



# Performance variability and affordance perception: practice effects on perceptual judgments for walking versus throwing

Christina Hospodar<sup>1</sup> · John Franchak<sup>2</sup> · Karen Adolph<sup>1</sup>

Received: 28 April 2023 / Accepted: 1 July 2023

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

## Abstract

To judge whether an action is possible, people must perceive “affordances”—the fit between features of the environment and aspects of their own bodies and motor skills that make the action possible or not. But for some actions, performance is inherently variable. That is, people cannot consistently perform the same action under the same environmental conditions with the same level of success. Decades of research show that practice performing an action improves perception of affordances. However, prior work did not address whether practice with more versus less variable actions is equally effective at improving perceptual judgments. Thirty adults judged affordances for walking versus throwing a beanbag through narrow doorways before and after 75 practice trials walking and throwing beanbags through doorways of different widths. We fit a “success” function through each participant’s practice data in each task and calculated performance variability as the slope of the function. Performance for throwing was uniformly more variable than for walking. Accordingly, absolute judgment error was larger for throwing than walking at both pretest and posttest. However, absolute error reduced proportionally in both tasks with practice, suggesting that practice improves perceptual judgments equally well for more and less variable actions. Moreover, individual differences in variability in performance were unrelated to absolute, constant, and variable error in perceptual judgments. Overall, results indicate that practice is beneficial for calibrating perceptual judgments, even when practice provides mixed feedback about success under the same environmental conditions.

**Keywords** Affordance · Variability · Practice · Performance · Perception · Judgment

## Introduction

### Effects of practice on perceiving affordances

A popular adage in skill learning is “practice makes perfect.” But what kind of practice? Successful motor performance requires people to perceive possibilities for action—what Gibson (1979) termed “affordances”—by detecting moment-to-moment relations between the environment and their own bodies and motor skills (Chemero 2003; Stoffregen 2003; Warren 1984). For example, to shoot a basketball through

the hoop, players must accurately perceive the relations between features of the environment (distance and angle to the hoop, location and movements of other players, etc.) and properties of their body and skills (height, strength, agility, fatigue, etc.). The ability to accurately judge what one can do depends on immediately available perceptual information (e.g., eye height, balance control) generated by exploratory movements. If the critical information is not immediately available, then further exploration is needed. Practice performing the action may be required because action performance generates rich and varied information that can calibrate perception (Franchak 2017, 2020). Although it makes intuitive sense that practice leads to improvements in performance, many everyday actions are relatively rare and unpracticed, yet require accurate perception of affordances nonetheless (e.g., slipping through a narrow doorway). Moreover, actions range in their inherent variability (e.g., throwing a ball through a narrow doorway is likely more variable than walking through a narrow doorway), but effects of performance variability on perceptual judgments

---

Communicated by Melvyn A. Goodale.

✉ Karen Adolph  
karen.adolph@nyu.edu

<sup>1</sup> Department of Psychology, New York University, New York, NY, USA

<sup>2</sup> Department of Psychology, University of California, Riverside, CA, USA

are unknown. Is practice equally effective for improving perception of affordances for more and less variable actions?

Experimental manipulation of practice variability did not affect people's perception of affordances for walking through narrow doorways while holding a horizontal bar or while rolling themselves in a manual wheelchair (Yasuda et al. 2014). Participants' judgments were equally accurate after practice with doorways near the limits of their abilities as after practice with doorways considerably larger and smaller than the limits of their abilities. A similar experimental manipulation of practice variability for walking sideways through doorways while wearing a backpack did not affect participants' perception of affordances (Franchak and Somoano 2018). However, prior work did not consider that different actions might inherently produce different levels of performance variability, and thus lead to natural differences in the variability of practice—with outcomes sometimes successful and sometimes not under the same body–environment conditions.

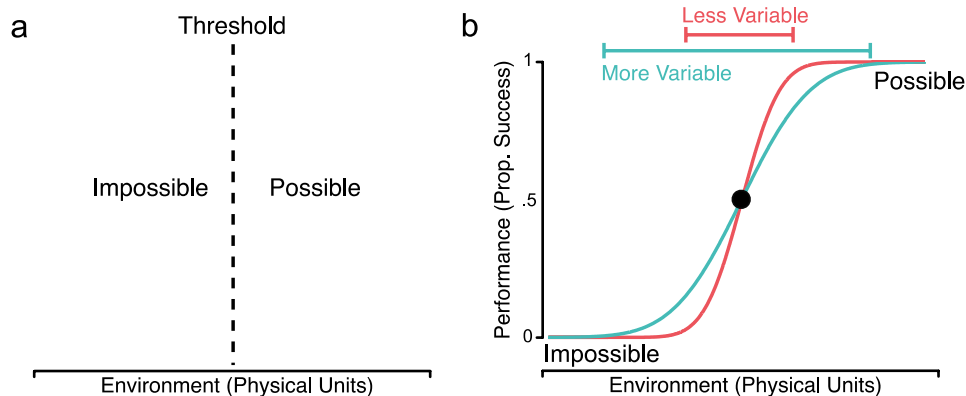
### Measuring performance variability

Performance variability is important because the same action performed in the same environment does not always yield the same outcome. Even professional basketball players with years of concerted practice are never perfectly consistent in even in the most consistent aspects of the game—shooting free-throws. Indeed, top professionals in the National Basketball Association make free-throws only 90% of the time, and other players score only 40% of free-throws (Sports-Reference-LLC 2023).

However, many researchers treat affordances as binary categories—as if actions were invariably possible or impossible at each environmental increment (Warren

1984; Warren and Whang 1987). Thus, the affordance is marked by a threshold, boundary, or *pi* ratio, such that performance outcome has no variability at any environmental unit (Fig. 1A). The “threshold” is the transition point that marks the boundary at which an action shifts from possible to impossible, that is, from always to never successful, and the slope of the function (i.e., performance variability) is assumed to be 0. With the binary approach, researchers characterize the affordance based on the upper or lower limits of participants' abilities (e.g., the farthest distance at which they can land a basket, the narrowest doorway they can walk through).

But affordances are likely better described by continuous, sigmoid functions such that performance outcomes vary probabilistically across repeated trials at the same increment (Franchak and Adolph 2014a; Fig. 1B). On an open basketball court, success for professionals is nearly perfect right next to the hoop, and becomes increasingly less likely at increasingly farther distances, until it eventually drops to zero. With the probabilistic approach, researchers characterize the affordance based on the consistency of participants' success (i.e., the limit of what participants can do at some level of consistency, say 50% success as in Fig. 1B). In this case, the threshold is an estimated interval where performance outcome decreases from one point to another along the inflection of the function. Performance variability can be measured by the slope of the success function, such that steeper slopes (pink slope in Fig. 1B) indicate variable outcomes across a narrower range of increments, and flatter slopes (blue slope in Fig. 1B) indicate variable outcomes across a wider range of increments. Of course, the threshold provides useful information, because a person might be consistently good or bad at performing an action such as shooting unimpeded free-throws. But for two players with the



**Fig. 1** Treatment of affordances. **a** Binary categories. The threshold divides the environment into regions where actions are uniformly impossible (0% success) or possible (100% success), such that performance outcomes have no variability at any environmental unit. **b** Probabilistic functions. Success is variable along the curve, and the

threshold is an estimated interval along the inflection of the function (here, 50%). Performance variability can be estimated based on the slope of the curve, with steeper slopes (pink curve) indicating less variable performance and flatter slopes indicating more variable performance (blue curve)

same shooting threshold, the one with lower performance variability is the better shot.

