# Starting from Zero: Task Learning in Completely Naïve Individuals

Aaron Wang[*,1], Lauren Williams[*,1], Oyindamola Ogunlade[1], Jorge Morales[#,2,3], and Kishore V. Kuchibhotla [#,1,4,5]

[1] Department of Psychological and Brain Sciences, Johns Hopkins University
[2] Department of Psychology, Northeastern University
[3] Department of Philosophy, Northeastern University
[4] Department of Neuroscience, Johns Hopkins University
[5] Department of Biomedical Engineering, Johns Hopkins University

[*] Authors contributed equally
[#] Co-senior authors.

Please address correspondence to jorge.morales@northeastern.edu and kkuchib1@jhu.edu

## Abstract

Humans often learn about the world by observing, imitating, and receiving explicit instruction from others[1–4]. Even when learning via trial-and-error or reflection, human adults often do so immersed in a relevant context[5–11]. But what happens when most of this scaffolding is removed? How effectively can humans bootstrap learning in an information- and context-impoverished environment? In eight pre-registered experiments, we sought to understand how effectively humans learn from 'zero' in a series of simple stimulus recognition tasks that subjects easily learned when provided with explicit instructions. When no explicit guidance was provided, fewer than half of the subjects learned the task despite receiving trial-by-trial feedback and monetary reward. Surprisingly, providing partial, explicit instructions about the structure of the task—a ubiquitous pedagogical approach—did not improve performance nor reduce individual variability. Instead, explicit instructions that constrained the action space (i.e., valid key presses) partially recovered performance but still pointed to action inhibition as particularly challenging to learn without explicit instructions. Our results suggest that individual differences emerge in impoverished environments without reflecting underlying capacity, and that a major driver of learning, even for the simplest of tasks, is a combination of luck and the nature and quality of scaffolding available to the learner.

*Keywords: instructions, reinforcement learning, go/no-go task*

## Introduction

Imagine the challenge of inventing the Calculus from scratch, rather than learning it in an advanced class. While arguably only two people in history—Leibniz and Newton—succeeded at the former, millions have succeeded since via instruction. Why is some learning so much easier when we are instructed compared to when we must figure it out for ourselves? May this be a feature of all human learning and not just of complicated challenges like learning Calculus?

Receiving explicit guidance—anything from asking others for help to online tutorials and formal training—is one of the most efficient ways to incorporate new knowledge and new actions into one's repertoire[1–4]. However, explicit information or cooperative, knowledgeable agents are not always available. In conjunction and sometimes instead of explicit guidance, humans also utilize exploration[12–16], trial and error[5–8], prior knowledge[14,17–19], and logical inference[9–11] to learn. Moreover, explicit instructional information sometimes can affect our learning strategies[12,20,21] and our actions[22–25]. Being able to bootstrap learning is, therefore, a crucial capacity to navigate the world. In everyday life, this happens more or less seamlessly in familiar conditions where context or affordances from the environment guide our learning behavior. For example, if you find yourself looking for the exit in an unfamiliar building, you are likely to figure out that the protruding object at mid-height attached to the big rectangular standing slab is a doorknob. But how do we learn 'from zero' when previous knowledge is of little to no use—like when trying to learn how to program for the first time without any useful tutorials at hand or when navigating a social interaction with serious language barriers and completely different social norms? How do we learn when an information-impoverished context provides little to no guidance as to which previous knowledge or actions are even relevant?

Laboratory experiments with non-human animals may offer insight into learning in 'zero knowledge' conditions. In behavioral experiments, researchers train animals to perform novel tasks (e.g., fixating on a stimulus, finding the solution to a maze, pressing a lever under the right circumstances, etc.)[26–30]. To do so, animals are often placed in unfamiliar settings with an impoverished and evolutionary irrelevant environment where they are meant to unmask artificial task structures and action sequences. The gold standard to achieve this is via stimulus-reward associative learning. In this way, animals from different species can learn complex stimulus-action rules and task-reward structures without explicit instruction[31–38].

How do humans learn a new task in an information-deprived environment without the benefit of explicit guidance? How well and with which strategies may we learn under these circumstances? Explicit instruction, often provided in linguistic form, is so central to our learning practices that the answers to these questions are far from obvious.

## Results

We trained subjects on a go/no-go (GNG) task where they pressed the spacebar in response to certain shapes and withheld from pressing any key in response to other

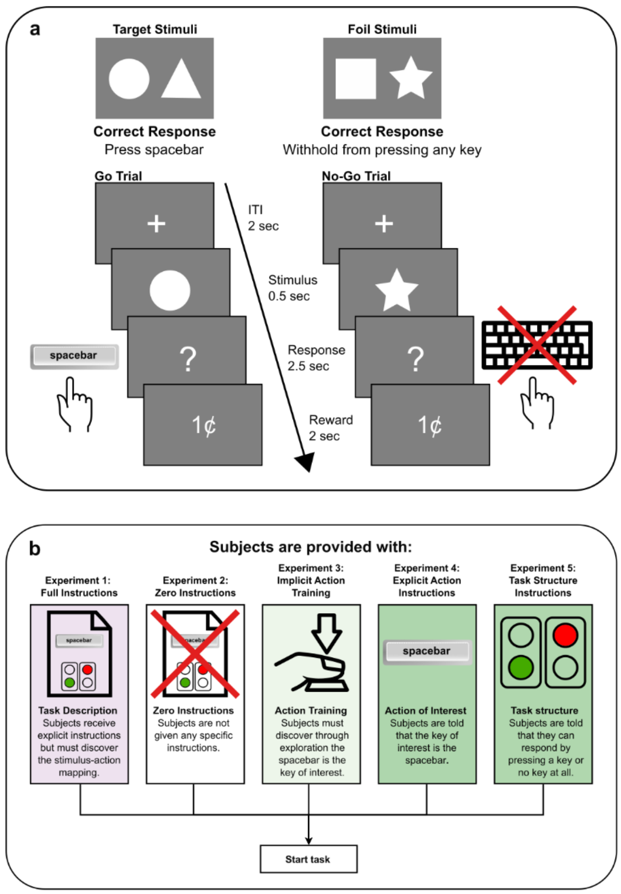



**Figure 1. Experimental Design. (a)** Subjects were first presented with a fixation cross followed by the stimulus. When the question mark appeared, subjects had to press spacebar after a target stimulus was presented (left stream) and withhold from pressing any key when a foil stimulus was presented (right stream). Subjects received 1¢ for correct responses and 0¢ for incorrect responses. The stimulus-action mapping was counterbalanced across subjects. **(b)** Subjects in all experiments were told that they would be presented with opportunities to earn rewards and their goal was to discover how to do so. Subjects in Experiment 1 received full, explicit instructions while subjects in Experiment 2 were not given any specific instructions. In Experiments 3-5, subjects received varying levels of partial instructions.

shapes (**Fig. 1**). We first confirmed that subjects provided with a complete set of explicit instructions rapidly learned this task, reaching expert performance between trial 9 to 24, with little inter-subject variability (**Figs. 2a, 2c**). Subjects reached near-perfect performance (**Fig. 2b**) with 93% of subjects having at least one block of 100% accuracy. Subjects exhibited both high hit rates (pressing the spacebar on target trials; M=95% correct, SD=0.22, **Fig. 2b**) and high correct rejection rates (not pressing any key on foil trials; M=99% correct, SD=0.06, **Fig. 2b**).

