# Meta-strategy learning in physical problem-solving: the effect of embodied experience

Kelsey R. Allen*
Department of Brain and Cognitive Sciences, MIT and Center for Brains, Minds, and Machines

Kevin A. Smith*
Department of Brain and Cognitive Sciences, MIT and Center for Brains, Minds, and Machines

Laura-Ashleigh Bird
Department of Psychology, Durham University

Joshua B. Tenenbaum
Department of Brain and Cognitive Sciences, MIT and Center for Brains, Minds, and Machines

Tamar R. Makin†
Institute of Cognitive Neuroscience, University

Dorothy Cowie†
Department of Psychology, Durham University

'Embodied cognition' suggests that our experience in our bodies – including our motor experiences – shape our cognitive and perceptual capabilities broadly. Much work has studied how differences in the physical body (either natural or manipulated) can impact peoples' cognitive and perceptual capacities, but often these judgments relate directly to those body differences. Here we focus instead on how natural embodied experience affects what kinds of abstract physical problem-solving strategies people use in a virtual task. We compare how groups with different embodied experience – children and adults with congenital limb differences versus those born with two hands – perform on this task, and find that while there is no difference in overall accuracy or time to complete the task, the groups use different *meta-strategies* to come to solutions. Specifically, both children and adults born with limb differences take a longer time to think before acting, and as a result take fewer overall actions to reach solutions to physical reasoning problems. Conversely, the process of development affects the particular actions children use as they age regardless of how many hands they were born with, as well as their persistence with their current strategy. Taken together, our findings suggest that differences in embodied experience drive the acquisition of different meta-strategies for balancing acting with thinking, deciding what kinds of actions to try, and deciding how persistent to be with a current action plan.

Everyday experience is both constrained and enabled by the bodies we inhabit. Taller people can reach further, while stronger people can lift heavier objects without assistance. 'Embodied cognition' (Wilson, 2002) suggests that such constraints play a fundamental role in shaping our cognitive and perceptual experiences broadly. Many versions of embodiment theory suggest that these effects reach beyond the types of experiences we have (e.g., how imagining a scene is intimately tied to our visual experiences), and into the way that we reason about those experiences. In support of this view, researchers have shown that by manipulating people's bodies and the actions they have access to, they can similarly manipulate their perceptual (Makin, Wilf, Schwartz, & Zohary, 2010; Aglioti, Cesari, Romani, & C., 2008; Maimon-Mor, Johansen-Berg, & Makin, 2017) and cognitive capacities (Hagura, Haggard, & Diedrichsen, 2017). Here we take a different approach: we ask whether a *lifetime* of differences in embodied experience can affect the ways that people think about and plan to act in the world, even when their capabilities for action are made equal in the current moment.

Researchers have studied the effects of embodied experience on cognition by investigating the perceptual and motor capabilities of individuals born with limb differences, as those individuals have lifelong differences in the way that they can interact with artifacts and objects relative to people born with two hands. However, the tasks used to study these capabilities often require judgments related to absent body parts, and therefore differences in behavior might be driven by differences in available information. For example, while people with limb differences are slower to judge whether a picture is of a left or right hand (Maimon-Mor, Schone, Moran, Brugger, & Makin, 2020), other studies highlight that

Correspondence should be addressed to Kelsey R. Allen (krallen@mit.edu). ∗ indicates KRA and KAS contributed equally to this work. † indicates TRM and DC contributed equally to this work.

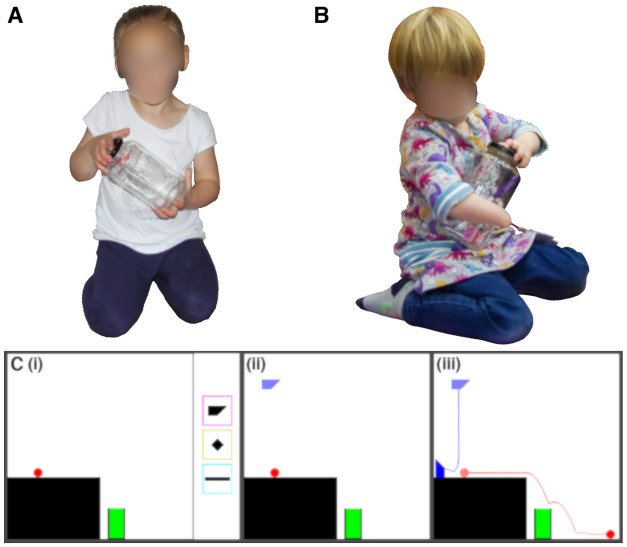


*Figure 1.* Having fewer than two hands changes the cost of interacting with everyday objects, like opening a jar. With two hands, opening a jar can be accomplished by using one hand to stabilize the jar while the other one twists the lid (**A**). With a single hand (**B**), opening a jar can be accomplished by using one's arm and torso to stabilize the jar. The Virtual Tools game (**C**) equalizes action costs for individuals with different types of limbs by creating a virtual action space. (i) A participant selects a tool from the right-hand side of the screen and places it somewhere in the scene (ii). Once placed, physics is "turned on" and the participant can see the result of their action (iii); the blue line represents the observed object motion trajectories).

individuals who were born without any hands show similar response biases as those born with two hands, despite having profoundly different embodied experiences (Vannuscorps & Caramazza, 2016; Vannuscorps, Pillon, & Andres, 2012). People with missing hands, however, lack first-person experience of that hand, and therefore have a subset of the relevant experience that people with both hands do.

We are interested in whether long-term differences in embodiment might affect more general cognitive capabilities, even when the capabilities tested are divorced from particular body differences. We therefore study differences in "meta-strategies" for action, such as how persistent to be (Leonard, Lee, & Schulz, 2017), how to adapt motor plans to one's own levels of motor variability and uncertainty (Harris & Wolpert, 1998; Gallivan, Chapman, Wolpert, & Flanagan, 2018), or how to navigate the trade-offs between planning and acting (Dasgupta, Smith, Schulz, Tenenbaum, & Gershman, 2018). Studies on these meta-strategies suggest that people make action decisions (both implicit and explicit) based on the expected costs and benefits of those actions (Gallivan et al., 2018). Thus, if people with congenital limb differences have learned that actions are in general more costly – because, perhaps, it is more difficult to use artifacts designed for people with two hands (see Figure 1 for an example) – we might expect that they will differ in how they approach physical problems generally, even when everyone is placed in a situation where action costs are equated.

To test the influence of embodied experience on meta-strategy learning, we studied behavior in a *virtual* physical problem-solving task where all participants had equal capabilities to interact with the world, despite having different real-world embodied experience. We chose participant groups to represent a diverse range of embodied experience: children (5-10-year-olds) and adults born with two hands, and age-matched children and adults born with limb differences. Children have strictly less embodied experience than adults, while individuals with limb differences have dramatically different kinds of experience. By using a virtual task with simple manipulation inputs, we control for manipulation capabilities and instead study how embodied experience affects the cognitive processes that support planning and reasoning for action more generally.

For the virtual physical problem-solving task, we use the recently introduced Virtual Tools game (Allen et al., 2020), which breaks the link between manipulation and physical problem-solving. The task requires people to use virtual objects as tools to solve a physical problem (such as knocking the red ball into the green goal, Fig. 1C) in a computerized environment using a single hand to control a mouse. Puzzles in the Virtual Tools game have the additional advantage of being equally unfamiliar to children and adults: all the 'tools' and puzzles are virtual and it is very unlikely for children or adults to have interacted with objects like these previously. Thus, unlike real-world objects which may be more familiar to adults than children, the use of the virtual tools should not be driven by direct experience with similar objects.

While we did not find any differences in overall performance between people with and without limb differences, we found distinctions in the *meta-strategies* they used to solve these puzzles: participants with limb differences took fewer actions to solve the puzzles, but spent more time per action. This was not simply due to the cost of taking actions, as participants with and without limb differences were equally good at controlling the interface. We also found similar patterns in both children and adults. This suggests that a history of embodied experience affects the abstract meta-strategies people use to solve new problems, and that these meta-strategies may be learned early in development.