### Are performance variability and judgment error related?

The evidence linking performance variability with perceptual judgments is mixed. For example, adults show lower performance variability while walking sideways through narrow doorways compared to walking sideways along narrow ledges, and their perceptual judgments are similarly less variable for doorways than ledges (Comalli et al. 2013). However, falling off the ledge has a larger penalty for error than getting stuck in a doorway, so both performance and judgment variability could reflect participants' risk tolerance rather than a true relation between performance variability and perceptual judgments. Likewise, adults' gait modifications are more variable while walking sideways through narrow doorways compared to ducking under overhead barriers, and their perceptual judgments about passable openings are similarly more variable for horizontal doorways (Franchak et al. 2012). However, it is unclear whether differences in performance variability drive differences in perceptual judgments, or if judging possibilities for passage through horizontal openings is perceptually more difficult than judging passage through vertical openings.

Moreover, error in perceptual judgments can be characterized in different ways (Wagman et al. 2001). Absolute error characterizes deviation from the performance threshold without considering the direction of the error (i.e., overly optimistic versus under confident). Constant error takes the direction of the error into account, despite the risk that summary scores can be misleading if the magnitude of constant error is cancelled due to differences in direction. Variable error measures the consistency of judgments. The three types of error provide different information and can have different relations. For example, judgments can be accurate with low variability (typically the best case scenario as when all the darts cluster around the bullseye), accurate with high variability (scattered around the bullseye), inaccurate with low variability (clustered around a point distant from the bullseye), or inaccurate with high variability (scattered around a point distant from the bullseye). Possibly, different aspects of perceptual judgments may be affected by the underlying variability of the action.

### Current study

Practice with novel walking and throwing tasks reduces error in perceptual judgments (Franchak 2017; Franchak and Adolph 2014b; Franchak and Somoano 2018; Franchak et al. 2010; Labinger et al. 2018; Zhu and Bingham 2010). But

what happens when practice is variable due to inherent variability in performance? Do people's perceptual judgments improve equally well after practicing actions with more and less variable outcomes?

We designed two tasks that eliminated potential confounds in prior work by equating penalties and task dimensions across tasks and separating perceptual judgments from measures of performance in the procedure. We measured performance variability as participants walked sideways and threw a beanbag through a doorway varying in width, in both cases without touching the edges of the doorway. To test effects of practice on judgment error, participants judged passable doorways before and after performing each task.

We expected the throwing task to be more variable for two reasons. For walking, body size is the primary limiting factor; whereas for throwing, action mechanics are the limiting factor (Fajen 2007). Since a person's action mechanics are more variable than their body size, throwing should lead to more variable performance. Moreover, people likely have more relevant experience with walking than throwing—everyday life occasionally obliges sideways navigation through tight spaces (e.g., sidling between chairs in a crowded classroom or slipping between closing subway doors), but rarely entails underhand throwing (apart from avid cornhole players).

Our primary aim was to test whether people's perceptual judgments improve after short-term practice performing the walking and throwing tasks, and if so, whether effects of practice differ by task. Based on prior work, we predicted that judgment error would decrease after practice for both tasks, but we were agnostic about whether practice effects would differ by task. Natural differences in performance variability could be beneficial, detrimental, or neutral for perceptual judgments.

Our secondary aim was to test relations between performance variability and the three types of judgment error within and between tasks. We predicted that people would display larger performance variability and larger judgment error for throwing compared to walking. We predicted that, within tasks, performance variability would correlate with variable judgment error, but we were agnostic about the remaining inter-correlations.

## Methods

### Preregistration and data sharing

Prior to data collection, we preregistered the procedure and analysis plan on AsPredicted.com (<https://aspredicted.org/du9zh.pdf>). "Judgment variability" in the preregistration is now termed "variable error" for clarity. As outlined in the preregistration, confirmatory analyses consisted of paired

*t*-tests comparing walking vs. throwing for performance variability and judgment error (absolute error and variable error); one correlation per task (performance variability  $\times$  variable error); and two two-way ANOVAs (pre- vs. posttest, walking vs. throwing) on judgment error (absolute error and variable error). For all outcome measures, we report results with and without outliers defined by the interquartile range (IQR);  $Q1 - 1.5 \times IQR$  or  $Q3 + 1.5 \times IQR$ . Additional analyses (e.g., effects of practice variability on constant error) were exploratory.

Power analyses included in our preregistration (input parameters: effect size ( $d$ ,  $\rho$ ) = 0.5, alpha (one-sided) = 0.05, power = 0.80) indicated that a sample of 27 participants was sufficiently powered to detect medium effect sizes, as estimated based on the results of Franchak et al. (2012); we recruited 30 participants to exceed a medium effect size.

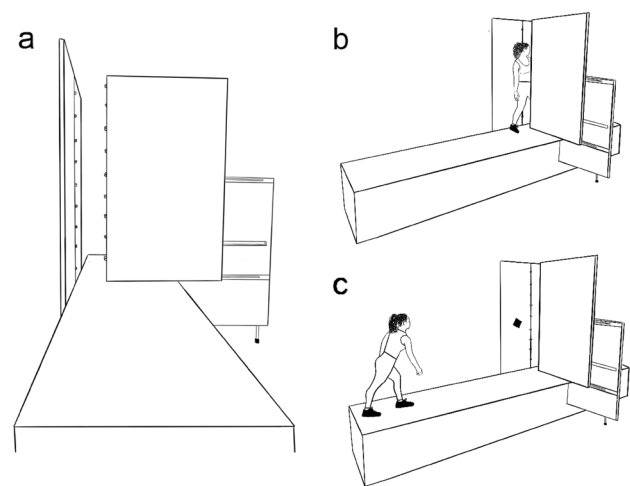
With participants' permission, videos of each session are openly shared with authorized investigators in the Databrary digital library (<https://nyu.databrary.org/volume/1448>). Exemplar video clips, flat file processed data, and analysis code are publicly available in the Databrary volume.