**Providing zero instructions resulted in decreased performance and increased learning variability across subjects.**

In contrast, when subjects were placed in an information-poor environment without *any* explicit guidance, 60% (36/60) of individuals failed to master the task despite receiving trial-by-trial feedback and monetary reward. Moreover, there was far greater inter-subject variability compared to Experiment 1 (Experiment 2, SD=0.3 vs. Experiment 1, SD=0.13). Even when considering maximum performance in any block of trials, subjects given zero instructions performed far worse than those given full, explicit instructions (**Fig. 2b**, M=69% correct, SD=0.28; overall % correct one-way ANOVA: $F(4, 295)=21.08$, $p<0.0001$, post-hoc Exp. 1 vs Exp. 2: $p<0.0001$, 95% CI=[-0.39, -0.16]) and reached their subject-specific peak performance much later (trials 33-48 versus trials 9-24). Only 42% of subjects had at least one block of 100% (**Fig. 2a**). This overall lower performance was driven by lower hit rates with large variance (**Fig. 2b**, M=55% correct, SD=0.48; hit rate one-way ANOVA: $F(4, 295)=25.49$, $p<0.0001$; post-hoc Exp. 1 vs Exp. 2: $p<0.0001$, 95% CI=[-0.57, -0.22]). Interestingly, correct rejection rates were slightly, but non-significantly, lower compared to subjects given full instructions (**Fig. 2b**, M=83% correct, SD=0.36; CR rate one-way ANOVA: $F(4, 295)=9.12$, $p<0.0001$; post-hoc Exp. 1 vs Exp. 2: $p=0.09$, 95% CI=[-0.33, 0.01]). Subjects were aware of how well they learned the task: their performance correlated with their explicit confidence (max performance vs. confidence: $r=0.746$, $p=7.68e-12$; explicitly reported knowledge of shape-action contingencies vs. confidence: $r=0.84$, $p=7.41e-17$; Pearson's correlation test).

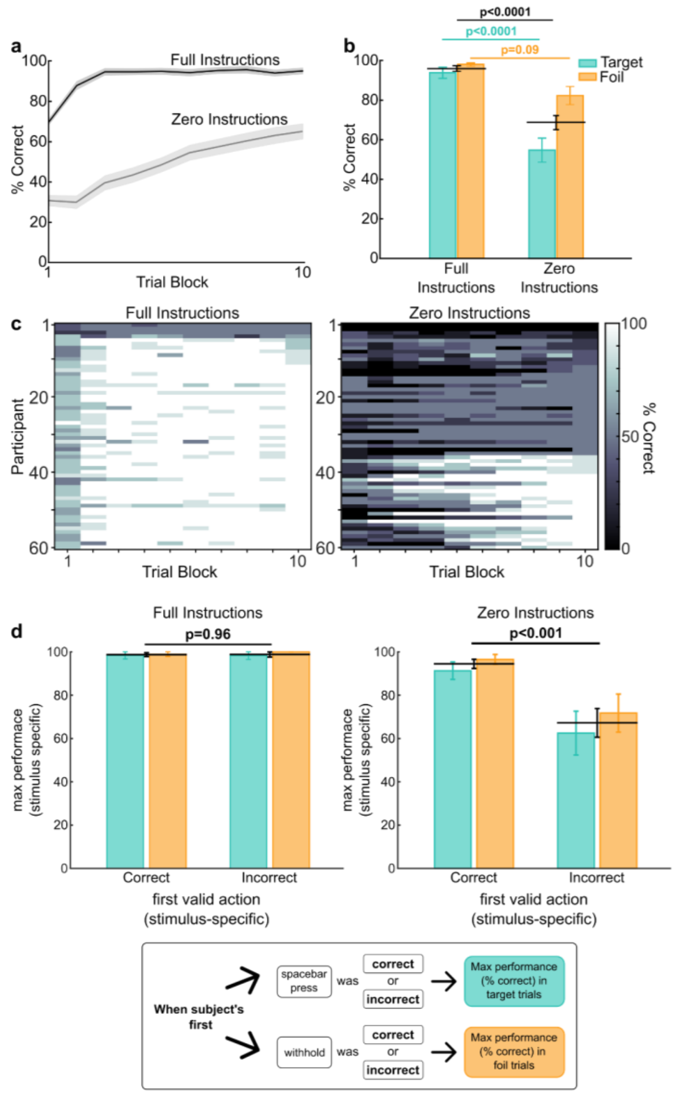



**Figure 2. Majority of subjects fail to learn in the absence of explicit instructions. (a)** Average percent correct values from trial block 1 to trial block 10 (8 trials per block) for GNG Full Instructions (Exp. 1) and GNG Zero Instructions (Exp. 2). **(b)** Percent correct averaged across each subject's maximum performance block, both overall (black line) and separated by trial-type (target trials: green bar, foil trials: orange bar). **(c)** Individual Learning Trajectories for Experiments 1-2. All subjects were sorted by final block percent correct, from least to most accurate. **(d)** Maximum performance (% correct) in target (green bar), foil (orange bar), and all trials (black line) when the subjects' first valid action (either a spacebar press or a withhold) was correct or incorrect. Full instructions: spacebar correct (n=31) vs. incorrect (n=29), withhold correct (n=47) vs. incorrect (n=13). Zero Instructions: spacebar correct (n=23) vs. incorrect (n=22), withhold correct (n=36) vs. incorrect (n=23). Mean ±SEM.

What might explain the poor performance and high inter-subject variability in subjects provided with zero instructions? We conducted a detailed analysis of error types across subjects and found three primary drivers. First, subjects often pressed keys other than the spacebar (**Fig. 3b**, 'go errors', M=0.09, SD=0.25 and 'no-go errors', M=0.1, SD=0.27; **Extended Data Fig. 1b**). This suggests that subjects extensively explored the action space but failed to pinpoint the spacebar as the key of interest and/or failed to realize that refraining from pressing any key was a valid option. Second, many subjects chose to stop pressing any keys at all. These subjects typically pressed many, distinct keys for many trials but then subsequently stopped pressing any key (**Extended Data Fig. 1b**), even for target trials, when compared to subjects provisioned with complete instructions (**Fig. 3b**, 'miss', M=0.35, SD=0.44; miss rate one-way ANOVA: $F(4, 295)=16.39$, $p<0.0001$; post-hoc Exp. 1 vs Exp. 2: $p<0.0001$, 95% CI=[0.14, 0.47]. This likely reflects that these subjects were either content with a 50% reward rate or, alternatively, gave up on the task due to frustration. Either interpretation, however, suggests that the subjects found the task extremely difficult. Third, we focused on subjects who have performed a valid action during the task (i.e. pressed the spacebar or withheld from pressing any key) at least once. When these subjects were rewarded on their first valid action (i.e. pressed the spacebar for the first time on a target trial or did not press any key for the first time on a foil trial) they were more likely to learn the relevant action (**Fig. 2d**, $p<0.001$, Wilcoxon rank sum test). One possibility is that those who were rewarded on their first valid action had a more targeted strategy of using spacebar presses and withholding. However, this was likely not the case as these subjects executed the valid action on similar trial numbers compared to subjects who were not rewarded on their first valid action (**Extended Data Fig. 2c**, difference in first spacebar press trial number when first spacebar press is correct vs. incorrect: $p=0.84$, difference in first withhold trial number when first withhold is correct vs. incorrect: $p=0.42$; Wilcoxon rank sum test). When considered together, all three drivers suggest that in the absence of explicit instructions, chance—and not strategy or capacity—governs task success.

