Seeing and understanding epistemic actions

Sholei Croom[a], Hanbei Zhou[a], Chaz Firestone[a,b]

[a] *Department of Psychological & Brain Sciences, Johns Hopkins University*
[b] *Department of Philosophy, Johns Hopkins University*

Abstract

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, 5 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.

Email addresses: `scroom1@jhu.edu` (Sholei Croom), `chaz@jhu.edu` (Chaz Firestone)

In press, Proceedings of the National Academy of Sciences *July 28, 2023*

Introduction

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 (Blakemore and Decety, 2001; Hafri et al., 2013; Runeson and Frykholm, 1983; Isik et al., 2018). 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 (Blakemore and Decety, 2001), agency (Hafri et al., 2013), deception (Runeson and Frykholm, 1983), confidence (Patel et al., 2012), belief (Baker et al., 2017), preference (Baker et al., 2017), and value (e.g., inferring that someone who accepts significant costs to achieve a goal likely values that goal highly; Liu et al., 2017).

Independent of the inferences we might 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, beyond actions with physical or pragmatic aims, we also perform actions with other goals. For example, we might act to communicate with others (e.g., waving or pointing; Royka et al., 2022), to signal physical or social characteristics (e.g., assuming an aggressive posture, or imitating; Powell and Spelke, 2018), or even to be creative and act 'for its own sake' (e.g., dancing; Schachner and Carey, 2013).

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 also have physical consequences, they are undertaken in service of a different goal: acquiring information.

Epistemic actions pervade our lives, and recognizing them does too (e.g., inferring that a meandering visitor to a college campus is seeking directions, or that a friend who repeatedly 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 action classes, despite some work investigating epistemic actions as actually performed by actors (Kirsh and Maglio, 1994; Burton et al., 1990; though see Droop and Bramley, 2022; Varga et al., 2021). This raises an intriguing 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 (Figure 1). First, we filmed volunteers playing a 'physics game': Objects were hidden inside 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 (Burton et al., 1990; Siegel et al., 2021) suggests that players should succeed at this task — i.e., perform well at guessing the number and shape of the hidden objects.

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 corresponded to 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 estimation task 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 will create such information-revealing interactions, whether the actors understand these dependencies, etc. If participants succeed, this would suggest that naive observers can recognize an agent's epistemic intent, simply by observing the kinematics of their actions.

Results

Experiment 1 filmed 16 naive participants ('players') completing the box-shaking game. Players approached an opaque box (7.25x7.25x5in) and guessed some 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') were given a different task: to determine which videos came from a player's Number round and which from their Shape round. On each trial, observers saw two

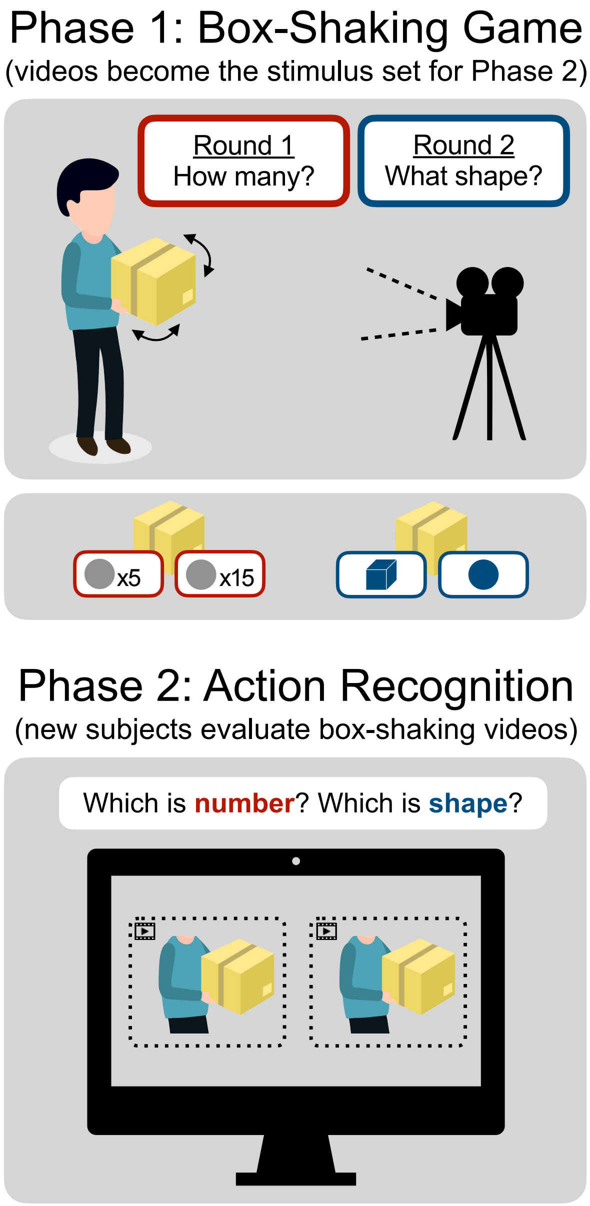
Phase 1: Box-Shaking Game
(videos become the stimulus set for Phase 2)
Round 1
How many?
Round 2
What shape?
x5
x15
Phase 2: Action Recognition
(new subjects evaluate box-shaking videos)
Which is number? Which is shape?