## Results

### Equivalences in group characteristics

To ensure that the game equalized action capabilities between participants with congenital limb differences (LD) and those born with no limb differences (NLD), and between children and adults, all participants completed a motor pretest before proceeding to the main experiment (Fig. S1).

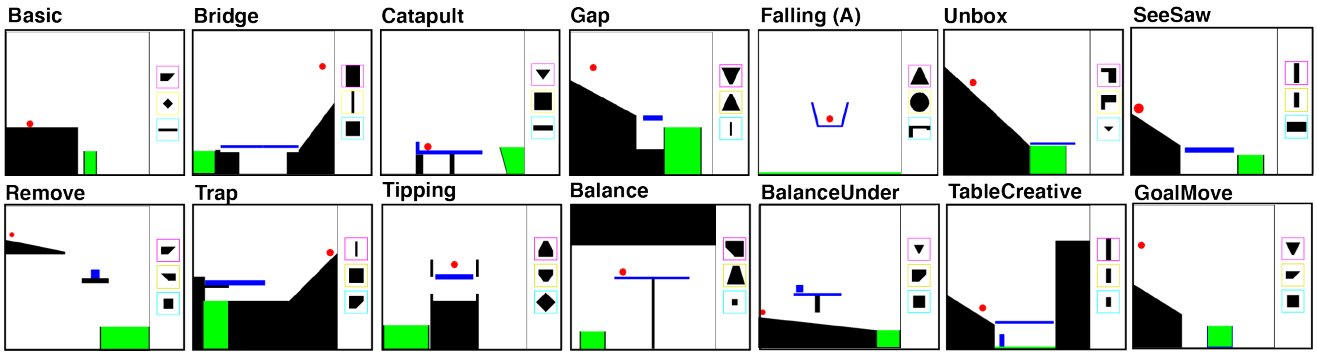


*Figure 2*. The fourteen levels of the Virtual Tools game (Allen, Smith, & Tenenbaum, 2020) that participants played. These cover a wide variety of physical action concepts including "balancing", "launching", "catapulting", "supporting" and "tipping". To play the game, please see `https://sites.google.com/view/virtualtoolsgame`.

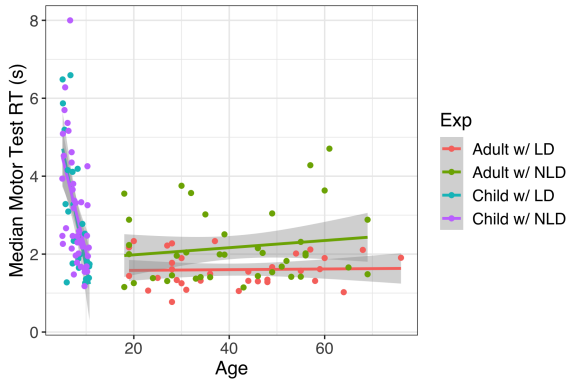


*Figure 3*. Performance on the motor task, specifically the motor reaction time compared with participant age across groups. Participants with limb differences were slightly *faster* on this task, suggesting that differences in mouse control cannot explain the additional time they take per action.

The motor pre-test consisted of ten trials where participants clicked on a star in the center of the screen, triggering the appearance of a circle in the periphery, which they were instructed to click on. We measured participants' reaction time as the time between the star and circle click, as well as position error, measured as pixels between their click and the center of the circle.

Children could control the cursor, with an average pixel error of $7.65px$ (95% CI=[6.74, 8.55]) and reaction time of $3.04s$ (95% CI=[2.67, 3.41]), albeit less accurately and more slowly than adults (error: $2.92px$, 95% CI=[2.52, 3.31]; RT: $1.91s$, 95% CI=[1.73, 2.09]). However, children's control improved with development, reaching adult-like levels by the time they were 9-10 years old (see Fig. S3).

Importantly, individuals with and without limb differences performed comparably in both motor error and in reaction time. While there was a difference in median click-time between participants with and without limb differences, participants *with* limb differences were slightly *faster* (by $356ms$, 95% CI=[14, 698]; $\chi^2(1) = 4.12$, $p = 0.044$), though they clicked marginally further away from the target (by $0.79px$, 95% CI=[−0.14, 1.72]; $\chi^2(1) = 2.78$, $p = 0.098$). There was no interaction found between age category and embodiment for either motor speed ($\chi^2(1) = 1.76$, $p = 0.19$) or error ($\chi^2(1) = 0.01$, $p = 0.98$). Note that the differences found in motor control are relatively inconsequential for the Virtual Tools game – $0.79px$ additional error would have little effect on $600x600px$ game screens, and an extra $356ms$ would be hard to detect with an average time between actions of over $10s$. Thus we find that the existence of limb differences (or not) did not affect participants' motor control capabilities. We therefore expect that both groups should have a level playing field for interacting with the Virtual Tools game; however, to control for any individual differences we include participants' median motor reaction time as a covariate in all performance analyses.

## Performance across groups

Given that participants with and without limb differences are matched on motor and cognitive measures, we next study whether overall performance on the Virtual Tools game differs across groups. We measure overall performance in two ways: as the solution rate (hereafter, accuracy) across all game levels (see Fig. 2), as well as the total time taken to solve those levels.

We find gross differences in accuracy between children and adults (adults: 85%, children: 77%; $\chi^2(1) = 11.2$, $p = 0.0008$; Figure 4), but no effect of hand embodiment ($\chi^2(1) = 1.21$, $p = 0.27$), nor any interaction between age and limb group ($\chi^2(1) = 0$, $p = 0.99$). We further find that treating age as a continuous variable additionally predicted success in a way that differed between children and adults ($\chi^2(2) = 12$, $p = 0.0025$), with children's accuracy improving with age (log-odds increase per year: 0.290, 95% CI = [0.038, 0.543]), and adults' performance getting worse (log-odds decrease per year: 0.029, 95% CI = [0.006, 0.053]).

The time to solve these levels also shows a similar pattern of results: overall, children are slower than adults to find a solution ($\chi^2(1) = 40.0$, $p = 2.6 * 10^{-10}$; Fig. 5B),

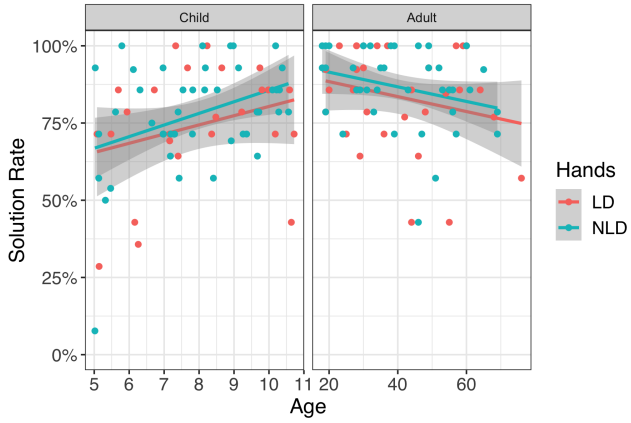


*Figure 4*. Solution rate (percentage of trials solved by each participant) as a function of age for children (left) and adults (right) with limb differences (LD) and with no limb differences (NLD). Grey areas represent standard error regions on the regression lines. Children's accuracy improves as they age (log-odds increase per year: 0.290, 95% CI = [0.038, 0.543]), and adults' performance worsens (log-odds decrease per year: 0.029, 95% CI = [0.006, 0.053]). There is no effect of hand embodiment on solution rates ($\chi^2(1) = 1.21,\ p = 0.27$).

but we do not find evidence that limb differences affect the time to solve the levels ($\chi^2(1) = 0.41,\ p = 0.52$), nor is there an interaction between limb group and age category ($\chi^2(1) = 0.22,\ p = 0.64$). However, we do find a similar pattern of how continuous age measures impact solution time: there is an effect that differed between children and adults ($\chi^2(2) = 44.5,\ p = 2.2 * 10^{-10}$), with adults slowing down with age (on average taking an additional 1.16$s$ per year, 95% CI=[0.87, 1.46], $\chi^2(1) = 59.4,\ p = 1.3 * 10^{-14}$, and children becoming numerically but not statistically faster with age (on average taking 2.64$s$ less per year, 95% CI=[−1.63, 6.90], $\chi^2(1) = 1.47,\ p = 0.23$).

