Available online at www.sciencedirect.com

ScienceDirect

Physics of Life Reviews 24 (2018) 67–80

PHYSICS of LIFE reviews

www.elsevier.com/locate/plrev

Review

Seeing mental states: An experimental strategy for measuring the observability of other minds

Cristina Becchio [a,b,*], Atesh Koul [a,b], Caterina Ansuini [a], Cesare Bertone [b], Andrea Cavallo [b,a]

[a] *C'MoN, Cognition, Motion and Neuroscience Unit, Fondazione Istituto Italiano di Tecnologia, Via E. Melen 83, 16152 Genova, Italy*
[b] *Department of Psychology, University of Turin, Via Po 14, 10123 Torino, Italy*

Received 25 July 2017; received in revised form 29 September 2017; accepted 1 October 2017
Available online 6 October 2017
Communicated by J. Fontanari

Abstract

Is it possible to perceive others' mental states? Are mental states visible in others' behavior? In contrast to the traditional view that mental states are hidden and not directly accessible to perception, in recent years a phenomenologically-motivated account of social cognition has emerged: *direct social perception*. However, despite numerous published articles that both defend and critique direct perception, researchers have made little progress in articulating the conditions under which direct perception of others' mental states is possible. This paper proposes an empirically anchored approach to the observability of others' mentality – not just in the weak sense of discussing relevant empirical evidence for and against the phenomenon of interest, but also, and more specifically, in the stronger sense of identifying an experimental strategy for measuring the observability of mental states and articulating the conditions under which mental states are observable. We conclude this article by reframing the problem of direct perception in terms of establishing a definable and measurable relationship between movement features and perceived mental states.

Keywords: Mental states; Observability; Direct perception; Kinematics; Modeling

---

1. Seeing mental states: the current debate

1.1. The unobservability principle

How is knowledge of other minds possible? *How-possible questions* of this nature arise when something that seems impossible nevertheless happens. These are, therefore, *obstacle-dependent* questions: we ask how knowledge of x is possible when there appears to be an insuperable obstacle preventing knowledge of x [1]. In the case of knowledge of other minds, the apparent obstacle is the supposed opacity of other minds, such as the idea that we can never have direct knowledge of another's mental state [2]. This idea, called the 'Unobservability Principle' [3], 'Unobservability

---

[*] Corresponding author at: C'MoN Unit, Fondazione Istituto Italiano di Tecnologia, Center for Human Technologies, Via E. Melen 83, 16152 Genova, Italy.

E-mail address: cristina.becchio@iit.it (C. Becchio).

https://doi.org/10.1016/j.plrev.2017.10.002
1571-0645/

Thesis' [4], or 'Principle of Imperceptibility' [5,6], albeit generally unexpressed, has and continues to inform the hypotheses and frameworks behind much of the research in social cognition [7–9]. According to the 'Unobservability Principle', we *never* actually see another's mental states; that is, mental states, qua being intracranial phenomena, are perceptually inaccessible to everyone but their owner.

Accepting this principle motivates the question of how one might access another person's mind.

> *People do not have direct information about others' mental states and must therefore base their inferences on whatever information about others' mental states they do have access to. This requires a leap from observable behavior to unobservable mental states that is so common and routine that people often seem unaware that they are making a leap* [10].

A logical corollary to the 'Unobservability Principle' is the idea of a mechanism to infer others' invisible mental states [11], such as a theory of mind [12]. As Gallagher [13] puts it: "I cannot see into your mind, hence, I have to devise some way of inferring what must be there". According to this account, "one of the most important powers of the human mind is to conceive of and think about itself and other minds. Because the mental states of others are completely hidden from the senses, they can only ever be inferred" [11].

1.2. The direct social perception thesis

This view of the mind as an unobservable phenomenon has recently been challenged by the 'direct social perception' thesis. Proponents of this thesis argue that it is sometimes possible to *directly* perceive the mental states of others [13–17]. Often, we do not need to infer mental states through observing a target's behavior. We just *see* the mental states with the same immediacy and directness as we perceive ordinary objects. "If I directly see my car I do not ordinarily have to make an inference on the basis of what I see that it is my car", writes Gallagher [13]; "I do not see red mass, shape, and color, and then try to piece all of that together to make it add up to my car. I simply and directly see my car" [13]. With the same immediacy, we may sometimes see another person's emotions or intentions.

In this increasingly popular view, the alleged obstacle blocking the possibility of direct knowledge of other minds – their fundamental hiddenness – is eschewed [18]. The problem of other minds dissolves, as does the need of inferential processes [3]. Inferring a hidden set of mental states is simply unnecessary, seeing that "expressive behavior is saturated with the meaning of the mind; it reveals the mind to us" [19].

But how would mental states actually be revealed [3]? What degree of visual presence would they exhibit [20]? Which mental states would be visible and which would not? Despite numerous articles that defend and critique direct perception, there has been little progress toward answering these questions and articulating the conditions under which direct perception of mental states is, or would be, possible [21–24]. The claim that mental states are 'observable' thus far remains just as speculative as the claim that they are 'unobservable'.

2. Our approach

As a way out of this impasse, this review proposes an *empirically anchored approach to the problem of the* (*un-*) *observability of others' mentality* – not just in the weak sense of discussing relevant empirical evidence for and against the observability of others' mental states, but also in the stronger sense of *identifying an experimental strategy for determining whether and to what extent a given mental state is observable*.

Surprisingly, one aspect largely neglected by both proponents and opponents of the direct social perception thesis is the *availability of information* to perceive mental states (see Box 1). We suggest herein that an apt characterization of the observability of others' mental states requires one to quantitatively assess the mentalistic information available in the observable behavior. Put simply, to probe observers' basic capacity to *perceive* mental states, one must first demonstrate the availability of mentalistic information in the observed behavioral patterns (*Step 1*). Having demonstrated that information about mental states is encoded in behavioral patterns, one can then use rigorous quantitative behavioral techniques to test the perceptual efficiency of this information, i.e., the usefulness of this information for perception (*Step 2*). Using modeling techniques, one can then identify the specific features that observers use to detect mental states, and establish a measurable relationship between those features and the observability of the mental states (*Step 3*). Finally, one can manipulate the observability of the mental states (i.e., increasing or decreasing the visibility of mental states) by modifying the parameters of the observed movements (*Step 4*).

In this article, we illustrate this strategy via analysis of a prominent example from the mental state perception literature: the perception of others' intentions from subtle variations in movement kinematics [25]. First, we briefly review the phenomenon of interest as it appears, or fails to appear, and identify some of the pitfalls that might account for apparent contradictory findings in the literature. Then, drawing on data from recent and decisive studies [26], we delineate the experimental steps to overcome these pitfalls and quantitatively characterize the visibility of others' intentions. The focus of this paper is on perceiving intentions; however, as discussed in last part of the article, the approach herein could be used to probe the observability of any mental state instantiated into a behavioral pattern, ranging from emotions to motives. We conclude by proposing an alternative conception of direct social perception, where 'direct' no longer defines the nature of perception (inferential vs. direct), but rather the perceptual efficacy of available mentalistic information.