**Implicitly constraining the action space does not improve performance**

The previous experiment points to extensive exploration of the action space as a major driver of poor performance in the zero instructions task. We sought to control this exploration through implicit spacebar training, a shaping procedure often used in animal learning to link responses to outcomes. Specifically, subjects were not given any explicit instructions but instead experienced a pre-task shaping phase in which they learned the response-outcome association (pressing the spacebar results in a small monetary reward). This shaping procedure successfully constrained the action space during subsequent task learning but, surprisingly, it did not improve performance compared to receiving zero instructions (**Fig. 3a**, M=74% correct, SD=0.24; post-hoc Exp. 3 vs Exp. 2 (overall % correct one-way ANOVA): $p=0.83$, 95% CI=[-0.07, 0.16]). Subjects given implicit spacebar training exhibited a significant increase in spacebar presses for both target and foil trials, which indicates that implicit training generally increased spacebar presses in a non-discriminative manner (hits: M=85% correct, SD = 0.33; post-hoc Exp. 3 vs Exp. 2 (hit rate one-way ANOVA): $p<0.0001$, 95% C.I.=[0.12, 0.48]; false alarms: (M=0.3, SD=0.39, false alarm rate one-way ANOVA: $F(4, 295)=13.4$, $p<0.0001$; post-hoc Exp. 3 vs Exp. 2: $p<0.01$, 95% C.I.=[0.03, 0.30]).

Interestingly, while subjects rarely pressed alternate keys on target trials, some continued to do so on foil trials resulting in similar no-go error rates (M=0.08, SD=0.21) compared to subjects who received zero instructions (M=0.1, SD=0.27; no-go error one-way ANOVA: $F(4, 295)=5.1$, $p<0.001$; post-hoc Exp. 3 vs Exp. 2: $p=0.96$, 95% CI=[-0.19, 0.02]). Implicit spacebar training thus had little impact on limiting action exploration for subjects during foil trials (**Fig. 3b-c**, **Extended Data Fig. 1c**). These subjects preferred to explore other key presses rather than exploring behavioral inhibition, suggesting that *not* pressing the spacebar (i.e. behavioral inhibition) is less likely to be explored than pressing an alternate key. Subjects benefited from implicit instructions in an extremely narrow manner—increases in that specific action (i.e. spacebar presses)—but remain unable to construct an accurate model of the task to enable improved performance.

**Explicit instructions constraining the action space partially restore task performance**

Given the failure of implicit spacebar training to improve overall performance despite partially constraining the action space, we next sought to understand how explicit spacebar training would fare. In Experiment 4, we provided subjects with written guidance that the only valid key to press was the spacebar. Subjects showed significant performance improvement compared to those given zero instructions (Exp. 2) as well as subjects given implicit action training (Exp. 3) (**Fig. 3a**, M=87% correct, SD=0.2; post-hoc Exp. 4 vs Exp. 2 (overall % correct one-way ANOVA): $p<0.001$, 95% C.I.=[0.06, 0.29]; post-hoc Exp. 4 vs Exp. 3 (overall % correct one-way ANOVA): $p=0.01$, 95% C.I.=[0.02, 0.24]). Subjects never pressed a key other than the spacebar for target trials and rarely did so on foil trials (M=0.02, SD=0.13). Interestingly, while the overall rate of incorrect responses on target trials was dramatically reduced compared to subjects who received zero instructions (miss rate: M=0.02, SD=0.09, post-hoc Exp. 4 vs Exp. 2 (miss rate one-way ANOVA): $p<0.0001$, 95% C.I.=[-0.50,- 0.17]; go-error: M=0, post-hoc Exp. 4 vs Exp. 2 (go-error rate one-way ANOVA): $p=0.03$, 95% C.I.=[-0.18, -0.004]), the overall rate of incorrect responses for foil trials remained similar due to the comparatively higher false alarm rate (i.e., spacebar press for foil stimuli; **Fig. 3b**, M=0.23, SD=0.37; post-hoc Exp. 4 vs Exp. 2 (false alarm rate one-way ANOVA): $p<0.01$, 95% C.I.=[0.03, 0.30]). Overall, these data demonstrate that explicit instructions are more effective than implicit response-outcome training in constraining the action space.

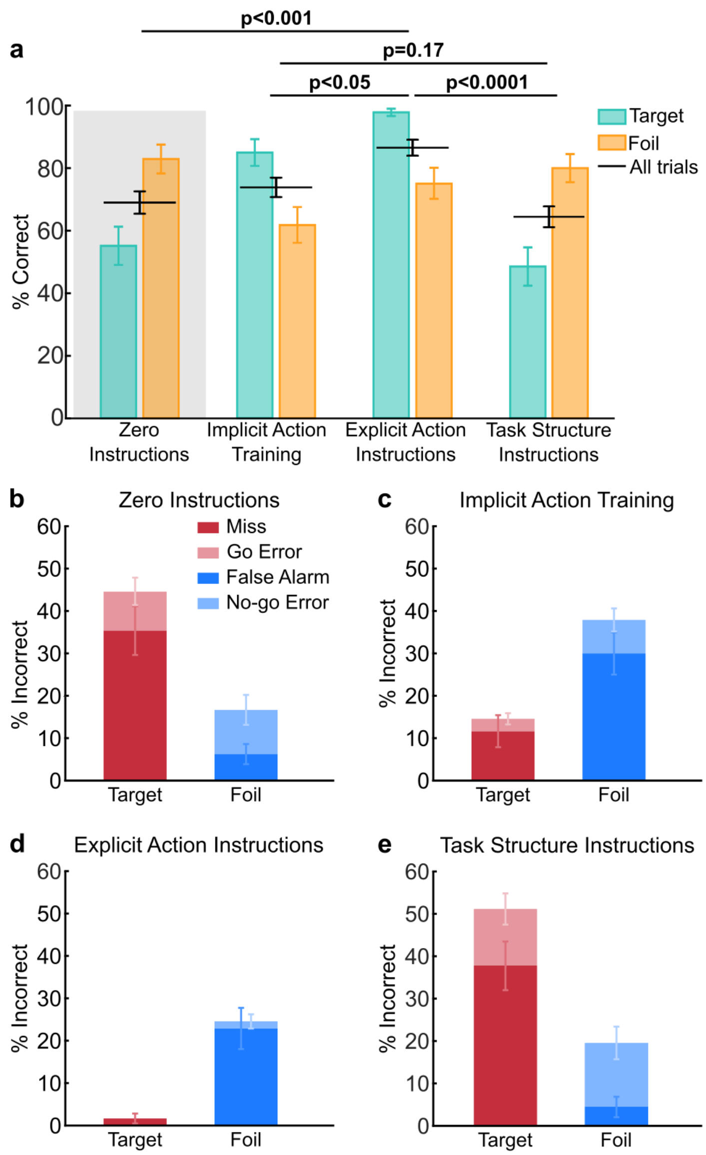



**Figure 3. The nature of instructions impacts overall performance and the underlying error types. (a)** Percent correct averaged across each subject's maximum performance block, both overall (black line) and separated by trial-type (target trials: green bar, foil trials: orange bar). When compared with subjects given zero instructions (Exp. 2), only explicit action instructions improved overall performance. Implicit action training improved hit rates but decreased correct reject rates. **(b)-(e)** Errors were classified into four types. Errors for target trials (red) could be due to inaction (misses) or incorrect key press (go errors). Errors for foil trials (blue) could be due to incorrect spacebar presses (false alarms) or other key presses (no-go errors). Mean ±SEM.

### Explicit instructions outlining the task structure had minimal impact on subject behavior

Given the effectiveness of explicit action instructions for improving performance, we next sought to test whether explicit instructions regarding the structure of the task would improve performance by increasing the likelihood of withholding. In Experiment 5, subjects were provided with written instructions to respond by either pressing or not pressing a key, without telling them which key to press. Surprisingly, these subjects performed the worst out of the three partial instruction experiments, with overall performance (M=65% correct, SD=0.26; post-hoc Exp. 5 vs Exp. 2 (overall % correct one-way ANOVA): p=0.77, 95% C.I.=[-0.16, 0.06]), hit rates (M=49% correct, SD=0.48; post-hoc Exp. 5 vs Exp. 2 (hit rate one-way ANOVA): p=0.84, 95% C.I.=[-0.24, 0.11]), and correct reject rates (M=80% correct, SD=0.35; post-hoc Exp. 5 vs Exp. 2 (correct reject rate one-way ANOVA): p=0.99, 95% C.I.=[-0.20, 0.14]) on-par with subjects given zero instructions (**Fig. 3a**).