Thus we find that development and aging cause noticeable changes in overall performance on the Virtual Tools game, but do not find that participants with or without limb differences are any better at the game. Nonetheless, participants might achieve an overall similar level of performance but do so in different ways; we therefore next consider whether participants with or without limb differences might demonstrate distinctions on more detailed performance metrics.

## Differences in embodiment

We first investigate whether there is a difference in the number of actions that participants with and without limb differences took to solve each level, and hence the time between each action. The participants with limb differences took fewer actions on average to come to a solution than the participants with no limb differences (0.175 fewer actions, 95% CI=[0.024, 0.326]; $\chi^2(1) = 5.19,\ p = 0.023$; Fig. 5A). Conversely, they took more time for each action, including thinking more before the first action (3.79$s$ more, 95% CI=[1.99, 5.59]; $\chi^2(1) = 17.0,\ p = 3.7 * 10^{-5}$; Fig. 5C), and between all subsequent actions (2.80$s$ more on average, 95% CI=[1.51, 4.10]; $\chi^2(1) = 17.9,\ p = 2.3 * 10^{-5}$; Fig. 5D). Again, we found differences by age (number of actions: $\chi^2(1) = 6.51,\ p = 0.011$; time to first action: $\chi^2(1) = 13.4,\ p = 0.00025$; time between actions: $\chi^2(1) = 41.7,\ p = 1.1 * 10^{-10}$), but no evidence for an interaction between age and limb differences for any of these measures (number of actions: $\chi^2(1) = 2.07,\ p = 0.15$; time to first action: $\chi^2(1) = 0.10,\ p = 0.76$; time between actions: $\chi^2(1) = 0.16,\ p = 0.69$).[1]

We further investigated whether there were differences in first actions taken by participants with and without limb differences, using the action classification methodology of (Allen et al., 2020); however, we did not find statistically reliable effects. For qualitative differences between groups see Fig. 7, and the Supplemental Materials for further details.

Together, these results suggest that individuals born with limb differences learn a different meta-strategy for physical problem-solving: they learn to rely more on thinking about the problem and less on gathering information from their actions to solve physical problems.

## Differences over development

Because we found a strong effect of age on overall performance, we consider how this is driven by changes in the actions that children and adults use.

First, we ask whether children's performance can be differentiated from adults *even from the first action*. We apply a leave-one-out classification analysis similar to Allen et al. (2020): for each participant, we form probability distributions over the actions taken by all other members of their age category and all actions taken by members of the other category, then calculate the relative likelihood that the tool placement in the first attempt for a given level was a member of the correct group (see Supplementary Materials for details). If this measure is on average reliably above chance on a trial,

[1]When testing for the effects of limb differences in children and adults separately, we find reliable differences in placement timing in both children (first placement: 4.13$s$, 95% CI=[0.96, 7.30], $\chi^2(1) = 6.52,\ p = 0.011$; between placements: 2.57$s$, 95% CI=[0.31, 4.84], $\chi^2(1) = 4.96,\ p = 0.026$) and adults (first placement: 3.29$s$, 95% CI=[1.32, 5.26], $\chi^2(1) = 10.7,\ p = 0.0011$; between placements: 2.63$s$, 95% CI=[1.26, 4.01], $\chi^2(1) = 14.2,\ p = 0.00017$). However, while we find that adults with two hands use more actions than adults with limb differences (0.286 additional placements, 95% CI=[0.047, 0.526], $\chi^2(1) = 5.49,\ p = 0.019$), children with two hands take numerically more actions, but this does not reach statistical significance (0.053 additional placements, 95% CI=[−0.140, 0.246], $\chi^2(1) = 0.29,\ p = 0.59$). Thus while we can claim that participants born with two hands overall took more actions, we do not have enough evidence to discriminate whether this is because these differences are consistent, or whether the distinction grows through development.

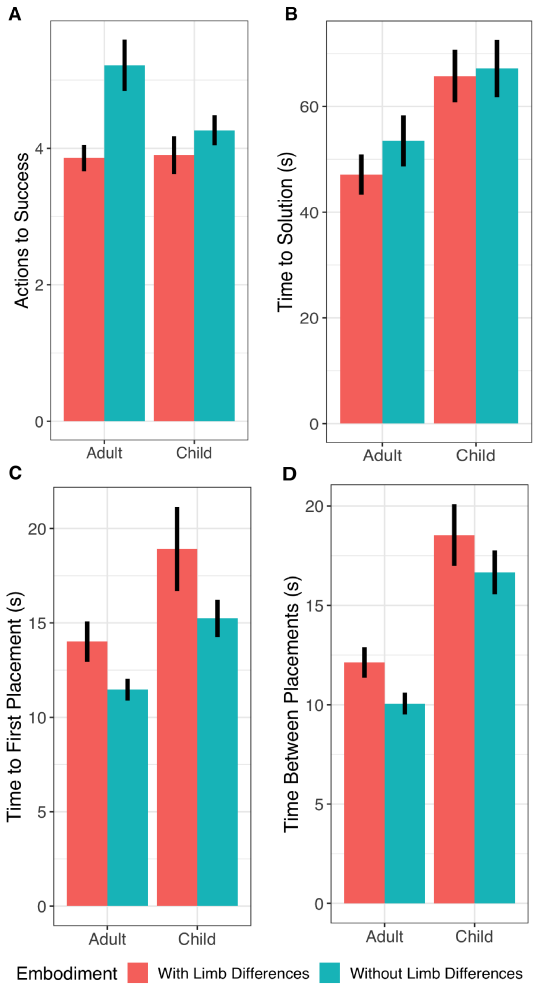


*Figure 5.* The efficiency of finding solutions measured by number of placements (**A**), and time as measured in seconds (**B**), time to first placement (**C**), and time between placements (**D**). Error bars represent 95% confidence intervals. Participants with and without limb differences did not reliably differ on time to success, but participants with limb differences solved the levels in fewer placements, and took more time until the first and between placements.

this suggests that children and adults are beginning their solution search in different ways. Indeed, we find that over all levels, children's and adults' first actions can be differentiated (born with two hands: $t(84) = 4.91,\ p = 4.5*10^{-6}$; born with limb differences: $t(57) = 3.05,\ p = 0.0035$). Thus, children do not merely take more time and actions but choose to do similar things – instead they often choose different actions to adults from the outset.

But what makes children's first actions different from adults'? As can be seen in Fig. 6 and Fig. 7, adults are often more likely to place tools high above the objects they are trying to interact with, while children often drop tools from much closer. Quantifying this, we found that children tend to place the tools closer to the objects that they intend to move or support than adults do (average vertical distance in children: 74px, adults: 91px; $\chi^2(1) = 5.20,\ p = 0.023$; Figure 6C). We do not find evidence that this propensity changes with age in years ($\chi^2(2) = 1.59,\ p = 0.45$). This is not simply due to a tendency to place tools nearer to objects in general, as there is no reliable difference in horizontal distance between tools and the nearest object (children: $31.3px$, adults: $30.7px$; $\chi^2(1) = 0.65,\ p = 0.42$). It is therefore possible that these vertical differences might be driven by differences in real-world experience between children and adults, perhaps because of additional experience with objects falling under gravity, or because adults have better motor control when dropping real-world objects.