BOX 1. Ecological psychology and beyond

The operational approach described here to defining direct social perception has a kindred spirit as well as a precedence in the ecological approach to perception-action as originally conceived by Gibson [27] and further developed by Shaw, Turvey, and Mace [28]. Ecological psychology is grounded in the assumption of the availability of regularities in the properties of the world as the basis of reliable information for perceiving-acting systems ([29] for review). Accordingly, much of the research agenda for ecological psychology is aimed at identifying the sources of information that specify these regularities [30]. Our approach complements and extends the ecological psychology agenda by positing that specificational information is also available to perceive at least some of the mental states of other people. Put simply, regularities in kinematic pattern specify the performer's mental states. The challenge, then, becomes one of determining whether and to what extend perceivers can or do utilize information specifying mental states [see also [22]].

3. The observability of others' intentions: a current controversy

Is it possible to understand others' intentions by simply observing their movements? Previous studies, including some from our own laboratory, have yielded conflicting results [31–38]. Some studies indicate that observers can identify and use early differences in movement kinematics to discern intentions and subjective states [31,33,36–39]. For instance, Sartori and colleagues [36] showed that in a binary choice design observers were able to judge whether the agent's intent in grasping the object was to cooperate or compete by only using available kinematic information. In a follow-up study, Manera et al. [33] demonstrated that observers have no trouble identifying cooperative and competitive intentions, even from relatively degraded point-light displays of grasping actions. However, the same study also found that performance dropped to chance level when participants discriminated cooperative actions from individual-oriented actions [33]. Other investigators had difficulties replicating Sartori et al.'s [36] initial findings. Using a somewhat different procedure, Naish et al. [34] varied the duration of the observed movements so as to estimate how much of the action participants needed to see in order to correctly predict the action unfolding (to eat or to place). The result was that observers were unable to anticipate the intention to eat or to place until they had seen at least part of the post-grasp kinematics.

Why did Naish et al.'s participants show less perceptual sensitivity to intention-related information than Sartori et al.'s? Apart from methodological differences, one problem in interpreting these patterns of results arises from a general lack of data about availability of stimulus information. All the above studies relied almost exclusively on indirect evidence of intention-related information in the observed movements; none measured this information, resulting potentially in a set of 'pitfalls' that we enumerate below.

Pitfall #1: Studies do not consider the availability of intention-information. Somewhat surprisingly, some studies did not even consider the problem of quantifying intention-related information in the observed movements (e.g., [37,38]). Insofar as observers can detect intention, this may be negligible. In such cases, as Runeson [40] suggested, it is tempting "to use perception as a measure device", treating the human perceptual system as a practical device for measuring complex kinematics invariants. However, from a logical perspective, using perception to measure kinematic information means confounding the availability of stimulus information with its usefulness for perception. The limits of this approach become apparent when evaluated against the inability to detect intention. In such cases, it would indeed be impossible to determine whether observers lack the ability to pick up intention-related information (despite its availability), or whether such information is simply unavailable in the first place.

Pitfall #2: Studies assume the availability of intention-information. Other studies assumed, rather than measured, the availability of stimulus information. Commenting on the negative results by Naish et al., for example, Catmur [41] recently emphasized that "this inability to acquire intention information from kinematics is perhaps surprising since there were reliable differences between the kinematic profiles of the two types of actions". However, Naish et al. did not collect kinematic data for *the movements being viewed*. Rather, they obtained kinematic data within an independent action execution study by testing 17 participants. Having demonstrated that grasp-to-eat and grasp-to-place actions result in reliable kinematic differences, they then filmed three actors performing these movements and had participants view the corresponding clips. As such, the researchers did not directly assess intention-related differences in *the movements being viewed*, but rather assumed them based on the results obtained in a separate and much larger sample. The problem with this strategy is that it neglects a ubiquitous, often unwanted characteristic of motor performance: motor variability [42]. Performing a movement repeatedly does not result in the same motor output on every attempt. "Practice", as Nikolai Bernstein summarized his theory, "is repetition without repetition" [43]. In performing the same task, outputs of the motor system vary quite substantially from one trial to another, as well as from one individual to the next. One can thus not assume that all movements performed with a given intention are informative to the same extent. It is thus not clear whether Naish et al. have sufficient evidence to conclude "that most people are not able to detect subtle kinematic differences between, specifically, pre-grasp movements with different subsequent outcomes".

Pitfall #3: Studies only consider average kinematic differences. Caution is needed when interpreting kinematic differences averaged over trials and subjects. Take, for instance, the work by Manera et al. [33] discussed above. The findings in their experiment appear to suggest that observers were unable to report kinematic differences between cooperative and individual-oriented movements; but again, this conclusion might be spurious. The rationale for questioning this conclusion is that observers were not exposed to the entire set of movements ($n = 160$), for which mean differences are duly reported, but only to a subset of randomly selected movements ($n = 60$). Owing to variations of single trials, the selection procedure might have narrowed the kinematic differences between cooperative and individual-oriented movements. Thus, as the authors recognize, one cannot be certain of the amount of intention information available to judge the intention.

4. Measuring the observability of intentions

The strategy we propose to overcome these pitfalls and quantify the observability of intentions combines rigorous kinematic and quantitative behavioral techniques with modeling and classification analysis. As the implementation of these techniques has been recently reviewed elsewhere [44] in the following section, we only provide a general conceptual overview of the methods and their function in determining whether, and to what extent, intentions are observable.

4.1. Step 1: Quantify the availability of intention information

Following Marteniuk et al.'s seminal work [45], a plethora of studies have documented the influence of intention on human grasping parameters. The logic of these studies has been to manipulate the intention while keeping constant the to-be-grasped object. Ansuini et al. [46], for example, asked participants to reach towards and grasp a bottle so they might accomplish one of four possible actions: pouring, displacing, throwing, or passing. Their analysis of reach-to-grasp kinematics revealed that when participants grasped the bottle with the intent to pour, both the middle and the ring fingers were more extended than in all the other considered intentions. The authors took this as evidence for the influence of intention on hand-preshaping. Other studies have reported similar effects in the domain of social and communicative intention [47–49]. For instance, one study showed that participants' maximal finger aperture is larger and the peak grip closing velocity decreases when they grasp the object with the intention to hand it to another person, compared to intending to place it on a concave support [47]. Table 1 provides a non-exhaustive list of studies demonstrating modulation of grasping parameters by intentions.

Table 1
Studies demonstrating modulation of reach-to-grasp parameters by intention. n.a. not applicable.