Figure 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?

videos (without audio, and without faces visible) from the same player and made a two-alternative forced-choice judgment about which was which. Demonstrations are available at https://perceptionresearch.org/epistemicaction.

Observers succeeded: Mean accuracy was 76.2%, $t(99) = 21.24, p < 0.0001, d = 2.12$ (Figure 2). Moreover, this success was pervasive: Only 4/100 observers produced numerically below-chance performance, and nearly all actors (14/16) produced shaking motions that elicited above-chance discrimination in observers. This experiment thus 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) depended on players' success (at determining the number or shape of the hidden objects). The box-shaking game was adjusted to elicit a higher error-rate 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. 100 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; $t(99) = 16.97, p < .0001, d = 1.70$. 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 1–2, 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 (at least for the objects used here) — 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. 18 new players and 100 new observers participated.

Observers succeeded again, despite the identical contents: 63.69% accuracy, $t(98) = 8.52; p < .0001, d = 0.86$. Thus, success in this task goes beyond mere differences in shaking movements afforded by the contents of the box.

Experiments 1–3 gave observers many details about the box-shaking task, including a video reenactment of the instructions, the precise quantities and shapes in-

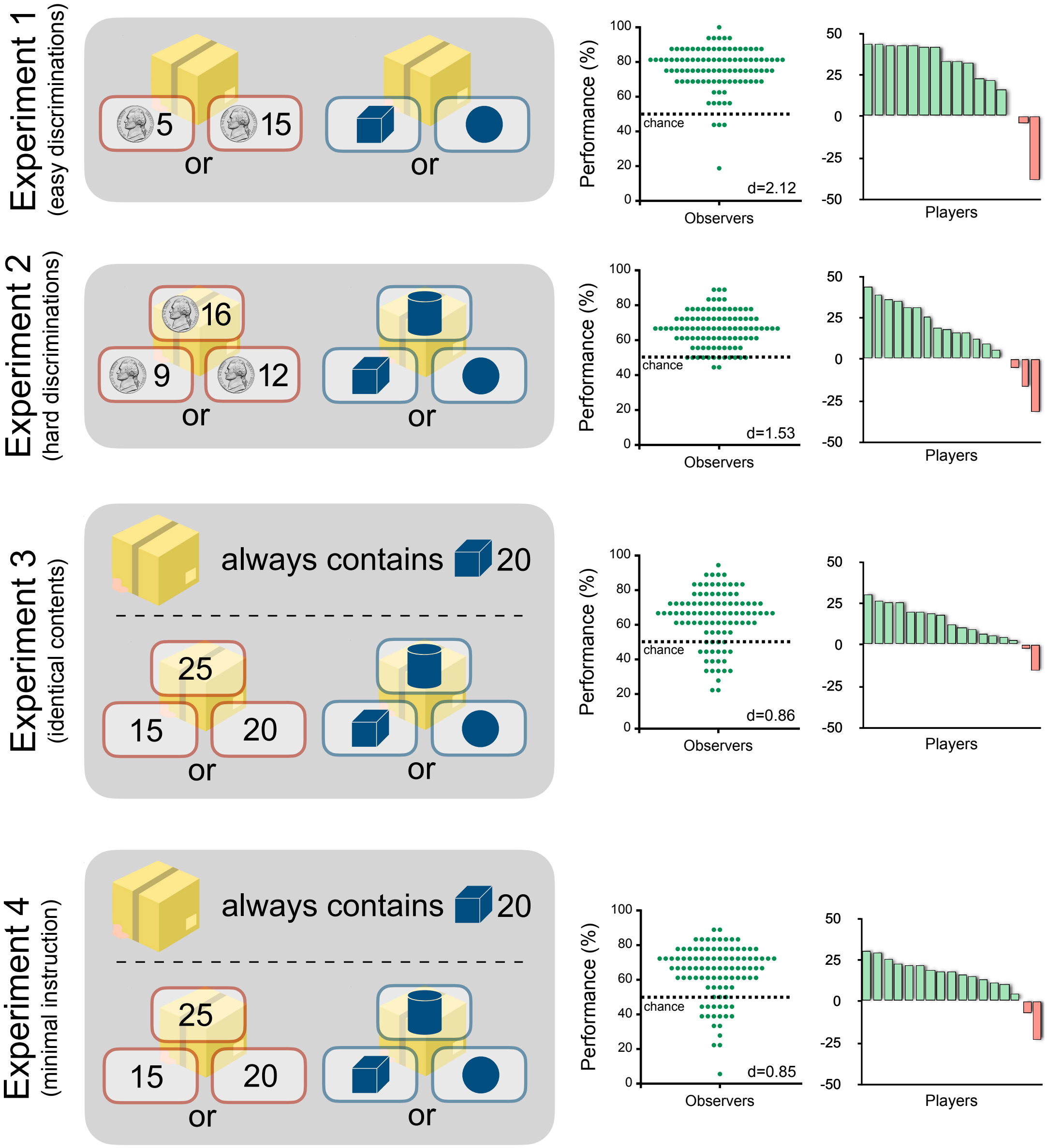

Figure 2: Observers reliably determined which round they were shown — in other words, they could tell who was shaking for number and who was shaking for shape. This pattern arose across observers (with a strong majority performing above chance) and players (with a strong majority producing shaking motions that elicited above-chance performance in observers).

volved, 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' size/material/weight — just two sentences telling observers 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 closely matched Experiment 3: 63.72% accuracy, $t(99) = 8.54, p < .0001; d = 0.85$.