Another unexpected result was that instructions regarding the task structure had little effect on the rate of correct rejections compared to zero instructions (**Fig. 3a**). Despite being told that not pressing a key was a valid response, many subjects still explored pressing different keys during both target and foil trials (**Fig. 3e**). Moreover, subjects in this experiment pressed multiple distinct keys per block, creating a distribution that closely resembles the key presses of subjects given zero instructions (**Extended Data Figs. 1b, 1e**). These data suggest that, for non-learners, despite being explicitly told that not pressing a key was a valid option, these subjects did not internalize behavioral inhibition into their exploratory strategy. This can also be seen in Exp. 1 where significantly more subjects, even when given full, explicit instructions, press spacebar on the first trial but do not try, on average, to withhold until around trial 6 (**Extended Data Fig. 2b**, $p<0.0001$, Wilcoxon rank sum test).

### A two-response task complements core results

Lastly, we tested whether our results were unique to a go/no-go paradigm due to the requirement for exploration via behavioral inhibition. We trained subjects on a two-response task. The structure of this task mirrored the existing go/no-go paradigm, except that the 'no-go' component was replaced with a return key press, requiring the subject to learn the mapping of the stimuli to their respective key press (the spacebar or the return key). Subjects who participated in this task were given either full instructions (Exp. 6), zero instructions (Exp. 7), or implicit action training (Exp. 8). Subjects given full instructions in the two-response mapping task (Exp. 6) easily learned the task with little variability between subjects (**Fig. 4**). Subjects reached near-perfect overall accuracy in their maximum performance block (**Fig. 4a-b,** M=97% correct, SD=0.11). Subjects also reached almost perfect performance on both spacebar and return trials (**Fig. 4b,** spacebar: M=99% correct, SD=0.03; return: M=95% correct,

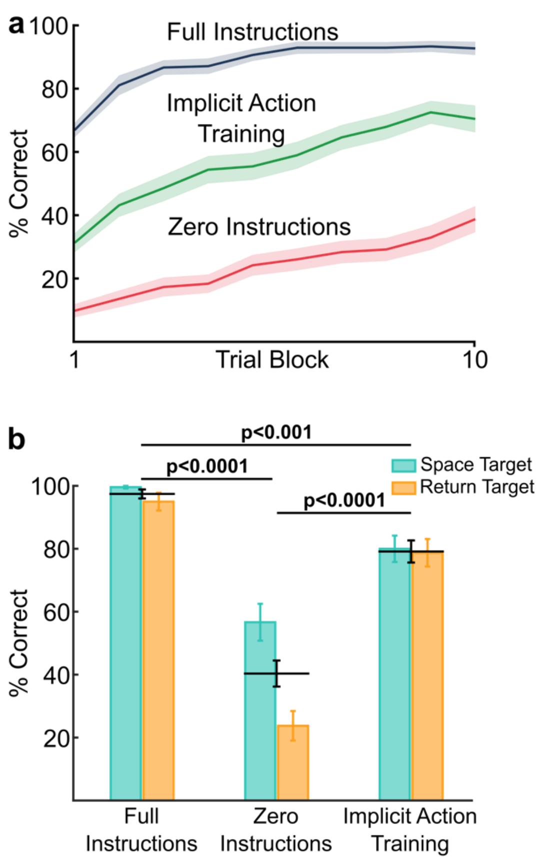



**Figure 4. Replacing withholding with a return key press recapitulates our previous results. (a)** Average percent correct values from trial block 1 to trial block 10 (8 trials per block) for the two-response mapping task with Full Instructions (Exp. 6), Zero Instructions (Exp. 7), and Implicit Action Training (Exp. 8). **(b)** Percent correct averaged across each subject's maximum performance block, both overall (black line) and separated by trial-type (spacebar target trials: green bar, return key target trials: orange bar). **(c)-(e)** Individual Learning Trajectories for Experiments 6-8. Individual Learning Trajectories for Experiments 6-8. All subjects were sorted by final block percent correct, from least to most accurate. Mean ±SEM.

SD=0.22). In contrast, subjects failed to adequately learn the two-response mapping task when given zero instructions (**Fig. 4a-b, d**, Exp.7; M=40% correct, SD=0.32; overall % correct one-way ANOVA: $F(2, 177)=80.9$, $p<0.0001$; post-hoc Exp. 6 vs

Exp. 7: $p<0.0001$, 95% C.I.=[-0.68, -0.46]). Subjects given zero instructions had significantly lower spacebar hit rates (M=57% correct, SD=0.45; spacebar hit rate one-way ANOVA: $F(2, 177)=26.65$, $p<0.0001$; post-hoc Exp. 6 vs Exp. 7: $p<0.0001$, 95% C.I.=[-0.57, -0.29]) and lower return key hit rates (M=24% correct, SD=0.36; return hit rate one-way ANOVA: $F(2, 177)=85.19$, $p<0.0001$; post-hoc Exp. 6 vs Exp. 7: $p<0.0001$, 95% C.I.=[-0.57, -0.29]) compared to subjects given full instructions (**Fig. 4b**). Subjects also, on average, had a higher spacebar hit rate compared to the return hit rate; this could potentially be because the spacebar is larger and more centrally located on the keyboard, making it more salient over the return key.

Subjects in Experiment 8 were given implicit action training on both the spacebar and the return key, gaining implicit information about both actions of interest. Here, we see that subjects given implicit action training in the two-response mapping task were able to recover a large proportion of performance lost when zero instruction were given. Subjects given implicit action training showed a significant increase in overall accuracy (M=79% correct, SD=0.27; post-hoc Exp. 8 vs Exp. 7 (overall % correct one-way ANOVA): $p<0.0001$, 95% C.I.=[0.28, 0.50]), spacebar hit rate (M=80% correct, SD=0.32; post-hoc Exp. 8 vs Exp. 7 (spacebar hit rate one-way ANOVA): $p<0.001$, 95% C.I.=[0.10, 0.37]), and return key hit rate (M=79% correct, SD=0.34; post-hoc Exp. 8 vs Exp. 7 (return hit rate one-way ANOVA): $p<0.0001$, 95% C.I.=[0.42, 0.68]), compared to subjects given zero instructions (**Fig. 4a-b**). This increase in performance in Experiment 8 is in stark contrast to the effect of implicit action training in the go-no-go task (Experiment 3). In Experiment 8, subjects were likely able to internalize that there were two possible alternatives (spacebar and enter) from the implicit training while in the go/no-go task, subjects had to infer that not pressing any key was, indeed, a valid option. However, these subjects still fell well short of subjects given full instructions in overall accuracy (post-hoc Exp. 8 vs Exp. 6 (overall % correct one-way ANOVA): $p<0.001$, 95% C.I.=[-0.28, -0.07]), spacebar hit rate (post-hoc Exp. 8 vs Exp. 6 (spacebar hit one-way ANOVA): $p<0.01$, 95% C.I.=[-0.33, -0.06]), and return key hit rate (post-hoc Exp. 8 vs Exp. 6 (return hit rate one-way ANOVA): $p<0.05$, 95% C.I.=[-0.30, -0.03]), indicating that while the implicit information gained from key press training improved performance and provided subjects with a more complete model of the task, explicit instructions are still necessary to reach peak performance (**Fig. 4a-b**).

