An integrative framework for the human sense of control

Alireza Modirshanechi[1,2,*], Peter Dayan[2,3,†], and Eric Schulz[1,†]

[1] Helmholtz Munich, Munich, Germany
[2] Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[3] University of Tübingen, Tübingen, Germany

[*] Corresponding author: alireza.modirshanechi@helmholtz-munich.de
[†] These authors contributed equally to this work

Abstract

Being empowered and having a sense of control are central determinants of human well-being. However, there is currently no agreed-upon definition for the sense of control. We offer a prospective characterization as '*a priori* readiness to perform tasks that will be announced in the future.' We show that this highlights three underlying factors that variously underpin previous conceptions of control: (i) action availability, (ii) *certain* achievability of potentially desired outcomes, and (iii) *possible* achievability of potentially desired outcomes. We formalize this characterization with a measure of controllability that precisely balances these factors via a single parameter. Our proposed measure accurately predicts the great majority of human preferences in two novel, theory-guided experiments, where participants prepare themselves to collect future rewards or avoid future punishments. Notably, our formulation captures substantial differences in participants' preference for certainty versus possibility. Our findings collectively provide instrumental insights into how individuals sense control.

Introduction

In everyday life, we constantly navigate uncertain environments and attempt to steer the course of events in our favor. Consider a familiar situation: planning your weekend. You may have several activities in mind—some reliably available, others more uncertain. At times, all your options might be accessible; at others, only a subset may be offered, or availability may shift unexpectedly. Even when the number of possible outcomes stays the same, the structure of the environment, its predictability, and how directly your choices influence what happens, can drastically alter your experience of control. Across tasks as mundane as choosing a weekend activity or as consequential as planning a career, our sense of control shapes how we decide, how we learn, and how we feel. This sense is not just pleasant, it is motivational, empowering, and even essential. It determines how much effort we invest, how resilient we are in the face of setbacks, and how we interpret the consequences of our actions. But what does it actually mean to be in control? And what factors determine people's sense of control?

Notions of 'controllability' and 'empowerment' have played a central role in both neuroscience and psychology[1–5], as well as in computer science and machine learning[6–8]. From an evolutionary perspective, the desire for a sense of control has been described as a psychological and biological necessity[1], with far-reaching implications for our general well-being[1–3] and developmental trajectories[4,5]. In artificial intelligence, a parallel drive for control has motivated the design of agents capable of acquiring complex skills with no external rewards or supervision[6,7]. Despite this convergence, however, there is currently no consensus on (i) the objective question of what it means to be in control of an environment, or (ii) the subjective question of when *we*, as humans, experience a sense of control.

From an objective viewpoint, several engineering and mathematical measures of controllability have been proposed. These measures focus on different environmental factors. For example, conventional control-theoretic measures — concerned with deterministic, continuous environments — quantify controllability in terms of the diversity, or more precisely, the dimensionality, of states that an agent can reach by taking appropriately chosen actions[9,10]. Alternative information-theoretic measures — more often concerned with discrete, stochastic environments — quantify a similar notion by measuring the diversity of reachable next states through the maximum state entropy[11]. Yet other measures take a somewhat orthogonal approach and, using the notion of channel capacity[12], quantify controllability regarding how distinctly the agent's available actions influence the environmental state[13–15].

From a subjective viewpoint, experimental quantities related to objective controllability consistently influence human behavioral[16–18] and physiological responses[19,20]. For instance, objective factors related to environmental controllability have been found to enhance motivation[21], increase happiness[22], and improve resilience to aversive events[23]. These findings provide strong evidence that humans dynamically infer *some* quantities related to the environment's controllability and adapt their behavior accordingly — in line with longstanding psychological theories[24–27].

However, the relationships both within and between the objective and subjective measures remain underexplored. In particular, the absence of a clearly formalized, unifying notion of controllability has made it impossible to characterize the environmental factors that govern the human sense of control precisely — or indeed describe crisply how these factors might differ across individuals.

Here, we offer a precise, yet interpretable, formulation of controllability as *a priori* readiness for future tasks. This perspective unifies existing notions within a coherent, integrative framework and can be applied both to studying the human sense of control and to designing machine learning agents. We focus on the former, and across two theory-guided, large-scale experiments, demonstrate that our formulation accurately explains the human subjective sense of control. We reveal that this sense is shaped by the interplay between action availability and the possible and certain reachability of potentially desirable outcomes. Moreover, we show substantial variability in how individuals weigh these different factors in their preferences. Thus, our work provides a solid foundation for future studies of controllability and the sense of control in both biological and artificial agents.

Results

We motivate our theoretical framework by developing our opening example into a thought experiment (Fig. 1A-B). Imagine your task is to pick an activity for the weekend, and you have four options in mind: going to a painting class, the circus, the theater, or a sports bar (Fig. 1A). Depending on where you live, only some activities may be accessible. For example, imagine three cities represented by c_1, c_2, and c_3 (Fig. 1A). In c_1, all activities are available every weekend (Fig. 1A1), while in c_2, there is only a painting class and a theater (Fig. 1A2). In c_3, on the other hand, there is only a sports bar and a circus, but on randomly chosen weekends, the circus venue hosts a painting class instead (Fig. 1A3). This means you cannot know beforehand whether you can attend the painting class (or even the circus) on a specific weekend. Given these scenarios, how much control would you feel, in each city, over the course of events on your weekend? It is reasonable to assume a maximal sense of control in c_1, where all desired outcomes are reliably achievable; but how should we rank c_2 and c_3?

To answer this question, we can model the thought experiment within a sequential decision-making framework[28–30], where your apartment in a given city is considered as the current state s and each activity as a potential next state s'. Then, each city's characteristics can be summarized by a corresponding transition probability $p(s'|s, a)$, denoting the likelihood by which your action a (arrows in Fig. 1A) takes you from your current state s to a subsequent state s'. Given this formalization, we can use different controllability measures to rank different cities. All existing mathematical measures assign the first city, c_1, the maximum value of controllability–consistent with common sense. However, different measures diverge in how they rank c_2 and c_3 (Methods). Some measures, such as Klyubin empowerment[13], assign equal control to c_2 and c_3, since in both cases, you have two available actions that have mutually exclusive outcomes (Fig. 1B2; also see Proposition 2). Other measures, such as maximum outcome entropy[11], predict a higher sense of control in c_3 than in c_2, since there is at least a chance of accessing the painting class if desired (Fig. 1B3). These differing predictions reflect fundamentally distinct perspectives on conceptualizing and formalizing the sense of control; but it remains unclear which perspective best matches our subjective sense of control.

We aim to propose an integrative framework that unifies these different perspectives and enables testing them against experimental data.