Finally, Experiment 5 explored epistemic action understanding beyond the forced-choice context of Experiments 1–4. After watching box-shaking 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 learn something about the contents, observers were asked about Number videos separately from Shape videos. An exploratory analysis 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, even without prompting. Moreover, 74% of observers answered the follow-up questions by invoking number or shape, and were more likely to invoke the correct property for the correct video than vice versa; $\chi^2(1, N = 96) = 16.23, p < .0001$). Thus, observers were sensitive to more fine-grained epistemic goals, even without predetermined responses to choose from.

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 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. Droop and Bramley, 2022; Varga et al., 2021), 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 (Blakemore and Decety, 2001; Hafri et al., 2013; Runeson and Frykholm, 1983; Isik et al., 2018; Patel et al., 2012; Baker et al., 2017; Liu et al., 2017) — the present work goes further in demonstrating that epistemic goals can be inferred from visual observation. Moreover, our observers showed sensitivity to finer-grained content of these goals, in ways that go beyond simpler forms of perceptual knowledge attribution (e.g., understanding that seeing leads to knowing). These findings also complement recent work investigating other action classes, including communicative (Royka et al., 2022) and affiliative (Powell and Spelke, 2018) actions, as well as actions taken 'for their own sake' (Schachner and Carey, 2013).

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, with players often shaking up-and-down for Number and tilting side-to-side for Shape; precisely characterizing such patterns — including their stability over task constraints such as the material(s) of the objects and the box — could be an interesting challenge for computer vision systems operating on human kinematic data (Kong and Fu, 2022). A model that formalizes how observers leverage physical knowledge to infer epistemic intent could also contribute to computational work on intuitive mentalizing (Baker et al., 2017; Liu et al., 2017).

