PNAS BRIEF REPORT | PSYCHOLOGICAL AND COGNITIVE SCIENCES OPEN ACCESS

Seeing and understanding epistemic actions

Sholei Croom[a,1], Hanbei Zhou[a], and Chaz Firestone[a,1]

Edited by Susan Gelman, University of Michigan, Ann Arbor, MI; received March 27, 2023; accepted July 27, 2023

Many actions have instrumental aims, in which we move our bodies to achieve a physical outcome in the environment. However, we also perform actions with epistemic aims, in which we move our bodies to acquire information and learn about the world. A large literature on action recognition investigates how observers represent and understand the former class of actions; but what about the latter class? Can one person tell, just by observing another person's movements, what they are trying to learn? Here, five experiments explore *epistemic action understanding*. We filmed volunteers playing a "physics game" consisting of two rounds: Players shook an opaque box and attempted to determine i) the number of objects hidden inside, or ii) the shape of the objects inside. Then, independent subjects watched these videos and were asked to determine which videos came from which round: Who was shaking for number and who was shaking for shape? Across several variations, observers successfully determined what an actor was trying to learn, based only on their actions (i.e., how they shook the box)—even when the box's contents were identical across rounds. These results demonstrate that humans can infer epistemic intent from physical behaviors, adding a new dimension to research on action understanding.

social perception | theory of mind | action recognition | intuitive physics

Beyond recognizing objects, faces, and scenes, we also recognize the actions of other people. Accordingly, a large literature explores how we represent and understand actions such as walking, reaching, pushing, lifting, eating, chasing, and following (1–4). Such understanding is important for anticipating others' behaviors (e.g., where they might move next) and may also inform richer inferences about underlying attitudes and mental states, such as intention (1), agency (2), deception (3), confidence (5), belief (6), preference (6), and value (e.g., inferring that someone who accepts significant costs to achieve a goal likely values that goal highly; 7).

Independent of the inferences we draw on their basis, actions like walking, reaching, eating, etc. share a common feature: They are instrumental, or pragmatic—actions whose primary aim is some physical outcome in the environment (retrieving something, moving somewhere, etc.). However, we also perform actions with other goals. For example, we act to communicate with others (e.g., waving or pointing; 8), to signal physical or social characteristics (e.g., assuming an aggressive posture, or imitating; 9), or even to be creative and act "for its own sake" (e.g., dancing; 10).

Among these broader action classes, an important and understudied example concerns *epistemic* or information-seeking actions. For example, someone might press on a door to figure out whether it is locked, dip their toe into a pool to gauge its temperature, or shake a box to determine its contents (e.g., a child wondering if a wrapped-up present contains Lego blocks or a teddy bear). Moreover, the content of one's epistemic goal may guide how one fulfills it; for example, one might shake a box differently to determine the number of objects inside than to determine their shape, texture, or weight. While such actions have physical consequences, they ultimately serve a different goal: acquiring information.

Epistemic actions pervade our lives, and recognizing them does too (e.g., inferring that a meandering campus visitor is seeking directions, or that a friend who checks shallow drawers and trays is looking for something small, like keys or earrings). However, they have not received the same scientific attention as other actions, despite some work investigating epistemic actions as actually performed by actors (11, 12; see also 13, 14). This raises a question: Can observers tell, just by watching how someone's body moves, what that person is trying to learn?

The Present Experiments: Can You See What I Want to Know? Here, we investigate *epistemic action understanding*, by exploring a case study of an epistemic behavior: learning about an object by manipulating it with one's hands. Our experiments consisted of two phases (Fig. 1). First, we filmed volunteers playing a "physics game": Objects were hidden in an opaque box, and players guessed what was inside only by shaking it. Importantly, the game had two rounds: 1) guessing the number of objects inside the box; 2) guessing the objects' shape. Previous work (12, 15) suggests that players should succeed—i.e., perform well at guessing the number and shape of the hidden objects.

