 'Eat', and 'Clean'. For non-object directed intentions, participant responses were similarly divided into lower-order intentions, outlining the physical movement of the body

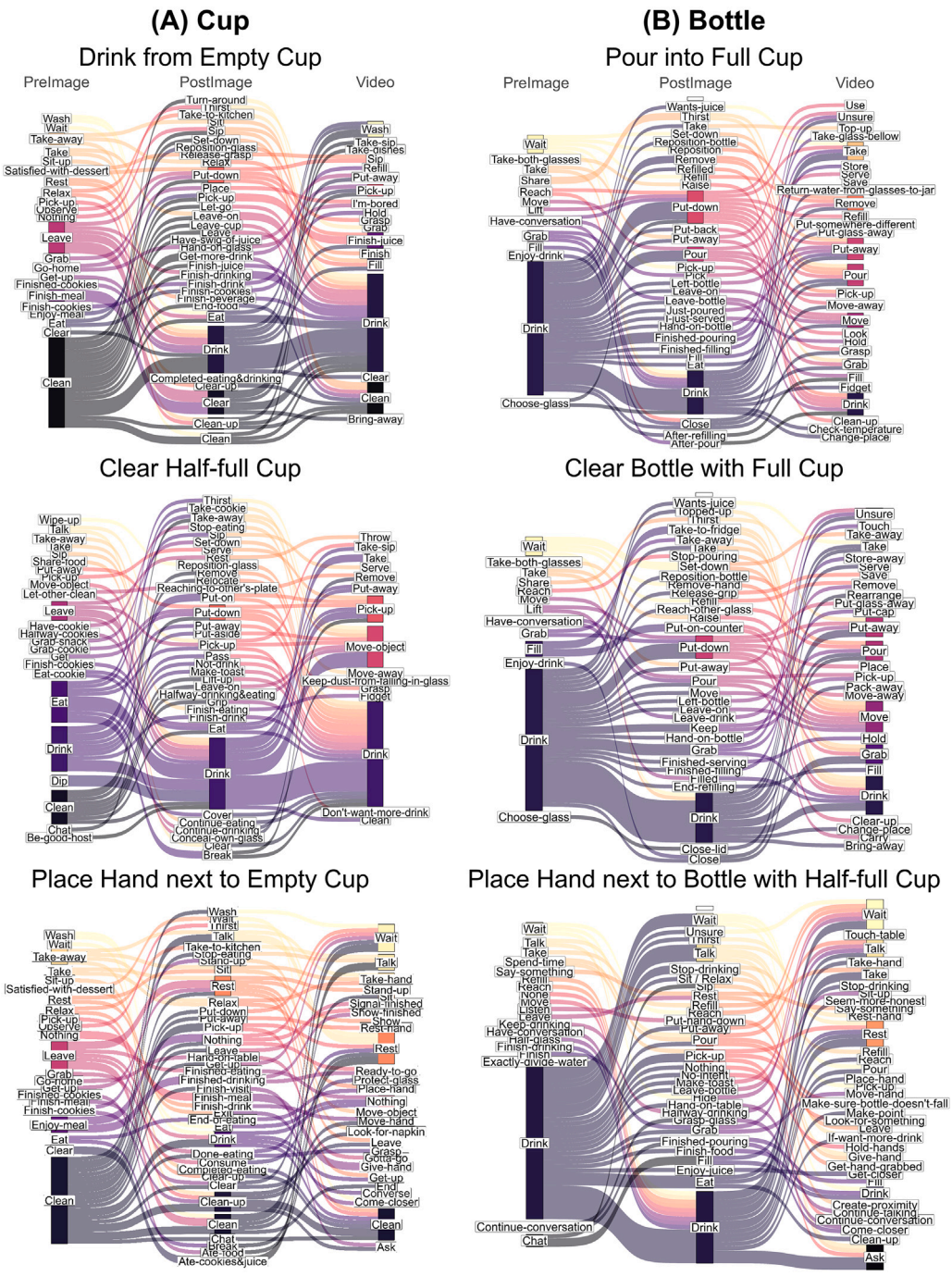


Fig. 2. Sankey diagrams illustrating the main differences in responses to pre-image, post-image and video stimuli for (A) Cup and (B) Bottle stimuli. These diagrams illustrate variations in the convergence of predicted intentions when participants were presented with action kinematics versus still images.

(e.g., standing up, moving closer, moving away) collected under the 'Move Body' cluster, and higher-order intentions, indicating the presumed purpose behind the body movement (e.g., to rest, talk, leave), sorted into four clusters: 'Rest', 'Talk', 'Leave', and 'Signal'. Finally, three other clusters were identified that covered responses that indicated that the participant was unsure of the intention, that there was no intention visible, or any other random response that did not fit in any other cluster: 'Unsure', 'None', and 'Other'.

2.5.4. Secondary coding and proportion analysis

The same experimenter and paid research assistant (still blind to the purpose of the study) then used this classification schema to classify each of the verbs and verb statements identified from the participants' responses. A second frequency analysis was then performed to identify the number of times a response fell within each cluster for each condition (action intention *x* environmental context) and object (cup, bottle, spoon). The frequency data were then converted to proportions of (frequency of each verb/total frequency per condition) for analysis and plotting (see Figs. 3 and 4).

2.5.5. Statistical modelling

Based on a preliminary assessment of the data collected, two different sets of statistical analyses were conducted. The first, analysed the effect of stimulus type on the variability of participants' responses, measured by the frequency of unique verbs, using a generalized linear model (GLM) with a Poisson distribution and a log-link function for

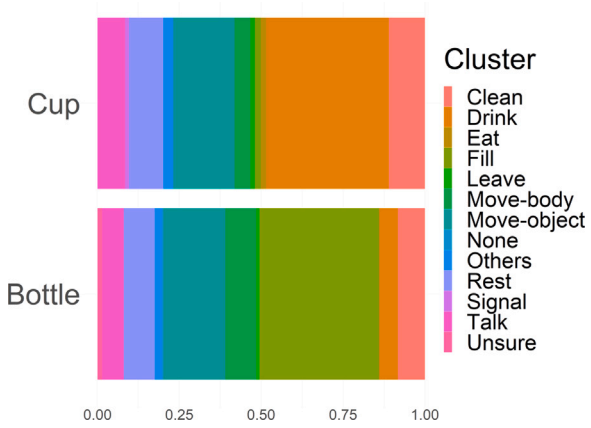


Fig. 3. Frequency analysis displaying the proportion of clusters per item. Proportions reveal a consistent distribution of clusters across all conditions. The most frequently predicted cluster consistently corresponds to the item's primary action possibility.

each object. An offset term, $log(n)$, was included to account for the number of participants across each level, allowing the model to estimate rates while accounting for the total number of participants.