Future work might also explore the developmental trajectory of these abilities, asking when epistemic action understanding emerges ontogenetically and whether it arises along with other mentalizing abilities (Varga et al., 2021; Gergely and Csibra, 1997). Another natural extension would be to explore other epistemic actions, such as those mentioned in the Introduction (e.g., inferring what kind of object 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 makes explicit that what someone is attempting to know, not just what someone is attempting to do, may be an important (and neglected) aspect of human cognition worth exploring in many domains.

Methods

Supplementary methods appear in SI Appendix. Studies were approved by the JHU IRB; observers and players provided informed consent.

References

Baker, C. L., Jara-Ettinger, J., Saxe, R., and Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, 1(4):0064.

Blakemore, S.-J. and Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, 2(8):561–567.

Burton, G., Turvey, M., and Solomon, H. Y. (1990). Can shape be perceived by dynamic touch? *Perception & Psychophysics*, 48(5):477–487.

Droop, S. and Bramley, N. R. (2022). Inferring epistemic intention in simulated physical microworlds. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 44.

Gergely, G. and Csibra, G. (1997). Teleological reasoning in infancy: The infant's naive theory of rational action: A reply to premack and premack. *Cognition*, 63(2):227–233.

Hafri, A., Papafragou, A., and Trueswell, J. C. (2013). Getting the gist of events: recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General*, 142(3):880–905.

Isik, L., Tacchetti, A., and Poggio, T. (2018). A fast, invariant representation for human action in the visual system. *Journal of Neurophysiology*, 119(2):631–640.

Kirsh, D. and Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive science*, 18(4):513–549.

Kong, Y. and Fu, Y. (2022). Human action recognition and prediction: A survey. *International Journal of Computer Vision*, 130(5):1366–1401.

Liu, S., Ullman, T. D., Tenenbaum, J. B., and Spelke, E. S. (2017). Ten-month-old infants infer the value of goals from the costs of actions. *Science*, 358(6366):1038–1041.

Patel, D., Fleming, S. M., and Kilner, J. (2012). Inferring subjective states through the observation of actions. *Proceedings of the Royal Society B: Biological Sciences*, 279(1748):4853–4860.

Powell, L. J. and Spelke, E. S. (2018). Human infants' understanding of social imitation: Inferences of affiliation from third party observations. *Cognition*, 170:31–48.

Royka, A., Chen, A., Aboody, R., Huanca, T., and Jara-Ettinger, J. (2022). People infer communicative action through an expectation for efficient communication. *Nature Communications*, 13(1):4160.

Runeson, S. and 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.

Schachner, A. and Carey, S. (2013). Reasoning about 'irrational'actions: When intentional movements cannot be explained, the movements themselves are seen as the goal. *Cognition*, 129(2):309–327.

Siegel, M. H., Magid, R. W., Pelz, M., Tenenbaum, J. B., and Schulz, L. E. (2021). Children's exploratory play tracks the discriminability of hypotheses. *Nature Communications*, 12(1):3598.

Varga, B., Csibra, G., and Kovacs, A. (2021). Infants' interpretation of information-seeking actions. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 43.

SI Appendix: Materials and Methods

Here, we provide detailed descriptions of the methods, analyses and results for each of the five experiments reported in the main text. Readers can also experience all experiments for themselves at https://perceptionresearch.org/epistemicaction/.

General Methods

Open Science Practices

Sample sizes, exclusion criteria, analyses, and key experimental parameters reported here were pre-registered for Experiments 1–4. Experiment 5 was exploratory. Data, analyses, stimuli and pre-registrations are publicly available at https://osf.io/wndkg/.

Participants

Players

All players in the box-shaking game were members of the Johns Hopkins University community. Experiment 1 recruited 16 players and Experiments 2 and 3 recruited 18 players each (52 distinct players total), with the same players' videos from Experiment 3 used in Experiments 4 and 5. Videos from 3 additional players were used for instructional purposes during each experiment, but were not shown during experimental trials. Players across all experiments varied by gender, race and body-type.

Observers

All observers were adults recruited from the online platform Prolific. Experiments 1–5 each recruited 100 different observers (500 observers total).