Author affiliations: [a]Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD 21218

Author contributions: S.C., H.Z., and C.F. designed research; S.C. and H.Z. performed research; S.C. analyzed data; and S.C., H.Z., and C.F. wrote the paper.

The authors declare no competing interest.

Copyright © 2023 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]To whom correspondence may be addressed. Email: scroom1@jhu.edu or chaz@jhu.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2303162120/-/DCSupplemental.

Published November 20, 2023.

Downloaded from https://www.pnas.org by 23.142.224.21 on November 20, 2023 from IP address 23.142.224.21.

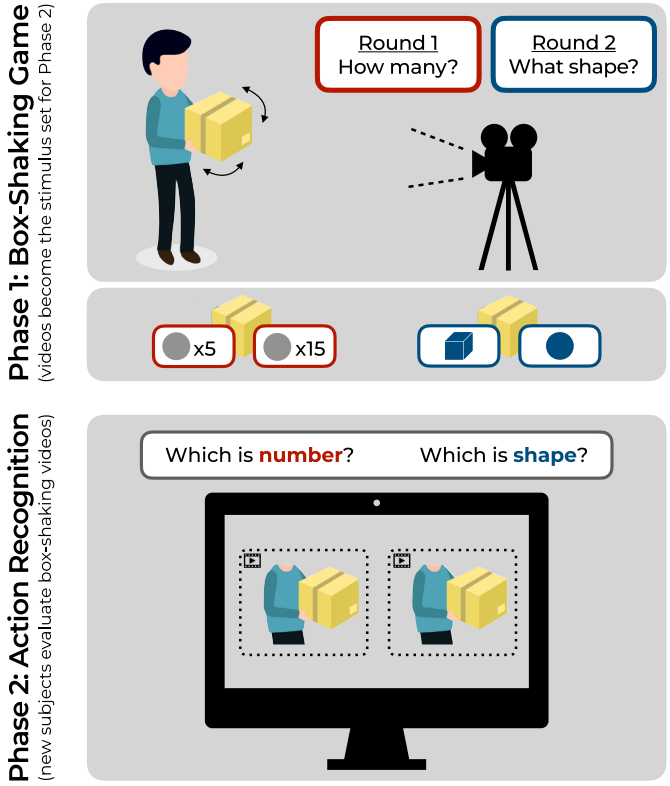

Fig. 1. *Top*: Players were filmed trying to determine the contents of a box (specifically, the number or shape of the objects inside), only by shaking it. Later experiments vary the box's contents. *Bottom*: Observers watched these videos and judged which came from which round: Who was shaking for number and who was shaking for shape?

The present contribution arises from the second phase. Independent participants watched videos of the physics game and were given a new task: to determine which videos came from which round—i.e., who was shaking for number and who was shaking for shape. This task requires many layers of physical and psychological reasoning: Determining which behaviors correspond to which round requires understanding which properties can be detected by which interactions (e.g., what information is revealed when objects hit the side of a box), which box-shaking strategies create such interactions, whether the players understand these dependencies, etc. If participants succeed, this would suggest that naive observers can recognize an agent's epistemic intent simply from the kinematics of their actions.

Results

Experiment 1 filmed 16 naive participants ("players") completing the box-shaking game. In a task adapted from Siegel et al. (15), players approached an opaque box (7.25 × 7.25 × 5in) and guessed a property of its contents, only by lifting and shaking it. In the Number round, the box contained several coins (US nickels), and players guessed whether there were 5 or 15. In the Shape round, the box contained a geometric solid (diameter = 2in), and players guessed whether it was a sphere or cube. Contents and round order were counterbalanced across players. As expected, this task was easy: 100% of players answered correctly in both rounds.

Next, these videos were uploaded online, where 100 naive participants ("observers") completed a different task: determining which videos came from the Number round and which from the Shape round. On each trial, observers saw two videos of the same player (without audio or visible faces), and simply judged which was which. Demonstrations are available at https://perceptionresearch.org/epistemicaction.