For this first analysis, an increased number of verbs indicated divergence in the type and variance of intentions predicted, while a reduction in the number of verbs indicated convergence in both type and variance. This analysis was complemented with the Sankey diagrams, which presented the proportions, types and number of different verbs (see Fig. 2 for illustrative examples, and the Supplementary Material (Figures) for all Sankey diagrams).

Although the models included main effects and higher-order interaction effects for *stimuli type*, *context*, and *intention* per object, our primary interest was the main effect of stimulus type which determines the degree to which kinematic information (i.e., from videos) influenced participants intention predictions compared to observing the actors end-state configuration alone (i.e., post-images); see below for more details. Where significant effects of stimulus type were observed, pairwise contrasts between the different stimuli types were estimated using the *emmeans* package in R (Lenth, 2023).

The second statistical analysis then examined the effects of intention and context on responses to *video* stimuli only, aiming to uncover how the combined manipulations of kinematic and contextual information influence participants' responses. Responses to videos were rearranged into five clusters that capture the majority of participant responses (i.e., 'Drink OR Pour OR Eat', 'Clean', 'Move-object', 'Non-object directed', 'Other'). Generalized logistic mixed random effects models (GLME) were used to analyse the responses from the five clusters as a function of intention and context for each object. All models were defined with maximally defined random-factor structures, including random intercepts for each subject (Barr et al., 2013), and specified a binomial distribution with a log-link function. Specifically, we evaluated evidence for the effects of Intention, Context, and their interaction on the likelihood that participants' responses would fall within any of the five clusters. The maximum likelihood estimation method was implemented in these analyses using the *lme4* package in R (Bates et al., 2015), and p-values for the main effects and overall interaction were estimated using the *car* package (Fox & Weisberg, 2019). We then used the *emmeans* package to estimate the mean probabilities per condition for each cluster based on our models (Lenth, 2023).

All of the above analyses were performed using R. All data and all analysis codes used for the study (including descriptions of any modelling decisions made) can be found on OSF (https://osf.io/4dj9r/) (Alhasan et al., 2024).

3. Results

Complimentary findings were observed across all three objects, with all three objects indicating that the variety and types of participants' predictions would systematically vary as a function of the stimuli type (pre-image, post-image, and video-clip) and the corresponding kinematic-in-context information the different stimuli entailed. For brevity, results for the cup and bottle analyses are presented here, with the results for the spoon stimuli provided in the OSF (Alhasan et al., 2024). Furthermore, although summary diagrams and plots that best illustrate the key findings are presented here, detailed Sankey diagrams, word clouds, and histogram plots for all conditions and objects are provided in Supplementary Material (Figures).

3.1. Stimuli type: Pre-image, post-image and video-clip differences

In this section, we present our key findings regarding the main effect of stimuli type and its estimated pairwise comparisons. To maintain brevity and focus on the most salient results, we outline only the critical results here. Full details of the models and comprehensive results can be found in Supplementary Material (Stimuli Type Analysis).

3.1.1. Cup

Our analysis of the convergence and divergence of responses, characterized by the number of unique verbs per condition, revealed significant main effect of stimuli type ($X^2(2) = 21.66, p < .0001$). Estimated pairwise comparisons revealed that, overall, **post-image** stimuli elicited more varied responses than both **pre-image** ($p = .003$) and **video** ($p < .0001$) stimuli. This indicates that, overall, the number of unique verbs participants responded with was larger for the post-image stimuli compared to pre-image and video. For example, in the condition where the actor reached for an empty cup to drink (Fig. 2A top), 32 verbs were used to describe the actors intention when they viewed post-images compared to 22 for pre-images and 18 for video-clips. Similarly, when the actor reached for a half-full cup to clear (Fig. 2A middle), there were 34 verbs for post-images compared to 24 and 15 for pre-images and video-clips respectively. Also, when the actor placed their hand next to an almost empty cup (Fig. 2A bottom), there were 37 verbs for post-images compared to 22 and 27 for pre-images and videos respectively. Therefore, our results indicate that, overall, for the cup scenario the information available from the end-states decreased the specificity of the intention as illustrated by the larger number of possible actions predicted compared to the information available from the context only stimuli or the stimuli including kinematics.

Our estimated pairwise comparisons further reveal a non-significant difference between the **pre-image** and **video** stimuli ($p = 0.266$). However, although the difference in the number of unique verbs between pre-image and video stimuli did not reach significance, our Sankey diagrams reveal a difference in the most predicted type of intention (intention with the highest proportion). For example, for the condition where the actor reached for an empty cup to drink (Fig. 2A top), although responses consisted of comparable numbers of unique verbs between pre-image and video stimuli (22 and 18 respectively), there was an evident change in the most predicted intention from 37.10% 'Clean' predictions for pre-images to 40.32% 'Drink' for video-clips. Similarly, when the actor reached for the half-full cup to clear (Fig. 2A middle), the most predicted intention changed from 19.35% 'Drink', 17.74% 'Eat' and 14.52% 'Clean' for pre-images to 45.16% 'Drink' and 17.74% 'Move-object'. This indicates that kinematic information specified a different type of intention than context alone.

3.1.2. Bottle

Similar trends were noticed in the analysis of the convergence and divergence of responses for the bottle, with a significant main effect of stimuli type ($X^2(2) = 24.107, p < .0001$). Estimated pairwise comparisons revealed that, overall, **post-image** ($p = 0.0001$) stimuli and **video** ($p = 0.0012$) stimuli elicited more varied responses than **pre-image** stimuli. This indicates that the number of unique verbs participants responded with was larger for the post-image and video stimuli compared to pre-image. For example, in the condition where the actor grasped the bottle to pour into an almost full cup (Fig. 2B top), the number of verbs for post-image and video stimuli were 31 and 28 respectively, compared to 13 for pre-image. Similarly, when he grasped the bottle to clear from the table next to full cups (Fig. 2B middle), there were 33 and 26 verbs for the post-image and video stimuli respectively, compared to 13 for the pre-image stimuli. Also, when the actor placed their hand next to the bottle with half-full cups (Fig. 2B bottom), there were 30 and 34 verbs for post-image and video respectively, compared to 20 for pre-image. Therefore, our results indicate that, overall, for the bottle scenario the information available from the end-states and kinematics decreased the specificity of the intention as illustrated by the larger number of possible actions predicted compared to the information available from the context only.

Our estimated pairwise comparisons further reveal a non-significant difference between the **post-image** and **video** stimuli ($p = 0.729$). Again, although the difference in the number of unique verbs between post-image and video stimuli did not reach significance, our Sankey diagrams reveal a difference in the most predicted type of intention. For example, for the condition where the actor placed their hand next to the bottle with half-full cups (Fig. 2B bottom), although responses consisted of 30 and 34 unique verbs for post-images and video-clips respectively, there is an evident change in the predominantly predicted intention from 27.43% 'Drink' for post-image to 11.29% 'Wait', 9.68% 'Rest' and 9.68% 'Ask' for videos. Similarly, when the actor grasped the bottle to pour into an almost full cup (Fig. 2B top), there is an evident change in the predominantly predicted intention from 19.87% 'Drink' and 14.75% 'Put-down' for post-image, to 9.84% 'Drink', 9.84% 'Pour' and 9.84% 'Put-away' for videos. This, again, indicates that kinematic information specified a different type of intention than the end-state grasp configuration.