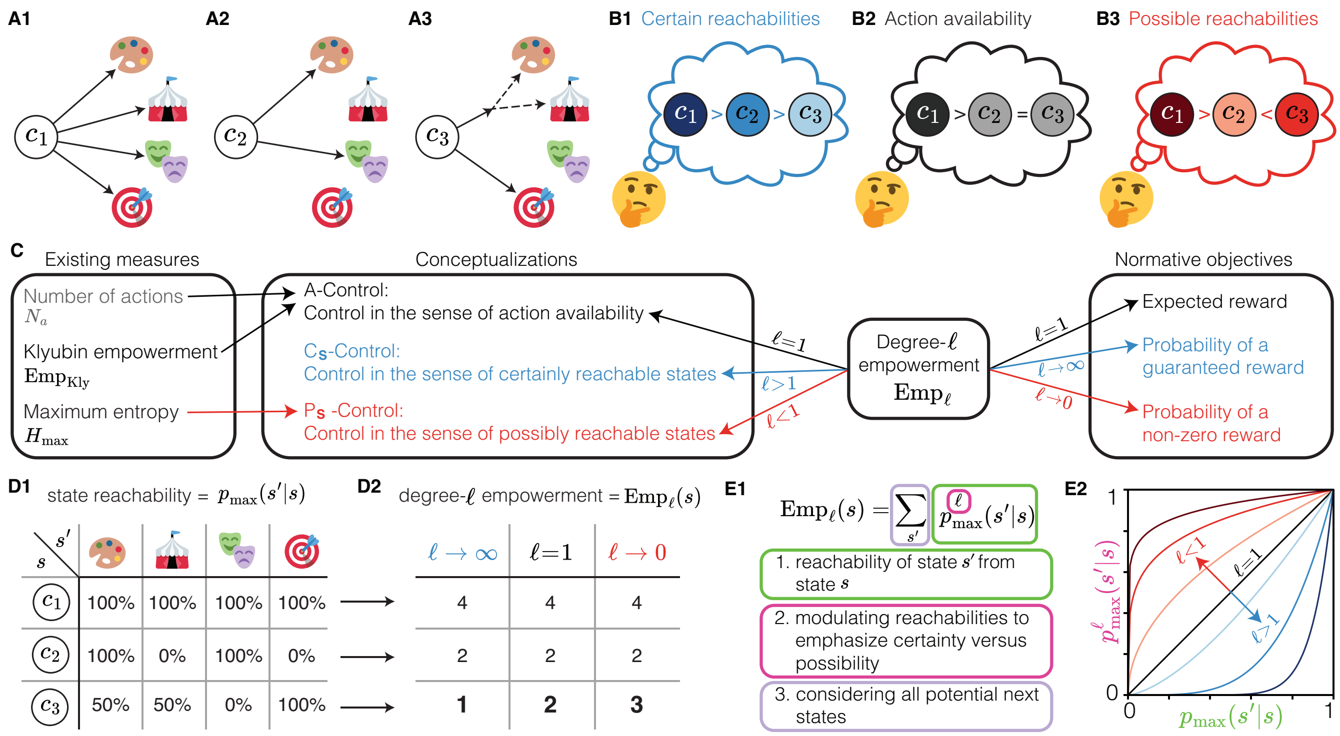

Figure 1: **An integrative framework for the sense of control. A**. Schematic of the thought experiment. We consider three cities as current states (c_1, c_2, and c_3) and four weekend activities (the painting class, the circus, the theater, and the sports bar) as potential next states. Solid arrows indicate available actions; dashed arrows indicate possible random transitions. **B**. Different conceptualizations of the sense of control yield different rankings of the cities shown in A. **C**. Framing the sense of control as '*a priori* readiness for to-be-announced tasks' bridges distinct conceptualizations and normative objectives via Emp_ℓ (Methods). **D**. Varying ℓ leads to predictions of Emp_ℓ for the rankings in A, aligning with seemingly conflicting views in B. **D1**. Reachability of each activity in the three cities. **D2**. Corresponding Emp_ℓ values for extreme choices of ℓ. Bolded values highlight how changes in ℓ affect the Emp_ℓ value assigned to c_3. **E**. Although derived from an abstract, top-down perspective, Emp_ℓ has a straightforward interpretation. **E1**. For a given state s, $\mathrm{Emp}_\ell(s)$ aggregates a modulated version of the reachabilities of all possible next states s'. **E2**. The modulation depends on the value of ℓ: increasing (decreasing) ℓ downplays (amplifies) the contribution of lower reachabilities to Emp_ℓ. *Credit: Emojis adapted from ref.* [31]*, under a CC BY 4.0 license.*

Sense of control as *a priori* readiness

We begin with the intuitive assumption that the sense of control in a state s reflects the degree of *a priori* readiness to reach a desired state s', where s' will be known only in the future. In other words, an agent senses control if it feels ready to immediately reach any state s' as soon as it is determined, for example, based on an internal decision or an announcement by an external party. In the thought experiment of Fig. 1A, this perspective suggests that we feel a greater sense of control in a city where we feel more 'prepared' to pursue *any* activity we might wish to do, despite not knowing in advance which activity it will be.

To formalize this idea, we assume that agents know (or believe they know) the actual dynamics of their environment (see Discussion); e.g., they know that the circus building sometimes hosts a painting class on random weekends. We also assume that there are no *a priori* extrinsic rewards associated with specific environmental states; e.g., there is no *a priori* preference for any of the weekend activities. Then, at some later point, a goal state g is randomly selected and revealed to the agent; for instance, going to the circus is chosen as the preferred activity for that weekend. The key question then becomes: 'In which state is an agent best prepared to reach g, once it is announced?'

This problem can be reframed within the sequential decision-making framework [28–30] by assigning a reward of 1 to the goal state g and 0 to all other states. In this framework, the answer to the question 'Where would an agent be most prepared?' depends on which reward statistics the agent seeks to optimize (i.e., its normative objective; Fig. 1C). Being prepared according to different objectives leads to preferences for conceptually distinct state features (Methods). If the agent aims to maximize expected reward, it prefers states with many available *distinct* actions (A-Control; see Proposition 1 and Corollary 1-2). On the other hand, if the goal is to maximize the probability of receiving a guaranteed reward, the agent favors states with many *certainly* reachable outcomes ($\mathrm{C_S}$-Control), whereas if the objective is to *minimize* the chance of *not* obtaining a reward, it prefers states with many *possibly* reachable outcomes ($\mathrm{P_S}$-Control).

Accordingly, the different normative objectives correspond to distinct conceptualizations of the sense of control that are all intuitively sensible, yet principally different (Fig. 1C). Our thought experiment further illustrates these distinctions (Fig. 1B). A-Control predicts that the agent experiences the same degree of control in c_2 and c_3 (Fig. 1B2), consistent with Klyubin empowerment[13]. In contrast, $\mathrm{P_S}$-Control predicts greater perceived control in c_3 than in c_2 (Fig. 1B3), similar to the maximum outcome entropy[11]. Conversely, $\mathrm{C_S}$-Control predicts a higher sense of control in c_2 than in c_3 (Fig. 1B1).

Emp_ℓ unifies and expands existing perspectives on the sense of control

Critically, we found that the three normative objectives correspond to special cases of a single, interpretable measure that we call degree-ℓ empowerment, Emp_ℓ (Methods; Fig. 1C–E). Intuitively, Emp_ℓ quantifies the sense of control in state s as how 'easily', on average, the agent can transition to other states in the environment, if it wishes to do so (Fig. 1E1):

$$\mathrm{Emp}_\ell(s) := \sum_{s'} p^\ell_{\max}(s'|s), \tag{1}$$

where $p_{\max}(s'|s)$ denotes the maximum probability $\max_a p(s'|s,a)$ of reaching s' from s in a single step; hence, we refer to $p_{\max}(s'|s)$ as the (single-step) *reachability* of s' from s (Discussion). The degree ℓ of the empowerment is a free parameter that modulates the reachabilities to emphasize certainty against possibility (Fig. 1E2). For $\ell > 1$, increasing ℓ progressively downplays the influence of small reachabilities. As a result, in the limit as $\ell \to \infty$, Emp_ℓ counts only *certain* reachabilities ($p_{\max}(s'|s) = 1$). For $\ell < 1$, on the other hand, decreasing ℓ progressively exaggerates the influence of small reachabilities. As a result, in the limit as $\ell \to 0$, Emp_ℓ counts any *possible* reachability ($p_{\max}(s'|s) > 0$), regardless of how small the reachability. Surprisingly, at the critical value $\ell = 1$, $\mathrm{Emp}_\ell(s)$ is proportional to the expected future reward and captures the distinctiveness of available actions in state s (Methods).

Consequently, Emp_ℓ connects and encompasses the distinct conceptualizations of the sense of control (Fig. 1C). Our thought experiment clearly illustrates this integration: Varying ℓ leads to different predictions from Emp_ℓ, aligning with perspectives that previously appeared contradictory (compare Fig. 1B and Fig. 1D). Additionally, Emp_ℓ also satisfies several theoretically desirable properties that one would intuitively expect from an appropriate measure of the sense of control (Theorem 1-2, and Proposition 3). In particular, Emp_ℓ aligns with psychological theories of self-efficacy[26] and perceived behavioral control[24], meaning that action-execution noise (i.e., reduced self-efficacy) naturally decreases $\mathrm{Emp}_\ell(s)$ for any value of ℓ (Theorem 2).

In conclusion, Emp_ℓ both *unifies* and *expands* existing perspectives on the sense of control, offering a principled foundation for understanding how humans experience control.

Experiment 1: *A priori* readiness to collect gold

To test whether Emp_ℓ captures the human sense of control, we designed an experimental paradigm that closely resembles our thought experiment in Fig. 1A. The experiment intended to identify which states participants preferred when they did not know their future task. Hence, we divided the experiment into two phases (Fig. 2A).

In the first phase, 71 participants played 12 rounds of a gold-collecting game (Fig. 2B). In each round, participants began at the center of a 3×3 virtual room and waited until a gold coin randomly appeared in one of the peripheral states (Fig. 2B). They then selected one (and only one) available action to reach the gold coin. If an action had a single endpoint (indicated by green disks in Fig. 2B), it deterministically moved the participant to that endpoint. If the action had multiple endpoints (e.g., as in Room 2 in Fig. 2C), it moved the participant to one of the endpoints at random, with equal probability.

In the second phase, participants were shown 132 pairs of rooms and asked to choose which room they would prefer for collecting gold coins (Fig. 2C). The experiment included 12 prespecified rooms (Supplementary Materials), varying in both the number of available actions (2, 3, 4, or 8) and the level of outcome stochasticity associated with each action. Depending on the room, the gold coin could appear in a location that was unreachable (Fig. 2D1). In some cases, the coin could alternatively appear in a location that was only probabilistically reachable (Fig. 2D2). This design allowed us to make trials where comparing different rooms resembled comparing different cities in the thought experiment; e.g., Room 1 and Room 2 in Fig. 2C qualitatively mirror the properties of c_2 and c_3 from Fig. 1A.

Importantly, in the second phase, participants were only asked to choose between two rooms; they were informed that the experimenter would play the gold-collecting game on their behalf by *always* selecting the most appropriate action. They were told that their monetary bonus would increase each time a gold coin was successfully collected for them. At the end of the experiment, participants completed two questionnaires: the Locus of Control (LOC) scale[32] and the Intolerance of Uncertainty (IUC) scale[33].