# Discussion

A significant portion of how we understand and cope with the world is acquired by observing, imitating, and receiving explicit instruction from others [39]. Even when we learn via direct observation and reflection, human adults typically do so while immersed in a meaningful context, with a wealth of prior knowledge about their environment. But what happens when this scaffolding is removed? How effectively can we bootstrap learning in an information and context-barren environment? Are there (minimal) kinds of information that support learning better than others?

Here, we demonstrate that robust individual differences during learning emerge in information-poor contexts that are completely absent when learners are provisioned with clear, explicit instructions. In seeking to understand how effectively humans learn from 'zero' in a series of simple stimulus recognition tasks, we found that fewer than half of the subjects learned these tasks without instructions despite receiving trial-by-trial feedback and monetary reward. We reasoned that providing explicit instructions about the structure of the task (i.e., the general rule governing the task: pressing a key or not pressing a key) could improve learning. This type of 'rule instruction' is a common pedagogical approach in which educators instruct based on conceptual rather than concrete scenarios. However, an unexpected result was obtained: explicit instructions about the task structure did not improve performance nor reduce individual variability. Instead, explicit instructions that constrained the action space (i.e., valid key presses) partially recovered performance but still pointed to action inhibition as particularly challenging to learn without explicit instructions. This is in consonance with previous results showing that action inhibition is harder than action [38,40] and, more generally, that doing nothing is a difficult exploratory state for humans [41,42]. Taken together, we find that individual variability in the learning of even extremely simple tasks is highly sensitive to the nature and form of explicit instructions. The nature of the instructional scaffolding heavily influence performance and learning differences.

These results suggest that information- and context -impoverished environments create the veneer of individual variability that does not necessarily reflect differences in underlying capacity. By recruiting many subjects (almost 500 across experiments), we show that individual differences in learning are largely driven by the nature of instructions rather than underlying capacity. These findings have major implications for how we interpret cognitive performance both in experimental and real-word settings. Moreover, our data show that small changes in the nature and content of instructions (in our case, explicit limits on action exploration) can lead to profound changes in final performance and cross-subject variability. Explicit instruction that limits action exploration improves performance and reduces variance far more than implicit behavioral shaping and explicit instructions regarding general task structures. This lesson is of utmost importance for how we understand and promote learning when individuals have little to no background knowledge, such as when learning a new subject or entering a radically different context. A direct, actionable instruction that reduces exploration produces more efficient learning outcomes compared to laying out a more general, albeit abstract and therefore not immediately actionable, picture of the task at hand.

Another implication is that luck plays a disproportionate role in learning outcomes. In the absence of any explicit guidance, subjects who's first allowed action was correct (which occurs by random chance), performed much better than subjects who despite performing the right action (i.e., pressing spacebar or withholding) were unlucky to do so at the wrong time (when the 'wrong' stimulus was on the screen). This finding reveals the vicissitudes of learning and how underperformers might lag behind for no fault of their own.

Unlike human infants who start learning about the world without any background and therefore rely on core principles and non-linguistic cues to accrue knowledge [43,44], human adults may favor learning from explicit linguistic instruction rather than purely from experience. In our study, we treat humans as we would non-human animals in a standard laboratory task [45–50]. We mimicked, as closely as possible, the task structure (including stimulus presentation, response window, feedback) of rodent learning tasks. In such a scenario, it is surprising that more than half of the human subjects were unable to learn the task (despite clear evidence that they were trying to). While there is ample evidence of human implicit learning in different domains (with and without feedback) [51,52] these data suggest that many humans do not immediately learn from trial-level feedback (especially when no other context is provided). Instead, subjects potentially adopt a 'model-based' strategy [34,53–57], continually testing abstract models of the world rather than privileging low-level feedback to guide future actions.

Human learning rarely takes place in a complete void. This is why many can learn Calculus even if only a couple humans have ever invented it. However, it is easy to underestimate the crucial role played by subtle background knowledge, explicit instruction, and luck in learning even the simplest of tasks. Future work should explore how these variables interact in more complex, real-world scenarios where historically disadvantaged backgrounds, ambiguous feedback, and/or hidden curricula are at play.

## Acknowledgements

We would like to thank J. Halberda and C. Firestone for their helpful comments on earlier drafts of the manuscript. This work was partially supported by an NIH PREP fellowship to LW (NIH R25GM109441).

## Methods

### Open science

All our experimental procedures and main analysis were preregistered (https://aspredicted.org/36bv9.pdf and https://aspredicted.org/2re4s.pdf). Raw data, stimuli, experimental and data analysis code can be accessed at https://osf.io/k3gdq/).

### Subjects

This study includes data from 480 participants (60 participants for each of the eight experiments) who were recruited online via the participant recruitment platform Prolific (www.prolific.co). We prescreened our participants to be located in the US, between the ages of 18 and 40, native English speakers, and have normal or corrected-to-normal vision. At the start of the experiment, all participants consented to being included in the research study. We excluded participants who were unable to complete the study and participants with corrupted or incomplete data files. Participants received monetary compensation for their completion of the study, including a base rate of $9.32/hr and additional bonus payments of $0.01 for every correct response throughout the task. All experiments were approved by the Johns Hopkins Homewood Institutional Review Board.

### Stimuli

At the beginning of each trial, a white fixation cross appeared in the middle of the screen, occupying 6% of the window's height and width. This was followed by the presentation of one of the stimuli. Stimuli consisted of simple, white geometric shapes (triangle, circle, star, square) presented against a mid-gray background. These shapes appeared in the center of the screen occupying 25% of the window's height and width. A white question mark occupying 10% of the window's height and width was centrally presented after stimulus offset, indicating that the subject's response was expected. A 1¢ or 0¢ white legend, occupying 10% of the window's height and width, was used to provide feedback on each trial for correct and incorrect responses, respectively.

### General procedure

All experiments presented a stimulus-action mapping task to subjects, where they must discover which keypress corresponded to which stimuli. All experiments presented the same stimuli, but the actions (keypresses) differ between Experiments 1-5 and Experiments 6-8. Experiments 1-5 followed a simple go/no-go task structure in which two of the four shape stimuli were associated with 'go' (pressing the spacebar on the keyboard) and the other two shapes were associated with 'no-go' (not pressing any key on the keyboard); stimulus-action assignments were counterbalanced across subjects. Trials began with a small white fixation cross displayed on the computer screen for 2 seconds followed by the presentation of a shape for 500 milliseconds. Next, a white question mark was displayed on the computer screen for 2.5 seconds during which the participant had to make their response: press the spacebar for 'go' trials or withhold pressing any key for 'no-go' trials. Participants then were provided feedback by receiving either a 1¢ or 0¢ reward for correct or incorrect responses, respectively (see Fig. 1a). A trial block consisted of 4 'go' trials and 4 'no-go' trials, and the task consisted of 10 trial blocks (80 trials in total). At the end of each trial block, the participants' total cumulative earnings were briefly displayed on the computer screen, and they were asked to press the return key on the keyboard to advance to the next trial block. At the end of the task, participants were given a brief multiple-choice survey that asked about their experience with completing online studies and their confidence in figuring out the task correctly. The survey also asked how participants responded when seeing each stimulus during the task; subjects indicated whether they pressed return, the spacebar, pressed another key besides the spacebar, did not press any key, or none of the above. Completing the experiment required approximately 10-15 minutes.