Observers succeeded: Mean accuracy was 76.2%, $t(99) = 21.24$, $P < 0.0001$, $d = 2.12$ (Fig. 2). This success was pervasive: Only 4/100 observers had numerically below-chance performance, and 14/16 actors elicited numerically above-chance discrimination in observers. This result provided initial evidence that participants inferred epistemic goals from motor behavior; the box-shaking dynamics allowed observers to determine what someone was trying to learn.

Experiment 2 further isolated epistemic intent by asking to what extent observers' success (at determining which video displayed which round) relied on players' success (at determining the number or shape of the objects). The box-shaking game was adjusted to elicit more errors in players, who discriminated between 9, 12, or 16 coins (Number), and a sphere, cylinder, or cube (Shape). As expected, player accuracy diminished: only 4/18 players guessed correctly in both rounds. One hundred new observers participated.

Observers succeeded again: 65.9% accuracy, $t(99) = 15.30$; $P < 0.0001$, $d = 1.53$. Crucially, all correctness subgroups elicited successful observer performance, including videos from players who answered incorrectly in both rounds; $t(99) = 16.97$, $P < 0.0001$, $d = 1.70$. Thus, epistemic action understanding does not require the action to be successful; merely attempting to acquire information produces behaviors that can signal one's epistemic goals.

In Experiments 1 and 2, players' epistemic intent (determining number vs. shape) was confounded with the box's contents (coins vs. one large object); could that explain observers' performance? Though this possibility may still implicate sophisticated action understanding—it is not trivial to infer the contents of a box based only on observed shaking behaviors—it would not implicate epistemic action understanding itself. Experiment 3 thus equated the box's contents across rounds. Here, the box always contained 20 small cubes; however, this was not disclosed to players, who were simply told that the Number round contained 15, 20, or 25 objects, and the Shape round contained spheres, cubes, or cylinders. Eighteen new players and 100 new observers participated.

Observers succeeded again, despite the identical contents: 63.69% accuracy, $t(98) = 8.52$; $P < 0.0001$, $d = 0.86$. Thus, success in this task goes beyond mere differences in shaking movements elicited by the contents of the box.

Experiments 1 to 3 gave observers many details about the box-shaking task, including a reenactment of the instructions, the precise quantities and shapes involved, and more; is epistemic action understanding possible only under such leading circumstances? Experiment 4 repeated Experiment 3 with dramatically diminished instructions: no reenactment, no candidate quantities/shapes, and no information about the objects' composition—just two sentences explaining that some players tried to determine the number of objects in a box and some tried to determine shape. Even with this minimal guidance, observer performance matched Experiment 3: 63.72% accuracy, $t(99) = 8.54$, $P < 0.0001$; $d = 0.85$.

Finally, Experiment 5 explored epistemic action understanding beyond a forced-choice context. After watching videos from Experiment 3, observers were asked "Why do you think these people were shaking the boxes?"; then, after affirming that players were trying to guess the contents, observers were asked about Number videos separately from Shape videos. Exploratory analyses revealed that 75% of observers answered the first question by invoking information-seeking, suggesting that these actions are readily interpreted in terms of epistemic goals. Moreover, 74% of observers answered the follow-up questions by invoking number or shape and more often invoked the correct property for the correct video than vice versa; $\chi^2(1, N = 96) = 16.23$, $P < 0.0001$). Thus, observers were sensitive to epistemic goals, even without preset responses.

Downloaded from https://www.pnas.org by 23.142.224.21 on November 20, 2023 from IP address 23.142.224.21.