Study	Type of intention	Reaching component	Grasping component
Ansuini et al., 2006	Individual (grasp-to-lift vs. grasp-to-place)	Movement duration	Finger angular excursions
Ansuini et al., 2008	Individual (grasp-to-pass, grasp-to-throw, grasp-to-place, and grasp-to-pour)	Movement duration	Finger angular excursions; Abduction angles
Armbruster & Spijkers, 2006	Individual (grasp-to-lift, grasp-to-throw away, and grasp-to-place)	Time to peak acceleration; Mean velocity;	Maximum grip aperture
Becchio et al., 2008a	Individual vs. Social (grasp-to-place vs. grasp-to-pass)		Maximum grip aperture; Peak grip closing velocity
Becchio et al., 2008b	Social (grasp-to-compete vs. grasp-to-cooperate)	Movement duration; Trajectory height; Deceleration time; Mean velocity	Maximum grip aperture; Time of maximum grip aperture
Crajè et al., 2011	Individual (grasp-to-lift vs. grasp-to-pour)	*n.a.*	Finger position at contact
Ferri et al., 2010	Individual vs. social (grasp-to-eat vs. grasp-to-feet)	Movement duration; Peak velocity; Deceleration time	
Flindall & Gonzales, 2013	Individual (grasp-to-eat vs. grasp-to-place vs. grasp-to-spit)	Peak velocity	Maximum grip aperture
Georgiou et al., 2007	Social (grasp-to-compete vs. grasp-to-cooperate)	Movement duration; Peak velocity; Deceleration time	Maximum grip aperture; Time of maximum grip aperture
Marteniuk et al., 1987	Individual (grasp-to-throw vs. grasp-to-lit)	Movement duration; Duration of deceleration and acceleration phase	*n.a.*
Naish et al., 2013	Individual (grasp-to-eat vs. grasp-to-place)	Peak acceleration; Trajectory length	Grasp position at contact
Quesque et al., 2013	Individual vs. social (grasp-to-move vs. grasp-to-pass)	Movement duration; Wrist displacement	
Sartori et al., 2009	Individual vs. social (grasp-to-lift vs. grasp-to-communicate)	Time of maximum trajectory height; Maximum wrist deviation	Finger opening velocity
Sartori et al., 2011	Individual (grasp-to-pour vs. grasp-to-move)	Movement duration	Maximum grip aperture; Time of maximum grip aperture; Fingers' position at contact
Schuboe et al., 2008	Individual (grasp-to-fill vs. grasp-to-place)	Movement duration; Peak velocity; Time to peak velocity; Deceleration time; Time to peak acceleration and deceleration	Grasp position at contact

These studies indicate that humans grasp objects in ways that reflect their intention. But does movement kinematics convey *specificational information* to discriminate intention? This goes beyond asking whether movements performed with different intentions differ on average on one or more kinematic parameters; it requires determining whether movement patterns specify intentions, i.e., whether information relevant to discriminate intention is available in the spatiotemporal patterns underlying the movements.[1]

---

[1] The term *information* is used here in the *specificational sense* of Turvey [62] and is thus reserved for those patterns that unambiguously specify intentions.

A straightforward way to measure the availability of specificational information is to submit data features to a *classification analysis*, such as linear discriminant analysis (LDA). Given a set of kinematic features, LDA can be used to find the linear combinations of features that provide the best discrimination between two or more intentions. If the kinematic features are informative about the intentions, then a high classification score is achieved. Fig. 1 shows the results of an LDA applied to reach-to-grasp movements of bottle performed with four possible intentions: pouring, placing, drinking, and passing.

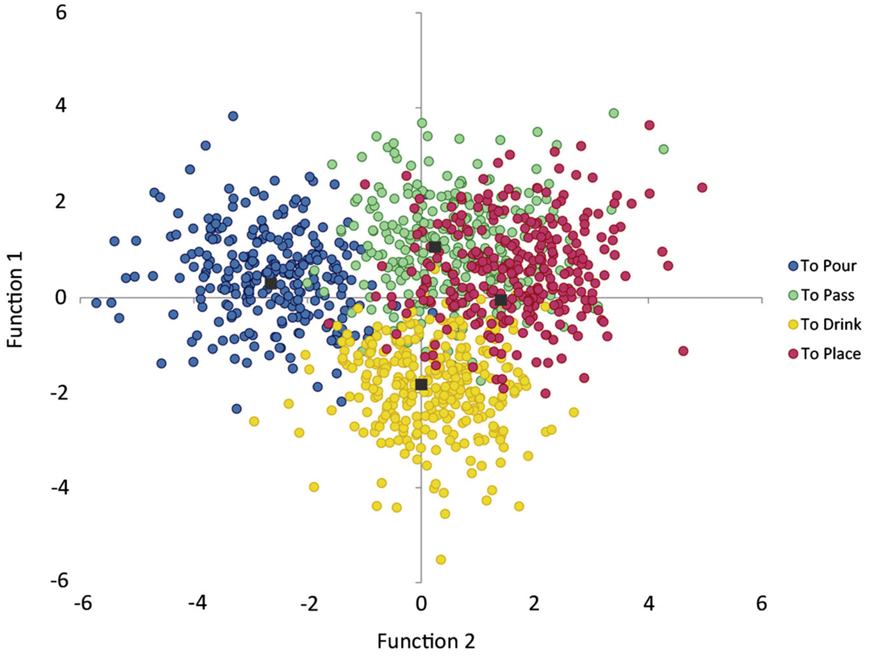

Fig. 1. Scatterplot of the two best linear discriminant functions in a linear discriminant analysis. Circles, labelled by the intention 'to place', 'to pour', 'to drink' and 'to pass', represent grasping movements. Grey squares represent the centroid of each intention. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

The scatterplot (axes represent the two best discriminant functions) represents movements as circles labelled by intention. The distance of each circle to the centroid (i.e., mean value based on the linear combination of the predictor variables) of each class reflects the probability that the movement belongs to the corresponding intention. In other words, this distance (known as Mahalanobis distance) is interpretable in terms of standard deviations from the centroid. A case located 2 Mahalanobis distance units from the centroid, for example, would have a less than 0.05 chance of belonging to that intention. Overall, classification score (0.79 in the portrayed example) provides a measure of intention-specific information available in the recorded movements.

4.2. Step 2: Determine the perceptual efficiency of intention information

As Runeson and Frykholm noted [50], the demonstration of availability of specificational information is, *per se*, not sufficient to establish its *perceptual efficiency*, that is, the usefulness of the available information for perception. Indeed, despite being available in the stimulus, intention information might nevertheless remain invisible to observers. For one thing, information might reside in parametrical variations that are non-discriminable by human observers – that is, variations too small to be perceived. Observers may also lack the necessary attunement to pick up the available intention information. Provided that specificational information is available in the movements to be used as stimuli, the second step is thus to probe observers' sensitivity to this information.