All experiments were approved by the Homewood Institutional Review Board of Johns Hopkins University. Players and observers gave informed consent, and were monetarily compensated for their participation.

Stimuli

All stimuli in the experiment were videos of players completing our box-shaking game. Videos were recorded using a Canon EOS M50 mounted on a tripod at 1080p and 27 frames per second**.** Each video was presented to observers without audio, and with only the torso and arms of the player visible. The box used in all experiments was 7.25 x 7.25 x 5 inches. Two pieces of photo-reflective tape in the shape of a cross were adhered to the front face of the box to facilitate perception of pitch and yaw.

Procedure

Box-shaking game

Players were instructed to shake the box and guess some property of the contents inside (number or shape). They were told in advance that there would be two rounds, but were only told the property of interest before that particular round. Players were instructed to shake in whichever way felt most natural for answering the question they were asked, as long as they kept the front of the box within frame. Players were also encouraged not to shake for longer than 10 seconds, though this limit was not enforced. Each player completed both rounds of the game; the order of the rounds, as well as the contents of the box, were counterbalanced across players.

Action-recognition task

Observers in Experiments 1–3 were introduced to the rules of the box-shaking game, including the two properties of interest from the two rounds (number and shape). They were then given the following instruction:

> *Your goal is to figure out which round is which. In other words, you will pick out which videos show someone trying to guess the number of objects inside the box, and which videos show someone trying to guess the shape of an object inside the box.*

On each trial, a number video and a shape video from the same player appeared side-by-side on a single display. Each video was edited such that it began with the box sitting on a table in front of the player, after which the player could be seen picking up the box and shaking it to guess its contents. The videos concluded when the volunteer returned the box to the original location on the table in front of them. Only the torso and arms of the players were visible in each video, and they were played without audio. Observers could play the videos as many times as they liked, but had to play each video at least once in order to give their response. The videos had an average duration of 8 seconds; further information about video length is available in our data archive, and analyses of video length appear below.

Experiments 4 and 5 varied the instructions and response method; see below for more detail.

Experiment 1: Easy Discriminations

This experiment presents observers with videos of players making easy discriminations in the box-shaking game. 'Easy' here is reflected in players' success rate in guessing the contents of the box after shaking. In Experiment 1, 100% of players correctly guessed the contents of the box in both the shape round and the number round.

Box-Shaking Game

In Experiment 1, the contents of the box in each round were as follows: in the number round, the box could contain either 5 or 15 coins (US nickels); in the shape round, the box could contain either of two solid wooden shapes, a sphere (2in diameter) or cube (2in length).

Analyses and Results

Mean observer accuracy was 76.2%, t(99)=21.24, p<0.0001, d=2.12. Thus, observers could tell, on average, who was shaking for number and who was shaking for shape.

In a supplementary analysis, we asked whether *video length* could be the source of observers' successful performance. For example, if it turns out that players tended to shake longer for shape than for number, and observers had some insight or intuitions about this relationship, then perhaps observers could succeed just by picking "shape" for the longer of the two videos on each trial. To rule out this possibility, we divided players into two groups: Those who shook longer for number than for shape, and those who shook longer for shape than for number. Results showed that observer performance was significantly above chance for *both* of these player groups' videos, t(99)=22.82, p<0.0001, d=2.28; t(99)=7.28, p<0.0001, d=0.73. This means that observers' success cannot be explained merely by video length.

Experiment 2: Hard Discriminations

This experiment presents observers with videos of players making hard discriminations in the box-shaking game. 'Hard' here is reflected in players' success rate in guessing the contents of the box after shaking. Whereas all players in Experiment 1 answered correctly on both rounds, only 4/18 (22%) players in Experiment 2 correctly guessed the contents of the box in both rounds; 12/18 (67%) players correctly guessed the contents of the box in only one round; and 2/18 (11%) players guessed incorrectly in both rounds.

Box-Shaking Game

In Experiment 2, the contents of the box in each round were as follows: in the number round, the box could contain either 9, 12 or 16 coins (US nickels); in the shape round, the box could contain either of three solid wooden shapes, a sphere (2in diameter), a cube (2in length) or a cylinder (2in diameter, 2in height).