3.2. Kinematic (video) context differences and classification analysis

The above findings suggest that the kinematic information (video) influenced participants' judgements by resulting in both different types of intentions and different numbers of verbs being reported, reflecting either a narrower or broader range of perceived action possibilities, when compared to responses to context-only (pre-image) and context and end-state information (post-image). Consequently, given that the combination of kinematic, end-state and contextual information provided different responses, the subsequent sections of the results focus on the predictions from the video stimuli, as it is more resonant with our day-to-day experience.

We were predominantly interested in identifying the main trends in participants' responses to the nine different (intention x context) conditions to identify how they compare to the actor's actual intention and what they inform us about how participants make these predictions. As previously detailed under the Response Coding and Statistical Analysis section, participants responses to video stimuli were clustered into 13 total clusters by semantic meaning as illustrated in Fig. 3, namely: 'Move Object', 'Move Body', 'Drink', 'Fill', 'Eat', 'Clean', 'Leave', 'Rest', 'Signal', 'Talk', 'Unsure', 'None', 'Other'. A frequency analysis was then performed to identify the clusters that dominated the responses to each condition. Additionally, GLME models were used to analyse the effects of both the actor's intention and the environmental context on the probability that participants would respond with one of the following responses: 'Drink (for cup)/Pour (for bottle)', 'Clear', 'Non-object-directed', 'Move-object', 'Other'.

Before detailing these results, however, it is important to note that if one defines the actor's manipulated intentions as the "correct" response, and assess the participants' "accuracy" in predicting the object-directed intentions for the video condition (i.e. accuracy excluding the ambiguous intention to place-hand on table, and only considering the two object-directed intentions), the accuracy rates are found to be 47.11% for the cup stimuli and 33.54% for the bottle stimuli. In other words, participants were not highly accurate in predicting the actor's manipulated object-directed intentions even for the video stimuli. However, as evident in Fig. 3, this analysis also revealed that participants' responses demonstrate an overarching pattern that spans across items. It is evident that participants often judged the actor's intentions in terms of the objects' primary action possibilities or their designed function. More specifically, in the cup condition, the most commonly reported verb predicting intentions was 'Drink', comprising 37% of all responses. Similarly, in the bottle condition, the most commonly reported intention was 'Fill/Pour', comprising 36% of responses. Recall that 'Drink' and 'Fill/Pour' were classified here as higher-order object-directed intentions. Interestingly, the second most commonly reported verb (intention) for each object also consistently fell within the 'Move-object' cluster, comprising 19% of responses on average.

In the following sections, the results of the frequency analyses and GLME models for each item are detailed.

3.2.1. Cup

Frequency Analysis

When the actor reached for a full or half-full cup with the intention to either drink or clear, the predominant intention prediction was 'Drink' regardless of the actor's enacted intention. However, more specifically and as evident from Fig. 4a, when the actor's intention was to drink, 89% of responses indicated 'Drink' and predictions fell within only 2 clusters of verbs used overall to describe the actor's intention for both full and half-full cups. In contrast, when the actor's intention was to clear, only 46% to 48% of responses indicated 'Drink' and predictions fell within 4 to 6 clusters of verbs. Furthermore, when the actor's intention was to clear, around 37% of responses indicated 'Move-object' which refers to 'what' the actor was doing rather than 'why' (e.g. grab, grasp, pick-up, hold). In contrast, only 11% of responses indicated 'Move-object' when the actor's intention was to drink. This observation suggests that there was a lower probability of making predictions of higher-order intentions when the actor's kinematics reflected the intention to clear with both full and half-full cups compared to the actor's intention to drink. Together, these results indicate that although the most probable prediction under these conditions was to 'Drink', participants were also sensitive to the different kinematics and to the different action possibilities afforded when the actor intended to clear. However, the kinematic and contextual information provided by the actor's intention to clear full or half-full cups at grasp did not clearly specify an alternative higher-order intention.

Furthermore, when the actor reached for an almost empty cup to either drink or clear the predominant intention predictions aligned with the actor's enacted intention. Specifically, when the actor's intention was to drink, 53% of responses indicated 'Drink' and predictions fell within 5 clusters of verbs. Whereas, when the actor's intention was to clear, 42% of responses indicated 'Clean' and predictions fell within 9 clusters of verbs. This indicates that the most predicted intentions aligned with the actor's actual intention. However, when the actor's intention was to clear the almost empty cup, 33% of responses indicated 'Move-object' compared to 15% when the intention was to drink. The higher percentage of lower-order intention predictions and wider range of clusters again indicate a reduced probability of making higher-order intention predictions. This further indicates that there was less

Table 1
Result of GLME model for the effect of Intention and Context for Cup.

Cluster	Intention		Context	
	$X^2(2)$	*p*-value	$X^2(2)$	*p*-value
Drink	47.23	<.001*	24.63	<.001*
Move-object	20.21	<.001*	0.53	0.766
Clean	<0.001	0.999	<0.001	0.999
Non-object-directed	33.36	<.001*	0.00	1.000
Other	0.11	0.945	<0.001	0.999

* $p < 0.05$.

Table 2
Result of GLME model for the effect of Intention and Context for Bottle.

Cluster	Intention		Context	
	$X^2(2)$	*p*-value	$X^2(2)$	*p*-value
Pour/Fill	42.43	<.001*	40.69	<.001*
Move-object	5.54	.063	8.89	0.011*
Clean	1.14	0.564	13.83	0.001*
Non-object-directed	27.79	<.001*	<0.001	0.999
Other	3.50	0.174	9.11	0.010*

* $p < 0.05$.

intention specifying information when the kinematics reflected the intention to clear with an almost empty cup compared to the intention to drink. It is also evident that participants predicted a larger number of action possibilities when the cup was almost empty compared to when the cup was either full or half-full and the actor reached to the cup to either drink or clear. This indicates that the kinematic and contextual information provided by the actor's intention to drink or clear empty cups did not specify the actor's intention as much as the relative conditions with full or half-full cups.

Finally, when the actor placed his hand on the table next to the cup, the average number of clusters featured in responses was larger than that observed when the actor's intention was to drink or clear, across all contexts (full, half-full, or empty cups). The predominant intention prediction, however, aligned with the actor's non-object directed nature of intention. More specifically, when the actor's intention was to place hand-on-table with a full cup, 32% of responses indicated 'talk' and predictions ranged across 9 clusters of verbs. Additionally, when the cup was either half-full or almost empty, 29% to 34% of responses respectively indicated 'rest', and predictions ranged across 10 and 8 clusters of verbs. The equally wide range of intention prediction clusters across contexts demonstrate that intentions were under-specified by the kinematics in all three contexts, making it less likely that responses converge onto a specific intention since these conditions allow for a wide range of action possibilities. However, it is evident that the kinematics were indeed informative of the directionality of the actor's intention. We notice that 74.67% of predictions on average were non-object directed when the actor's intention was to place hand on table compared to only 1.67% when the actor's intention was to drink or clear across all contexts.