One approach to examining the usefulness of intention-related information for perception is to manipulate the information available through temporal occlusion [26]. Imagine one is observing an agent grasping a bottle with the action occluded at the time the fingers contact the bottle: is she intending to drink or to pass the bottle to another person?

Decisions of this sort require one to continuously evaluate inflowing information, and so are best described by sequential sampling models [51–55]. These models, including the drift-diffusion model (DDM; [56]), are centered on the assumption that sequentially sampling and accumulating noisy evidence to a decision criterion produces discrimination judgments [56,57]. For binary choices, the DDM decomposes behavioral data (RTs and choice) into four main parameters mapped onto latent psychological processes: boundary separation a for response caution, drift rate v for speed of accumulation, starting point z for *a-priori* response bias, and non-decision time t_0 for stimulus encoding and response execution latencies [58]. When no prior information is available, the distance from the starting point z to the decision criterion a is the same for the two alternative choices. At the same time, the strength of the observed signal determines the rate of evidence accumulation. The mean rate of approach to the decision criterion, i.e., the drift rate v, indicates the relative amount of information per time unit that is absorbed.

Fig. 2 provides a schematic illustration of the DDM applied to a perceptual case in which participants have to categorize two mental states: intention A and B. The process begins at the starting point z. Intention information is collected over the unfolding of the movement until evidence points with sufficient clarity to one intention – the process reaches the decision criterion at a. The pattern in Fig. 2 denotes that in most cases, the process reaches the upper threshold, indicating a positive drift rate and leading to the choice of the correct intention.

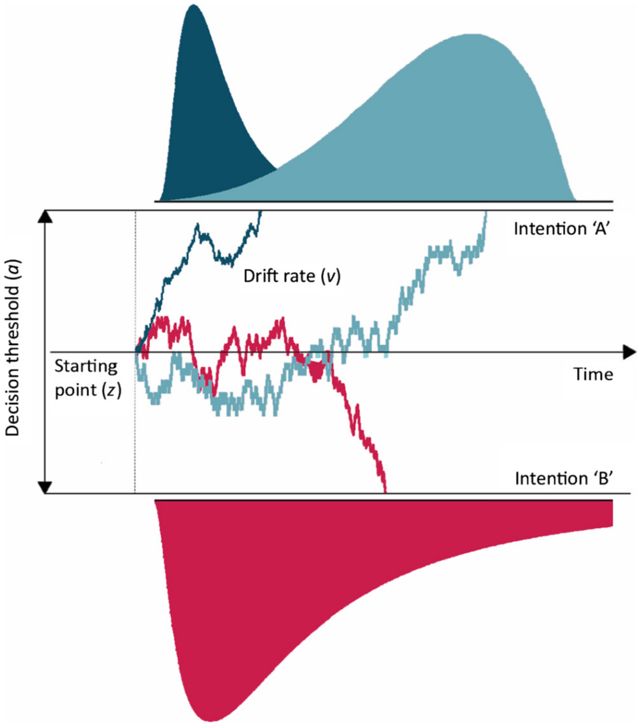

Fig. 2. Schematic of Drift-Diffusion Model. The two decision boundaries represent Intention 'A' and Intention 'B' in an intention discrimination task. The drift rate v represents mean sensory evidence per unit of time. The magnitude of v is determined by the kinematic features of the observed movements. The diffusion process starts at a starting point between the two boundaries (denoted as a proportion of a by z) until the accumulated evidence reaches one of two boundaries. If the correct boundary is hit (blue sample paths), the model makes a correct decision. Because of noise, the model may sometime hit the incorrect boundary (red sample path). (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

4.3. Step 3: Identify the features observers use to detect intentions

Accumulator models such as the DDM provide a measure of the relative intention-related information gathered over time, and are thus well-suited for making graded quantitative assessments of perceptual efficiency. What they do not provide, however, is an understanding of the specific features that observers utilize to judge intention.

One way to relate intention judgements to the expression of specific kinematic features is to create a Classification and Regression Tree (CART) model [26]. This approach is also known as *rule induction* because it permits classification or prediction of future observations based on a set of decision rules. Is the agent being observed grasping the bottle with the intent to drink or to pass? Through a series of yes/no questions, a CART analysis can be used to find the *kinematic features* that separate the data into movements that observers either classify as ‘grasp-to-pour’ or ‘grasp-to drink’.

A diagram tree displays the result, as Fig. 3 demonstrates. By traversing any given path from the root node to any leaf, classification trees of this sort can transform easily to *if . . . then . . .* rules, where the *if* statement defines a