Experiments 6-8 followed a modified structure in which the 'no-go' response was replaced by a 'return' key press. This resulted in a two-response mapping learning task, where one pair of shapes was associated with pressing the spacebar and another pair was associated with pressing the return key on the keyboard; these stimulus-action assignments were also counterbalanced across subjects. Experiments 6-8 followed the same general procedure and were approximately the same length as experiments 1-5. Participants were asked to press the 'a' key on the keyboard instead of the return key to advance between trial blocks to reduce bias in learning the proper stimulus-response mapping.

**Go/no-go task procedures (Fig. 1b)**

*General instructions:* Subjects in all experiments were explicitly informed that their performance could earn them up to 50% extra payment. They were also informed about the relevant time period for earning such a payment (while the question mark was on the screen) and that their goal was to figure out how to earn it. The rest of each experiment proceeded as follows:

*Experiment 1—Full instructions.* Subjects received full, explicit instructions about the nature of the task and how to navigate it. They were informed that a reward was associated with pressing the spacebar when some shapes appeared ('go' or 'target' trial) and not pressing any key when some other shapes appeared ('no-go' or 'foil' trial) and that their goal was to discover which shapes corresponded to which contingency. Subjects were also explicitly told that they needed to press the spacebar or withhold from pressing any key when the question mark appeared. Subjects were informed about the presence of feedback. Subjects saw examples of the relevant experimental screens, and they performed four practice trials before starting the experiment. No feedback about the shape-response association was provided during the practice trials, but if subjects pressed a key different from the spacebar, they were reminded that the only allowable actions were pressing the spacebar or no key at all.

*Experiment 2—Zero instructions.* Except for the general instructions, subjects received no further information about the task or its goal, and the experimental trials started promptly.

*Experiment 3—Implicit action training.* In addition to the general instructions, before starting the experiment subjects saw a blank screen with a question mark and were required to discover that pressing the spacebar was the action of interest by exploring different actions with their keyboard or mouse. Subjects received feedback via the presentation of '1¢' on their screen after each spacebar press. After subjects pressed the spacebar five times (not necessarily consecutively), the experiment began.

*Experiment 4—Explicit action instructions.* In addition to the general instructions, subjects were explicitly told that the key that mattered in the task was the spacebar.

*Experiment 5—Task structure instructions.* In addition to the general instructions, subjects were told that they could respond by pressing a key on the keyboard or not pressing a key at all. Subjects were not provided with information revealing that the spacebar was the key of interest.

**Two-response mapping task**

*Experiment 6—Full instructions.* Subjects received instructions identical to subjects in Experiment 1 except that subjects were tasked with discovering which shapes required them to press the spacebar and which shapes required them to press the return key (instead of not pressing any key) to earn a reward.

*Experiment 7—Zero instructions.* Subjects received instructions identical to subjects in Experiment 2.

*Experiment 8—Implicit action training.* Subjects received instructions identical to subjects in Experiment 3 except that subjects implicitly trained for both the spacebar and the return key and were required to press each key five times before advancing to the experiment.

### Data Collection & Analysis

All experiments were designed in PsychoPy v2021.1.4 and launched online using Pavlovia, an online behavioral experiment platform, to host the study. Prolific, an online research subject recruitment platform, was used to recruit subjects; subjects who were recruited were directed via a URL to complete the study. The primary experimental data was collected from Pavlovia and additional demographic data was collected from Prolific. Unless otherwise, the number of subjects in each group for all statistical analyses is n=60.

All analyses were performed using MATLAB R2021a and, unless otherwise stated, all statistical values are calculated using each subject's maximum performance block (i.e. the block with the highest performance throughout the experiment); if a subject had two or more blocks of their maximum performance, we used the first of their maximum performance blocks. We use maximum performance block instead of total average performance to distinguish performance from block to block, allowing us to investigate learning trajectories. Note that the use of final block performance instead of maximum block performance produces similar statistical results.

We use hits to denote pressing the spacebar when the target stimulus is presented (Exps. 1-5); in Experiments 6-8, spacebar hits denote correctly pressing the spacebar for target stimuli and return key hits denote correctly pressing the return key for foil stimuli. Correct rejections are foil trials where subjects correctly withhold from pressing any key (Exps. 1-5). False alarms denote spacebar presses when the foil stimuli is presented, and misses are target trials where the subject does not press any key (Exps. 1-5). Additionally, go errors refer to key presses other than spacebar during target trials and no-go errors refer to key presses other than spacebar during foil trials. Unless otherwise noted, all statistical analyses were one-way Analysis of Variance (ANOVA) tests, and all post-hoc tests were Tukey's honestly significant difference [HSD].

## References

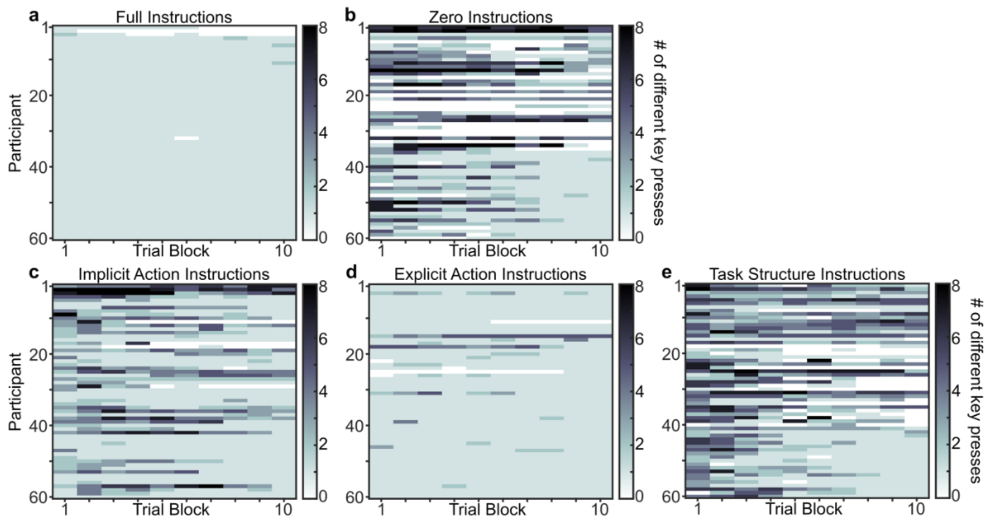

1. Csibra, G. & Gergely, G. Natural pedagogy. *Trends Cogn. Sci.* **13**, 148–153 (2009).
2. Gergely, G., Egyed, K. & Király, I. On pedagogy. *Dev. Sci.* **10**, 139–146 (2007).
3. Koenig, M. A. & Harris, P. L. Preschoolers Mistrust Ignorant and Inaccurate Speakers. *Child Dev.* **76**, 1261–1277 (2005).
4. LeFevre, J.-A. & Dixon, P. Do Written Instructions Need Examples? *Cogn. Instr.* **3**, 1–30 (1986).
5. Callander, S. Searching and Learning by Trial and Error. *Am. Econ. Rev.* **101**, 2277–2308 (2011).
6. Hull, C. L. Simple trial and error learning: A study in psychological theory. *Psychol. Rev.* **37**, 241–256 (1930).
7. Starch, D. A demonstration of the trial and error method of learning. *Psychol. Bull.* **7**, 20–23 (1910).
8. Young, H. P. Learning by trial and error. *Games Econ. Behav.* **65**, 626–643 (2009).
9. Gijsen, S., Grundei, M. & Blankenburg, F. Active inference and the two-step task. *Sci. Rep.* **12**, 17682 (2022).
10. Keenan, T., Ruffman, T. & Olson, D. R. When do children begin to understand logical inference as a source of knowledge? *Cogn. Dev.* **9**, 331–353 (1994).
11. Sodian, B. & Wimmer, H. Children's Understanding of Inference as a Source of Knowledge. *Child Dev.* **58**, 424 (1987).
12. Bonawitz, E. *et al.* The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition* **120**, 322–330 (2011).
13. Cohen, J. D., McClure, S. M. & Yu, A. J. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philos. Trans. R. Soc. B Biol. Sci.* **362**, 933–942 (2007).
14. Gijlers, H. & de Jong, T. The relation between prior knowledge and students' collaborative discovery learning processes. *J. Res. Sci. Teach.* **42**, 264–282 (2005).
15. O'Neill, D. K., Astington, J. W. & Flavell, J. H. Young Children's Understanding of the Role That Sensory Experiences Play in Knowledge Acquisition. *Child Dev.* **63**, 474 (1992).
16. Stahl, A. E. & Feigenson, L. Observing the unexpected enhances infants' learning and exploration. *Science* **348**, 91–94 (2015).
17. Kaplan, A. S. & Murphy, G. L. Category learning with minimal prior knowledge. *J. Exp. Psychol. Learn. Mem. Cogn.* **26**, 829–846 (2000).
18. Murphy, G. L. & Allopenna, P. D. The locus of knowledge effects in concept learning. *J. Exp. Psychol. Learn. Mem. Cogn.* **20**, 904–919 (1994).
19. Shapiro, A. M. How Including Prior Knowledge As a Subject Variable May Change Outcomes of Learning Research. *Am. Educ. Res. J.* **41**, 159–189 (2004).
20. Hammadou, J. Interrelationships among Prior Knowledge, Inference, and Language Proficiency in Foreign Language Reading. *Mod. Lang. J.* **75**, 27 (1991).