Emp_ℓ accurately predicts human behavior

We used Bayesian model selection[36,39] to compare three hypotheses regarding participants' room preferences (Methods). The first, simplest hypothesis assumed that participants prefer rooms with a higher number of available actions (N_a), regardless of how the actions' endpoints are arranged (e.g., see ref.[22]). The second hypothesis proposed that preferences are driven by the Emp_ℓ value of the central state, with ℓ treated as a participant-specific free parameter. The third and most flexible hypothesis modeled participants' behavior using the general Bradley–Terry model[34,35], in which each of the 12 rooms was assigned a separate 'subjective utility' value as a participant-specific free parameter. This model subsumes any rational strategy, including those based on N_a or Emp_ℓ, but its increased flexibility is penalized in Bayesian model selection[40]; our model-recovery[37] analysis confirms this expected trade-off (Supplementary Materials). Finally, we also included a random,

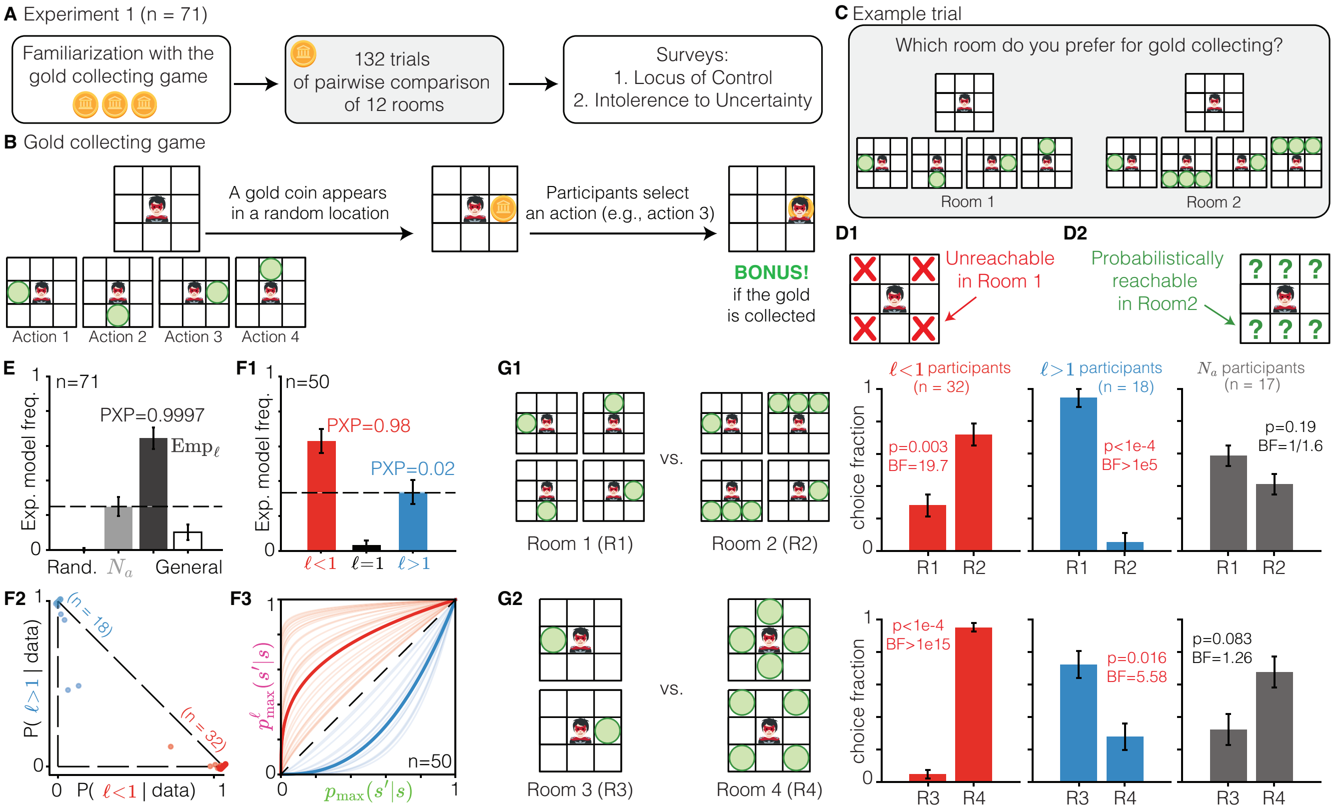

Figure 2: **Emp_ℓ accurately captures individual differences in preference for certainty versus possibility.** **A**. Experiment 1. Participants first completed a familiarization phase with the gold-collecting game (see B), followed by 132 room-selection trials (see D) and, finally, the LOC[32] and IUC[33] surveys. **B**. Gold-collecting game. Participants waited in the center of a 3×3 virtual room until a gold coin appeared in one of the peripheral states. They then selected a single action to reach and collect the coin. **C**. In each room-selection trial, participants chose which of the two presented rooms they preferred for collecting gold coins. **D**. Rooms differed in the certainty and possibility of reaching the gold coin. Depending on the room, some locations were entirely unreachable (e.g., Room 1 in C; **D1**) or only probabilistically reachable (e.g., Room 2 in C; **D2**). **E**. Model comparison results evaluating participants' preferences against (i) the random, null model, (ii) the number of available actions (N_a), (iii) Emp_ℓ, and (iv) the general Bradley-Terry model[34,35] (Methods). Expected model frequency reflects the expected proportion of participants best explained by each model. PXP: Protected Exceedance Probability[36]. Error bars: Posterior standard deviation (SD). See Supplementary Materials for model-recovery analyses[37]. **F**. Emp_ℓ captures individual differences among participants. **F1**. Expected model frequencies comparing fits for $\ell < 1$, $\ell = 1$, and $\ell > 1$ (see Fig. 1C). **F2**. Model probability simplex showing individual participants (data points); most participants were best explained (probability ~1) by either $\ell < 1$ or $\ell > 1$. **F3**. Individual variabilities in the extent of reachability modulation (as in Fig. 1E2), based on the posterior expected value $\hat{\ell}$. Thick lines represent group averages. See Supplementary Materials for model and parameter recovery[37]. **G**. Fraction of choices of different participant groups favoring Room 1 over Room 2 (G1) and Room 3 over Room 4 (G2). Error bars: The standard error of the mean (SEM). Red p-values: Statistically significant effects (FDR controlled at 0.05)[38]. Red Bayes Factors (BF): Strong evidence for the alternative hypothesis (BF $\geq$ 3). *Credit: Emojis adapted from ref.*[31]*, under a CC BY 4.0 license.*

null model to validate that our exclusion criteria had effectively removed inattentive, randomly behaving participants (Methods and Supplementary Materials).

The choices of the majority of the 71 participants were best explained by Emp_ℓ ($n = 50$), followed by the N_a model ($n = 17$), and the general utility model ($n = 4$; see Fig. 2E and Supplementary Materials); we refer to the 50 participants whose behavior was best explained by Emp_ℓ as 'Emp_ℓ participants.' The poor performance of the general utility model is particularly important as it rules out the possibility that some untested measure of the sense of control could substantially