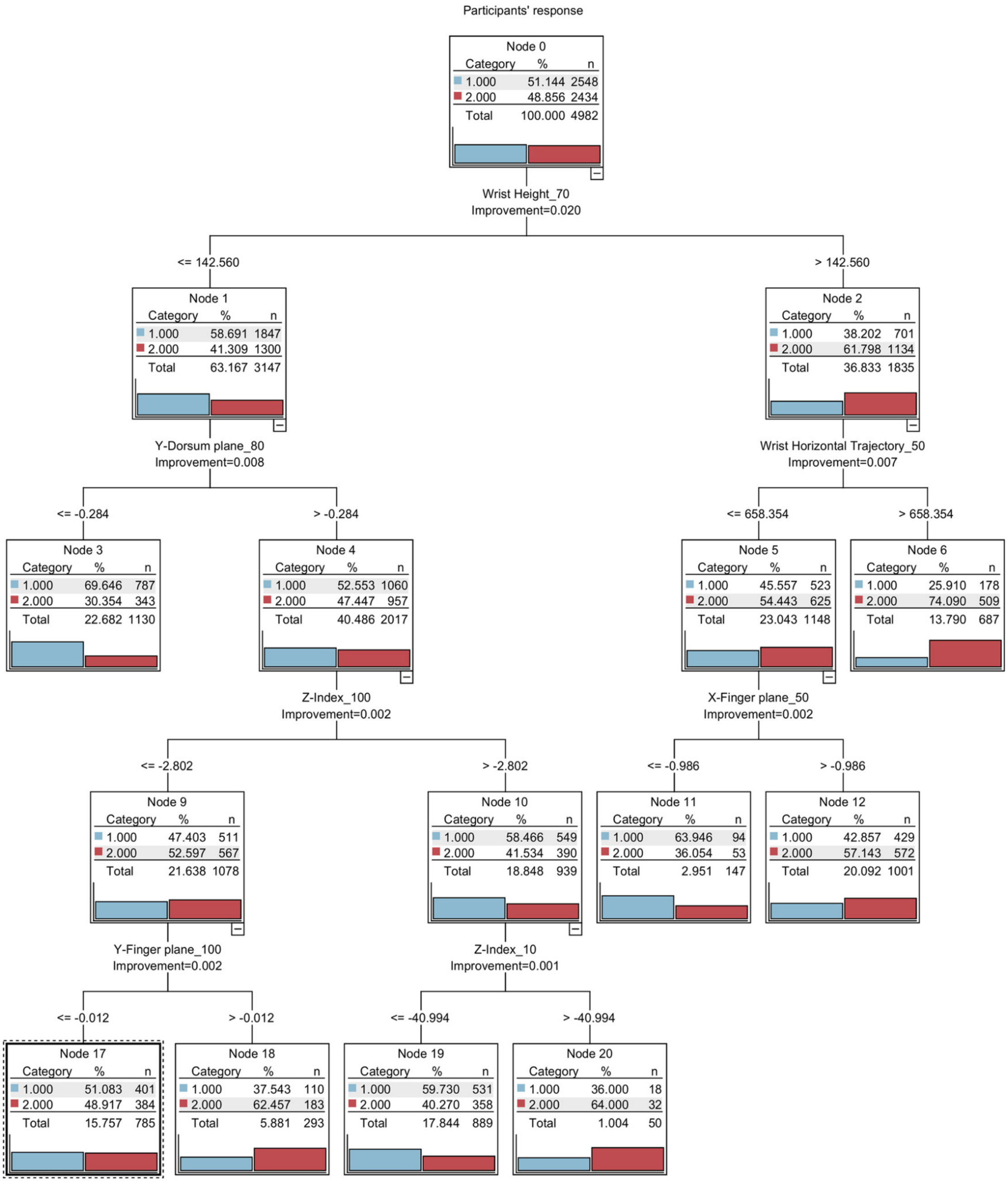

Fig. 3. CART model for predicting intention choice (to-pour vs. to-drink) from the observation of grasping movements. The model is generated using participants’ responses as outcome and kinematics features of the observed movements as predictors. Each node provides the total number of trials within the node and the number of ‘to-pour’ and ‘to-drink’ choices. For example, node 3 represents high probability of ‘to-pour’ choice (category 1 – light blue), while node 6 represents high probability of ‘to-drink’ choice (category 2 – red). Adapted from [26].

partition of a set of kinematic features, and the *then* corresponds to the predicted intention choice. For example, *if* wrist height positioned at 70% of movement duration is lower than 14 cm from the table surface and the dorsum plane is more flexed than −0.30 cm, *then* intention choice by an observer is expected to be 'to pour' with 0.74 probability. This form enables the computation of two types of predictions: i) a point prediction about the intention choice, and, ii) a distributional prediction about the probability associated with the point prediction. Using a CART model, we can thus not only predict an observer's intention choice (e.g., whether the observer will judge the movement to be 'to pour' or 'to drink'), but also the likelihood of the predicted choice (i.e., how likely the observer will report that the observed movement is 'to pour' or 'to drink').

4.4. Step 4: Alter the observability of intentions

As illustrated in *Step 3*, CART provides a direct tool with which to map intention choice to the expression of specific kinematic features. But how do we know that observers are actually *using* those features? The ultimate demonstration would be the precise *manipulation* of intention visibility based on those kinematic features. If observers use the subset of kinematic features identified by CART to classify intention, then presenting participants with movements that express these features should aid in intention discrimination. Conversely, presenting movements that do not express these features should make judgement more difficult.

One way to test these predictions is to combine the probability of the predicted choice derived from CART with the actual movement intention, and use this information to select, for each intention, two subsets of movements:

- High-informative movements, i.e., movements for which the predicted probability of correct choice is high (e.g., 'to pour' with 0.70 probability);
- Low-informative movements, i.e., movements for which the predicted probability of correct choice is low (e.g., 'to pour' with 0.50 probability).

Intention observability, as measured by drift rate (see *Step 2*), should vary as a function of movement informativeness, being significantly higher for high-informative movements compared to low-informative movements. As shown in Fig. 4, this result is precisely what is found, which indicates that modifying the kinematic parameters of the movements being viewed can directly affect intention observability.

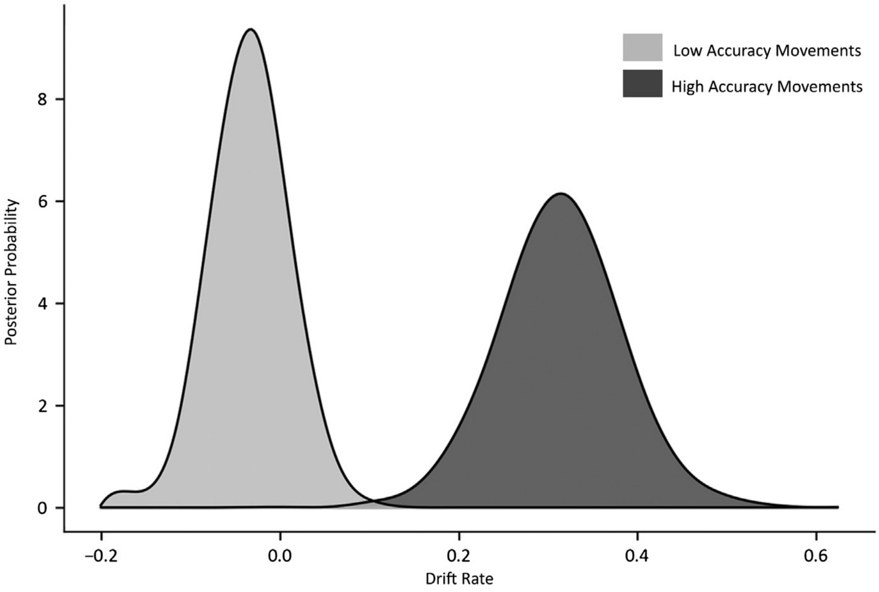

Fig. 4. Drift rate as function of CART selection. Drift rate is significantly larger for movements for which the CART model predicts higher classification accuracy.

5. The 'Observability Principle'

Are intentions observable in others' movements? Debate on this question has been primarily focused on untestable theoretical considerations on the observability of mental states (for an overview of the arguments, see [18]). The operational approach we have developed here allows one to translate the abstract concept of 'observability' into a quantifiable property – a property that one can not only measure, but also manipulate. Summarizing the above reported findings, observability (o) could be described formally as follows:

$$o = f(Ib + E)$$

where o is a function of the product between the behavioral specification of an intention (the quantified available information, I) and the usefulness of I for perception (perceptual efficiency, b) plus the error (E). One might represent this product as a sum of individual contributions of multiple sources of information (I_i) and their corresponding perceptual efficiencies (b_i):

$$o = f(I_1b_1 + I_2b_2 + I_3b_3 + I_4b_{4\ldots.+}I_nb_n + E)$$

In the case of intention-from-movement understanding, these sources would correspond to information from multiple kinematic features, and their respective efficiencies.