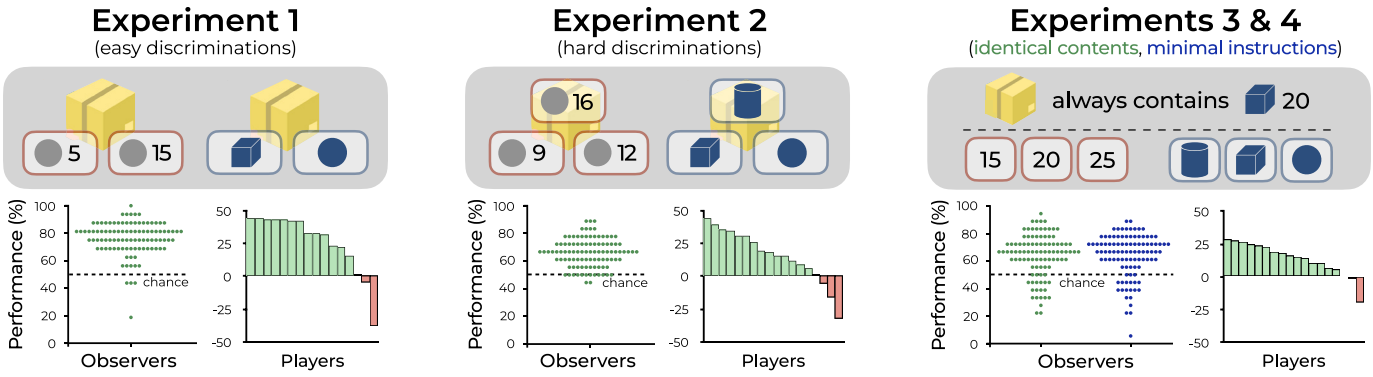

Fig. 2. Observers reliably determined which round they were shown—i.e., who was shaking for number and who was shaking for shape. This pattern arose across observers (with most performing above chance) and players (with most producing shaking motions that elicited above-chance performance in observers).

Discussion

The present work explored *epistemic action understanding*: Across hundreds of participants and several variations, naive observers inferred what information another person was attempting to acquire, only by observing their motor behavior directed towards a box. This pattern arose for players who correctly and incorrectly guessed the box's contents, with diminished information about players' task, and with the box's contents equated. These results were robust: Most players produced shaking motions that differed systematically across rounds (number vs. shape), and most observers successfully determined which shaking motions corresponded to which epistemic goals. Finally, though our task placed constraints on both players and observers, these effects nevertheless emerged under fairly naturalistic conditions: Our experiments used real-life videos of ordinary people genuinely attempting to learn something (rather than, e.g., trained actors, synthetic animations, point-light displays, or photographs; cf. 13, 14), and our observers were naive participants without any training or feedback.

These results suggest that observers can visually recognize not only what someone wants to do but also what someone wants to know. While it has been demonstrated that observers can infer someone's instrumental or pragmatic goals from their behavior—and make further inferences about higher-level mental states (1–7)—the present work goes further in demonstrating that epistemic goals can be inferred from visual observation. Moreover, our observers showed sensitivity to the finer-grained content of these goals beyond simpler forms of perceptual knowledge attribution (e.g., understanding that seeing leads to knowing). These findings complement recent work investigating other action classes, including communicative (8) and affiliative (9) actions, as well as actions taken "for their own sake" (10).

This work opens the door to future research on epistemic action understanding. Visual inspection of the box-shaking videos suggests clear strategic differences by round: Players often shook up and down for Number and tilted side to side for Shape. Precisely characterizing such patterns—including their stability over variables such as the material of the objects and box—could be an interesting challenge for computer vision systems operating on kinematic data (16). A model formalizing how observers use physical knowledge to infer epistemic intent could also inform computational work on intuitive mentalizing (6, 7).

Future work might explore the developmental trajectory of these abilities, asking when epistemic action understanding emerges ontogenetically and whether it arises alongside other mentalizing abilities (14, 17). Another natural extension is to explore other epistemic actions, such as those in the Introduction (e.g., inferring someone is seeking by the size and shape of the containers they search). Finally, while the present work explored sensitivity to one epistemic goal rather than another (determining number vs. shape), one could also probe sensitivity to epistemic goals as opposed to pragmatic goals—e.g., navigating an environment to reach a destination vs. to scout the terrain.