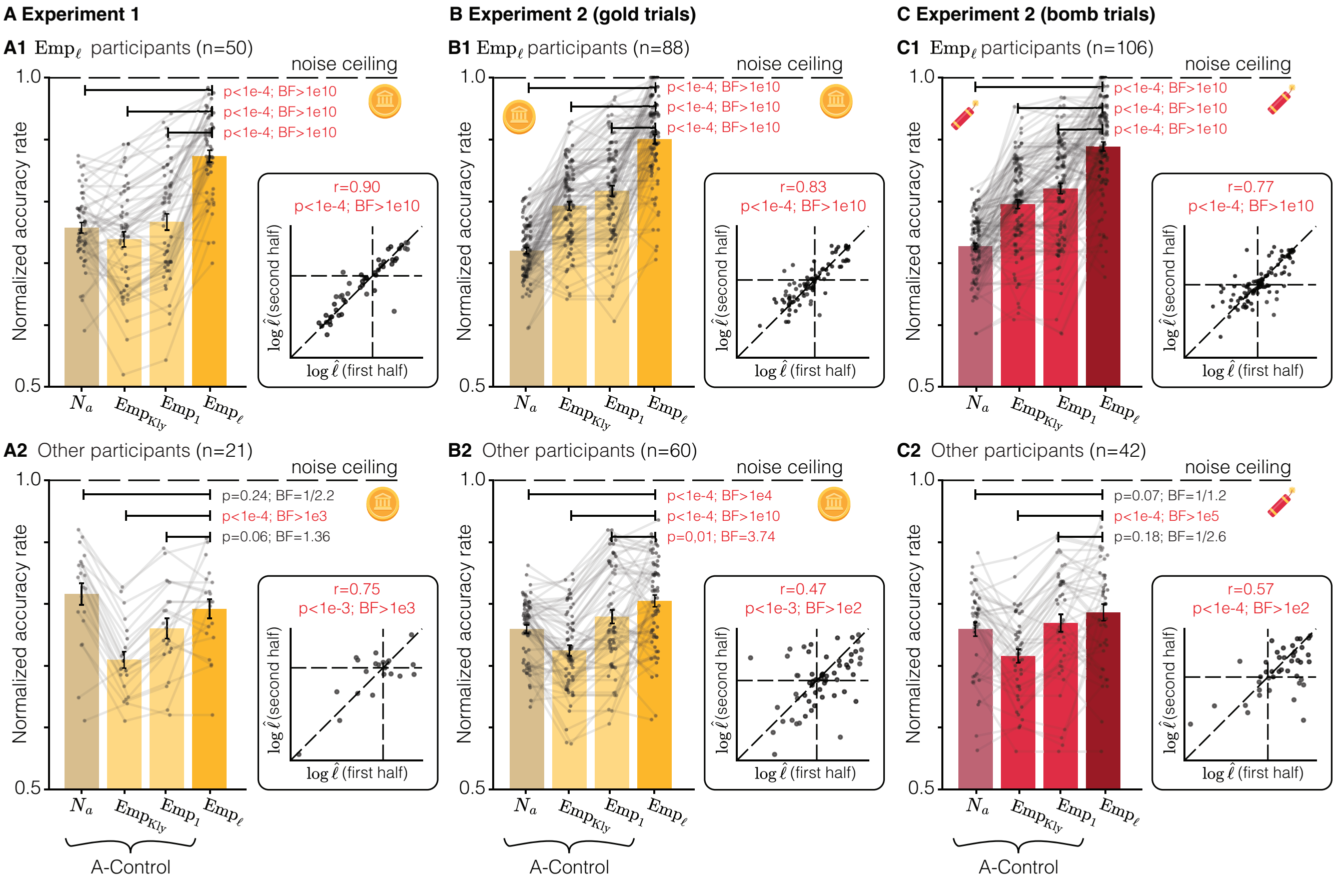

Figure 3: **Emp$_\ell$ predicts participants' choices with high accuracy.** Normalized accuracy rates of different models in Experiment 1 (A), as well as in the gold trials (B) and bomb trials (C) of Experiment 2. For each participant, normalized accuracy was computed as the cross-validated model accuracy (Eq. 11) divided by the split-half noise ceiling (Eq. 14), which accounts for the inherent unpredictability due to noisy actions (see Methods for details). As baselines, we also included the accuracy rates for the three measures of A-Control: Mere number N_a of actions, Klyubin empowerment [13] Emp_{Kly}, and the degree-1 empowerment Emp_1 (i.e., Emp_ℓ for $\ell = 1$). Emp_ℓ reaches an almost 90% normalized accuracy rate for the Emp_ℓ participants (A1-C1) and an almost 80% for the rest (A2-C2). Error bars: SEM Single points: Individual participants. Inset: Posterior expected values of $\log \hat{\ell}$ estimated separately for the first and second halves of the data (see Methods). Red p-values: Statistically significant effects (False Discover Rate, FDR, controlled at 0.05) [38]. Red BF: Strong evidence for the alternative hypothesis ($\text{BF} \geq 3$). *Credit: Emojis adapted from ref.* [31]*, under a CC BY 4.0 license.*

outperform Emp_ℓ in explaining participants' choices. To further assess this, we evaluated the predictive accuracy of Emp_ℓ using 2-fold cross-validation [41] (Methods; Fig. 3A). Emp_ℓ achieved an average normalized accuracy of 87% for the 50 Emp_ℓ participants–with individual scores ranging from 70% to 98% (Fig. 3A1). This corresponds to Emp_ℓ successfully predicting approximately 90% of participants' *predictable* choices, where predictability was determined by a split-half consistency analysis (Methods). Moreover, even for the other 21 participants whose behavior was best explained by N_a or the general utility model, Emp_ℓ achieved an almost 80% normalized accuracy rate (Fig. 3A2). This was not significantly different from the maximum accuracy rate achieved by the mere number N_a of actions.

Emp_ℓ captures individual preferences for certainty versus possibility

We next focused on the Emp_ℓ participants and analyzed the posterior expectation $\hat{\ell}$ of the empowerment degree to determine which conceptualization of the sense of control (Fig. 1C) best aligns with their preferences (see Supplementary Materials for parameter recovery[37]). We found substantial individual differences (Fig. 2F–G), with promisingly high split-half consistency (Fig. 3A1, inset). Approximately two-thirds of participants preferred *possible* reachability, while one-third preferred *certain* reachability (Fig. 2F1–F2, G1–G2). Within each group, we additionally observed considerable variability in the strength of these preferences (Fig. 2F3 and Fig. 1E2); nonetheless, the group-averaged behavior closely aligned with the qualitative predictions of the corresponding conceptualizations (Fig. 2G).

Notably, we did not observe any participants whose behavior was best explained by the intermediate case of $\ell = 1$, which corresponds to a more complex notion of A-Control beyond the mere number N_a of actions (Fig. 2F1–F2; see also Fig. 2E); this was despite the model-recovery analysis confirming that our procedure could reliably distinguish $\ell = 1$ from both $\ell < 1$ and $\ell > 1$ (Supplementary Materials). The poor empirical performance of Emp_ℓ at $\ell = 1$ suggests that some well-known measures of A-Control, such as Klyubin empowerment $\mathrm{Emp}_{\mathrm{Kly}}$[13], do not align with the preferences exhibited by human participants in our experiment (see Discussion). This is further supported by the substantially lower normalized accuracy rate of $\mathrm{Emp}_{\mathrm{Kly}}$ and Emp_1 compared to Emp_ℓ with varying ℓ (Fig. 3A).

In conclusion, we found strong quantitative (Fig. 2F) and qualitative (Fig. 2G) evidence for individual differences in participants' preferences for certainty versus possibility. While Emp_ℓ effectively captures these differences, existing measures of the sense of control do not account for such a certainty-possibility dichotomy.

Experiment 2: *A priori* readiness to collect gold or avoid bombs

We derived Emp_ℓ from the premise that the sense of control reflects *a priori* readiness to reach a desired goal state that has yet to be announced. However, the correspondence between Emp_ℓ and different normative objectives (Fig. 1C) holds for any outcome distribution that does not impose *a priori* preferences over specific subsets of states (Theorem 1). In this context, it is particularly interesting to consider loss-based outcome distributions. Decades of research in behavioral economics have shown that human behavior can differ substantially in the loss domain, even when choices can be exactly remapped from equivalents in the gain domain[42–44].

To test this, we designed loss-based trials that exactly mirror gain-based ones by assigning an equal punishment (i.e., a 'reward' of -1) to all states except a yet-to-be-designated safe state g. Then, Emp_ℓ would be the *a priori* readiness to *avoid* all states except yet-to-be-announced g (Corollary 3). In a second experiment, we then examined whether participants' preferences for certainty versus possibility would change when the task was framed as avoiding bombs rather than collecting gold ($n = 148$; Fig. 4A–C).

The experiment included a gold block and a bomb block, where the block order was randomized across participants (Fig. 4A). Specifically, 71 participants completed the gold block first, thereby replicating Experiment 1 in this part, and 77 participants completed the bomb block first. The procedure for the gold block was identical to Experiment 1 (Fig. 2A–D), with one minor difference: Only 9 of the 12 prespecified rooms were used in Experiment 2 to shorten the total experiment