## Participants and procedure

We recruited 30 participants (18 women, 12 men) from the New York City area through word-of-mouth ( $M$  age = 25.7 years,  $SD$  = 3.3 years.). Participants reported their race (14 Asian, 1 Black or African-American, 1 Hawaiian or other Pacific Islander, 12 White, 2 chose not to answer), ethnicity (6 Hispanic/Latinx, 24 not Hispanic/Latinx), and highest level of education completed (3 high school diploma, 15 Bachelor's degree, 12 Master's degree). All participants were right-handed and had normal or corrected-to-normal vision. Participants received \$20 for participation. All followed our instructions and completed both tasks. All participants provided informed consent prior to participation.

### Adjustable doorway

As shown in Fig. 2 and in the exemplar videos (<https://nyu.databrary.org/volume/1448>), participants performed both tasks on a raised wooden walkway (4.9 m long  $\times$  1.0 m wide  $\times$  0.6 m high). We created an adjustable doorway (located  $\sim$  3 m from the start of the walkway) from a stationary wall (1.2 m long  $\times$  1.7 m high) attached to one side of the walkway and an adjustable perpendicular wall (1.1 m wide  $\times$  1.9 m high) attached to the other side of the walkway. The moving wall adjusted in 0.1-cm increments, creating doorways that varied from 0 to 74 cm in width. To more easily determine when participants erred by touching the sides of the doorway with their body or the beanbag and to increase the performance variability of both tasks,



**Fig. 2** Walking and throwing tasks. **a** Participant's view. A sliding, rectangular panel perpendicular to a stationary wall created doorways 0- to 74-cm wide. The doorway (edge of sliding panel and middle of stationary wall) was lined with small bells. In both tasks, participants' goal was to pass through the doorway (by walking or throwing a beanbag) without touching the edges of the doorway and thereby causing the bells to jingle. **b** Walking task. Participants started 2 m from the doorway and walked through the doorway by turning sideways as they neared the opening (facing away from the stationary wall). **c** Throwing task. Participants stood 2 m from the doorway and threw the beanbag underhand through the doorway

we sewed small “jingle bells” (20 cm apart) on flat elastic ribbons and attached the ribbons to the sides of the doorway as in Franchak (2020). One ribbon wrapped fully around the stationary door and the other ribbon wrapped fully around the perpendicular moving door, such that the bells fully lined the edges of the doorway (Fig. 2A). Small pieces of cardboard under the ribbons at the top and bottom of the doorways enabled the ribbon and bells to vibrate with a light touch. The back wall was white and uniform so that participants had no extraneous visual cues about the width of the doorway.

### Video recording

To set the size of the doorway for each trial and to record the size of the doorway on video, a small webcam mounted to the moving doorway recorded a ruler fixed to the doorway. Two fixed cameras recorded participants from opposite ends of the walkway, one from the perspective of the participant facing the doorway and the other from the landing platform. A fixed camera mounted on the ceiling recorded an overhead view. All four camera views (see exemplar videos, <https://nyu.databrary.org/volume/1448>) were mixed into a single frame and digitally captured by a computer using a video capture card.

## Walking and throwing tasks

In both tasks, participants stood approximately 2 m from the adjustable doorway. In the walking task, participants walked forward for about 1.5 m, turned sideways with their back to the stationary wall, and then walked sideways through the doorway (Fig. 2B and exemplar video for the walking task, <https://nyu.databrary.org/volume/1448>). After moving past the doorway, the experimenter widened the doorway and participants returned to the starting location. Participants were instructed to maintain the same pace across all trials.

In the throwing task, participants threw the beanbag underhand, but could hold the beanbag, however, they wished (Fig. 2C and exemplar video for the throwing task, <https://nyu.databrary.org/volume/1448>). We used regulation “cornhole” bags (15 × 15 cm, ~450 g). An assistant collected bean bags between trials, and resupplied bean bags when participants ran out.

## Procedure for perceptual judgments and practice trials

We first used the apparatus to measure participants’ front-to-back body width. They stood with their back flat against the stationary wall and pulled the moving door toward themselves until it touched their body (disregarding the bells along the sides of the doorway). We recorded body width using the ruler webcam on the doorway.

Task order (walking or throwing) was counterbalanced across participants and gender. Exemplar videos (<https://nyu.databrary.org/volume/1448>) show the procedure for each task. *Pretest perceptual judgments*: For the first task, participants first completed eight pretest judgment trials using the method of adjustment: For each trial, the experimenter slowly opened (four trials) or closed (four trials) the door until participants told her to stop at the width they thought they could perform successfully 50% of the time. *Practice trials*: Next, participants completed 75 practice trials for the same task. For both tasks, participants were instructed to attempt every trial, no matter the width of the doorway. After each trial, the research assistant provided feedback by declaring aloud the outcome of the trial as “yes” (successful passage without touching the edge of the doorway causing the bells to jingle) or “no” (unsuccessful passage; for the throwing task, hitting the stationary wall before passing through the doorway also counted as a failure). Doorway width varied systematically according to an adaptive, computer-based staircase procedure using the Escalator ([https://github.com/JohnFranchak/escalator\\_toolbox](https://github.com/JohnFranchak/escalator_toolbox)) and Palamedes (<https://palamedestoolbox.org>) toolboxes. The software determined the doorway width for each trial based on the outcome (successful or not) of the prior trial, and the experimenter set the doorway to the correct width. *Posttest perceptual judgments*: Finally, participants completed eight

posttest judgment trials, using the same procedure as in pretest. The entire procedure was repeated for the second task (8 pretest judgment trials, 75 performance trials, 8 posttest judgment trials).