21. Williams, J. J. & Lombrozo, T. Explanation and prior knowledge interact to guide learning. *Cognit. Psychol.* **66**, 55–84 (2013).
22. Desmedt, O. *et al.* Contribution of Time Estimation and Knowledge to Heartbeat Counting Task Performance under Original and Adapted Instructions. *Biol. Psychol.* **154**, 107904 (2020).
23. Doll, B. B., Jacobs, W. J., Sanfey, A. G. & Frank, M. J. Instructional control of reinforcement learning: A behavioral and neurocomputational investigation. *Brain Res.* **1299**, 74–94 (2009).
24. Kelly, V. E., Eusterbrock, A. J. & Shumway-Cook, A. The Effects of Instructions on Dual-Task Walking and Cognitive Task Performance in People with Parkinson's Disease. *Park. Dis.* **2012**, e671261 (2012).
25. Plummer, P., Grewal, G., Najafi, B. & Ballard, A. Instructions and skill level influence reliability of dual-task performance in young adults. *Gait Posture* **41**, 964–967 (2015).
26. Gubner, N. R., Wilhelm, C. J., Phillips, T. J. & Mitchell, S. H. Strain Differences in Behavioral Inhibition in a Go/No-go Task Demonstrated Using 15 Inbred Mouse Strains. *Alcohol. Clin. Exp. Res.* **34**, 1353–1362 (2010).
27. Han, Z., Zhang, X., Zhu, J., Chen, Y. & Li, C. T. High-throughput automatic training system for odor-based learned behaviors in head-fixed mice. *Front. Neural Circuits* **12**, (2018).
28. Han, Z. *et al.* Awake and behaving mouse fMRI during Go/No-Go task. *NeuroImage* **188**, 733–742 (2019).
29. Holfoth, D. P., Neilans, E. G. & Dent, M. L. Discrimination of partial from whole ultrasonic vocalizations using a go/no-go task in mice. *J. Acoust. Soc. Am.* **136**, 3401–3409 (2014).
30. Moore, S. & Kuchibhotla, K. V. Slow or sudden: Re-interpreting the learning curve for modern systems neuroscience. *IBRO Neurosci. Rep.* **13**, 9–14 (2022).
31. Arican, C., Bulk, J., Deisig, N. & Nawrot, M. P. Cockroaches Show Individuality in Learning and Memory During Classical and Operant Conditioning. *Front. Physiol.* **10**, 1539 (2020).
32. Carandini, M. & Churchland, A. K. Probing perceptual decisions in rodents. *Nat. Neurosci.* **16**, 824–831 (2013).
33. Dolan, R. J. & Dayan, P. Goals and Habits in the Brain. *Neuron* **80**, 312–325 (2013).
34. Gershman, S. J. & Niv, Y. Learning latent structure: carving nature at its joints. *Curr. Opin. Neurobiol.* **20**, 251–256 (2010).
35. Hanks, T. D. & Summerfield, C. Perceptual Decision Making in Rodents, Monkeys, and Humans. *Neuron* **93**, 15–31 (2017).
36. Newsome, W. T. & Pare, E. B. A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J. Neurosci.* **8**, 2201–2211 (1988).
37. Rosenberg, M., Zhang, T., Perona, P. & Meister, M. Mice in a labyrinth show rapid learning, sudden insight, and efficient exploration. *eLife* **10**, e66175 (2021).