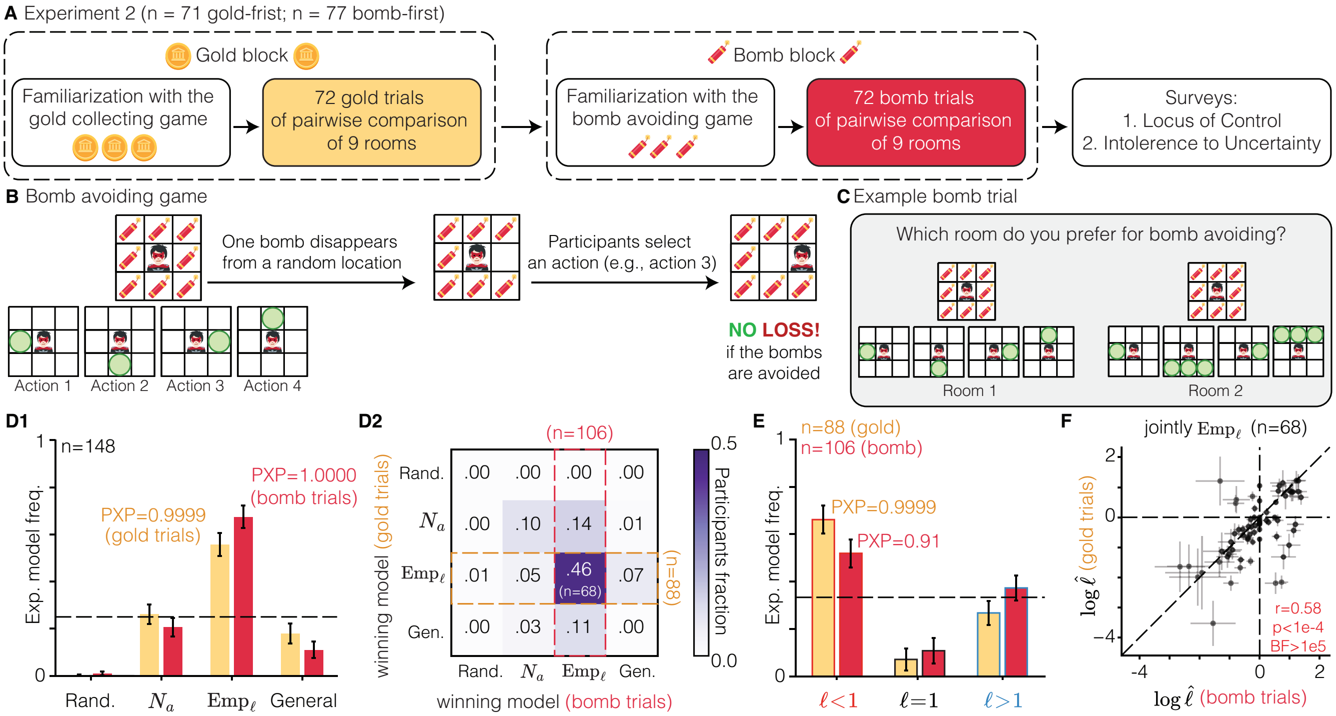

Figure 4: **Individual differences in preference for certainty versus possibility remain stable across gain and loss domains. A**. Experiment 2 included a gold block and a bomb block. The gold block replicated the first two phases of Experiment 1 (Fig. 2A), using only 9 of the 12 predefined rooms (Supplementary Materials). The bomb block followed the same structure but focused on bomb avoidance rather than gold collection. Block order was randomized across participants. Participants completed the LOC[32] and IUC[33] surveys at the end of the experiment. **B**. Bomb avoiding game. Participants waited in the center of a 3×3 virtual room filled with bombs until one of the bombs randomly disappeared. They then selected a single action to reach the resulting safe location. **C**. In room-selection trials, participants chose the room they would prefer to avoid bombs. **D**. Model comparison results evaluating participants' preferences against (i) the random, null model, (ii) the number of available actions (N_a), (iii) Emp_ℓ, and (iv) the general Bradley-Terry model[34,35] (Methods). **D1**. Expected model frequencies, separately for the gold trials (yellow) and bomb trials (red). Same format as Fig. 4E. **D2**. Fraction of participants best explained by each model. **E**. Expected model frequencies comparing $\ell < 1$, $\ell = 1$, and $\ell > 1$ (see Fig. 1C), separately for the gold trials (yellow) and bomb trials (red). This analysis includes only participants whose preferences were best explained by Emp_ℓ (highlighted row and column in D2). Same format as Fig. 4F1. **F**. Relationship between the posterior expected values of $\log\hat{\ell}$ in the gold and bomb trials. This analysis includes only participants whose preferences were best explained by Emp_ℓ in both blocks (dark cell in D2). Red p-values: Statistically significant effects (FDR controlled at 0.05)[38]. Red BF: Strong evidence for the alternative hypothesis (BF $\geq$ 3). Error bars: Posterior SD. *Credit: Emojis adapted from ref.*[31]*, under a CC BY 4.0 license.*

time (Supplementary Materials). This resulted in 72 room selection trials per block (Methods), for a total of 144 trials across both blocks. The bomb block followed the same overall structure as the gold block, with one key difference: During the familiarization phase, participants played 12 rounds of a bomb avoiding game (Fig. 4B), and in the room selection phase, they chose which room they would prefer for avoiding bombs (Fig. 4C). In each round of the bomb avoiding game, participants began in the center of a 3×3 virtual room filled with bombs and waited until one of the eight bombs randomly disappeared (Fig. 4B). They then selected one (and only one) available action to move to the newly revealed safe location.

As in Experiment 1, participants were only asked to choose between pairs of rooms. They were informed that the experimenter would play both the gold-collecting and bomb-avoiding games on their behalf by always selecting the most appropriate action. In the bomb block, participants were told that they would lose one gold coin each time we failed to escape the bombs for them. To ensure that no participant ended up in debt, we provided all participants with 50 extra gold coins at the beginning of each bomb block (Methods). At the end of the experiment, all participants completed the LOC and IUC surveys.

Emp_ℓ accurately predicts human behavior in both gold and bomb trials

We repeated the Bayesian model-selection procedure used in Fig. 2E, applied separately to the gold and bomb trials. For both trial types, the choices of the majority of the 148 participants were best explained by Emp_ℓ, followed by N_a and the general utility model (Fig. 4D). The expected model frequencies were highly similar across gold and bomb trials (Fig. 4D1) and closely matched those observed in Experiment 1 (Fig. 2E). Consistent with these results, Emp_ℓ also predicted approximately 90% of the *predictable* choices of the Emp_ℓ participants (Fig. 3B1–C1) and 80% of the *predictable* choices of the other participants (Fig. 3B2–C2). Together, these findings demonstrate that the core results from Experiment 1 in the reward-seeking context generalize robustly to a punishment-avoiding context.

However, we observed a slightly higher model frequency for Emp_ℓ in the bomb trials than the gold trials (exceedance probability > 0.95; Fig. 4D1). Notably, the preferences of nearly 50% of participants were best explained by Emp_ℓ in both trial types. Still, an additional 25% of participants showed Emp_ℓ-consistent preferences exclusively in the bomb trials, while only 12% of participants showed Emp_ℓ-consistent preferences exclusively in the gold trials (Fig. 4D2). This pattern suggests that humans may be more inclined to follow Emp_ℓ in punishment-avoiding contexts than in reward-seeking ones. We found no substantial difference between participants who completed the gold block first and those who started with the bomb block (Supplementary Materials).

Certainty–possibility preferences are stable across gold and bomb trials

We next analyzed $\hat{\ell}$ separately for participants whose preferences were best explained by Emp_ℓ in the gold and bomb trials. As in Experiment 1, we observed substantial individual differences (Fig. 4E–F and Supplementary Materials), along with a promisingly high split-half consistency (Fig. 3B1–C1, inset). The overall distribution of participants' preferences for certainty versus possibility was similar across trial types and consistent with Experiment 1 (Fig. 4E vs. Fig. 2F1). Nevertheless, we still observed a slightly higher frequency of certainty preference in the bomb trials (exceedance probability > 0.95; Fig. 4E). This modest shift toward certainty in the loss domain aligns qualitatively with findings from behavioral economics[42], though the effect was notably smaller than might be expected. Crucially, among participants whose preferences were best explained by Emp_ℓ in both trial types, we found a strong positive correlation between their $\hat{\ell}$ values in the gold and bomb trials (Fig. 4F).

In conclusion, these findings demonstrate that (i) preferences for certainty versus possibility are relatively stable across both reward-seeking and punishment-avoiding contexts, and (ii) Emp_ℓ reliably captures these individual differences.

Origins of individual differences in certainty–possibility preferences

Finally, we investigated potential sources of the observed individual differences in preferences for certainty versus possibility. One hypothesis was that these differences might be rooted in some personality traits. To test this, we examined correlations between participants' $\hat{\ell}$ values and their responses to the LOC and IUC surveys (36 items in total). We found no significant correlations between $\log \hat{\ell}$ and the raw survey scores (Supplementary Materials). We then conducted a Principal Component Analysis (PCA) on the survey responses and identified three main principal components (Supplementary Materials). However, none of these components consistently correlated with $\log \hat{\ell}$ across both experiments. These findings suggest that the traits measured by the LOC and IUC surveys may be distinct from the underlying factors driving participants' behavioral preferences in our paradigm (see Discussion).

Discussion

From planning a weekend to navigating major life decisions, our experiences are shaped by how much control we feel we have over future outcomes. Over the past decades, the sense of control has played a pivotal role in shaping psychological theories of learned helplessness[25,45,46], self-efficacy[26], agency[2,47,48], and autonomy[49]. This subjective perspective highlights how people perceive, anticipate, and adjust to the objective controllability of their environments. In parallel, computer scientists and engineers have attempted to formalize notions of controllability[9,10,50] and empowerment[7,8,13] from an objective viewpoint, designing artificial agents that can flexibly manipulate their environments in pursuit of goals. Yet the absence of a single formal, unifying framework has made it challenging to characterize the environmental factors that shape the human sense of control or are essential for machine learning agents. This has left the relationships and inter-dependencies between objective and subjective viewpoints insufficiently understood.

In this work, we offer such a unifying framework. We formalized the sense of control as the degree of *a priori* readiness to act in future tasks, regardless of the task itself. To formalize this concept, we introduced a novel, interpretable measure of environmental controllability, Emp_ℓ, which captures how the structure of an environment admits flexible implementation of future tasks.

This formulation unifies a wide range of existing constructs by showing that both objective and subjective perspectives on control reflect different statistics of readiness that an agent might optimize before its task is known. The degree ℓ in Emp_ℓ characterizes this continuum: when $\ell = 1$, the measure reflects pure action availability (A-Control); values of $\ell > 1$ prioritize certainty in reaching outcomes ($\mathrm{C_S}$-Control), while $\ell < 1$ emphasize maintaining possibility ($\mathrm{P_S}$-Control). Importantly, across two large-scale, theory-guided experiments, we demonstrated that measuring the sense of control via Emp_ℓ accurately predicts human preferences, providing a principled account of how environmental structure shapes the subjective sense of control. These findings offer a foundation for a unified understanding of control, one that bridges objective and subjective perspectives and connects insights across machine learning and psychology.