We may in principle apply this formulation – which we intend to be provocative, rather than definitive – not only to intentions, but also to any mental state instantiated into a specific pattern of behavior, including emotions, motives, and even desires. Are others' emotions visible? Are motives in performing a given action visible to others? If we are to accept the above formulation of observability, then the process for answering these questions will involve the following steps:

Step 1. Quantify behavioral specification of the mental state, being the specificational information available to discriminate the mental state in patterns of behavior.
Step 2. Assess the perceptual efficiency of this information by probing observers' sensitivity to the available mentalistic information.
Step 3. Identify the specific features that observers use to detect the mental states.
Step 4. Put the identified features to the test by manipulating the visibility of the mental states.

In this process, each step presupposes the results of the previous step to be available. Importantly, the process does not assume the observability of mental states, but rather is meant as a strategy to quantify the observability of those states. As a result of *Steps 1–4*, a mental state will be quantified as more or less visible. When information to discriminate a mental state is not available in the pattern of behavior (*Step 1*) or, despite being available, is not useful for perception (*Step 2*), the mental state will be invisible to human observers. In either of these two cases, understanding what the other person thinks, desires, or believes requires "a leap from observable behavior to unobservable mental states", as Epley and Waytz [10] stated. Thus, the proposed approach does not deny that some mental states are invisible, but rather contends that unobservability is something that can be measured. This has not only theoretical significance but potential impact in application, the range of applications spanning from social signal processing to human-robot interaction (e.g., making humanoid robots' movements more predictable for human partners, see BOX 2).

BOX 2. From fundamental principles to applications: the case of robotics

Implementing a theory of mind for a robot has been a long-lasting dream for robotics researchers [59]. While this remains a challenge, the ‘Observability Principle’ suggests that a robot would often need less than a full-blown theory of mind to take part in human social dynamics. The ability to pick-up mentalistic information in behavioral patterns would already allow a robot to read a range of mental states, i.e., those mental states that are instantiated in discriminable patterns of behavior.

What is more, reverse engineering of the ‘Observability Principle’ may inform the design of robotic motion to improve human-robot interaction. If we are to build human-like robots that can interact naturally with people, our robots must not only be able to read the mental states of human agents, they also must be able to express behaviors that are ‘readable’ for human agents. The simplest solution would appear to incorporate human-like patterns of behavior in the controller of humanoid-robots; for example, by simply retargeting human motion capture data to robots. Due to constraints of robot mechanics (e.g., the degrees of freedom differ in number or location on the kinematic structures of robots and humans), however, this often does not produce human-like patterns [60].

The ‘Observability Principle’ suggests that instead of retargeting human motion capture data, one could use *Steps 3 and 4* in the above described process to identify a minimal set of kinematic parameters that the robotic motion should satisfy (Fig. 5). Consider the CART model discussed above. In addition to the point and distributional predictions, CART analysis also provides a variable importance score, reflecting the contribution each variable (i.e., kinematic feature) makes in predicting the intention choice by observers. We predict that optimizing robotic movements to satisfy kinematic features ranked as most important should improve detection accuracy of robot motions, thus making it easier for a human partner to anticipate the robot’s intention, i.e., to anticipate what the robot will do next. Following a similar logic, other mental states, such as emotion and attention, could be made observable in the robotic behavior.

The methods described in *Step 2* could then be used to define a metric for the evaluation of quality in robotic motion, and validate the optimization of the robotic motion based on the selected features. As illustrated in Fig. 5, the process of feature selection could be reiterated (including other features ranked as important) until discrimination of robotic motion by human observers reaches the desired accuracy.

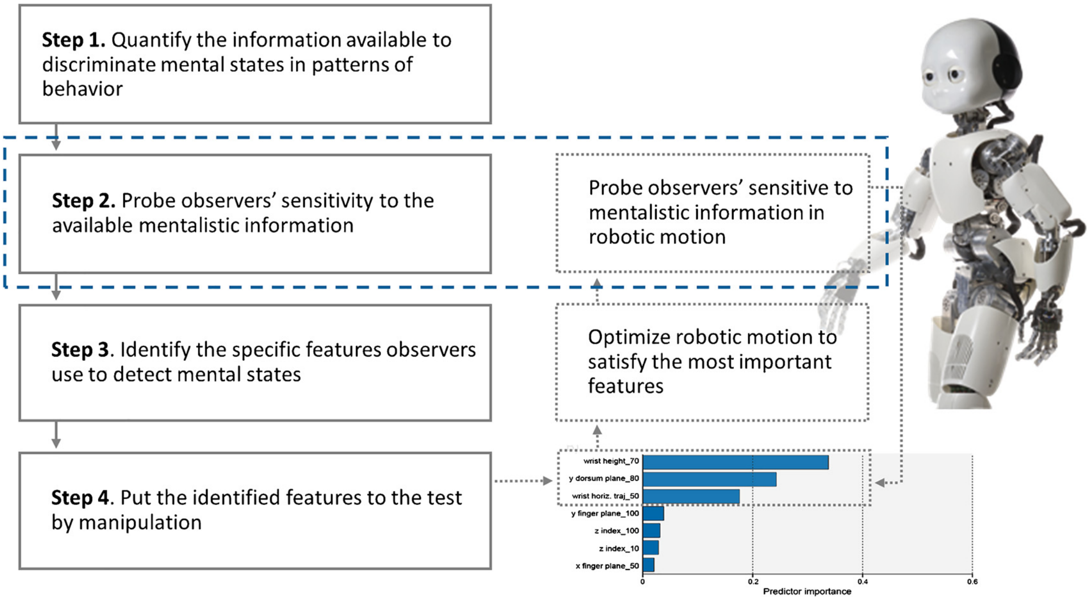

Fig. 5. Experimental strategy to measure and manipulate the observability of mental states. Experimental steps to determine whether and to what extent mental states are observable in patterns of behavior are shown in boxes surrounded by a solid line. Application to design of robotic motion is shown in boxes surrounded by a grey dotted line. The robotic platform ICub (courtesy of ICub facility) is represented on the right side. The horizontal bar plot (lower right) shows the variable score importance obtained from the CART model in Step 3 and validated in Step 4. The blue dashed line indicates commonality of methods.

6. A new look at 'direct' perception

In line with the Gibsonian tradition that defines directness as not being supplemented by inferential processes [30], the standard way of framing the direct social perception thesis has focused on the nature of perception. Is perception smart enough to enable us to grasp what others are thinking or doing without the addition of some mediating 'mindreading', inferential mechanism [13]? If the perceptual processes involved in social cognition were 'dumb', perception would need to be supplemented with interpretation sustained by extra-perceptual inferential processes before an individual could figure out another's emotions and intentions [61]. On the contrary, if perception were 'smart', then what one sees would already make sense. In other words, one could non-inferentially see others' mental states.