Effects of Intention and Context. To quantify the effect of Intention and Context the data was re-clustered into five main clusters: three clusters presented the actor's actual intention (Drink, Clean, Non-object-directed), lower-order 'what' intentions (Move-object), and any other intention (Others) (see Fig. 5a for illustration). Using GLME models we analysed the effect of Intention and Context on the probability that participants' responses would fall into each cluster. As can be discerned from an inspection of Table 1, this analysis revealed that the probability that participants would predict the actor's intention as Drink, Move-object or perform a non-object directed action varied as a function of the actor's intention and kinematics. This is illustrated in Fig. 4a by higher Drink responses when the actor's intention was to Drink compared to Clear or place hand-on-table, higher Move-object responses when the actor's intention was to Clear compared to Drink or place Hand-on-table, and higher non-object-directed responses when the actor's intention was to place hand-on-table compared to Drink or Clear. The contextual manipulation also only had a significant effect on the probability that participants would make Drink predictions. This is evident in the higher Drink predictions when the cup was full and half-full compared to when it was almost empty. The probability that participants would make Clean or any Other prediction did not differ significantly as a function of either intention or context. Additionally, the probability that participants would predict an alternative object-directed higher order intention other than Drink was significantly low (difference between the probability of Drink vs. Clean responses $X^2(1) = 92.57, p < .001$; Drink vs. Other responses $X^2(1) = 110.54, p < .001$).

3.2.2. Bottle

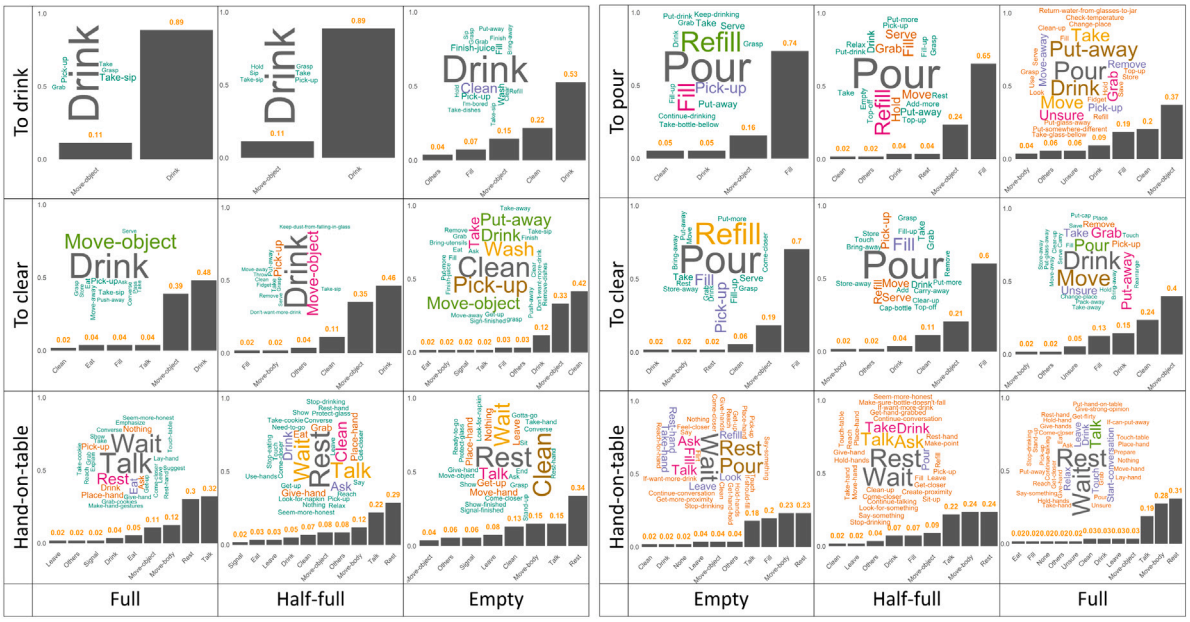
Frequency Analysis

When the actor reached for the bottle next to almost empty or half-full cups with the intention to either Pour or Clear, the predominant intention prediction was 'Fill/Pour' regardless of the actor's enacted intention. Specifically, as illustrated in Fig. 4b, when the actor's intention was to Pour, 74% and 69% of responses indicated 'Fill/Pour' when the cups were empty or half-full, respectively, and predictions fell within 4 clusters of verbs overall. Additionally, when the actor's intention was to Clear, 70% and 59% of responses indicated 'Fill/Pour' when the cups were empty or half-full, respectively, and predictions fell within 6 clusters of verbs overall. The almost equally high percentages as well as the similar number of clusters across conditions indicate an equally high probability of judging the intention as being to 'Fill/Pour' across these four conditions regardless of the actor's intention.

When the actor reached for the bottle next to almost full cups to either pour or clear, the predominant intention prediction was 'Move-object' regardless of the actor's enacted intention. Specifically, when the actor reached for the bottle next to full cups, 37% and 40% of responses indicated 'Move-object' when the actor's intention was to pour or clear, respectively, and predictions fell within 7 clusters of verbs overall. This observation indicates that participants judged that the full cups context specifies other action possibilities than 'Fill/Pour' but that the intention specifying information provided by both context and kinematics was not enough for most to make predictions of higher-order ('why') intentions when the cups were almost full and the kinematics reflected the actor's intention to either pour or clear. It is worth noting, however, that the second most predicted intention was to 'Clean', where 20% and 24% of responses indicated 'Clean' when the actor's intention was to pour or clear, respectively.

Finally, when the actor placed his hand on the table next to the bottle, the number of clusters featured in responses was larger on average than that seen when the actor's intention was to pour or clear over all contexts - i.e. empty, half-full or full cups. However, as previously reported for the Cup stimuli, the predominant intention prediction here also aligned with the actor's non-object directed nature of intention. More specifically, when the actor's intention was to place hand-on-table next to bottle, 23%, 24% and 31% of responses indicated 'Rest' when the cups were empty, half-full and full respectively, and predictions ranged across 12 to 9 clusters of verbs. The various clusters of intention prediction observed across the three contexts indicate that kinematics did not provide enough information to specify intentions in all three contexts. As a result, it is less probable for responses to converge onto a specific intention, as these conditions allow for a wide range of possible actions. However, it is clear that the kinematics provided some information about the intended direction of the actor's actions. We notice that around 73.67% of predictions were non-object directed compared to only around 1.5% when the actor's intention was to pour or clear across all contexts.