Analyses and Results

Mean observer accuracy was 65.9%, t(99)=15.30; p<0.0001, d=1.53. Thus, observers could tell, on average, who was shaking for number and who was shaking for shape. Crucially, performance was significantly above chance in all correctness subgroups, including videos from players who answered incorrectly in both rounds; t(99)=16.97, p<.0001, d=1.70. This result suggests that epistemic action understanding does not require the action itself to be successful.

We again verified that video length could not explain the results. Indeed, both groups (number longer than shape, and shape longer than number) produced significantly above-chance performance in observers, t(99)=11.56, p<.0001, d=1.15; t(99)=8.76, p<.0001, d=0.88.

Experiment 3: Same Objects

This experiment controls for the contents of the box itself, by always including the very same objects in the box in both rounds of the box-shaking game.

Box-Shaking Game

Here, the box always contained 20 small wooden cubes (1in length).

In the Number round, players were told that "the box contains some wooden objects of the same size, and we want you to tell us how many objects are in the box. There could be 15, 20, or 25 objects in the box".

In the Shape round, players were told: "the box contains some wooden objects of the same size, and we want you to tell us what the shapes of the objects are. All the objects have the same shape, and that shape could be either cubes, spheres, or cylinders".

Analyses and Results

Mean accuracy was 63.69%, t(98)=8.52; p<.0001, d=0.86. Thus, success in this task goes beyond mere differences in shaking movements afforded by the contents of the box, reflecting epistemic intent *per se*.

We once again verified that video length could not explain the results. Indeed, both groups (number longer than shape, and shape longer than number) produced significantly above-chance performance in observers, t(98)=3.81, p<.001, d=0.38; t(98)=7.60, p<.0001, d=0.76.

Experiment 4: Minimal Instructions

This experiment is identical to Experiment 3 except that it dramatically reduces the information given to observers about the box-shaking task.

Experiments 1–3 provided observers with (a) a video reenactment of the instructions given to players; (b) information about the precise quantities and shapes that the players were discriminating between (e.g., informing observers that players had to determine whether there were 9, 12, or 16 coins in Experiment 2); and (c) sample photographs of the box's possible contents, which revealed the size, material, and various other properties of the objects contained inside the box.

By contrast, Experiment 4 completely eliminated (a), (b), and (c), instead providing observers only the following paragraph of instruction:

> *In this experiment, you will see videos of people trying to figure out what's inside a box by shaking it. Each trial will have two videos from the same person: One video in which the person is trying to figure out the **number** of objects in the box, and one video in which the person is trying to figure out the **shape** of the objects in the box. Your job is to figure out which video is which. Which videos show someone shaking for **number** and which videos show someone shaking for **shape**?*

Analyses and Results

Mean accuracy was 63.72%, t(99)=8.54, p<.0001; d=0.85. These results, essentially identical to those of Experiment 3, reveal that leading or heavy-handed instructions to observers are not critical (and may contribute little, if anything) to overall success in this task.

We once again verified that video length could not explain the results. Indeed, both groups (number longer than shape, and shape longer than number) produced significantly above-chance performance in observers, t(99)=5.27, p<.0001, d=0.53; t(99)=7.96, p<.0001, d=0.80.

Experiment 5: Free Responses

Experiment 5 asked whether epistemic action understanding can arise outside of the forced-choice context of Experiments 1–4. Whereas all previous experiments were pre-registered, this experiment was exploratory.

Procedure

Observers saw six videos from Experiment 3 (one Number video and one Shape video from each of three players). Observers were then asked two free-response questions, designed to determine (1) whether the videos were readily interpreted in terms of information-seeking, even without prompting; and (2) whether observers would generate "number" and "shape" as properties of interest, and associate them with the correct videos at rates reliably above chance.

Question 1. After watching all six videos, observers were asked:

> *Why do you think these people were shaking the boxes? What do you think their goal was? Please answer in at least three sentences. This isn't a trick question or anything like that; just say what you think.*

The purpose of Question 1 was to determine whether observers interpreted the box-shaking behaviors in terms of *some* epistemic goal (regardless of what that goal might be), rather than (or in addition to) more pragmatic or instrumental goals.