An interesting implication of the approach described here is that it does not require perception to be that smart for it to be direct. As long as mentalistic information is available in the stimulus, even not-so-smart perception [13] is capable of perceiving mental states.

Based on this framework, we propose to reconceptualize the notion of 'direct' perception as follows:

> *Perception of a mental state from a given behavior is 'direct' insofar as the features of the observed behavior predict the mental state an observer will perceive.*

A key advantage of this formulation is that it turns the issue of 'direct' perception into one that is empirically addressable. If a definable and measurable relationship between movement features and perceived mental states can be established, then perception may be qualified as 'direct'. In this case, direct would no longer define the nature of perception, but instead would relate to the perceptual efficacy of available mentalistic information.

Acknowledgements

The research was funded by a grant of the European Research Council under the European Union's Seventh Framework Programme (FP/2007–2013)/ERC Grant Agreement n. 312919 to CB.

References

[1] Cassam Q. The possibility of knowledge. Oxford: Oxford University Press; 2007.
[2] Hyslop A. Other minds, vol. 246. Springer Science & Business Media; 2013.
[3] Krueger J. Seeing mind in action. Phenomenol Cogn Sci 2012;11:149–73. http://dx.doi.org/10.1007/s11097-011-9226-y.
[4] Overgaard S. The unobservability thesis. Synthese 2017;194:743–60.
[5] Gallagher S, Varga S. Social constraints on the direct perception of emotions and intentions. Topoi 2014;33:185–99. http://dx.doi.org/10.1007/s11245-013-9203-x.
[6] McNeill WES. Embodiment and the perceptual hypothesis. Philos Q 2012;62:569–91. http://dx.doi.org/10.1111/j.1467-9213.2012.00065.x.
[7] Leslie AM, Friedman O, German TP. Core mechanisms in "theory of mind". Trends Cogn Sci 2004;8:528–33. http://dx.doi.org/10.1016/j.tics.2004.10.001.
[8] Karmiloff-Smith BA. Beyond modularity: a developmental perspective on cognitive science. Int J Lang Commun Disord 1994;29:95–105. http://dx.doi.org/10.3109/13682829409041485.
[9] Gopnik A, Wellman HM. Why the child's theory of mind really is a theory. Mind Lang 1992;7:145–71.
[10] Epley N, Waytz A. Mind perception. In: Fiske ST, Gilbert DT, Lindzey G, editors. The handbook of social psychology. 5th ed. New York: Wiley; 2009. p. 498–541.
[11] Leslie A. Children's understanding of the mental world. In: Gregory RL, editor. The Oxford companion to mind. Oxford: Oxford University Press; 1987. p. 139–42.
[12] Leudar I, Costall A. Against theory of mind. London: Palgrave Macmillan; 2009.
[13] Gallagher S. Direct perception in the intersubjective context. Conscious Cogn 2008;17:535–43. http://dx.doi.org/10.1016/j.concog.2008.03.003.
[14] Ratcliffe M. 'Folk psychology' is not folk psychology. Phenomenol Cogn Sci 2006;5:31–52. http://dx.doi.org/10.1007/s11097-005-9010-y.
[15] Reddy V. How infants know minds. Harvard University Press; 2008.
[16] Hutto DD. ToM rules, but it is not OK. In: Leudar I, Costall A, editors. Against theory mind. London: Palgrave Macmillan; 2009. p. 221–38.
[17] Zahavi D, Gallagher S. The (in) visibility of others: a reply to Herschbach. Philos Explor 2008;11:237–44.
[18] Overgaard S. Other people. In: Zahavi D, editor. The Oxford handbook of contemporary phenomenology. Oxford: Oxford University Press; 2012. p. 460–79.
[19] Gallagher S, Zahavi D. The phenomenological mind: an introduction to philosophy of mind and cognitive science. Routledge (Taylor & Francis); 2008.

[20] Smith J. Seeing other people. Philos Phenomenol Res 2010;81:731–48. http://dx.doi.org/10.1111/j.1933-1592.2010.00392.x.

[21] Michael J, De Bruin L. How direct is social perception? Conscious Cogn 2015;36:373–5. http://dx.doi.org/10.1016/j.concog.2015.08.005.

[22] Bohl V. Continuing debates on direct social perception: some notes on Gallagher's analysis of "the new hybrids". Conscious Cogn 2015;36:466–71. http://dx.doi.org/10.1016/j.concog.2015.05.003.

[23] Spaulding S. On direct social perception. Conscious Cogn 2015;36:472–82. http://dx.doi.org/10.1016/j.concog.2015.01.003.

[24] Wiltshire TJ, Lobato EJC, McConnell DS, Fiore SM. Prospects for direct social perception: a multi-theoretical integration to further the science of social cognition. Front Human Neurosci 2015;8:1007. http://dx.doi.org/10.3389/fnhum.2014.01007.

[25] Ansuini C, Cavallo A, Bertone C, Becchio C. Intentions in the brain: the unveiling of Mister Hyde. Neuroscientist 2015;21:126–35. http://dx.doi.org/10.1177/1073858414533827.

[26] Cavallo A, Koul A, Ansuini C, Capozzi F, Becchio C. Decoding intentions from movement kinematics. Sci Rep 2016;6:37036. http://dx.doi.org/10.1038/srep37036.

[27] Gibson JJ. The ecological approach to visual perception. Boston: Hought Mifflin; 1979.

[28] Shaw RE, Turvey MT, Mace WM. Ecological psychology: the consequence of a commitment to realism. In: Weimer WB, Palermo DS, editors. Cognition and the symbolic processes, vol. 2. Hillsdale: Lawrence Erlbaum Associates, Inc.; 1982. p. 159–226.

[29] Petrusz SC, Turvey MT. On the distinctive features of ecological laws. Ecol Psychol 2010;22:44–68. http://dx.doi.org/10.1080/10407410903493152.

[30] Richardson MJ, Shockley K, Fajen BR, Riley MA, Turvey MT. Ecological psychology: six principles for an embodied-embedded approach to behavior. In: Calvo P, Gomila A, editors. Handbook of cognitive science: an embodied approach. Elsevier; 2008. p. 159–87.

[31] Becchio C, Cavallo A, Begliomini C, Sartori L, Feltrin G, Castiello U. Social grasping: from mirroring to mentalizing. NeuroImage 2012;61:240–8. http://dx.doi.org/10.1016/j.neuroimage.2012.03.013.

[32] Lewkowicz D, Quesque F, Coello Y, Delevoye-Turrell YN. Individual differences in reading social intentions from motor deviants. Front Psychol 2015;6:1175. http://dx.doi.org/10.3389/fpsyg.2015.01175.