Our framework also has significant implications for computer science and machine learning. Deep reinforcement learning (RL) in high-dimensional environments has used the sense of control and empowerment for learning controllable state representations[51,52] and self-supervised skill discovery[6,7,53]. Similarly, states with high controllability have been considered to be intrinsically rewarding[13,54–60] in relevant forms of RL[61–63]. However, the information-theoretic measures that typically underpin these works[6,7,13,51,53] have lacked the transparent and explicit links that Emp_ℓ enjoys to standard RL objectives, such as maximizing expected reward or minimizing the chance of failure. We suggest that Emp_ℓ may, therefore, be a valuable alternative building block for intrinsic control-driven RL.

Currently, however, a limitation of Emp_ℓ is its focus on the notions of '*immediate* readiness' and '*one-step* reachability.' In real-world settings, multiple intermediate actions are usually required to carry out tasks, and any meaningful notion of controllability must extend to multi-step horizons. However, the generalizations in classical control theory [9,10] or information theory [13] to multiple steps do not point in the same direction, and various possible assumptions remain highly debated [15,64]. Thus, a prominent question for future research is to examine extensions of Emp_ℓ to a multi-step setting. A second current limitation of Emp_ℓ is its assumption that an agent believes to be aware of the true dynamics of the environment. However, most real-world settings involve epistemic uncertainty, which must be accounted for by the agent to both quantify and seek control. Thus, another prominent question for future research is to account for the imperfect knowledge of the environment in Emp_ℓ and to model the interplay between control-seeking and different forms of information-seeking (e.g., see ref. [59,60]).

Empirically, we found substantial individual differences in participants' fitted values of ℓ. The lack of consistent associations between these values and standard trait-based LOC and IUC measures aligns with recent findings on the generally weak, or even absent, relationship between survey-based measures and cognitive task performance [65–67]. Thus, in future studies, it would be interesting to explore links with distinct cognitive capacities, such as working memory, executive function, or risk sensitivity, possibly through batteries of cognitive tasks (e.g., as in ref. [68]). It would also be important to examine the psychometric properties of ℓ, including its test-retest reliability [65–67] and generalizability across other contexts. Crucially, existing protocols that measure the human sense of control, whether by manipulating homogeneous action-outcome noise [16,18,19,21] (Theorem 2) or by relying on deterministic games [69,70], are inadequate to identify ℓ (Proposition 3), having left a key aspect of control sensitivity previously inaccessible.

An important advantage of a unified measure of the objective controllability of the environment is that it licenses richer examination of the dysfunctional psychological processes that might underlie biased subjective reports. This may aid our understanding of, for instance, how incorrect beliefs about the environment's dynamics based on over-generalization may make us fall into helplessness even when we can objectively control the environment [45], or why we may experience illusions of control when none exists [71], if we perceive the mere availability of actions as the ability to control, regardless of actual action-outcome contingencies [22].

In conclusion, we introduced a formal and integrative framework for the human sense of control, grounded in the notion of *a priori* readiness for unknown future tasks. By unifying divergent theoretical perspectives under a single, interpretable parameter, our measure reconciles previously conflicting views on controllability and provides a principled link between objective environmental structure and subjective experience. Across two experiments, we showed that our measure not only captures key behavioral regularities but also reveals stable individual differences in how people value certainty versus possibility. These findings open up new avenues for understanding the foundations of agency, individual variation in control sensitivity, and their relevance across domains. Together, our results mark a step toward a unified science of control, one that can bridge minds and machines, theory and data, and certainty and possibility.

Methods

Experimental procedure

Ethics statement. Both experiments were approved by the ethics committee of the University of Tübingen. All participants consented to participation through an online consent form and a second form for the data protection protocol, presented in the beginning of the experiments.

Participants. Participants were recruited from the Prolific platform ($n = 116$ for Experiment 1 and $n = 307$ for Experiment 2, before exclusions). There were only three prescreening conditions: (i) fluency in English, (ii) age between 18 and 65, and (iii) a Prolific approval rate of higher than 95% (Experiment 1) or 99% (Experiment 2). We did not impose any additional pre-screening conditions, such as those based on participants' nationality, country of residence, or educational background. Participants were paid a fixed participation fee plus a performance-contingent bonus.

Experiment 1. Participants first practiced moving around in one deterministic and one probabilistic room (without any gold coin). Then, they played 12 rounds of the gold-collecting game in the same deterministic and probabilistic rooms. This was followed by an instruction about the room-selection procedure and how their choice of rooms would influence their bonus. Then, they practiced four room-selection trials followed by a comprehension test. Importantly, in one comprehension question, participants were presented with the *perfect* room, which had eight deterministic actions, where *all* peripheral locations were *certainly* reachable (i.e., equivalent to c_1 in the thought experiment). They were asked, 'How could we increase the chance of gold-collecting in this room?' and had three options: 'By making actions probabilistic,' 'By adding more actions,' and 'The chance of collecting the gold is already maximum in this room.' They could continue the experiment only after selecting the correct option (i.e., the third option).

Participants performed 132 ($= 12\times11$) room selection trials where they compared every pair of rooms twice (from the 12 predefined rooms; Supplementary Materials), with randomized spatial arrangements. The pairs were presented in shuffled order. Trials were aborted with a warning (and no bonus) if participants failed to choose a room within 10 seconds; no participant had more than 5 aborted trials. One of the 12 predefined rooms in the experiment was the perfect room mentioned above; hence, there were 11 trials in which participants had to choose between the perfect room and one of the other 11 rooms. We defined 'perfect-room preference' as the fraction of these 11 trials where the participants chose the perfect room. We excluded participants whose perfect-room preference was lower than the 95% quantile chance level (i.e., 0.7 for 11 trials). The remaining participants were included in our analyses ($n = 71$; 35 Female, 36 Male; age 34 ± 12). The excluded and included participants had clearly different performance statistics, apart from those used for exclusion (Supplementary Materials). This provides further confirmation that the exclusion criterion removed participants who indeed were not paying attention during the experiment.

Experiment 2. For the 156 participants who completed the gold block first, the general procedure for the gold block was identical to Experiment 1, with 9 rooms instead of 12 (Supplementary Materials). After completing the gold block, these participants were told that they would receive an additional 50 gold coins as a prize for their efforts and were asked to 'protect their gold coins' by avoiding the bombs. For the participants who completed the bomb block first, the initial instructions were also similar to those in Experiment 1, but bomb avoiding was the primary task. Initially, they were given 50 gold coins and were asked to protect the coins by avoiding the bombs. After completing the bomb block, they were told that they could collect more gold coins as a prize. For both conditions, and for both bomb and gold blocks, participants were required to complete a comprehension test. We included the question about the prefect room in all comprehension tests.

All participants completed a total of 144 (=72×2) room selection trials. Trials were aborted with a warning (and no bonus) if participants failed to choose a room within 10 seconds; we excluded participants with more than 5 aborted trials in either of the gold or bomb blocks ($n = 4$). One of the 9 predefined rooms in the experiment was the perfect room mentioned above; hence, there were 8 trials in each block, where participants had to choose between the perfect room and one of the other 8 rooms. We excluded participants whose perfect-room preference was lower than the 95% quantile chance level (i.e., 0.75 for 8 trials) in either the gold or the bomb block. The remaining participants were included in our analyses ($n = 148$; 56 Female, 89 Male, 3 'Rather not say'; age 32±10). The excluded and included participants had clearly different performance statistics, apart from those used for exclusion (Supplementary Materials). This provides further confirmation that the exclusion criterion removed participants who indeed were not paying attention during the experiment.

Sample size estimation. The number of trials and the rooms were selected based on the model- and parameter-recovery analysis to dissociate different models and infer ℓ (Supplementary Materials). The number of participants was chosen to be more than 70; accordingly, we would have gotten a maximum SEM of 5-6% for estimating the fraction of participants in a specific group, i.e., $\sqrt{0.5^2/70} \approx 0.06$.

Additional exclusion for surveys. Participants completed the LOC[32] and IUC[33] surveys at the end of both experiments. The LOC survey quantifies how much an individual attributes the causal sources of events to themselves (8 questions; I-LOC), to chance (8 questions; C-LOC), and to powerful others (8 questions; P-LOC), whereas IUC quantifies an individual's degree of uncertainty intolerance and aversion (12 questions). In compliance with our ethics approval, the surveys did not influence participants' bonuses, and the participants could skip any question they wished. Participants who skipped any questions were excluded from the analysis of the survey data. Additionally, we included three trivial questions (e.g., asking whether the participant could read) as attention checks. We also excluded participants who failed more than one attention check from further analysis (see ref. [72]), leaving a total of 177 (56 for Experiment 1 and 121 for Experiment 2) participants for the analysis of the survey data.