Effects of Intention and Context. Similar to the Cup stimuli data, the data for the Bottle stimuli was re-clustered into five main clusters: three clusters related to the actor's actual intention (Pour/Fill, Clean, Non-object-directed), with the forth capturing to lower-order 'what' intentions (Move-object), and the fifth capturing any other intention



(a) Cup (b) Bottle

Fig. 4. Frequency analysis of clusters accompanied by word clouds of coded verbs per condition for (a) Cup and (b) Bottle. The analysis indicates that participants commonly predicted the primary action possibility of the grasped item as the intended action, provided that both kinematics and context specified its feasibility. The range of predictions broadened when either the kinematics (e.g., a hand placed on the table) or the context (e.g., reaching for an empty cup) suggested that the item's primary action was less likely. Additionally, the type of predictions—whether higher or lower order—varied across different conditions and items. For full sized result figures of all items, see Supplementary Material (Figures).

(Others) (see Fig. 5b for illustration). Overall, the probability of participants making Pour/Fill responses significantly changed as a function of Intention and Context. As can be seen from an inspection of Table 2, this analysis revealed that the probability that participants would predict the actor's intention as Pour/Fill, Move-object and Clean varied as a function of the contextual manipulation. This is illustrated in Fig. 4b by higher Pour/Fill responses when the context included almost empty and half-full cups compared to full cups, as well as higher Move-object and Clean responses when the context included almost full cups compared to empty or half-full cups. The manipulation of the actor's intention only had an effect on the probability that participants would make Pour/Fill and non-object-directed predictions. This is evident by the higher Pour/Fill predictions when the actor intended to pour or clear table compared to when the actor placed Hand-on-Table, and more non-object-directed predictions when the actor placed hand-on-table compared to when the actor intended to pour or clear. The probability that participants would make any Other prediction only differed significantly as a function of context. Additionally, the probability that participants would predict an alternative object-directed higher order intention other than Pour/Fill was significantly low (difference between the probability of Pour/Fill vs. Clear responses $X^2(1) = 110.1, p < .001$; Pour/Fill vs. Other responses $X^2(1) = 88.47, p < .001$).

4. Discussion

The current study investigated how individuals predict others' intentions by analysing free-text responses to reach-to-grasp movements. The interplay between kinematics and contextual information was first examined to determine whether predictions were based solely on the end-state hand configuration within a context, or whether the hand's movement (kinematics) also informed intention predictions. The findings revealed that when kinematic information was presented alongside end-state and contextual information, the resultant intention predictions of participants diverged significantly from those where only the latter two types of information were available. However, further analysis revealed notable variability in the number and type of intention predictions across different conditions and objects, indicating that the extent to which kinematic information specified the actor's intended action was influenced by the context (including the interaction object; i.e., cup, bottle). Additionally, the most commonly predicted intention did not consistently align with the actor's actual intended action. Instead, the predictions of participants reflected two main semantic clusters: higher-order 'why' predictions and lower-order 'what' predictions. Notably, responses within the 'why' cluster reflected the primary action possibility of the contextual object (e.g., 'drink' for a cup and 'pour' for a bottle).

Our GLME analysis of the five main semantic clusters per object indicated that the impact of the actor's intention and context on responses varied between the different objects. This variation reflects discrepancies in the level of intention specificity in the presented kinematics-in-context information between objects when considered at the moment of grasp. For the bottle, kinematic information had a lesser impact on participants' predictions compared to the cup. This finding was further supported by the frequency analysis, which showed that participants' responses were consistent with the action possibilities most functionally afforded by the kinematics-in-context. For instance, when the kinematics of the actor's object-directed intentions were distinct but equally allowed for the item's primary action, the predicted intentions were almost identical between the two kinematic conditions within each context. For example, whether the actor grasped the middle of the bottle to pour or the top to clear, both actions similarly afford pouring from the bottle, resulting in nearly identical patterns of responses. Similarly, when distinct contexts clearly afforded the item's primary action possibility, the predicted intentions were almost identical between the contexts for each distinct kinematics condition. For example, whether the cup was empty or half-full, both contexts allow pouring from the bottle, resulting in nearly identical patterns of responses. Additionally, when the kinematics indicated a different action possibility to the item's primary action by making it less likely (e.g. kinematics specifying the intention to clear the cup from the table by grasping it around the top rim), but the context allowed for the items' primary action possibility (e.g. full and half-full cups), most people still predicted that the actor intends to use

Fig. 5. Frequency analysis of re-clustered data illustrates the five main clusters: those representing the actor's actual intentions (Drink OR Pour, Clean, Non-object-directed (NOD)), lower-order 'what' intentions (Move-object), and all other intentions (Others) for various conditions involving (a) Cup and (b) Bottle. This analysis highlights the variability in responses based on the actor's intention and context. It also shows that while previous analyses indicated a wider diversity of responses when the actor placed a hand on the table, these responses were primarily non-object directed.

the item for its primary functional purpose (refer to Fig. 4 for the full pattern of results per item). However, in these conditions more people chose to respond with a lower-order prediction of 'what' the actor was doing rather than 'why' they were doing it. Furthermore, when the kinematics or context indicated that the item's primary action possibility was not likely (e.g., the actor placed their hand on the table or the context better specified the intention to clear the item), the range of predictions was highly variable. Thus, our results provide evidence that when an observer is asked to predict the intention of another actor, their responses reflect the most functional action possibilities or *affordances* defined by the whole actor-environment context. Moreover, when the kinematics-in-context information is ambiguous and maps to many action possibilities (affordances), a corresponding multitude of intention predictions are generated in response.

By "affordances" we mean the relational action possibilities defined by the perception-action capacities of an agent within a given physical and informational context (Gibson, 1979; Stoffregen, 2003; Turvey et al., 1981; Warren, 1984). Extensive research has demonstrated that individuals can accurately identify affordances relevant to their own actions (van der Kamp et al., 1997; Mark, 1987; Ramenzoni, Riley, Shockley, & Davis, 2008; Warren, 1984) and, of most relevance here, those of other observed individuals (Ramenzoni, Riley, Davis, et al., 2008; Ramenzoni, Riley, Shockley, & Davis, 2008; Richardson et al., 2007; Stoffregen et al., 1999; Wagman et al., 2018). Crucially, the detection of affordances, including in contexts with multiple action possibilities relies on both kinematic and contextual information (Mark, 2007; Ramenzoni, Riley, Shockley, & Davis, 2008). This interplay specifies which affordances can or should be enacted. Therefore, the ability of an observer to predict another's intentions fundamentally depends on whether the actor-in-context information clearly specifies one or more affordances. Consistent with the findings reported here, if the information made available specifies one actionable affordance over others, most observers will perceive that as the actor's intention (Mark, 2007; Richardson et al., 2007). However, if the information specifies multiple affordances or no clear realizable action possibility at all, observers will report a diverse range of intentions.