We also investigated whether there was any difference in how children's actions evolved through the course of a single trial as compared to adults, again using similar methodology as introduced in Allen et al. (2020). We tested for differences in exploration behavior: would children be more or less likely to stick with similar actions to what they had just tried, or attempt something new? To measure this behavior, we used nonparametric clustering to group all actions across all participants within a single level type, in order to classify actions as belonging to different "strategies" in a data-driven way (see Figure 6A for examples). We could then assign each action to one of these strategies, and ask whether children or adults were more likely to switch strategies between actions. Children were in fact more likely to try new strategies than adults (children: 39% strategy switches, adults: 33%; $\chi^2(1) = 10.2,\ p = 0.0014$), suggesting that their lower accuracy might be due to either increased exploration of inefficient actions, or giving up on promising strategies early.

## Discussion

We asked participants, both with and without limb differences, to play the Virtual Tools Game to measure their physical problem-solving capabilities in novel situations where motor capabilities were equalized between groups. Both children and adults with congenital limb differences solved tasks in the game differently from their peers born with two hands – they relied more on *thinking* and less on *acting* to arrive at solutions. Across development we found that overall performance in the game increases, even though all children were above the age where they would be expected to be able to use tools in simple situations (Keen, 2011). The reason for this performance increase seems to be driven by two factors: first, children were more likely to "switch" strategies relative to adults, and second, children considered different kinds of actions relative to adults, notably placing tool objects much closer to other objects in the scene.

Within embodied cognition, we give one of the first demonstrations of natural embodiment affecting a high-level cognitive task unrelated to body or hand representations. This has several implications for embodied cognition and tool use. We found minimal differences in the specific kinds

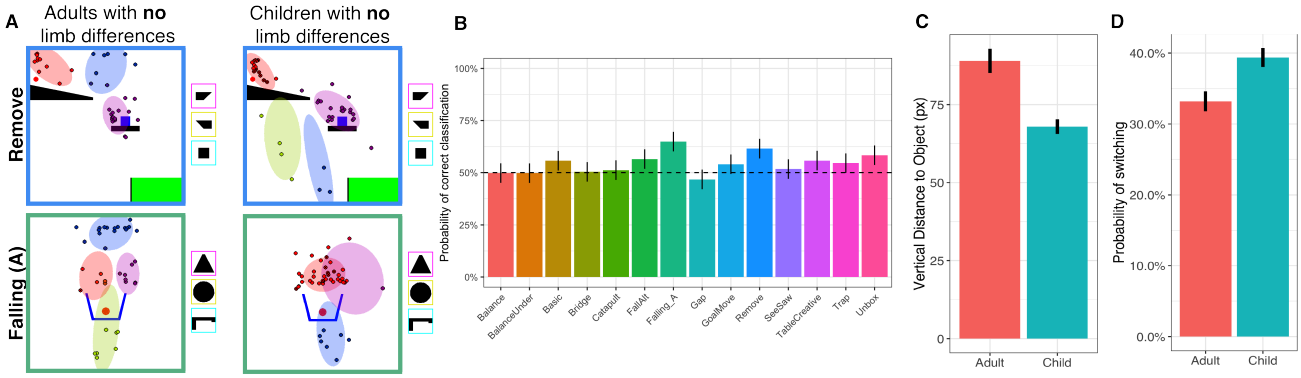


*Figure 6*. First placement comparisons across children and adults born with two hands. (**A**) An example of fitting a Dirichlet process mixture model to the first placements of adults (top) and children (bottom) for two tasks. (**B**) We tested whether we could classify participants' age group based on their first placement. Bars represent average classification accuracy across participants for each trial, with 95% confidence intervals on that estimate. (**C**) The average vertical distance between the tool placement and the nearest moveable object. Points represent the average for an individual participant, while bars represent the average across age categories. (**D**) The proportion of actions that belong to different placement clusters, indicating a "switch" in strategy.

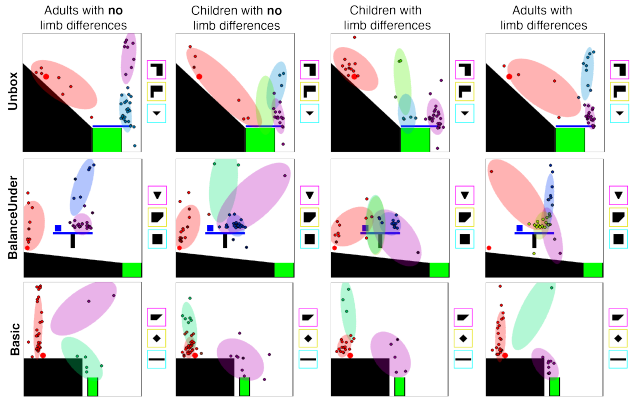


*Figure 7*. A comparison of the first placements for children and adults with and without limb differences. Colored clusters determined by Dirichlet Process Mixture Model shown for each group. Clustering was performed using Scikit-Learn (Pedregosa et al., 2011).

of actions used by individuals with and without limb differences across both children and adults, suggesting that the process underlying physical problem-solving is not dependent on similar kinds of manipulation experience. However, we also found that both children and adults born with limb differences spent more time thinking about virtual physical problems instead of taking actions. We propose that individuals born with different limbs may learn at a very young age just how costly action can be; many actions that come naturally to their peers with two hands might be very difficult to execute, and fixing a failed action might add even more cost. Naturally, this would lead to an increased reliance on planning before acting. What is striking is that this learned *meta-strategy* extends to a task in which actions are not any more costly – individuals with and without limb differences were well matched in their control of the mouse to play the game.

We suggest that these meta-strategies are learned through experience similarly to how people can learn to take more efficient motor actions over time (Huang, Kram, & Ahmed, 2012). The difference here is that the target of learning is not the motor plan itself, but *when to deploy* those motor plans. This would suggest that people with two intact hands should be able to learn similar meta-strategies with the appropriate cost structures. If the structure of the game were changed to make actions more costly in general – e.g., by requiring multiple clicks to select or place a tool – we would expect that all participants would learn to think more and act less. While people's information sampling has been shown to be sensitive to the costs of obtaining that information (Juni, Gureckis, & Maloney, 2016; Jones et al., 2019) and motor cost manipulations have been shown to affect the efficiency of motor reaching actions (Summerside, Shadmehr, & Ahmed, 2018), it has not previously been shown that motor differences directly affect the meta-strategies that people employ. By focusing on these types of utility manipulations, future work could further explore how these meta-strategies develop.

This finding also suggests a connection between two different approaches to understanding human tool use. The "manipulation-based" approach, more closely aligned with embodied cognition, suggests that tool use is supported by sensorimotor knowledge related to tool manipulation (Buxbaum, 2001; Buxbaum & Kalénine, 2010; Watson & Buxbaum, 2014; Gonzalez Rothi, Ochipa, & Heilman, 1991; van Elk, van Schie, & Bekkering, 2014), while the "reasoning-based" approach suggests that tool use is supported by physical knowledge which allows more generic physical planning and action (Allen et al., 2020; Osiurak & Badets, 2016; Osiurak & Reynaud, 2020). These theories, supported by neuro-imaging and lesions studies, have led researchers to suggest that there are distinct cognitive systems supporting different kinds of tool knowledge (Orban & Caru-

ana, 2014; Goldenberg & Spatt, 2009). However, our results suggest a connection between the two systems: by its virtual nature and novel objects, the Virtual Tools game must rely on reasoning-based systems for tool use, yet we find that manipulation capabilities affect how this reasoning is used. Thus the development of the reasoning-based system is grounded in the embodied way that we interact with the world.