Question 2. Next, observers were shown the same videos again. The three Number videos had been outlined in blue, and the three Shape videos had been outlined in red. Observers were told that the box contained hidden objects that the players were trying to learn about (though most observers had already reached that conclusion), and that the people in the blue and red videos were asked certain questions about the objects in the box. Observers were then asked:

> *What do you think the people in the* ***blue*** *videos were asked about the objects in the box? You could answer like this: 'They were trying to figure out _____ '*

as well as:

> *What do you think the people in the* ***red*** *videos were asked about the objects in the box? You could answer like this: 'They were trying to figure out _____ '".*

The purpose of this question was to see whether observers would mention number and/or shape in their responses, and if they would be more likely to invoke those properties for the correct videos than for the incorrect videos.

<u>Analysis and Results</u>
As noted in the main text, this experiment was exploratory. However, the approach we took to coding observers' free-response answers was algorithmic, informed by a pilot study. Using responses from that study, we chose key terms and phrases that indicated information-seeking goals (for Question 1) and Number/Shape (for Question 2), then searched the text of the free-response answers for these terms.

Question 1. The phrases we took to indicate information-seeking goals included "determine", "figure out", "find out", "guess", and synonyms of these phrases (more detail about this analysis appears in our data archive). Indeed, a strong majority of observers (75.0%) used one or more of these terms in their free responses. Representative examples include:

- *It looks like they are trying to guess what is in the box. It reminds me of Christmas eve when everyone is trying to see what present they got. Their goal was to correctly guess what is in the box by shaking it around trying to get hints.*
- *They are probably trying to figure out what kind of object is inside of the box.*
- *They were trying to get an idea of what was inside.*
- *I think they are shaking the boxes to figure out what type of object is in it.*
- *Identify what the box may contain, such as coins, some candies, balls, among other things.*
- *They were trying to find out what was inside. They were shaking it to listen while also inspecting the sides. They were hoping to be able to guess.*
- *They are trying to find out what is in the box. It could be how many objects are in the box. It could also be what shape is the object in the box.*

However, 25.0% of observers did not invoke epistemic notions such as these. Representative examples of non-epistemic responses include:

- *To me the people seemed to be carefully shaking the boxes around. Maybe its possible they were trying to mix together whatever was inside. Its like they didnt want to forcefully mix up the contents but do it gently.*
- *There was one thing that kept popping up in my mind. It looks to me like they are shaking cracked eggs. It looks like they are mixing up the egg yolks and whites.*

Question 2. The phrases we took to indicate information-seeking about Number included "number", "how many", "amount", particular numbers, and synonyms of these phrases; the phrases we took to indicate information-seeking about Shape included "shape", "dimension", properties of shapes (e.g., "round"), and synonyms of these phrases (more detail about this analysis appears in our data archive). Indeed, a strong majority of observers (74.0%) mentioned these terms in their free responses about the blue and red videos. Representative examples include:

Number

- *They were trying to figure out how many objects were in the box.*
- *I think maybe the people in the blue videos were told to guess the amount of items in the box.*
- *They were trying to figure out how much was in the box.*
- *They were trying to figure out the number of objects in the box.*

Shape

- *They were trying to figure out what shape the objects are in the box.*
- *they were trying to figure out the shape of items inside the box.*
- *They were trying to figure out what the shape of the objects were by shaking it.*
- *They were trying to figure out the shape of the objects in the box (e.g. cubes or spheres).*

A minority of observers (26.0%) mentioned neither number nor shape in their responses. Representative examples of those answers include:

- *if the object was liquid or solid.*
- *What the object in the box weighs.*
- *how to get the object to come out through the hole by rolling the object around inside the box.*

Of course, some observers mentioned shape for the Number videos, and some observers mentioned number for the Shape videos (just as some observers answered incorrectly under the forced-choice conditions of Experiments 1–4). However, and crucially for our purposes, observers were much more likely than chance to invoke number for the Number videos and shape for the Shape videos, as revealed by a chi-square test over the frequencies of each response for each video: $\chi 2(1, N=96)=16.23$, $p<.0001$). Thus, observers demonstrated sensitivity to the particular content of players' epistemic goals, even under free-response questioning.