Adopting the view that predicting others' intentions involves the perception of affordances aligns with recent research showing that individuals often identify objects based on their functional affordances (Cosentino, 2019; Kalénine et al., 2016; Lee et al., 2013; Roche & Chainay, 2017). When someone perceives an object like a hammer, they not only see it as a physical item, but also understand its use for hammering things — linking it to the intentions behind its design and typical use (aka its functional affordance). Our findings further emphasize the role of individuals' attunement (Gibson, 1966; McArthur & Baron, 1983) to an object's functional affordances, which plays a key role in determining predictions of an actor's intentions. Notably, in the current study we designed some contexts to be potentially ambiguous (Amoruso et al., 2016), allowing for multiple interpretations of the actor's intentions. However, our results show that these contexts were not generally perceived as ambiguous — subject responses primarily reflected the functional affordance when the context afforded it. It was only in contexts where the likelihood of enacting the object's functional affordance was reduced (e.g., reaching for an almost empty cup), or where the kinematic information was ambiguous (e.g., placing a hand on a table) that a wider spectrum of intentions were observed. This pattern suggests that in these conditions, kinematics-in-context information was generally consistent with a wide range of potential action possibilities.

Furthermore, while contexts indicating a reduced likelihood of enacting the functional affordance led to a broader range of predictions, there was also considerable variability in the type of intentions predicted between the different items, with more precise higher-order intention predictions for the cup than for the bottle. This variation suggests that the specificity of an alternative higher-order intention, although low overall, also differed between items. Our findings contrast with those of Amoruso et al.'s (2016) study, where participants' responses were generally more aligned with the context (e.g., they report more 'drink' responses when reaching for a full cup, more 'clear' responses when reaching for an empty cup, and around 50% 'drink' and 'clear' responses when the cup was half-full).

It is important to note that although our findings suggest that individuals might struggle to recognize higher-order intentions beyond an item's functional affordance, everyday observations often contradict this. For example, a caregiver can frequently predict whether a child intends to spill water from a cup rather than drink it. This is because affordances are typically embedded in much richer spatial and temporal contexts, which extend beyond the limitations of experimental settings (Alhasan et al., 2025; Heft, 1989, 2003; McArthur & Baron, 1983). Thus, the specificity of object-directed intentions often requires

considering a wider spatio-temporal context (e.g., grasping a cup at a kitchen table might imply different intentions than grasping a cup in a minimally furnished room). If individuals are accurately identifying others' affordances, their predictions should align more closely with the alternative action possibilities specified by the additional spatio-temporal context.

Such wider spatio-temporal contexts could also encompass social constraints that further influence our predictions of the actions available to an agent in a given environment (Baggs, 2021; Baron, 2007; Becchio et al., 2010; Heft, 2003; Marsh et al., 2006; McArthur & Baron, 1983; Schmidt, 2007). For instance, an individual may opt for a glass of juice closer to themselves rather than to another person, reflecting social norms that shape our predictions of other's intentions (McArthur & Baron, 1983). Therefore, a complete understanding of what drives a person's actions may require knowledge of both their physical capabilities (physical laws) and the social contexts that guide or restrict their behaviour (social rules) (also see Alhasan et al. 2025 for further discussion of this point and its relevance to the present findings). Thus, future studies should examine how information within richer spatio-temporal and social contexts might influence our ability to predict others' intentions.

It is also essential to examine how variability in movement kinematics, both within and between individuals, interacts with spatio-temporal and social contexts to shape intention predictions. The many-to-many relationship between kinematics and intentions introduces variability in the kinematics associated with the same intention both across and within individuals (Scaliti et al., 2023). However, our findings emphasize that it is the relationship between movement kinematics and the action possibilities they afford – within a given context – rather than variability in kinematics or context alone, that drives these predictions. While our study provides initial insights into this relationship, presenting each condition only once limited our ability to evaluate how natural variations in kinematics influence participants' predictions. Future research should address this by including multiple recordings of each condition and involving multiple actors to investigate how the informativeness of kinematics about action possibilities varies both across and within individuals in the same context.

Finally, adopting the perspective that predicting others' intentions involves perceiving affordances requires a shift in experimental focus. Rather than examining the relative influence of specific kinematic versus contextual information on intention prediction, the emphasis should be on identifying higher-order kinematic-in-context information – namely, the spatiotemporal patterns of behaviour within dynamic agent-environment systems – that specify the most beneficial or functional action possibilities available to an observed other. This information could exist in ongoing or past behaviour across short and long timescales (Stoffregen & Wagman, 2024; Wagman et al., 2018). While it may be true that we cannot directly know about another individual's intentions, we can know about (perceive) the affordances that are available, goal relevant, or preferable to a co-actor or observed other. Future research should therefore prioritize identifying the information that specifies and constrains the action possibilities most likely or appropriate for someone to enact. We expect that such research will pave the way for a much deeper understanding of how observers predict intentions in complex, real-world scenarios.

5. Conclusion

The current study enriches the existing literature by elucidating how kinematic and contextual information together influence our ability to predict others' intentions. By employing free-text responses, we discovered that participants' predictions, while generally inaccurate, align with the affordance landscape shaped by the actor's movements within a specific context. This contrasts with outcomes from standard two-alternative forced-choice paradigms, as a diverse range of predictions emerged when kinematics or context suggested that the object's functional affordance was less likely. Furthermore, these predictions rarely converged on any specific higher-order intention beyond the object's functional affordance. This may reflect the simplified contexts and limited information provided about potential alternative higher-order intentions in our stimuli, which might be more clearly specified in broader spatio-temporal contexts. Future experiments will be needed to explore this possibility. Overall, our findings suggest that conventional methods might significantly oversimplify how humans understand and predict intentions. Future research, including our own, must embrace more complex experimental paradigms to more fully reveal the information that specifies what an observed other might, could or will do in the future, and the degree to which the perception of such action possibilities (affordances) is the foundation of judging or predicting the intentions of others in natural settings.

CRediT authorship contribution statement

Ayeh Alhasan: Writing – review & editing, Writing – original draft, Visualization, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Eyal Karin:** Writing – review & editing, Visualization, Formal analysis. **Nathan Caruana:** Writing – review & editing, Methodology, Conceptualization. **Emily Cross:** Writing – review & editing, Supervision, Methodology, Conceptualization. **David Kaplan:** Writing – review & editing, Validation, Supervision, Investigation. **Michael J. Richardson:** Writing – review & editing, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used WriteFull and ChatGPT-4o to check for correct grammar and to improve readability. Both AI tools were used for purely grammatical assistance and all presented scientific material and writing is original. After using this tool/service, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the publication.

Acknowledgements

This research is supported by an Australian Government Research Training Program (RTP) Scholarship and the Australian Research Council Future Fellowship, Richardson, FT180100447.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.cognition.2025.106122.

Data availability

All stimuli (including videos), data and all analysis codes used in this study are publicly available at https://osf.io/4dj9r/ (Alhasan et al., 2024).