Developmentally, our results extend existing knowledge about children's problem-solving and tool use. While even preschoolers (Gopnik, Sobel, Schulz, & Glymour, 2001) or infants (Baillargeon, Li, Gertner, & Wu, 2011) can understand cause and effect, our task involves reasoning about the specific effects of the virtual tools on their environment. By 2-3 years old, children can use and select known tools (Keen, 2011). By 4 years, children reliably copy adults' use of novel tools (Carr, Kendal, & Flynn, 2016). Yet until 8-9 years, children rarely innovate new tools, for example bending a pipe cleaner to use as a hook (Beck, Apperly, Chappell, Guthrie, & Cutting, 2011). Recent explanations of this striking dissociation between tool use and tool innovation highlight the significant cognitive demands of innovation, including creativity, attentional control, inhibition, and planning (Rawlings & Legare, 2020).

Children's performance in our game is likely driven by some of these same skills. Like complex tool-use, or even tool innovation, the game requires children to explore a large solution space where they must strike an effective balance between exploration and exploitation (Gopnik, 2020). On the one hand, children must avoid perseverating on a single solution (Rawlings & Legare, 2020). The current study accords with work suggesting that perseveration is not high at this age (Cutting, Apperly, & Beck, 2011). On the other hand, children must not explore the space too much, losing track of promising solutions as they go. Our results suggests that this is the more likely trap for children, who tended to switch solutions more often than adults, with lower success rates to match. We suggest therefore that the central difficulty with tool innovation and other physical problem-solving tasks at this age may precisely be the need to search through large solution spaces, where children's natural curiosity and propensity for exploration (Oudeyer & Smith, 2016) may come at the cost of some short-term gains in solution-finding.

Our work opens up new directions for the role of embodiment and development in physical problem solving and tool use. By studying adults and children born with and without limb differences using an online game instead of motor manipulation experiments, we equalized the actions available for each group, and could more carefully investigate differences between populations at a very fine-grained level. We found that being born with a different body does not change the fundamental ways in which people try to act on the world, but it can change the meta-strategies people learn in order to plan and act efficiently.

| Age Category | Affected Limb(s) | | |
|---|---|---|---|
| | Right Arm | Left Arm | Both |
| Children | 0.32 | 0.44 | 0.24 |
| Adults | 0.42 | 0.55 | 0.03 |

Table 1

*Proportion of children and adults with limb differences who have an affected right or left arm, or both arms.*

## Participants

We recruited a total of 145 participants across four groups: 40 adults without limb differences (Adult-NLD), 35 adults with limb differences (Adult-LD), 45 children with no limb differences (Child-NLD), and 25 children with limb differences (Child-LD). LD and NLD participants were well matched for age (Child-LD mean: 7.91yo, sd: 1.84; Child-NLD mean: 7.94yo, sd: 1.74; Adult-LD mean: 40.7yo, sd: 15.5; Adult-NLD mean: 41.2yo, sd: 15.2). Adults were matched on education level, and we tested for similarities in cognitive capabilities by performing IQ tests on a subset of the children from both the LD and NLD groups. We assessed both Raven's matrices measures of spatial IQ (), and BPVS as a measure of verbal IQ. 20/25 Child-LD and 34/43 Child-NLD participants were tested. All scores were within normal range (lowest for Ravens 85, highest 135, lowest for BPVS 75, highest 134. Ravens means: Child-LD: 113, Child-NLD: 111 for Ravens. BPVS means Child-LD: 105, Child-NLD: 108), and 2-sided unpaired t-tests show no evidence for differences between these populations (Ravens: $t(40) = 0.5$, $p = 0.6$, BPVS: $t(50) = 0.8$, $p = 0.4$). Details of participants' limb differences can be found in Table 1 and Supplemental Materials.

Children with limb differences were recruited through the BOLDkids database of volunteer families and via Reach and Limbo (charities supporting children with upper limb differences and their families), while two-handed children were recruited through a university-affiliated developmental cognition facebook page.

We excluded two adults with limb differences from analysis as they were amputees rather than having congenital limb differences. We also excluded results from two two-handed children: one due to a data recording error, and one who provided unreliable motor test data due to continuously clicking rather than attempting the task.

## Experiment

The experiment progressed through two stages main: motor pre-test, and Virtual Tools game. After each of these phases was a short questionnaire.

All participants were given the same experiment, with only three exceptions that differed between children and adults: (1) children received simplified instructions for all phases, (2) adults played one additional Virtual Tools level that we removed from the children's experiment due to ex-

cessive challenge, and (3) adults were given a more extensive questionnaire that included additional questions about the strategies they had used and video games they had played before.

**Motor pre-test.** The motor pre-test was used to measure participants' facility with controlling the mouse cursor. In this phase, each trial would begin with a star in the center of a 600x600px area on the screen. Once the star was clicked, a circle of radius 10px would appear in a random position either 150px or 250 px from the center of the screen, and participants were instructed to click on the circle as quickly and accurately as possible. However, when the circle was visible a click anywhere on the screen would end the trial, so that we could appropriately capture speed-accuracy trade-offs without worrying that people might not notice mis-clicks. Participants all completed 10 motor test trials, with five circles appearing 150px away, and five appearing 250px away.

On each trial we measured (a) the time between the circle appearing and when the participant clicked on it, and (b) the distance (in px) between the center of the circle and the mouse click. As a measure of participants' motor facilities, we took the median of both of those measures across all 10 trials; we used the median to avoid skew from outlier trials (e.g., if the participant accidentally clicked or was distracted on a trial), and found in pilot testing that this was a relatively stable measurement.

**Virtual Tools game.** Following Allen et al. (2020), at the beginning of this phase, participants were given instructions about how the game functioned, then three introductory trials: one which required them to place tools without an objective, and two simple levels that they were required to solve but that were not analyzed.

Following this, participants would be given the analyzed trials to solve. In each of these trials, there would be a goal condition (e.g., "get the red object into the green goal area") that they needed to accomplish by placing a single tool somewhere on the screen such that it did not overlap with other objects or illegal areas. Participants could attempt to place tools as many times as they wished, but were required to reset the world back to its initial state between attempts. Participants could move onto the next trial once they had accomplished the goal, or 60 seconds had passed. On each trial, we recorded the actions that each participant took (which tool they selected, and where they placed it) and at what time, as well as if they had solved the level or not.

Adult participants were given 15 different trials to solve: the 14 shown in Fig 1C, and one additional level: Spiky (see Allen et al. (2020)). Because of the low solution rate of adults on this level (18%), we were concerned that it might frustrate children and cause attrition, so only had children play the other 14 trials, and removed the Spiky level from analysis.

**Questionnaire.** After both the motor test and Virtual Tools task, participants were given a short questionnaire. After the motor task we asked what device participants were using to control the mouse, and, for participants with limb differences, whether they were using their intact or non-intact limb. After the Virtual Tools task, we asked whether participants had changed how they were controlling the mouse. Additionally, for the adult participants we included the questions asked in Allen et al. (2020), including about participant age, gender, prior video game experience, and free-form responses about strategies they had used on the task.

### Statistical methods

For aggregate performance analyses, we analyzed data at the trial level, using summary statistics including (a) whether the trial was solved, (b) times to the first action, the solving action, and the average time between actions, and (c) how many actions were taken over the course of the trial. For the later two metrics, we conditioned our analyses only on *successful* trials, as we were interested in the mental processes that led to solutions, and not processes that might be indicative of frustration or perseverance; however, analyzing all trials produced a qualitatively similar pattern of results (see Supplemental Materials for further details).

We modeled all statistical analyses as (generalized) linear mixed effect models using the 'lme4' package in R (Bates, Mächler, Bolker, & Walker, 2015). We treated accuracy as a binomial response, time measures as having Gaussian error, and placements as a Poisson process (using the number of non-solution placements as the dependent variable so that we could observe zero-placement outcomes). In all models, we assumed random intercepts for participants and trials.

Additionally, we included two covariates in our analyses. For all analyses, we used the median motor test response time as a covariate, as we had hypothesized that motor facility might cause better performance. We selected response time instead of error because the two measure were somewhat correlated ($r = 0.34$), and in pilot analyses we found that adding a second motor measure explained very small amounts of additional variance in performance over just a single measure. Finally, because we observed such large effects of age on performance, for all analyses testing the difference between one- and two-handed participants, we included age as a covariate, allowing its effects on performance to differ for children and adults (treated as an age by child/adult interaction).