Modeling and analyzing experimental data

General setting. We denote the participant's response at trial t by $y_t \in \{1, 2\}$ and the pair of presented rooms by $x_t = (x_{t,1}, x_{t,2}) \in \{1, \ldots, N_r\}^2$, where N_r is the total number of rooms (i.e., 12 for Experiment 1 and 9 for Experiment 2). We denote the data of participant n by $D_n = \{(x_t^{(n)}, y_t^{(n)})\}_{t=1}^{T_n}$, where T_n is the total number of completed room selection trials by participant n. If there were no aborted trials, then $T_n = 132$ for Experiment 1 and $T_n = 72$ for each block of Experiment 2.

Model likelihood formulation. Given a model m and its parameters θ_m, the probability that a participant selects the first presented room (i.e., $y_t = 1$) is

$$p(y_t = 1 | x_t, \beta_m, \theta_m, m) = \frac{e^{\beta_m \cdot V_m(x_{t,1};\theta_m)}}{e^{\beta_m \cdot V_m(x_{t,1};\theta_m)} + e^{\beta_m \cdot V_m(x_{t,2};\theta_m)}}, \tag{2}$$

where $V_m(x_{t,i}; \theta_m)$ is the 'value' assigned to room $x_{t,i}$ by model m, and $\beta_m \geq 0$ is the model inverse temperature that controls the sensitivity of decision-making to the model's assigned values. For the random and N_a models, there is no parameter θ_m, and we have

$$V_{\text{Rand}}(x_{t,i}) = 0 \quad \text{and} \quad V_{N_a}(x_{t,i}) = N_a(x_{t,i}), \tag{3}$$

where $N_a(x_{t,i})$ denotes the number of actions available in room $x_{t,i}$. For the Emp$_\ell$ model, we have $\theta_{\text{Emp}} = \ell$ and

$$V_{\text{Emp}}(x_{t,i};\ell) = \text{Emp}_\ell(x_{t,i}), \tag{4}$$

where $\text{Emp}_\ell(x_{t,i})$ is the Emp$_\ell$ value assigned to the central state in room $x_{t,i}$. For the general Bradley-Terry model[34,35], we have $\theta_{\text{Gen.}} \in \mathbb{R}^{N_r}$ and

$$V_{\text{Gen.}}(x_{t,i};\theta_{\text{Gen.}}) = \theta_{\text{Gen.},x_{t,i}}, \tag{5}$$

meaning that the value assigned to each room is considered as a free parameter[35]. Without loss of generality, we always assumed $\theta_{\text{Gen.},1} = 0$, implying that $\theta_{\text{Gen.},j}$ quantifies the differential value of room j *relative* to room 1.

As is the typical assumption in pairwise comparisons[35,44], we assumed independence between different trials. This implies that the total log likelihood of D_n is

$$\log p(D_n|\beta_m,\theta_m,m) = \sum_{t=1}^{T_n} \log p(y_t^{(n)}|x_t^{(n)},\beta_m,\theta_m,m). \tag{6}$$

Model selection. We first performed a non-hierarchical inference, considering different participants to be fully independent of each other. Specifically, we considered that a participant's data D was generated according to the following generative model (see[37,73,74]):

$$\begin{aligned} m &\sim \text{Uniform}\big(\text{Rand.}, N_a, \text{Emp}_\ell, \text{General}\big) \\ \beta_m|m &\sim \mathcal{N}_{\text{trunc}(0,\infty)}(0,10) \\ \theta_m|m &\sim p_m \\ D_n|\beta_m,\theta_m,m &\sim p(.|\beta_m,\theta_m,m), \end{aligned} \tag{7}$$

where $\mathcal{N}_{\text{trunc}(0,\infty)}(0,10)$ is a truncated normal distribution (between 0 and ∞, with mean zero and SD 10) acting as a rather uninformative prior for β_m; the prior distribution p_m for the parameters θ_m will be presented separately for each model. The data likelihood $p(.|\beta_m,\theta_m,m)$ is given by Eq. 6.

For the random and N_a models, there is no θ_m. For the Emp$_\ell$ model, we assumed $\theta_m = \ell$ with $p_m = \mathcal{N}_{\text{trunc}(0,\infty)}(0,5)$ as a rather broad prior. For the general model, we assumed that the components of $\theta_m \in \mathbb{R}^{N_r}$ are statistically independent, $\theta_{m,1} = 0$, and $\theta_{m,i} \sim \mathcal{N}(0,1)$; the idea is that θ_m captures the scaled preferences between rooms (with a prior variance of 1), whereas β_m scales the preferences to drive action-selection.

For each participant n, we used Markov Chain Monte Carlo (MCMC) methods to evaluate the posterior probability $p(m_n|D_n)$; all inferences were based on a Gibbs[75] combination of Hamiltonian Monte Carlo (for continuous variables) and Metropolis-Hastings algorithms (for discrete variables), as implemented in `Turing.jl`[75]. Model recovery analysis[37] showed that the model selection based on the generative model of Eq. 7 successfully recovered the true underlying generating model of the data, even when the prior distributions for parameters did not match the exact ones in Eq. 7 (see Supplementary Materials).

Noting that $\log p(m_n|D_n) = \log p(D_n|m_n) + \text{const.}$, we used $\log p(m_n|D_n)$ as the log-model evidence in the hierarchical generative model of ref.[39]: Assuming that $m_n = m$ with with some model frequency r_m that is shared across participants (see ref.[37,39]). The frequencies were assumed to have a Dirichlet prior with uniform mean and 1/4 concentration parameters, with 4 being the number of models. Fig. 2E and Fig. 2D1 show $\mathbb{E}[r_m|D_{1:N}]$ and $\sqrt{\text{Var}[r_m|D_{1:N}]}$, where $D_{1:N}$ is the concatenated data of all N participants; we have $N = 72$ for Experiment 1 and $N = 148$ for Experiment 2. PXP denotes $\mathbb{P}[r_m > r_{m'}\ \forall m' \neq m|D_{1:N}]$, with additionally accounting for the null hypothesis that all models have the same frequencies[36]. We assigned a winning model $\hat{m}_n$ to each participant, where we defined $\hat{m}_n = \arg\max_m \mathbb{P}(m_n = m|D_{1:N})$ (Supplementary Materials). Fig. 4D2 shows the fraction of the winning models. All the inferences for the hierarchical model were done through a second MCMC procedure using the Metropolis-Hastings algorithm (same as ref.[76]).

Inferring ℓ. For inferring ℓ, we only included the Emp_ℓ participants, i.e., those with $\hat{m}_n = \mathrm{Emp}_\ell$. For inferring the range of ℓ, we used the same 2-step model-selection procedure as above to infer a second model $m' \in \{\ell < 1,\ \ell = 1,\ \ell > 1\}$. For $\ell < 1$, we considered $\ell \sim \mathcal{N}_{\mathrm{trunc}(0,1)}(0.5, 1)$, and, for $\ell > 1$, we considered $\ell \sim \mathcal{N}_{\mathrm{trunc}(1,\infty)}(1, 2)$. Model recovery analysis[37] showed that the model selection based on these assumptions successfully recovered the true range of ℓ, even when the prior distributions for parameters did not match the exact ones we used (see Supplementary Materials). Fig. 2F1 and Fig. 4E show the expected model frequencies for this model selection, and Fig. 2E2 shows the probability of m'_n for individual participants. Given the winning model, $\hat{m}'_n$, a second inference was done to evaluate $p(\ell_n|D_n, \hat{m}'_n)$; the values in Fig. 2F3 and Fig. 4F corresponds to $\hat{\ell}_n := \mathbb{E}[\ell_n|D_n, \hat{m}'_n]$ and $\log \hat{\ell}_n := \mathbb{E}[\log \ell_n|D_n, \hat{m}'_n]$, respectively.

Statistical tests. All frequentist statistical tests were either a one-sample paired t-test (Fig. 2G and Fig. 3 outsets) or a correlation test (Fig. 3 insets and Fig. 4F). The correction for multiple hypotheses testing was done by controlling the FDR at 0.05[38] over all 31 null hypotheses that are presented in Fig. 2 (6), Fig. 3 (24), and Fig. 4 (1); p-value threshold: 0.040. All Bayes Factors (abbreviated BF in the figures) were evaluated using the Schwartz approximation[77] to avoid any assumptions on the prior distribution.