[33] Manera V, Becchio C, Cavallo A, Sartori L, Castiello U. Cooperation or competition? Discriminating between social intentions by observing prehensile movements. Exp Brain Res 2011;211:547–56. http://dx.doi.org/10.1007/s00221-011-2649-4.

[34] Naish KR, Reader AT, Houston-Price C, Bremner AJ, Holmes NP. 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. Exp Brain Res 2013;225:261–75. http://dx.doi.org/10.1007/s00221-012-3367-2.

[35] Quesque F, Delevoye-Turrell Y, Coello Y. Facilitation effect of observed motor deviants in a cooperative motor task: evidence for direct perception of social intention in action. Q J Exp Psychol 2016;69:1451–63. http://dx.doi.org/10.1080/17470218.2015.1083596.

[36] Sartori L, Becchio C, Castiello U. Cues to intention: the role of movement information. Cognition 2011;119:242–52. http://dx.doi.org/10.1016/j.cognition.2011.01.014.

[37] Sebanz N, Shiffrar M. Detecting deception in a bluffing body: the role of expertise. Psychon Bull Rev 2009;16:170–5. http://dx.doi.org/10.3758/PBR.16.1.170.

[38] Stapel JC, Hunnius S, Bekkering H. Online prediction of others' actions: the contribution of the target object, action context and movement kinematics. Psychol Res 2012;76:434–45. http://dx.doi.org/10.1007/s00426-012-0423-2.

[39] Patel D, Fleming SM, Kilner JM. Inferring subjective states through the observation of actions. Proc Biol Sci 2012;279:4853–60. http://dx.doi.org/10.1098/rspb.2012.1847.

[40] Runeson S. Perceiving people through their movements. In: Kirkcaldy B, editor. Individual differences in movement. Dordrecht: Springer Netherlands; 1985. p. 43–66.

[41] Catmur C. Understanding intentions from actions: direct perception, inference, and the roles of mirror and mentalizing systems. Conscious Cogn 2015;36:426–33. http://dx.doi.org/10.1016/j.concog.2015.03.012.

[42] Ting LH, Chiel HJ, Trumbower RD, Allen JL, McKay JL, Hackney ME, et al. Neuromechanical principles underlying movement modularity and their implications for rehabilitation. Neuron 2015;86:38–54. http://dx.doi.org/10.1016/j.neuron.2015.02.042.

[43] Bernstein NA. The co-ordination and regulation of movements: conclusions towards the study of motor co-ordination, biodynamics of locomotion. Oxford: Pergamon Press; 1967.

[44] Koul A, Becchio C, Cavallo A. PredPsych: a toolbox for predictive machine learning based approach in experimental psychology research. Accepted for publication in Behav Res Methods.

[45] Marteniuk RG, MacKenzie CL, Jeannerod M, Athenes S, Dugas C. Constraints on human arm movement trajectories. Can J Psychol 1987;41:365–78. http://dx.doi.org/10.1037/h0084157.

[46] Ansuini C, Giosa L, Turella L, Altoè G, Castiello U. An object for an action, the same object for other actions: effects on hand shaping. Exp Brain Res 2008;185:111–9. http://dx.doi.org/10.1007/s00221-007-1136-4.

[47] Becchio C, Sartori L, Bulgheroni M, Castiello U. Both your intention and mine are reflected in the kinematics of my reach-to-grasp movement. Cognition 2008;106:894–912. http://dx.doi.org/10.1016/j.cognition.2007.05.004.

[48] Becchio C, Sartori L, Bulgheroni M, Castiello U. The case of Dr. Jekyll and Mr. Hyde: a kinematic study on social intention. Conscious Cogn 2008;17:557–64. http://dx.doi.org/10.1016/j.concog.2007.03.003.

[49] Sartori L, Becchio C, Bara BG, Castiello U. Does the intention to communicate affect action kinematics? Conscious Cogn 2009;18:766–72. http://dx.doi.org/10.1016/j.concog.2009.06.004.

[50] Runeson S, Frykholm G. Kinematic specification of dynamics as an informational basis for person-and-action perception: expectation, gender recognition, and deceptive intention. J Exp Psychol Gen 1983;112:585–615. http://dx.doi.org/10.1037/0096-3445.112.4.585.

[51] Usher M, McClelland JL. The time course of perceptual choice: the leaky, competing accumulator model. Psychol Rev 2001;108:550–92. http://dx.doi.org/10.1037/0033-295X.108.3.550.

[52] Mazurek ME, Roitman JD, Ditterich J, Shadlen MN. A role for neural integrators in perceptual decision making. Cereb Cortex 2003;13:1257–69. http://dx.doi.org/10.1093/cercor/bhg097.

[53] Reddi BAJ, Asrress KN, Carpenter RHS. Accuracy, information, and response time in a saccadic decision task. J Neurophysiol 2003;90:3538–46. http://dx.doi.org/10.1152/jn.00689.2002.

[54] Smith PL, Ratcliff R. Psychology and neurobiology of simple decisions. Trends Neurosci 2004;27:161–8. http://dx.doi.org/10.1016/j.tins.2004.01.006.

[55] Bogacz R, Usher M, Zhang J, McClelland JL. Extending a biologically inspired model of choice: multi-alternatives, nonlinearity and value-based multidimensional choice. Philos Trans R Soc Lond B, Biol Sci 2007;362:1655–70. http://dx.doi.org/10.1098/rstb.2007.2059.

[56] Ratcliff R. A theory of memory retrieval. Psychol Rev 1978;85:59–108. http://dx.doi.org/10.1037/0033-295X.85.2.59.

[57] Gold JI, Shadlen MN. The neural basis of decision making. Annu Rev Neurosci 2007;30:535–74. http://dx.doi.org/10.1146/annurev.neuro.29.051605.113038.

[58] Ratcliff R, McKoon G. The diffusion decision model: theory and data for two-choice decision tasks. Neural Comput 2008;20:873–922. http://dx.doi.org/10.1162/neco.2008.12-06-420.

[59] Scassellati B. Theory of mind for a humanoid robot. Auton Robots 2002;12:13–24. http://dx.doi.org/10.1023/A:1013298507114.

[60] Sciutti A, Ansuini C, Becchio C, Sandini G. Investigating the ability to read others' intentions using humanoid robots. Front Psychol 2015;6:1362. http://dx.doi.org/10.3389/fpsyg.2015.01362.

[61] De Jaegher H. Social understanding through direct perception? Yes, by interacting. Conscious Cogn 2009;18:535–42. http://dx.doi.org/10.1016/j.concog.2008.10.007.

[62] Turvey MT. Intentionality: a problem of multiple reference frames, specificational information, and extraordinary boundary conditions on natural law. Commentary on: Sayre KM. Intentionality and information processing: an alternative model for cognitive science. Behav Brain Sci 1986;9:153–5.