## Acknowledgements

We thank Mathew Kollamkulam for collecting the data from the adult samples, Opcare, Limbo and Reach for assistance with participant recruitment, and all participating families and volunteers. The study was supported by a Wellcome Trust Senior Research Fellowship (215575/Z/19/Z), awarded to TRM. KRA, KAS, and JBT were supported by National Science Foundation Science Technology Center Award CCF-1231216; Office of Naval Research Mul-

tidisciplinary University Research Initiative (ONR MURI) N00014-13-1-0333; and research grants from ONR, Honda, and Mitsubishi Electric. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.

## References

Aglioti, S. M., Cesari, P., Romani, M., & C., U. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience*, *11*(9), 1109-1116.

Allen, K. R., Smith, K. A., & Tenenbaum, J. B. (2020). Rapid trial-and-error learning with simulation supports flexible tool use and physical reasoning. *Proceedings of the National Academy of Sciences*, *117*(47), 29302–29310.

Baillargeon, R., Li, J., Gertner, Y., & Wu, D. (2011). How do infants reason about physical events? In *The wiley-blackwell handbook of childhood cognitive development* (pp. 11–48). Wiley-Blackwell.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Beck, S. R., Apperly, I. A., Chappell, J., Guthrie, C., & Cutting, N. (2011, May). Making tools isn't child's play. *Cognition*, *119*(2), 301–306. doi: 10.1016/j.cognition.2011.01.003

Buxbaum, L. (2001). Ideomotor apraxia: A call to action. *Neurocase*, *7*(445), 458.

Buxbaum, L., & Kalénine, S. (2010). Action knowledge, visuomotor activation, and embodiment in the two action systems. *Annals of the New York Academy of Sciences*, *1191*, 201–218.

Carr, K., Kendal, R. L., & Flynn, E. G. (2016). Eureka!: What is innovation, how does it develop, and who does it? *Child development*, *87*(5), 1505–1519.

Cutting, N., Apperly, I. A., & Beck, S. R. (2011). Why do children lack the flexibility to innovate tools? *Journal of experimental child psychology*, *109*(4), 497–511.

Dasgupta, I., Smith, K. A., Schulz, E., Tenenbaum, J. B., & Gershman, S. J. (2018). Learning to act by integrating mental simulations and physical experiments. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society.* doi: 10.1101/321497

Gallivan, J. P., Chapman, C. S., Wolpert, D. M., & Flanagan, J. R. (2018). Decision-making in sensorimotor control. *Nature Reviews Neuroscience*, *19*(9), 519–534.

Goldenberg, G., & Spatt, J. (2009). The neural basis of tool use. *Brain*, *132*(6), 1645-1655. doi: 10.1093/brain/awp080

Gonzalez Rothi, L. J., Ochipa, C., & Heilman, K. M. (1991). A cognitive neuropsychological model of limb praxis. *Cognitive Neuropsychology*, *8*(6), 443–458.

Gopnik, A. (2020). Childhood as a solution to explore–exploit tensions. *Philosophical Transactions of the Royal Society B*, *375*(1803), 20190502.

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental psychology*, *37*(5), 620.

Hagura, N., Haggard, P., & Diedrichsen, J. (2017). Perceptual decisions are biased by the cost to act. *ELife*, *6*, 1-20.

Harris, C. M., & Wolpert, D. M. (1998). Signal-dependent noise determines motor planning. *Nature*, *394*(6695), 780–784.

Huang, H. J., Kram, R., & Ahmed, A. A. (2012). Reduction of metabolic cost during motor learning of arm reaching dynamics. *Journal of Neuroscience*, *32*(6), 2182–2190.

Jones, P. R., Landin, L., McLean, A., Juni, M. Z., Maloney, L. T., Nardini, M., & Dekker, T. M. (2019). Efficient visual information sampling develops late in childhood. *Journal of Experimental Psychology: General*, *148*(7), 1138.

Juni, M. Z., Gureckis, T. M., & Maloney, L. T. (2016). Information sampling behavior with explicit sampling costs. *Decision*, *3*(3), 147.

Keen, R. (2011). The development of problem solving in young children: A critical cognitive skill. *Annual review of psychology*, *62*, 1–21.

Leonard, J. A., Lee, Y., & Schulz, L. E. (2017). Infants make more attempts to achieve a goal when they see adults persist. *Science*, *357*(6357), 1290–1294.

Maimon-Mor, R. O., Johansen-Berg, H., & Makin, T. R. (2017). Peri-hand space representation in the absence of a hand – evidence from congenital one-handers. *Cortex*, *95*, 169-171.

Maimon-Mor, R. O., Schone, H. R., Moran, R., Brugger, P., & Makin, T. R. (2020). Motor control drives visual bodily judgements. *Cognition*, *196*, 104120.

Makin, T. R., Wilf, M., Schwartz, I., & Zohary, E. (2010). Amputees "neglect" the space near their missing hand. *Psychological Science*, *21*(1), 55-57.

Orban, G. A., & Caruana, F. (2014). The neural basis of human tool use. *Front Psychol*, *5*. doi: 10.3389/fpsyg.2014.00310

Osiurak, F., & Badets, A. (2016). Tool use and affordance: Manipulation-based versus reasoning-based approaches. *Psychological review*, *123*(5), 534.

Osiurak, F., & Reynaud, E. (2020). The elephant in the room: What matters cognitively in cumulative technological culture. *Behavioral and Brain Sciences*, 1–57.

Oudeyer, P.-Y., & Smith, L. B. (2016). How evolution may work through curiosity-driven developmental process. *Topics in Cognitive Science*, *8*(2), 492–502.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Rawlings, B., & Legare, C. H. (2020). Toddlers, tools, and tech: The cognitive ontogenesis of innovation. *Trends in Cognitive Sciences*.

Summerside, E. M., Shadmehr, R., & Ahmed, A. A. (2018). Vigor of reaching movements: reward discounts the cost of effort. *Journal of neurophysiology*, *119*(6), 2347–2357.

van Elk, M., van Schie, H., & Bekkering, H. (2014). Action semantics: A unifying conceptual framework for the selective use of multimodal and modality-specific object knowledge. *Physics of life reviews*, *11*(2), 220–250.

Vannuscorps, G., & Caramazza, A. (2016). Typical action perception and interpretation without motor simulation. *Proceedings of the National Academy of Sciences*, *113*(1), 86–91.

Vannuscorps, G., Pillon, A., & Andres, M. (2012). Effect of biomechanical constraints in the hand laterality judgment task: where does it come from? *Frontiers in Human Neuroscience*, *6*, 299.

Watson, C., & Buxbaum, L. (2014). Uncovering the architecture of action semantics. *Journal of Experimental Psychology: Human Perception and Performance*, *40*.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636. doi: 10.3758/BF03196322

## Supporting Information

### S1 Motor pre-test

The motor pre-test ensured that participants with and without limb differences had equal abilities to interact with the computer system, and therefore should be equally able to control the Virtual Tools game. The motor-test is shown in Figure S1. Participants were required to click a central star before clicking a colored circle in the periphery (either 150 or 250 px from the center of a 600x600px screen). They completed 10 rounds of this procedure.

We additionally include other measures on the motor test not presented in the main paper (Figure S3).

### S2 Participant Demographics

Please see Tables S1 for participant demographic information.