Beyond these directions, the present work illustrates that what someone is attempting to know, not just what someone is attempting to do, is an aspect of cognition worth exploring in many domains.

Methods

Supplementary methods appear in *SI Appendix*. Studies were approved by the Johns Hopkins Institutional Review Board; participants gave informed consent.

Data, Materials, and Software Availability. Anonymized raw and analyzed data are available on the Open Science Framework (https://osf.io/wndkg/) (18).

1. S.-J. Blakemore, J. Decety, From the perception of action to the understanding of intention. *Nat. Rev. Neurosci.* **2**, 561–567 (2001).
2. A. Hafri, A. Papafragou, J. C. Trueswell, Getting the gist of events: Recognition of two-participant actions from brief displays. *J. Exp. Psychol. General* **142**, 880–905 (2013).
3. S. Runeson, G. Frykholm, Kinematic specification of dynamics as an informational basis for person and action perception: Expectation, gender recognition, and deceptive intention. *J. Exp. Psychol. General* **112**, 585–615 (1983).
4. L. Isik, A. Tacchetti, T. Poggio, A fast, invariant representation for human action in the visual system. *J. Neurophysiol.* **119**, 631–640 (2018).
5. D. Patel, S. M. Fleming, J. M. Kilner, Inferring subjective states through the observation of actions. *Proc. R Soc. B* **279**, 4853–4860 (2012).
6. C. L. Baker, J. Jara-Ettinger, R. Saxe, J. B. Tenenbaum, Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nat. Hum. Behav.* **1**, 0064 (2017).
7. S. Liu, T. D. Ullman, J. B. Tenenbaum, E. S. Spelke, Ten-month-old infants infer the value of goals from the costs of actions. *Science* **358**, 1038–1041 (2017).
8. A. Royka, A. Chen, R. Aboody, T. Huanca, J. Jara-Ettinger, People infer communicative action through an expectation for efficient communication. *Nat. Commun.* **13**, 4160 (2022).
9. L. J. Powell, E. S. Spelke, Human infants' understanding of social imitation: Inferences of affiliation from third party observations. *Cognition* **170**, 31–48 (2018).
10. A. Schachner, S. Carey, Reasoning about 'irrational' actions: When intentional movements cannot be explained, the movements themselves are seen as the goal. *Cognition* **129**, 309–327 (2013).
11. D. Kirsh, P. Maglio, On distinguishing epistemic from pragmatic action. *Cogn. Sci.* **18**, 513–549 (1994).
12. G. Burton, M. T. Turvey, H. Y. Solomon, Can shape be perceived by dynamic touch? *Perception Psychophys.* **48**, 477–487 (1990).
13. S. Droop, N. R. Bramley, "Inferring epistemic intention in simulated physical microworlds" in *Proceedings of the Annual Meeting of the Cognitive Science Society* (2022), p. 44.
14. B. Varga, G. Csibra, A. Kovacs, "Infants' interpretation of information-seeking actions" in *Proceedings of the Annual Meeting of the Cognitive Science Society* (2021), p. 43.
15. M. H. Siegel, R. W. Magid, M. Pelz, J. B. Tenenbaum, L. E. Schulz, Children's exploratory play tracks the discriminability of hypotheses. *Nat. Commun.* **12**, 3598 (2021).
16. Y. Kong, Y. Fu, Human action recognition and prediction: A survey. *Int. J. Computer Vision* **130**, 1366–1401 (2022).
17. G. Gergely, G. Csibra, Teleological reasoning in infancy: The naive theory of rational action. *Trends Cogn. Sci.* **7**, 287–292 (2003).
18. S. Croom, H. Zhou, C. Firestone, Seeing and understanding epistemic actions. Open Science Framework. https://osf.io/wndkg/. Deposited 27 March 2023.

Downloaded from https://www.pnas.org by 23.142.224.21 on November 20, 2023 from IP address 23.142.224.21.