## Outcome variables

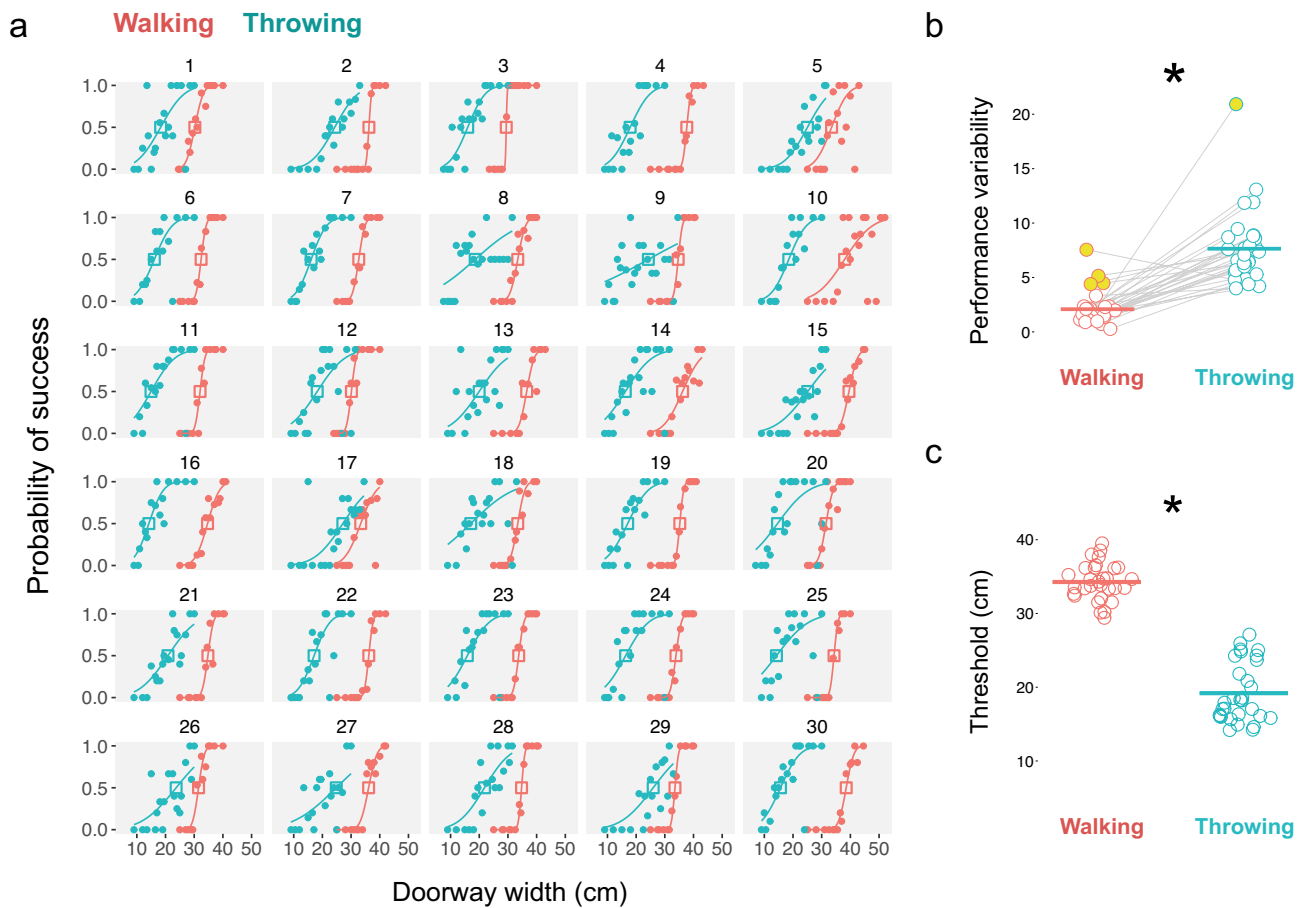
All outcome variables were computed for each participant in each task. As outlined in the preregistration, we fit a “success function” (using the *quickpsy* package in *R*) through the performance trials for each participant in each task, as described in Franchak and Adolph (2014a). We measured *performance variability* as the slope of the curve fit, with flatter slopes indicating greater performance variability, and steeper slopes indicating lower performance variability. We estimated the *threshold* of each curve fit based on the doorway width at which the probability of success was 50%. We used three different measures of judgment error. We calculated *absolute judgment error* based on the absolute difference between the participant’s threshold and their mean pretest or posttest judgment for each task. We calculated *constant judgment error* based on the signed difference between the participant’s threshold and their mean pretest or posttest judgment for each task. We calculated *variable judgment error* based on the standard deviations of each block of pretest and posttest judgment trials.

## Results

As described in our preregistration, we report outliers and analyzed the data with and without the outliers. For brevity, we only reported outlier analyses that changed the pattern of results. We report effect sizes for *t*-tests (Cohen’s  $d > 0.5$  = medium effect size;  $d > 0.8$  = large effect size) and for ANOVAs ( $\eta_p^2 > 0.06$  = medium effect size;  $\eta_p^2 > 0.14$  = large effect size).

### Throwing performance was more variable than walking performance

Walking and throwing were both relatively challenging tasks to perform—every participant displayed variable performance in each task. All participants (except for #10) displayed greater performance variability for throwing ( $M = 7.6$  cm,  $SD = 3.4$  cm) than for walking ( $M = 2.1$  cm,  $SD = 1.5$  cm), as shown by the slope of the individual curve fits in Fig. 3A, the group data in Fig. 3B, and the height of the boxes in Fig. 4A;  $t(29) = 7.81$ ,  $p < 0.001$ ,  $d = 2.14$ . Five participants displayed high-performance variability (4 outliers for walking and 1 for throwing—yellow symbols in Fig. 3B). Removing the outliers did not change the direction or significance of the results.



**Fig. 3** Performance variability. **a** Success at walking and throwing. Each panel shows one participant's data for walking (pink) and throwing (blue). Curve fits are based on sigmoid functions for each task. For all participants (except #10), performance variability was greater for throwing (flatter functions) than walking (steeper functions). Open symbols denote walking and throwing thresholds (50% probability of success). **b** Performance variability for walking and

throwing. Each pair of symbols is one participant's variability (slope of curve fits). Yellow symbols denote outliers. Horizontal lines show condition means. Asterisk denotes  $p < 0.001$ . **c** Thresholds for walking and throwing. For every participant, the walking threshold was larger than the throwing threshold. Each symbol is one participant's threshold (50% of success function). Horizontal lines show condition means. Asterisk denotes  $p < 0.001$

Walking required larger openings for passage than throwing as evidenced by larger thresholds for walking ( $M = 34.3$  cm,  $SD = 2.5$  cm) than throwing ( $M = 19.2$  cm,  $SD = 4.0$ );  $t(29) = 19.10$ ,  $p < 0.001$ ,  $d = 4.45$ , Fig. 3C. As predicted, walking thresholds were body-scaled, but throwing thresholds were not. Participants' front-to-back body width was correlated with their walking threshold ( $r = 0.52$ ,  $p = 0.003$ ), but not with their throwing threshold ( $r = 0.12$ ,  $p = 0.53$ ).