For details on specific limb differences in the children we tested, please refer to Table S2. Adult participants had less variability in their limb differences. Of the 33 adult participants, one had a bilateral limb difference (with a missing right arm at the shoulder, and left arm at the elbow), and 32 had unilateral limb differences, including:

- 3 transhumeral
- 3 at the elbow
- 18 transradial (one participant has a small residual digit)
- 8 at write (one participant has 5 small intact digits; one has a short thumb and small, non-jointed digit; one has two residual digits)

### S3 Differences in action choices

#### S3.1 Methodology

To test differences in action choices between groups, we apply a leave-one-out classification analysis similar to (Allen et al., 2020): for each participant, we form probability distributions over the actions taken by all other members of their embodiment category and all actions taken by members of the other category, then calculate the relative likelihood that the tool placement in the first attempt for a given level was a member of the correct group (see Methods and Materials for details). If this measure is on average reliably above chance on a trial, this suggests that participants with vs. without limb differences are beginning their solution search in different ways.

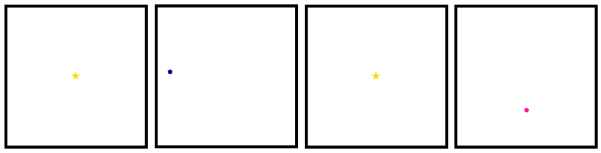

*Figure S1.* An example of two rounds of the motor pre-test. Participants clicked first the star, then a circle in the periphery, as quickly as possible.

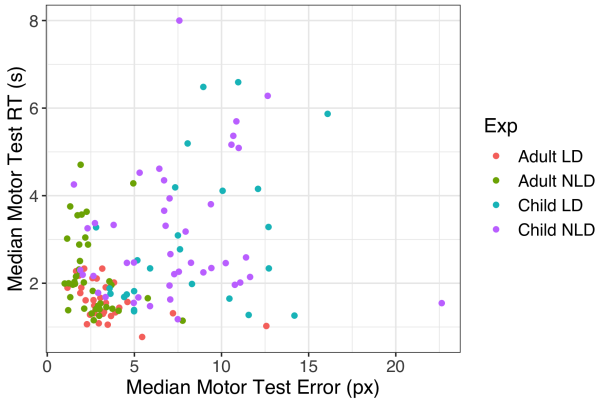


*Figure S2.* Comparison of Median Motor test reaction time with median motor test error for each participant across groups. LD = participant with limb differences, NLD = participant with no limb differences. Correlation is reasonably high, $r = 0.34$, so we use only the median motor test error as a covariate throughout analyses.

#### S3.2 Differences between LD and NLD participants

Overall, we found a trend towards being able to classify participants with and without limb differences (mean adult classification accuracy: 51.4%, 95% CI=[49.9, 52.9], $t(71) = 1.91$, $p = 0.060$; mean child classification accuracy:

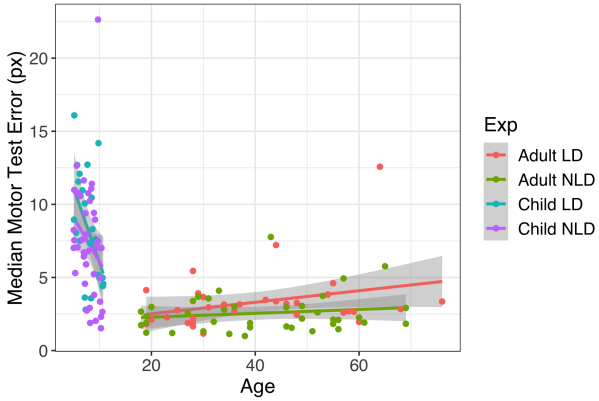


*Figure S3.* Median motor test error in pixels as a function of age. Children improve significantly as they age, while adults' performance is unaffected. No interaction was found between age and embodiment group ($\chi^2(1) = 0.01$, $p = 0.98$).

| Group | Handedness | | | Input Device | | |
|---|---|---|---|---|---|---|
| | Left | Right | Ambi. | Mouse | Touchpad | Other |
| Adult-LD | 0.45 | 0.55 | 0 | 0.61 | 0.36 | 0.03 |
| Adult-NLD | 0.08 | 0.92 | 0 | 0.55 | 0.45 | 0 |
| Child-LD | 0.36 | 0.64 | 0 | 0.32 | 0.68 | 0 |
| Child-NLD | 0.07 | 0.89 | 0.04 | 0.50 | 0.50 | 0 |

Table S1

*Demographics summaries for participants. LD: limb differences; NLD: no limb differences*

| **Participant** | **Limb Difference** | | | | **Dominant Limb** |
|---|---|---|---|---|---|
| | Left | Pincer | Right | Pincer | |
| 1 | Partial hand, 1 digit | No | Partial hand, 1 digit | No | R |
| 2 | 5 digits, limited function | Limited | Intact | Yes | R |
| 3 | Partial hand, 1 digit | No | Partial hand, 3 digits | Yes | R |
| 4 | Absent below elbow | No | Intact | Yes | R |
| 5 | Absent below elbow | No | Intact | Yes | R |
| 6 | Intact | Yes | Partial hand, 2 digits | Yes | L |
| 7 | Partial hand, 3 digits | No | Partial hand, 4 digits | No | R |
| 8 | Partial hand, no digits | No | Intact | Yes | R |
| 9 | Intact | Yes | Absent below elbow | No | L |
| 10 | Absent below elbow | No | Intact | Yes | R |
| 11 | Absent below elbow | No | Intact | Yes | R |
| 12 | Intact | Yes | Partial hand, no digits | No | L |
| 13 | Partial hand, 3 digits | Limited | Partial hand, 3 digits | Yes | R |
| 14 | 5 digits, non-functional | No | Intact | Yes | R |
| 15 | Intact | Yes | Partial wrist, no digits | No | L |
| 16 | Intact | Yes | Partial hand, 2 digits | Yes | L |
| 17 | 4 digits, non-functional | No | 4 digits, non-functional | No | R |
| 18 | Absent below elbow | No | Intact | Yes | R |
| 19 | Intact | Yes | Absent below elbow | No | L |
| 20 | Absent above elbow | No | Absent above elbow | No | L |
| 21 | Intact | Yes | Absent above elbow | No | L |
| 22 | Absent below wrist | No | Intact | Yes | R |
| 23 | Absent below elbow | No | Intact | Yes | R |
| 24 | Intact | Yes | Partial hand missing bones, 4 digits | Yes | L |
| 25 | Partial hand, 2 digits | No | Intact | Yes | R |

Table S2

*Details of children with a limb difference. Pincer: defined as the ability to grasp a small object between the thumb and index finger; Dominant limb: R=right, L=left*

52.3%, 95% CI=[49.5%, 55.1%, $t(70) = 1.62$, $p = 0.109$), but found only one task where the confidence interval on the estimated classification probability exceeded chance for both children and adults (see Table S3): BalanceUnder. This task requires preventing objects from falling by placing a tool as a counterweight to another object, which might be a strategy the participants with limb differences used more often in daily life, as they are have to rely on their residual arm (which is shorter than their intact arm) when manipulating objects bimanually. However, given that the estimated classification probability is not far from chance and that we cannot reliably classify actions in other trials that rely on balancing, future work would need to investigate particular differences in strategies learned from interaction with the environment.