References

Alhasan, A., Kallen, R. W., & Richardson, M. J. (2025). Intention prediction is affordance perception. *Ecological Psychology*, *37*(1), 21–35. http://dx.doi.org/10.1080/10407413.2024.2427403, arXiv:https://doi.org/10.1080/10407413.2024.2427403.

Alhasan, A., Karin, E., Caruana, N., Cross, E. S., Kaplan, D. M., & Richardson, M. J. (2024). *Kinematics in context: Predicting other's action intentions entails the perception of affordances*, http://dx.doi.org/10.17605/OSF.IO/4DJ9R. Dataset.

Amoruso, L., Finisguerra, A., & Urgesi, C. (2016). Tracking the time course of top-down contextual effects on motor responses during action comprehension. *Journal of Neuroscience*, *36*(46), 11590–11600.

Amoruso, L., Finisguerra, A., & Urgesi, C. (2018). Contextualizing action observation in the predictive brain: Causal contributions of prefrontal and middle temporal areas. *NeuroImage*, *177*, 68–78.

Amoruso, L., & Urgesi, C. (2016). Contextual modulation of motor resonance during the observation of everyday actions. *NeuroImage*, *134*, 74–84.

Azaad, S., & Sebanz, N. (2023). Predicting others' actions from their social contexts. *Scientific Reports*, *13*(1), 22047.

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.

Baggs, E. (2021). All affordances are social: Foundations of a gibsonian social ontology. *Ecological Psychology*, *33*(3–4), 257–278.

Baird, J. A. (1999). *Young children's understanding of the relation between actions and intentions* (Ph.D. thesis), University of Oregon.

Baldwin, D. A., & Baird, J. A. (1999). Action analysis: A gateway to intentional inference. In P. Haggard, Y. Rossetti, & M. Kawato (Eds.), *Early social cognition: understanding others in the first months of life* (pp. 215–240). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Baron, R. M. (2007). Situating coordination and cooperation between ecological and social psychology. *Ecological Psychology*, *19*(2), 179–199.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Becchio, C., Koul, A., Ansuini, C., Bertone, C., & Cavallo, A. (2018). Seeing mental states: An experimental strategy for measuring the observability of other minds. *Physics of Life Reviews*, *24*, 67–80.

Becchio, C., Sartori, L., & Castiello, U. (2010). Toward you: The social side of actions. *Current Directions in Psychological Science*, *19*(3), 183–188.

Betti, S., Finisguerra, A., Amoruso, L., & Urgesi, C. (2022). Contextual priors guide perception and motor responses to observed actions. *Cerebral Cortex*, *32*(3), 608–625.

Boria, S., Fabbri-Destro, M., Cattaneo, L., Sparaci, L., Sinigaglia, C., Santelli, E., Cossu, G., & Rizzolatti, G. (2009). Intention understanding in autism. *PLoS One*, *4*(5), Article e5596.

Caligiore, D., Pezzulo, G., Miall, R. C., & Baldassarre, G. (2013). The contribution of brain sub-cortical loops in the expression and acquisition of action understanding abilities. *Neuroscience & Biobehavioral Reviews*, *37*(10), 2504–2515.

Cavallo, A., Koul, A., Ansuini, C., Capozzi, F., & Becchio, C. (2016). Decoding intentions from movement kinematics. *Scientific Reports*, *6*(1), 37036.

Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press.

Cosentino, E. (2019). Artifacts and affordances. *Synthese*, *198*(Suppl 17), 4007–4026.

Creswell, J., & Clark, V. (2017). *Designing and conducting mixed methods research*. SAGE Publications.

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

Elsner, C., Falck-Ytter, T., & Gredebäck, G. (2012). Humans anticipate the goal of other people's point-light actions. *Frontiers in Psychology*, *3*, 120.

Finisguerra, A., Amoruso, L., & Urgesi, C. (2020). Beyond automatic motor mapping: New insights into top-down modulations on action perception. In *Modelling human motion: from human perception to robot design* (pp. 33–51). Springer.

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Thousand Oaks CA: Sage, URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston, USA: Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA, US: Houghton, Mifflin and Company.

Heft, H. (1989). Affordances and the body: An intentional analysis of Gibson's ecological approach to visual perception. *Journal for the Theory of Social Behaviour*, *19*(1), 1–30.

Heft, H. (2003). Affordances, dynamic experience, and the challenge of reification. In *How shall affordances be refined?* (pp. 149–180). Routledge.

Helm, F., Cañal-Bruland, R., Mann, D. L., Troje, N. F., & Munzert, J. (2020). Integrating situational probability and kinematic information when anticipating disguised movements. *Psychology of Sport and Exercise*, *46*, 1–8.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology*, *3*(3), Article e79.

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

Järvilehto, T. (1998). The theory of the organism-environment system: I. Description of the theory. *Integrative Physiological and Behavioral Science*, *33*, 321–334.

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

Kalénine, S., Wamain, Y., Decroix, J., & Coello, Y. (2016). Conflict between object structural and functional affordances in peripersonal space. *Cognition*, *155*, 1–7.

van der Kamp, J., Savelsbergh, G., & Smeets, J. (1997). Multiple information sources in interceptive timing. *Human Movement Science*, *16*(6), 787–821.

Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in Cognitive Sciences*, *15*(8), 352–357.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). The mirror-neuron system: a Bayesian perspective. *Neuroreport*, *18*(6), 619–623.

Kim, S., Yu, Z., & Lee, M. (2017). Understanding human intention by connecting perception and action learning in artificial agents. *Neural Networks*, *92*, 29–38.

Knoblich, G., & Sebanz, N. (2008). Evolving intentions for social interaction: from entrainment to joint action. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1499), 2021–2031.

Koul, A., Soriano, M., Tversky, B., Becchio, C., & Cavallo, A. (2019). The kinematics that you do not expect: Integrating prior information and kinematics to understand intentions. *Cognition*, *182*, 213–219.

Lee, C. l., Middleton, E., Mirman, D., Kalénine, S., & Buxbaum, L. J. (2013). Incidental and context-responsive activation of structure-and function-based action features during object identification. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 257–270.

Lenth, R. V. (2023). emmeans: Estimated marginal means, aka least-squares means. URL: https://CRAN.R-project.org/package=emmeans. R package version 1.8.6.

Lin, C. J., & Lukodono, R. P. (2021). Sustainable human–robot collaboration based on human intention classification. *Sustainability*, *13*(11), 5990.

Malle, B. F., Moses, L. J., & Baldwin, D. A. (2001). Introduction: The significance of intentionality. In *Intentions and intentionality: foundations of social cognition* (pp. 1–24). The MIT Press.

Manera, V., Schouten, B., Becchio, C., Bara, B. G., & Verfaillie, K. (2010). Inferring intentions from biological motion: a stimulus set of point-light communicative interactions. *Behavior Research Methods*, *42*(1), 168–178.

Mark, L. S. (1987). Eyeheight-scaled information about affordances: a study of sitting and stair climbing. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(3), 361.