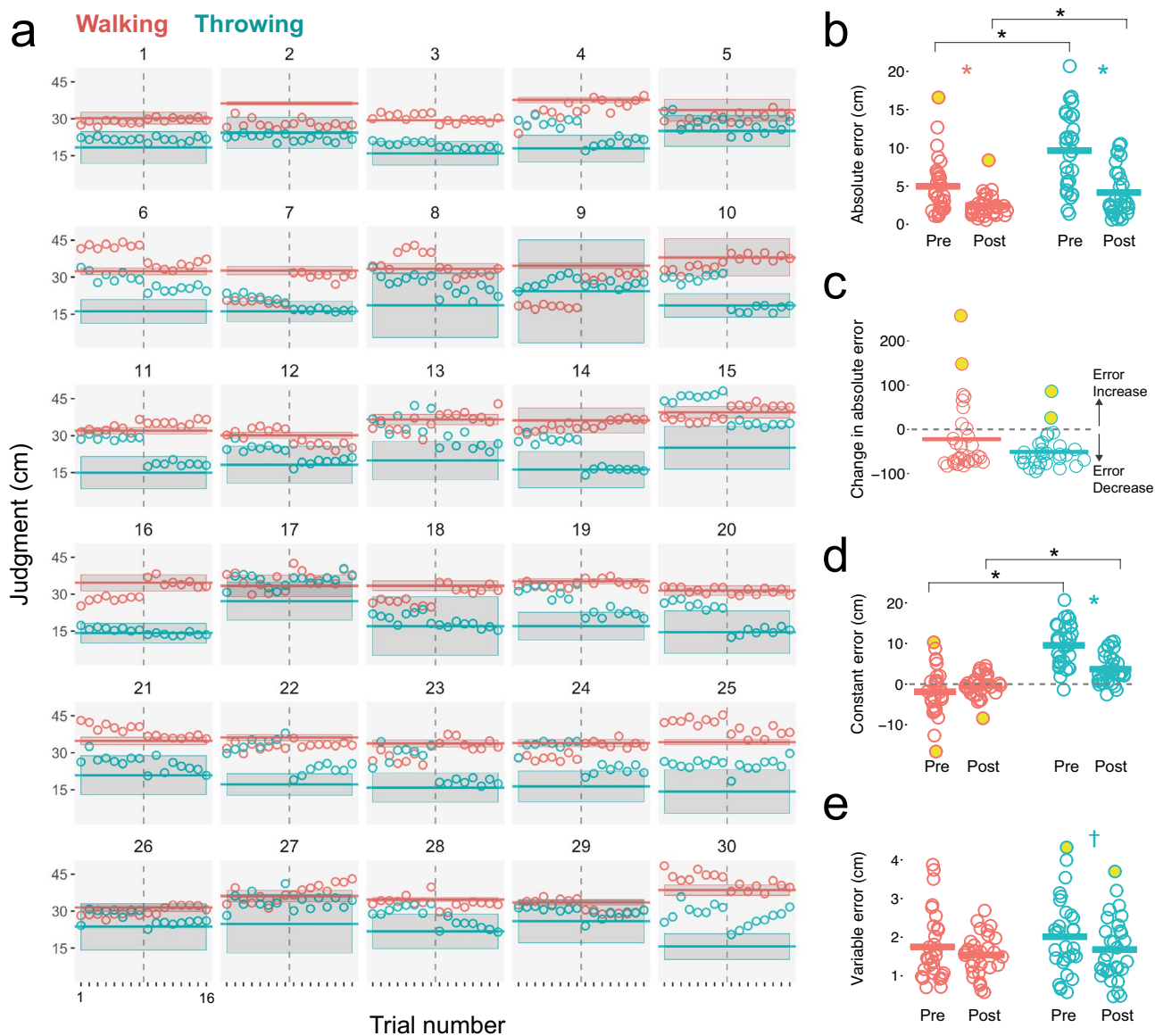
For throwing, thresholds and performance variability were correlated ( $r = 0.41$ ,  $p = 0.025$ ), suggesting that more skilled throwers—who could successfully throw the beanbag through narrower doorways—were also less variable in performance. However, for walking, thresholds and performance variability were not correlated ( $r = 0.29$ ,  $p = 0.12$ ), nor was normalized walking threshold (threshold normalized to body width) correlated with performance variability

( $r = 0.26$ ,  $p = 0.16$ ). That is, participants who could walk through smaller doorways were not less variable in their walking performance. Removing outliers did not change the pattern of results.

Performance variability was not correlated across tasks ( $r = -0.13$ ,  $p = 0.5$ ), suggesting that it is not driven by participants' overall motor control or motor planning abilities. Removing outliers did not change the results ( $r = 0.22$ ,  $p = 0.29$ ).

### Absolute error: greater for throwing, but reduced proportionally in both tasks

Absolute judgment error was greater for throwing than walking at both pretest and posttest, and practice reduced absolute error for both tasks (Fig. 4). A two-way repeated measures ANOVA on absolute error showed main effects



**Fig. 4** Perceptual judgments. **a** Judgments and performance at pretest and posttest for walking and throwing. Each panel shows one participant's data at each trial (symbols) for walking (pink) and throwing (blue) at pretest (trials 1–8) and posttest (trials 9–16). Vertical dashed line separates pretest from posttest. Horizontal lines indicate walking (pink) and throwing (blue) thresholds. Boxes indicate performance variability (slope of function in Fig. 2A) for walking (outlined in pink) and throwing (outlined in blue). **b** Absolute judgment error. Absolute error was larger for throwing than walking at both pretest and posttest. Error decreased from pretest to posttest for both tasks, but to a greater extent for throwing. Asterisks denote  $p < 0.01$ . **c**

Change in absolute error. Change in absolute error  $\left(\frac{\text{posttest error} - \text{pretest error}}{\text{pretest error}} \times 100\right)$  was similar for walking and throwing, particularly after excluding outliers. **d** Constant error. Constant error was greater for throwing than walking at both pretest and posttest. Error decreased from pretest to posttest only for throwing. Participants were overly optimistic about their walking abilities but under confident about their throwing abilities; that is, they choose doorways smaller than their ability for walking and doorways larger than their ability for throwing. Asterisks denote  $p < 0.001$ . **e** Variable error. Variable error was similar in both tasks at pretest and posttest. Sword denotes  $p = 0.06$

for task,  $F(1, 29) = 15.0, p < 0.001, \eta_p^2 = 0.34$ , and phase,  $F(1, 29) = 68.0, p < 0.001, \eta_p^2 = 0.70$ , and an interaction between task and phase,  $F(1, 29) = 8.2, p < 0.001, \eta_p^2 = 0.22$ , Fig. 4B. Paired  $t$ -tests with Bonferroni adjustments confirmed that at pretest, participants displayed larger errors for throwing ( $M = 9.6$  cm,  $SD = 5.1$ ) than

walking ( $M = 5.0$  cm,  $SD = 3.6$ );  $t(29) = 3.77, p < 0.001, d = 1.06$ . Likewise, at posttest, participants displayed larger errors for throwing ( $M = 4.2$  cm,  $SD = 3.2$ ) than walking ( $M = 2.5$  cm,  $SD = 1.6$ );  $t(29) = 2.84, p = 0.008, d = 0.66$ , suggesting they were better calibrated to walking than throwing. Moreover, although absolute error decreased at

posttest for both tasks,  $t(29) > 3.85$ ,  $p_s < 0.001$ ,  $d_s > 0.87$ , the decrease was larger for throwing than for walking. Removing outliers in absolute error (yellow symbols in Fig. 4B) did not change the pattern of results.

We found no evidence for order effects. Absolute judgment error for walking at pretest was similar for participants who ended with the walking task compared to participants who started with the walking task,  $t(21.45) = 1.31$ ,  $p = 0.20$ ,  $d = 0.48$ . Likewise, absolute judgment error for throwing at pretest was similar for participants who ended with the throwing task compared to participants who started with the throwing task,  $t(26.72) = 0.28$ ,  $p = 0.78$ ,  $d = 0.10$ . Thus, participants' judgments were unchanged by task order.