## S4 Additional tests for moderators

Here we test for the effect of other potential explanatory variables on our set of metrics describing performance on the tools game, including (1) participant gender, (2) the type of device participants used to control the game (mouse or touchpad), (3) whether participants were dominantly left- or right-handed, and (4) for the DL participants, what kind of limb differences they had. While we find possible impacts

| Trial Name | Adult: LD vs NLD | Children: LD vs NLD | NLD: Adult vs Children | LD: Adult vs Children |
|---|---|---|---|---|
| Balance | 52% [47%, 56%] | 50% [45%, 55%] | 50% [45%, 55%] | 47% [41%, 53%] |
| BalanceUnder | 58% [53%, 62%] | 51% [46%, 56%] | 50% [45%, 54%] | 60% [54%, 66%] |
| Basic | 53% [48%, 57%] | 50% [45%, 55%] | 56% [51%, 60%] | 53% [47%, 59%] |
| Bridge | 55% [50%, 60%] | 50% [46%, 55%] | 50% [46%, 55%] | 54% [48%, 59%] |
| Catapult | 47% [43%, 52%] | 53% [48%, 58%] | 51% [47%, 56%] | 56% [51%, 62%] |
| FallAlt | 48% [43%, 52%] | 58% [53%, 63%] | 57% [52%, 61%] | 61% [55%, 67%] |
| Falling_A | 51% [46%, 56%] | 60% [55%, 65%] | 65% [60%, 70%] | 55% [50%, 61%] |
| Gap | 47% [43%, 52%] | 51% [46%, 56%] | 47% [42%, 51%] | 48% [42%, 54%] |
| GoalMove | 53% [48%, 58%] | 47% [42%, 52%] | 54% [49%, 59%] | 47% [42%, 53%] |
| Remove | 52% [48%, 57%] | 51% [46%, 56%] | 62% [57%, 66%] | 51% [45%, 57%] |
| SeeSaw | 54% [50%, 59%] | 51% [46%, 56%] | 52% [47%, 56%] | 47% [41%, 53%] |
| TableCreative | 47% [42%, 51%] | 50% [45%, 55%] | 56% [51%, 60%] | 62% [56%, 68%] |
| Trap | 52% [48%, 57%] | 56% [51%, 61%] | 54% [50%, 59%] | 54% [49%, 60%] |
| Unbox | 50% [46%, 55%] | 53% [48%, 58%] | 58% [53%, 63%] | 50% [44%, 56%] |

Table S3

*Classification of first actions between different groups. Numbers in brackets represent bootstrapped 95% confidence intervals on classification percentages.*

of these variables on overall performance, in no cases do we find any interactions between these variables and embodiment, suggesting that they should not impact the main results of the paper.

### S4.1 Gender

For these analyses, we exclude 8 participants from analysis (3 adults with no limb differnces, 5 adults with limb differences) as their gender was unknown due to recording errors.

There is a main effect of gender on accuracy ($\chi^2(1) = 9.88$, $p = 0.0017$), with males slightly outperforming females (83% vs 77%), but no interaction with embodiment class ($\chi^2(1) = 0.53$, $p = 0.47$). We also find a small effect of gender on the first action time (males: 13.8$s$, females: 15.2$s$, $\chi^2(1) = 5.56$, $p = 0.018$), but again no interaction ($\chi^2(1) = 0.75$, $p = 0.39$. Beyond this we find no evidence for any main effects on our main performance variables (actions: $\chi^2(0.33) = 1$, $p = 0.57$, solution time: $\chi^2(1) = 0.32$, $p = 0.57$, time between actions: $\chi^2(1) = 1.19$, $p = 0.28$), nor any interactions between gender and embodiment (actions: $\chi^2(1) = 0.08$, $p = 0.77$, solution time: $\chi^2(1) = 0.54$, $p = 0.46$, time between actions: $\chi^2(1) = 0.00$, $p = 0.95$). Although there may be slight differences in how males and females perform this task overall, because we found no interactions with embodiment, the effect of embodiment itself does not appear to depend on participant gender.

### S4.2 Input device

For these analyses we excluded one adult participant with limb differences who did not use a touchpad or mouse for controlling their computer.

We find a main effect of device on time to solution ($\chi^2(1) = 7.96$, $p = 0.0048$), with participants using a mouse solving the levels slightly faster than participants using a touchpad (53.2$s$ vs 63.8$s$). However, we found no interaction between device and embodiment on solution time ($\chi^2(1) = 0.47$, $p = 0.49$), nor did we find any other main effects of device (accuracy: $\chi^2(1) = 0.04$, $p = 0.84$, number of actions: $\chi^2(1) = 3.33$, $p = 0.068$, first action time: $\chi^2(1) = 2.43$, $p = 0.12$, time between actions $\chi^2(1) = 1.96$, $p = 0.16$) or interactions between device and embodiment category (accuracy: $\chi^2(1) = 0.00$, $p = 0.96$, number of actions: $\chi^2(1) = 1.12$, $p = 0.29$, first action time: $\chi^2(1) = 0.14$, $p = 0.7$, time between actions $\chi^2(1) = 0.32$, $p = 0.57$). Thus participants' choice of device should not impact our tests of the effects of embodiment.

### S4.3 Hand laterality

For these analyses we exclude 1 child without limb differences who was ambidextrous.

We find no effect of hand laterality on any of our dependent variables – neither main effects (accuracy: $\chi^2(1) = 0.01$, $p = 0.91$, number of placements: $\chi^2(1) = 0.35$, $p = 0.55$, solution time: $\chi^2(1) = 0.00$, $p = 0.98$, first action time: $\chi^2(1) = 3.12$, $p = 0.077$, time between actions $\chi^2(1) = 0.78$, $p = 0.38$), nor interactions between laterality and embodiment category (accuracy: $\chi^2(1) = 0.05$, $p = 0.83$, number of placements: $\chi^2(1) = 2.00$, $p = 0.16$, solution time: $\chi^2(1) = 0.84$, $p = 0.36$, first action time: $\chi^2(1) = 0.80$, $p = 0.37$, time between actions $\chi^2(1) = 0.05$, $p = 0.83$).

## S5 Analysis of all trials

In the main body of the text, we studied the impact of age and embodiment on how participants solved *successful*

trials in order to avoid measuring effects of persistence or motivation. Here we re-run our main analyses on all trials to determine whether this choice might lead to any differences in analysis results. We find that in general, there is little difference between the success-conditioned and all-trials analyses, though one statistical test (overall number of placements by embodiment) does cross from statistically significant to barely not statistically significant. We discuss this difference and report all statistics below.

### S5.1 Number of total actions

While we found that participants with limb differences used fewer actions to get to a solution than participants without, the effect of embodiment on number of total actions is only marginally significant (participants with limb differences taking 0.127 fewer actions on average, 95% CI=[−0.020, 0.273]; $\chi^2(1) = 2.92,\ p = 0.088$).

We ask whether any differences in this analysis might be a function of differences in motivation, and so measure persistence as the ratio of total actions (regardless of success) to the number of actions taken on successful trials. We find that by this definition, participants with limb differences were numerically more persistent (adults: 145%, children: 147%) than participants with no limb differences (adults: 132%, children: 136%), though this difference was only marginally significant ($F(1, 135) = 3.80,\ p = 0.053$). Because of the marginal significance, we do not make strong claims about the persistence of LD vs. NLD participants, but the numerical difference is likely what drives the reduction in the measured effect of embodiment on total placements.

### S5.2 Overall time

Here, instead of testing the time it takes until people find a solution, we analyse the time until either a solution is found, or people decide to give up (after 60s). Similar to the success-conditioned analysis, we find that children take more time than adults ($\chi^2(1) = 37.8,\ p = 7.8 * 10^{-10}$), that there is no reliable effect of embodiment ($\chi^2(1) = 0.37,\ p = 0.55$), and that there is a difference by continuous age ($\chi^2(2) = 43.8,\ p = 3.1 * 10^{-10}$).

### S5.3 Time to first action

Just as with the success-conditioned results, we find that participants with limb differences took more time than participants with no limb differences to choose their first action (4.12$s$ more, 95%CI = [2.20, 6.03]; $\chi^2(1) = 17.7,\ p = 2.6 * 10^{-5}$), and that this differs by age ($\chi^2(1) = 12.4,\ p = 0.00043$), but no evidence of an interaction between age and embodiment ($\chi^2(1) = 0.32,\ p = 0.57$).

### S5.4 Time between actions

Similar to the success-conditioned analyses, we find that participants with limb differences spend more time thinking than participants with no limb differences between actions (2.33$s$ more, 95%CI = [1.04, 3.63]; $\chi^2(1) = 12.5,\ p = 0.00041$), and that this differs by age ($\chi^2(1) = 43.5,\ p = 4.2 * 10^{011}$), but no evidence of an interaction between age and embodiment ($\chi^2(1) = 0.04,\ p = 0.84$).