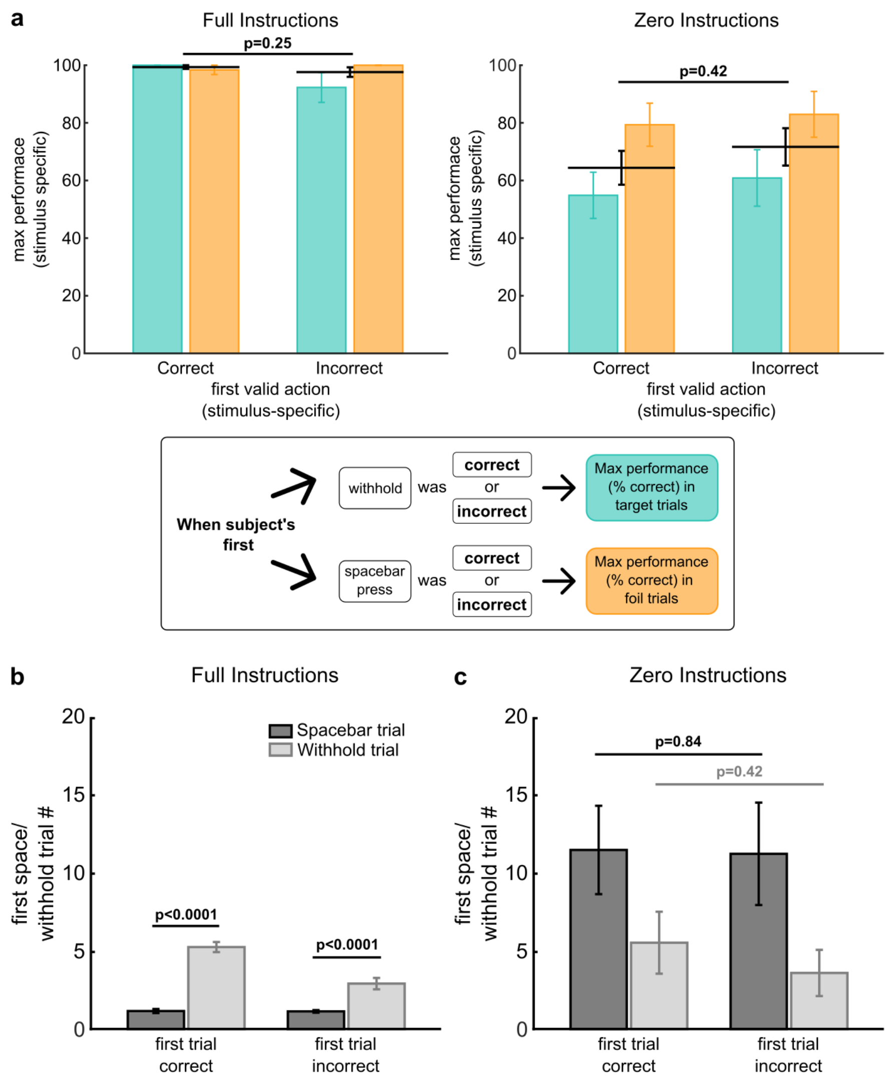
38. Kuchibhotla, K. V. *et al.* Dissociating task acquisition from expression during learning reveals latent knowledge. *Nat. Commun.* **10**, 2151 (2019).
39. Gariépy, J.-F. *et al.* Social learning in humans and other animals. *Front. Neurosci.* **8**, (2014).
40. Albayay, J., Castiello, U. & Parma, V. Task-irrelevant odours affect both response inhibition and response readiness in fast-paced Go/No-Go task: the case of valence. *Sci. Rep.* **9**, 19329 (2019).
41. Westgate, E. C. Why Boredom Is Interesting. *Curr. Dir. Psychol. Sci.* **29**, 33–40 (2020).
42. Wilson, T. D. *et al.* Just think: The challenges of the disengaged mind. *Science* **345**, 75–77 (2014).
43. Carey, S. The Origin of Concepts. *J. Cogn. Dev.* **1**, 37–41 (2000).
44. Kenderla, P. & Kibbe, M. M. Explore versus store: Children strategically trade off reliance on exploration versus working memory during a complex task. *J. Exp. Child Psychol.* **225**, 105535 (2023).
45. Delgado, M. R., Olsson, A. & Phelps, E. A. Extending animal models of fear conditioning to humans. *Biol. Psychol.* **73**, 39–48 (2006).
46. Kopec, C. D. & Brody, C. D. Controlling for temporal discounting shifts rats from geometric to human-like arithmetic bisection. 245092 Preprint at https://doi.org/10.1101/245092 (2018).
47. Meier, P. & Reinagel, P. Rats and humans differ in processing collinear visual features. *Front. Neural Circuits* **7**, (2013).
48. Shevinsky, C. A. & Reinagel, P. The Interaction Between Elapsed Time and Decision Accuracy Differs Between Humans and Rats. *Front. Neurosci.* **13**, (2019).
49. Smith, J. D., Beran, M. J., Redford, J. S. & Washburn, D. A. Dissociating uncertainty responses and reinforcement signals in the comparative study of uncertainty monitoring. *J. Exp. Psychol. Gen.* **135**, 282–297 (2006).
50. Sweis, B. M. *et al.* Sensitivity to "sunk costs" in mice, rats, and humans. *Science* **361**, 178–181 (2018).
51. Rmus, M., Ritz, H., Hunter, L. E., Bornstein, A. M. & Shenhav, A. Humans can navigate complex graph structures acquired during latent learning. *Cognition* **225**, 105103 (2022).
52. Watanabe, T., Náñez, J. E. & Sasaki, Y. Perceptual learning without perception. *Nature* **413**, 844–848 (2001).
53. Castro-Rodrigues, P. *et al.* Explicit knowledge of task structure is a primary determinant of human model-based action. *Nat. Hum. Behav.* **6**, 1126–1141 (2022).
54. Daw, N. D., Niv, Y. & Dayan, P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* **8**, 1704–1711 (2005).
55. Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. Model-Based Influences on Humans' Choices and Striatal Prediction Errors. *Neuron* **69**, 1204–1215 (2011).

56. Decker, J. H., Otto, A. R., Daw, N. D. & Hartley, C. A. From Creatures of Habit to Goal-Directed Learners: Tracking the Developmental Emergence of Model-Based Reinforcement Learning. *Psychol. Sci.* **27**, 848–858 (2016).
57. Shteingart, H. & Loewenstein, Y. Reinforcement learning and human behavior. *Curr. Opin. Neurobiol.* **25**, 93–98 (2014).

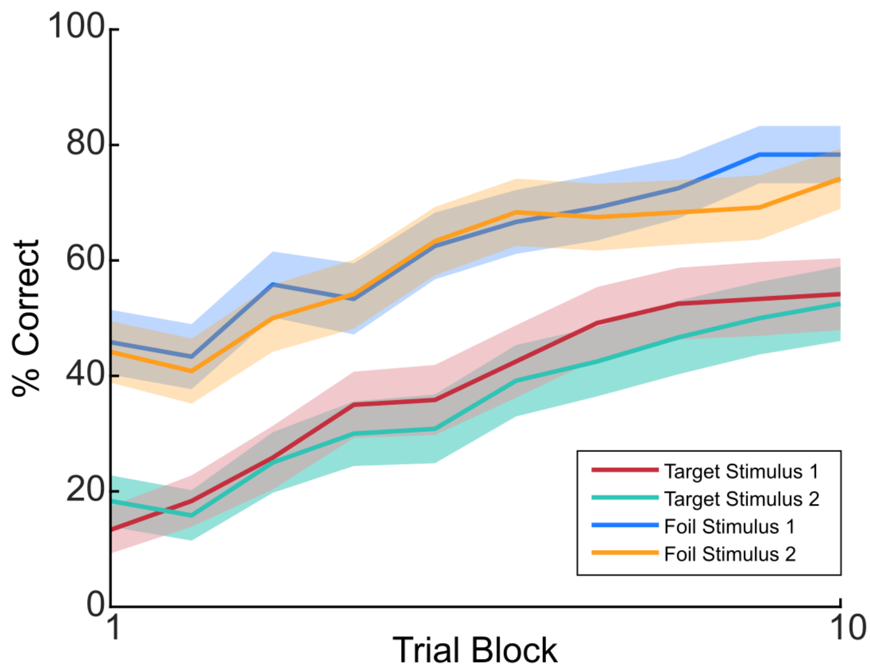
# Extended Data




**Extended Data Figure 1. Individual variability in key press behavior depends on nature of instructions. (a)-(e)** All subjects were sorted by final block percent correct, from least to most accurate. Most subjects given full **(a)** or explicit action **(d)** instructions in the GNG task pressed only one distinct key per block. In contrast, subjects given zero **(b)**, implicit action **(c)**, or task structure **(e)** instructions are shown to press multiple distinct keys per block. For some subjects, the action space was extensive, especially in **(b)**.




**Extended Data Figure 2. (a) The effect of a subject's first valid action on performance is specific to that action's appropriate trial type (target trials for spacebar presses and foil trials for withholds).** Maximum performance (% correct) in target (green bar), foil (orange bar), and all trials (black line) when the subjects' first valid action (either a spacebar press or a withhold) was correct (Exp. 1: n=78, Exp. 2: n=59) or incorrect (Exp. 1: n=42; Exp. 2: n=45). **(b) Subjects press the spacebar significantly earlier than they withhold when given full, explicit instructions.** Trial number when a subject in Exp. 1, on average, first presses the spacebar (dark grey bar) or withholds (light grey bar) depending on whether that action is correct or incorrect. Full instructions: spacebar correct (n=31) vs. incorrect (n=29), withhold correct (n=47) vs. incorrect (n=13). **(c) Subjects press the spacebar and withhold on similar trial numbers regardless of whether that action was correct or incorrect.** Trial number when a subject in Exp. 2, on average, first presses the spacebar (dark grey bar) or withholds (light grey bar) depending on whether that action is correct or incorrect. Zero Instructions: spacebar correct (n=23) vs. incorrect (n=22), withhold correct (n=36) vs. incorrect (n=23). Mean ±SEM.




**Extended Data Figure 3.** Subjects, on average, learn both target stimuli and both foil stimuli at similar rates. Average percent correct values from trial block 1 to trial block 10 (8 trials per block) for each of the 4 counter- balanced stimuli (2 target and 2 foil stimuli) in Exp. 2 (Zero Instructions). Mean ±SEM.