In accordance with the larger reduction of absolute errors for throwing, individual differences in error reduction were also more robust for throwing. For only 2 participants (#2, #17, Fig. 4A), absolute error slightly increased from pretest to posttest for throwing, but 8 participants (#2, #11, #12, #17, #20, #27, #28, #29) slightly increased absolute error from pretest to posttest for walking. Pretest and posttest errors were correlated for throwing ( $r = 0.57$ ,  $p < 0.001$ ), but not for walking ( $r = 0.23$ ,  $p = 0.21$ ), indicating consistent throwing errors even after practice. Removing outliers did not change the pattern of results.

Since pretest error was larger for throwing than walking, we compared the percent error change from pretest to posttest between tasks ( $\frac{\text{posttest error} - \text{pretest error}}{\text{pretest error}} \times 100$ ; Fig. 4C). Positive percentages indicate an increase in absolute error whereas negative percentages indicate a decrease, and numbers farther from 0 signify larger changes in either direction. On average, absolute error reduced 22% ( $SD = 79\%$ ) for walking and 51% ( $SD = 38\%$ ) for throwing. Error change did not differ between tasks;  $t(29) = 1.19$ ,  $p = 0.07$ ,  $d = 0.46$ . Removing 2 outliers with large increases in absolute error for walking and 2 outliers for throwing (yellow symbols in Fig. 4C) removed any potential differences between tasks;  $t(25) = 1.43$ ,  $p = 0.17$ ,  $d = 0.40$ . The similarity of change in absolute error between tasks suggests that the interaction revealed by the ANOVA was largely a result of larger error for throwing overall. Thus, participants reduced error in both tasks proportionally, suggesting they learned equally well from practice throwing and walking.

Absolute errors were unrelated across tasks, meaning that errors were a product of the task, not the person. Absolute errors for walking and throwing were not correlated at pretest ( $r = -0.19$ ,  $p = 0.31$ ) or posttest ( $r = 0.18$ ,  $p = 0.34$ ). Removing outliers for walking did not change the result for pretest ( $r = -0.08$ ,  $p = 0.67$ ), but the correlation between walking and throwing at posttest approached significance ( $r = 0.36$ ,  $p = 0.06$ ). Errors were correlated

from pretest to posttest for throwing ( $r = 0.57$ ,  $p < 0.001$ ), but not for walking ( $r = 0.23$ ,  $p = 0.21$ ); removing outliers did not change the result. Change in absolute error was not correlated between tasks ( $r = 0.11$ ,  $p = 0.55$ ), and removing outliers did not change the result.

### Constant error: walking judgments were overly optimistic, whereas throwing judgments were overly conservative

The signed difference between perceptual judgments and affordance thresholds provided a measure of constant judgment error (Fig. 4D). Positive constant error indicates that participants selected openings that were larger than what they really could do (overly conservative about their abilities); negative constant error indicates they selected openings that were smaller than what they really could do (overly optimistic about their abilities).

At pretest, participants were overly optimistic about their walking abilities ( $M = -1.9$  cm,  $SD = 5.9$  cm) and overly conservative about their throwing abilities ( $M = 9.5$  cm,  $SD = 5.3$  cm). At posttest, participants continued to err by selecting too-small openings for walking ( $M = -0.4$  cm,  $SD = 2.8$  cm), and they were still overly conservative about their throwing abilities ( $M = 3.7$  cm,  $SD = 3.7$  cm). The effect was more robust at the individual level for throwing than for walking (Fig. 4D). For throwing, only 1 participant (#2) had a negative constant error at pretest, and he continued to overestimate his throwing abilities at posttest; 3 additional participants (#10, #14, #16) also had negative constant errors at posttest; all other errors were positive (see Fig. 4A). In contrast, for walking, 18 participants had negative constant errors at pretest, and of those, 12 continued to display negative constant errors at posttest with 6 new participants also showing negative constant errors. A two-way repeated measures ANOVA on constant error followed the same pattern as absolute error: main effects for task,  $F(1, 29) = 130.7$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.82$ , and phase,  $F(1, 29) = 14.81$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.34$ , and an interaction between task and phase,  $F(1, 29) = 33.24$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.53$ , Fig. 4D. However, in contrast to absolute error, post hoc tests for constant error did not change from pretest to posttest for walking,  $t(29) = 1.60$ ,  $p = 0.12$ ,  $d = 0.30$ . Likely, error was cancelled due to differences in direction because participants were more likely to estimate both below and above their threshold for walking (Fig. 4A). But constant error differed between walking and throwing at both pretest,  $t(29) = 9.42$ ,  $p < 0.001$ ,  $d = 2.05$ , and posttest,  $t(29) = 7.83$ ,  $p < 0.001$ ,  $d = 1.19$ , and reduced from pretest to posttest for throwing,  $t(29) = 7.72$ ,  $p < 0.001$ ,  $d = 1.23$ . Removing outliers (yellow symbols in Fig. 4D) did not change the pattern of results.



### Variable error: similar between tasks and did not robustly change with practice

Every participant showed variability in judgments at pretest and posttest for both tasks (Fig. 4E). A two-way repeated measures ANOVA on variable error showed a main effect only for phase,  $F(1, 29) = 7.66$ ,  $p = 0.01$ ,  $\eta_p^2 = 0.20$ , not task,  $F(1, 29) = 2.60$ ,  $p = 0.12$ ,  $\eta_p^2 = 0.08$ , and no interaction between task and phase,  $F(1, 29) = 0.30$ ,  $p = 0.59$ ,  $\eta_p^2 = 0.01$ . Post hoc tests were not robust. For walking, variable error at pretest ( $M = 1.75$ ,  $SD = 0.89$ ) did not differ from posttest ( $M = 1.54$ ,  $SD = 0.53$ ),  $t(29) = 1.66$ ,  $p = 0.11$ ,  $d = 0.26$ . For throwing, variable error tended to decrease from pretest ( $M = 2.01$ ,  $SD = 0.97$ ) to posttest ( $M = 1.68$ ,  $SD = 0.82$ ),  $t(29) = 1.93$ ,  $p = 0.06$ ,  $d = 0.38$ . Removing outliers did not change the pattern of results. For walking, 19/30 participants showed reductions in variable error, and for throwing, 22/30 showed reductions (Fig. 4A). Thus, the main effect for phase revealed by the ANOVA is not robust.

Variable error at pretest and posttest were correlated for walking ( $r = 0.64$ ,  $p < 0.001$ ) and throwing ( $r = 0.46$ ,  $p = 0.011$ ), suggesting that participants were consistently variable in their judgments before and after practicing the task. Variable error for walking and throwing were correlated at pretest ( $r = 0.43$ ,  $p = 0.018$ ), but not at posttest ( $r = 0.25$ ,  $p = 0.18$ ). Removing outliers did not change the pattern of results.