Mark, L. S. (2007). Perceiving the actions of other people. *Ecological Psychology*, *19*(2), 107–136.

Mark, L. S., Ye, L., & Smart, L. J. (2015). Perceiving the nesting of affordances for complex goal-directed actions. In R. R. Hoffman, P. A. Hancock, M. W. Scerbo, R. Parasuraman, & J. L. Szalma (Eds.), *The cambridge handbook of applied perception research: vol. 1*, (pp. 547–567). Cambridge University Press.

Marsh, K. L., Richardson, M. J., Baron, R. M., & Schmidt, R. (2006). Contrasting approaches to perceiving and acting with others. *Ecological Psychology*, *18*(1), 1–38.

McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review*, *90*(3), 215.

McEllin, L., Knoblich, G., & Sebanz, N. (2018). Distinct kinematic markers of demonstration and joint action coordination? Evidence from virtual xylophone playing. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(6), 885.

Naish, K. R., Reader, A. T., Houston-Price, C., Bremner, A. J., & Holmes, N. P. (2013). To eat or not to eat? Kinematics and muscle activity of reach-to-grasp movements are influenced by the action goal, but observers do not detect these differences. *Experimental Brain Research*, *225*, 261–275.

Ortigue, S., Sinigaglia, C., Rizzolatti, G., & Grafton, S. T. (2010). Understanding actions of others: the electrodynamics of the left and right hemispheres. a high-density EEG neuroimaging study. *PLoS One*, *5*(8), Article e12160.

Ramenzoni, V., Riley, M. A., Davis, T., Shockley, K., & Armstrong, R. (2008). Tuning in to another person's action capabilities: perceiving maximal jumping-reach height from walking kinematics. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(4), 919.

Ramenzoni, V. C., Riley, M. A., Shockley, K., & Davis, T. (2008). An information-based approach to action understanding. *Cognition*, *106*(2), 1059–1070.

Richardson, M. J., Marsh, K. L., & Baron, R. M. (2007). Judging and actualizing intrapersonal and interpersonal affordances. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 845–859.

Richardson, M. J., Marsh, K. L., & Schmidt, R. (2010). Challenging the egocentric view of coordinated perceiving, acting, and knowing. *The Mind in Context*, 307–333.

Richardson, M. J., Shockley, K., Fajen, B. R., Riley, M. A., & Turvey, M. T. (2008). Ecological psychology: Six principles for an embodied–embedded approach to behavior. In P. Calvo, & T. Gomila (Eds.), *Handbook of cognitive science: an embodied approach* (pp. 161–188). Elsevier.

Rizzolatti, G., Cattaneo, L., Fabbri-Destro, M., & Rozzi, S. (2014). Cortical mechanisms underlying the organization of goal-directed actions and mirror neuron-based action understanding. *Physiological Reviews*, *94*(2), 655–706.

Roche, K., & Chainay, H. (2017). Is there a competition between functional and situational affordances during action initiation with everyday tools? *Frontiers in Psychology*, *8*, 1073.

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), 585–615.

Rutkowska, J. M., Meyer, M., & Hunnius, S. (2021). Adults do not distinguish action intentions based on movement kinematics presented in naturalistic settings. *Brain Sciences*, *11*(6).

Scaliti, E., Pullar, K., Borghini, G., Cavallo, A., Panzeri, S., & Becchio, C. (2023). Kinematic priming of action predictions. *Current Biology*, *33*(13), 2717–2727.e6. http://dx.doi.org/10.1016/j.cub.2023.05.055, URL: https://www.sciencedirect.com/science/article/pii/S0960982223006875.

Schmidt, R. C. (2007). Scaffolds for social meaning. *Ecological Psychology*, *19*(2), 137–151.

Searle, J. R. (1984). *Minds, brains and science*. Cambridge: Harvard University Press.

Sebanz, N., & Shiffrar, M. (2009). Detecting deception in a bluffing body: The role of expertise. *Psychonomic Bulletin & Review*, *16*(1), 170–175.

Spunt, R. P., & Adolphs, R. (2014). Validating the why/how contrast for functional MRI studies of theory of mind. *Neuroimage*, *99*, 301–311.

Stapel, J. C., Hunnius, S., & Bekkering, H. (2012). Online prediction of others' actions: the contribution of the target object, action context and movement kinematics. *Psychological Research*, *76*, 434–445.

Stoffregen, T. A. (2003). Affordances as properties of the animal-environment system. *Ecological Psychology*, *15*(2), 115–134.

Stoffregen, T. A., Gorday, K. M., Sheng, Y. Y., & Flynn, S. B. (1999). Perceiving affordances for another person's actions. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 120.

Stoffregen, T. A., & Wagman, J. B. (2024). Higher order affordances. *Psychonomic Bulletin & Review*, http://dx.doi.org/10.3758/s13423-024-02535-y.

Tahboub, K. A. (2006). Intelligent human-machine interaction based on dynamic bayesian networks probabilistic intention recognition. *Journal of Intelligent and Robotic Systems*, *45*, 31–52.

Teddlie, C., & Tashakkori, A. (2003). Major issues and controversies in the use of mixed methods in the social and behavioral sciences. *Handbook of Mixed Methods in Social and Behavioral Research*, *1*(1), 13–50.

Thompson, E. L., Bird, G., & Catmur, C. (2019). Conceptualizing and testing action understanding. *Neuroscience & Biobehavioral Reviews*, *105*, 106–114.

Turvey, M. T. (1992). Affordances and prospective control: An outline of the ontology. *Ecological Psychology*, *4*(3), 173–187.

Turvey, M. T., Shaw, R. E., Reed, E. S., & Mace, W. M. (1981). Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn (1981). *Cognition*, *9*(3), 237–304.

Uithol, S., Görgen, K., Pischedda, D., Toni, I., & Haynes, J. D. (2023). The effect of context and reason on the neural correlates of intentions. *Heliyon*, *9*(6).

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

Wagman, J. B., Stoffregen, T. A., Bai, J., & Schloesser, D. S. (2018). Perceiving nested affordances for another person's actions. *Quarterly Journal of Experimental Psychology*, *71*(3), 790–799.

Warren, W. H. (1984). Perceiving affordances: visual guidance of stair climbing. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 683.

Williams, A. M. (2009). Perceiving the intentions of others: how do skilled performers make anticipation judgments? *Progress in Brain Research*, *174*, 73–83.

Williams, A. M., & Jackson, R. C. (2019). *Anticipation and decision making in sport*. Routledge.

Yeshurun, Y., Carrasco, M., & Maloney, L. T. (2008). Bias and sensitivity in two-interval forced choice procedures: Tests of the difference model. *Vision Research*, *48*(17), 1837–1851.

Zhang, L., Liu, G., Han, B., Wang, Z., & Zhang, T. (2019). sEMG based human motion intention recognition. *Journal of Robotics*, *2019*(1), Article 3679174.