Variable error was related to absolute error for throwing, but not for walking (we did not correlate variable and constant error, given the necessary relations between absolute and constant error). For walking, absolute and variable error were not correlated at pretest ( $r = -0.18$ ,  $p = 0.35$ ) or posttest ( $r = 0.08$ ,  $p = 0.66$ ). For throwing, the correlation showed a trend at pretest ( $r = 0.32$ ,  $p = 0.08$ ) and reached significance at posttest ( $r = 0.53$ ,  $p = 0.003$ ). Removing outliers voided the correlation between absolute and variable error for throwing at pretest ( $r = 0.25$ ,  $p = 0.19$ ), but the correlation remained at posttest ( $r = 0.44$ ,  $p = 0.018$ ); walking correlations were unchanged.

Judgments were relatively stable across trials. For walking and throwing at pretest, and for walking at posttest, judgments did not change from the first to last two trials,  $t_s < 0.78$ ,  $p_s > 0.45$ ,  $d_s < 0.07$ . However, for throwing at posttest, judgments drifted upward,  $t(29) = 2.74$ ,  $p = 0.01$ ,  $d = 0.21$ , suggesting that participants' error increased.

### Judgment error was unrelated to performance variability

Performance variability did not relate to any measure of perceptual judgment error. Performance variability was not correlated with pretest or posttest absolute error for walking or throwing ( $r_s < 0.22$ ,  $p_s > 0.25$ ), and removing outliers

did not change the results. Performance variability was also not correlated with change in absolute error ( $r_s < 0.17$ ,  $p_s > 0.36$ ), and removing outliers did not change the results. Similarly, performance variability was not correlated with pretest or posttest constant error for walking or throwing ( $r_s < 0.17$ ,  $p_s > 0.36$ ), and removing outliers did not change the results. In addition, performance variability was not correlated with pretest or posttest variable error for throwing or walking ( $r_s < 0.22$ ,  $p_s > 0.25$ ). Removing outliers resulted in a positive correlation between performance variability and variable error for throwing at posttest,  $r = 0.38$ ,  $p < 0.05$ , but otherwise did not change the pattern of results.

## Discussion

We successfully designed two tasks with different inherent performance variability. Both were relatively novel and unpracticed for participants and both had similar demands (i.e., changing doorway width, successful passage without ringing the bells). Performance was uniformly more variable for throwing compared to walking. However, practice with the more variable throwing task was equally effective at calibrating perceptual judgments as practice with the less variable walking task. Although absolute error in perceptual judgments was larger for throwing at both pretest and posttest, reductions in absolute error were proportional after practice, suggesting that participants calibrated equally well in both tasks. Additional measures of judgment error were informative, suggesting that the three indices of error provide distinct information about perceptual judgments. Based on constant error, judgments for walking were overly optimistic and judgments for throwing were overly conservative. Variable error was similar between tasks and did not show robust changes with practice, suggesting stability in the variability of people's judgments. Moreover, individual differences in performance variability were unrelated to absolute, constant, and variable error in perceptual judgments.

### Why was performance more variable for throwing than for walking?

All actions are variable to some extent because people cannot execute movements in exactly the same way on repeated trials. Thus, variability in performance outcome (e.g., whether the beanbag jingled the bells or not) arises—at least in part—from moment-to-moment differences in the execution of movements. The outcome of throwing may be more variable compared to walking because throwing motions are more variable than walking movements. Throwing, for example, is a “launching” action, whereas walking is not (Cole et al. 2013; Day et al. 2015): Control in throwing ended the moment the beanbag left participants' hands,

whereas participants could continue to adjust their walking movements while passing through the doorway based on continual visual information about the edge of the doorway they were facing. Therefore, movement variability in throwing might have larger consequences on the outcome of the trial. Nonetheless, as Bernstein (1967, 1996) pointed out, people can achieve the same action outcome using very different movements (e.g., landing a basket using a jump shot versus hook shot), so there is no 1:1 correspondence between movement variability and action outcome (Button et al. 2003). Future research should directly assess relations between movement variability and performance variability to determine the extent to which differences in movement variability affect performance variability.

Performance variability was not driven by individual differences in participants' overall level of motor control or motor planning abilities. Performance variability was not correlated across tasks, and walking and throwing thresholds were not correlated. We did not find that some people were uniformly more variable or more skilled than others at performing our tasks. Although performance variability and thresholds were moderately correlated for throwing, they were not correlated for walking, suggesting that variability and skill level is task-specific, rather than a universal relation.

### **Why is more and less variable practice equally beneficial for reducing absolute error in perceiving affordances?**

Our data suggest that any practice is beneficial—at least when considering improvement as a proportion of the initial starting point. Although absolute error in perceptual judgments was nearly twice as large for throwing than walking at pretest, it decreased by half in both tasks after practice. Perhaps the more room for improvement, the more people improve, regardless of the variability of the practiced action.

But why were participants initially better at judging affordances in the walking task? We considered the possibility that outside the lab, people walk more than they throw. However, our tasks called for a level of precision that is not part of everyday experience: participants had to walk or throw without touching the edges of the doorway (jingling the bells)—a demand for precision that is not required for navigating the desks in a crowded classroom, slipping through a closing elevator door, or landing a basket. An alternative possibility is that people may be better calibrated to body-scaled than action-scaled affordances. Attending to the right variables and discriminating the relevant perceptual information—what Gibson termed “education of attention”—may be more difficult for action-scaled affordances (Fajen 2007; Fajen et al. 2008).

### **Why is constant error (e.g., overly optimistic versus overly conservative) task-specific?**

Participants erred in different directions for each task (overly optimistic judgments for walking abilities and overly conservative judgments for throwing). These task differences may arise from the dynamics of the action, the novelty of the action, or the variability of practice. For example, people underestimate their abilities for leaping, arm-swinging, crawling, and stepping, but they underestimate to a much larger degree for launching actions (leaping and arm-swinging) compared to non-launching crawling and stepping actions (Cole et al. 2013). Given that throwing is a launching action, dynamics of the action may influence the magnitude and/or direction of error. Further, in prior work, when asked to walk through a doorway while wearing a pregnancy pack, participants were also conservative about their abilities (whereas pregnant women were not), suggesting that the direction of error may be related to the novelty of the task (Franchak and Adolph 2014b). Practice variability may also affect the direction of judgments—after experiencing induced variability in reaching via a motorized orthopedic elbow brace, people were more conservative with their reachability estimates (Lin et al. 2021). But here, neither task showed a change in the direction of error after practice. Participants reduced the magnitude of the error, but continued to be overly optimistic for walking and overly conservative for throwing. Thus, more research is needed to understand the dynamics of the direction of error in affordance perception.

### **Where does variable error come from and what does it represent?**

Our results suggest that variable error is relatively stable within individuals. Variable error was similar between tasks and practice had no robust effect on variable error, despite large differences in absolute and constant error. Variable error was person-specific, at least before practice: variable error for walking and throwing was correlated at pretest, and variable error was correlated at pretest and posttest for both walking and throwing. Thus, our results indicate stability in the variability of people's perceptual judgments, indicating that people may differ in their sensitivity to doorway width or may use different strategies for judging passable doorways (e.g., intentionally casting a wider net by selecting different doorway widths versus trying to select the same increment repeatedly). Although several decades of work focused on absolute and/or constant error, researchers know little about the sources of variable error.

Furthermore, variable error showed a different pattern than absolute error, suggesting that absolute and variable errors are separable constructs. Absolute error was

different in throwing versus walking at both pretest and posttest, and reduced from pretest to posttest for both throwing and walking. In contrast, variable error did not differ between throwing and walking at pretest or posttest, with only a modest decrease from pretest to posttest for throwing and no decrease for walking. Moreover, absolute and variable errors were unrelated for walking at both pretest and posttest. For throwing, absolute error and variable error were weakly related at pretest, and more strongly related at posttest. Thus, our results suggest that absolute error and variable error provide different information, but their independence may be influenced by the nature of the task.

## Conclusion

Performance is inherently more variable for some actions than others. People cannot consistently throw a beanbag through the same fixed doorway width, but they can repeatedly walk through a fixed doorway width with relatively consistent outcomes. However, despite differences in performance variability, practice is equally effective for both tasks. Absolute error in perceptual judgments improves equally for throwing and walking. Thus, our study has implications for clinical and sport practice: even practice with large variability in performance outcomes is beneficial for calibrating perceptual judgments.

**Acknowledgements** We thank Yasmine Elasmr and Mert Kobas for their help with data collection.

**Author contributions** All the authors contributed to the study concept, design, and interpretation. CH managed the material preparation and data collection. CH and JF conducted the initial data processing. CH conducted the data analyses. CH wrote the manuscript with contributions from JF and KA. All the authors approved the final version of the manuscript for submission.

**Funding** This research was supported by a National Institute of Child Health and Human Development Grant (R01-HD033486) to Karen Adolph and a National Institute of Child Health and Human Development Grant (F31-HD107999) to Christina Hospodar.

**Data availability** With participants' permission, videos of each session are openly shared with authorized investigators in the Databrary digital library (<https://nyu.databrary.org/volume/1448>). Exemplar video clips, flat file processed data for analyses, and the code for the analyses are publicly available in the Databrary volume.

## Declarations

**Conflict of interest** The authors declare no conflicts of interest.

**Ethical approval** All the study procedures were approved by New York University's IRB (NYU IRB FY2016-825).

## References

- Bernstein NA (1967) The coordination and regulation of movements. Pergamon Press, Oxford
- Bernstein NA (1996) On dexterity and its development. In: Latash ML, Turvey MT (eds) Dexterity and its development. Erlbaum, Mahwah, pp 3–244
- Button C, Macleod M, Sanders R, Coleman S (2003) Examining movement variability in the basketball free-throw action at different skill levels. *Res Q Exerc Sport* 74:257–269
- Chemero A (2003) An outline of a theory of affordances. *Ecol Psychol* 15:181–195
- Cole WG, Chan GLY, Vereijken B, Adolph KE (2013) Perceiving affordances for different motor skills. *Exp Brain Res* 225:309–319
- Comalli DM, Franchak JM, Char A, Adolph KE (2013) Ledge and wedge: younger and older adults' perception of action possibilities. *Exp Brain Res* 228:183–192. <https://doi.org/10.1007/s00221-013-3550-0>
- Day BM, Wagman JB, Smith PJK (2015) Perception of maximum stepping and leaping distance: stepping affordances as a special case of leaping affordances. *Acta Physiol* 158:26–35
- Fajen BR (2007) Affordance-based control of visually guided action. *Ecol Psychol* 19:383–410
- Fajen BR, Riley MA, Turvey MT (2008) Information, affordances, and the control of action in sport. *Int J Sport Psychol* 40:79–107
- Franchak JM (2017) Exploratory behaviors and recalibration: what processes are shared between functionally similar affordances? *Atten Percept Psychophys* 79:1816–1829
- Franchak JM (2020) Calibration of perception fails to transfer between functionally similar affordances. *Q J Exp Psychol* 73:1311–1325
- Franchak JM, Adolph KE (2014a) Affordances as probabilistic functions: implications for development, perception, and decisions for action. *Ecol Psychol* 26:109–124
- Franchak JM, Adolph KE (2014b) Gut estimates: pregnant women adapt to changing possibilities for squeezing through doorways. *Atten Percept Psychophys* 76:460–472
- Franchak JM, Somoano FA (2018) Rate of recalibration to changing affordances for squeezing through doorways reveals the role of feedback. *Exp Brain Res* 236:1699–1711
- Franchak JM, van der Zalm D, Adolph KE (2010) Learning by doing: action performance facilitates affordance perception. *Vision Res* 50:2758–2765
- Franchak JM, Celano EC, Adolph KE (2012) Perception of passage through openings cannot be explained geometric body dimensions alone. *Exp Brain Res* 223:301–310
- Gibson JJ (1979) The ecological approach to visual perception. Houghton Mifflin, Boston
- Labinger E, Monson JR, Franchak JM (2018) Effectiveness of adults' spontaneous exploration while perceiving affordances for squeezing through doorways. *PLoS ONE* 13(12):e0209298
- Lin LPY, Plack CJ, Linkenauger SA (2021) The influence of perceptual-motor variability on the perception of action boundaries for reaching in a real-world setting. *Perception* 50:783–796
- Sports-Reference-LLC (2023) <https://www.basketball-reference.com>. In: Basketball statistics and history. Accessed 1 Apr 2023
- Stoffregen TA (2003) Affordances as properties of the animal-environment system. *Ecol Psychol* 15:115–134
- Wagman JB, Shockley K, Riley MA, Turvey MT (2001) Attunement, calibration, and exploration in fast haptic perceptual learning. *J Mot Behav* 33:323–327
- Warren WH (1984) Perceiving affordances: visual guidance of stair climbing. *J Exp Psychol Hum Percept Perform* 10:683–703
- Warren WH, Whang S (1987) Visual guidance of walking through apertures: body-scaled information for affordances. *J Exp Psychol Hum Percept Perform* 13:371–383

- Yasuda M, Wagman JB, Higuchi T (2014) Can perception of aperture passability be improved immediately after practice in actual passage? Dissociation between walking and wheelchair use. *Exp Brain Res* 232:753–764
- Zhu Q, Bingham GP (2010) Learning to perceive the affordance for long distance throwing: smart mechanism or function learning? *J Exp Psychol Hum Percept Perform* 36:862–875

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.