Normalized accuracy rates

General theory. Given a participant, we denote the true data distribution by p_d, i.e., $p_d(y|x) \in [0, 1]$ denotes the true probability with which a participant responds y when presented with the pair x of rooms. Similarly, a model of the participant is specified by p_m (as in Eq. 2). Given a pair x of rooms, the model predicts that the participant responds 1 if $p_m(1|x) > 0.5$. Otherwise, it predicts that the participants respond 0 if $p_m(0|x) > 0.5$. There will be a tie if $p_m(0|x) = 0.5$. Accordingly, the theoretical accuracy rate of the model is defined as[41,78]

$$\begin{aligned}\alpha_m :&= \mathbb{E}[\mathbb{1}_{p_m(y|x)>0.5}] \\ &= \underbrace{\sum_x p(x)}_{\text{averaged over all pairs}} \sum_y \underbrace{p_d(y|x)}_{\text{true probability of } y} \times \underbrace{\mathbb{1}_{p_m(y|x)>0.5}}_{\text{whether the model predicts } y}.\end{aligned} \tag{8}$$

where $\mathbb{1}$ is the indicator function (with the abuse of notation $\mathbb{1}_{0.5>0.5} := 0.5$), and $p(x)$ is the experiment-dependent distribution of the pairs of presented rooms. We can similarly define the accuracy rate α_d of the true model p_d as

$$\alpha_d := \mathbb{E}[\mathbb{1}_{p_d(y|x)>0.5}] = \sum_x p(x) \sum_y p_d(y|x) \mathbb{1}_{p_d(y|x)>0.5}. \tag{9}$$

It is straightforward to show that

$$\begin{aligned}\sum_y p_d(y|x) \mathbb{1}_{p_d(y|x)>0.5} &= \max\{p_d(0|x)\ ,\ p_d(1|x)\} \\ &\geq \sum_y p_d(y|x) \mathbb{1}_{p_m(y|x)>0.5}\end{aligned} \tag{10}$$

for any model m. Hence, we can recover a well-known result of statistical learning[41,78] that $\alpha_m \leq \alpha_d$ for any model m. This implies that α_d is the maximum possible accuracy rate which we could achieve only if we knew the true data distribution p_d (or, at least, the choices of x where $p_d(1|x) > 0.5$). Accordingly, the ratio α_m/α_d quantifies the fraction of predictable actions that model m can predict correctly. The rest of this section aims to find a cross-validated estimate of α_m and an upper-bound estimate of α_d.

Cross-validated estimation of model accuracy rate. Given a model m, the empirical accuracy rate $\hat{\alpha}_m$ is often defined as[41,78]

$$\hat{\alpha}_m := \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}_{p_m(y_t|x_t)>0.5}, \tag{11}$$

which is an unbiased estimation of the accuracy rate, i.e., $\mathbb{E}[\hat{\alpha}_m] = \alpha_m$. For each participant, model m can be any of our specified models that is *fitted* to the participant's data. We used two-fold cross-validation to find an unbiased estimate of α_m for such fitted models[79].

Since we have included each pair of rooms twice in the experiment, we can split the data D_n of participant n into two 'independent' replications. For the sake of notation, we define $N_{rr} = N_r(N_r - 1)/2$ to be the number of unique pairs of rooms and define $\{x_1, \ldots, x_{N_{rr}}\}$ as the set of unique pairs of rooms. For example, we have $N_{rr} = 66$ for Experiment 1 and $N_{rr} = 36$ for each condition of Experiment 2. Considering $y_{i,1}^{(n)}$ and $y_{i,2}^{(n)}$ to be the two responses of participant n to the presentation of the pair x_i, we split the data D_n into

$$D_{n,1} = \{(x_i, y_{i,1}^{(n)})\}_{i=1}^{N_{rr}} \quad \text{and} \quad D_{n,2} = \{(x_i, y_{i,2}^{(n)})\}_{i=1}^{N_{rr}}. \tag{12}$$

Then, for a given model m, we define

$$p_m(y|x) = \begin{cases} p(y|x, m, D_{n,2}) = \int p(y|x, m, \theta_m, \beta_m) p(\theta_m, \beta_m|m, D_{n,2}) d\theta_m d\beta_m, & \text{if } y \in D_{n,1} \\ p(y|x, m, D_{n,1}) = \int p(y|x, m, \theta_m, \beta_m) p(\theta_m, \beta_m|m, D_{n,1}) d\theta_m d\beta_m, & \text{if } y \in D_{n,2} \end{cases} \tag{13}$$

to be used in Eq. 11. To avoid any data contamination or assumption on ℓ, for the Emp_ℓ model, we used a broad prior $\ell \sim \mathcal{N}_{\mathrm{trunc}(0,\infty)}(0, 10)$ with no limitation of the model range. The inference was done using the Hamiltonian Monte Carlo algorithm as implemented in `Turing.jl`[75]. The data points in Fig. 3 show $\mathbb{E}[\log \ell_n|D_{n,2}]$ versus $\mathbb{E}[\log \ell_n|D_{n,1}]$.

Split half noise ceiling. We additionally used the splits in Eq. 12 to find a rough estimate of the maximum accuracy rate α_d as

$$\hat{\alpha}_d^+ := \frac{1}{2}\left(1 + \frac{1}{N_{rr}}\sum_{i}^{N_{rr}} \mathbb{1}_{y_{i,1}=y_{i,2}}\right). \tag{14}$$

Assuming that $y_{i,1}$ and $y_{i,2}$ are statistically independent, it is straightforward to show that $\hat{\alpha}_d^+$ is an unbiased estimate of an upper bound on α_d:

$$0.5 \leq \alpha_d \leq \mathbb{E}[\hat{\alpha}_d^+] \leq \frac{1+\alpha_d}{2} \leq 1. \tag{15}$$

The normalized accuracy rate in Fig. 3 refers to $\hat{\alpha}_m/\hat{\alpha}_d^+$, where $\hat{\alpha}_m$ is evaluated by 2-fold cross validation. The noise ceiling of 1 corresponds to a model m that achieves $\hat{\alpha}_m = \hat{\alpha}_d^+$.

Linking normative objectives to Emp_ℓ

General setting and notation. In this section, we show how normative objectives in Fig. 1C can be expressed in terms of Emp_ℓ. We consider an agent that interacts with an environment defined by a state space $\mathcal{S}$, a state-dependent action space $\mathcal{A}(s)$ at state s, and the probability $p(s'|s,a)$ for transition to state s' after taking action a in state s. We denote the number of states and actions by $N_s := |\mathcal{S}|$ and $N_a(s) := |\mathcal{A}(s)|$, respectively. Capital letters represent random variables; when no ambiguity arises, we omit the capital letter notation. The agent's policy in state s is specified by the probability $\pi_s(a) := \pi(a|s)$ of selecting action a.

Readiness for a to-be-announced goal state. To quantify the agent's sense of control in a given state s, we start from the notion that the sense of control is the degree of '*a priori* readiness for immediately collecting rewards or avoiding punishments,' by being sensitive to any to-be-announced extrinsic reward function $R_{\text{ext}}(s)$. Specifically, we first consider the case in which $R_{\text{ext}}(s) = \mathbb{1}_{s=G}$ rewards agents upon visiting a goal state G, where the goal state G is *a priori* unknown – considered as a random variable with a uniform distribution over $\mathcal{S}$. The idea is thus to ask 'Which state one should attempt to stay in to be best prepared to reach G in one step when the goal is announced?' *If* the goal state G was known, then we could find the immediate optimal 'value' of state s:

$$\begin{aligned} V_*(s;G) &= \max_{a\in\mathcal{A}(s)} \sum_{s'\in\mathcal{S}} p(s'|s,a) R_{\text{ext}}(s') \\ &= \max_{a\in\mathcal{A}(s)} p(G|s,a) =: p_{\max}(G|s), \end{aligned} \tag{16}$$

where $p_{\max}(G|s)$ is the immediate reachability of state G from s in one step. Hence, to be 'ready' to reach G when announced, the agent must locate itself in states with a high value of $V_*(s;G)$. However, since G is a random variable, $V_*(s;G)$ is also a random variable. Hence, being 'ready' to reach a goal state and, equivalently, the sense of control associated with s must be defined based on some statistics of $V_*(s;G)$.

One possibility is that the agent feels 'ready' when it is in a state with the highest *expected* value, where the expectation is taken over possible choices of the goal state G:

$$\mathbb{E}[V_*(s;G)] = \frac{1}{N_s} \sum_{s'\in\mathcal{S}} p_{\max}(s'|s) = \frac{\text{Emp}_1(s)}{N_s}, \tag{17}$$

where $\text{Emp}_1(s)$ is the degree-1 empowerment defined in Eq. 1. As is precisely stated in Proposition 1 below, this measure quantifies the number of effectively distinct actions; more intuitively, Corollary 1 shows that $\text{Emp}_1(s)$ can be seen as a regularized version of $N_a(s)$, where regularization accounts for overlap between the outcome distributions of different actions. Specifically, if there are only two available actions, then $\text{Emp}_1(s)$ can be expressed as a well-known distance of $p(.|s,a1)$ from $p(.|s,a2)$, explicitly quantifying their distinction (Corollary 2) Hence, being ready in terms of expected future rewards resembles maximizing a notion of A-Control (Fig. 1C).

Proposition 1. *For $\ell = 1$, we have*

$$\text{Emp}_1(s) = 1 + \sum_{a\in\mathcal{A}(s)} \Delta(a;s), \tag{18}$$

where $\Delta(a;s)$ measures how distinct a is from the other actions:

$$\Delta(a;s) := \frac{1}{N_a(s)} \sum_{a'\in\mathcal{A}(s)} \left[\sum_{s'\in\mathcal{S}^*(a;s)} |p(s'|s,a) - p(s'|s,a')| \right] \tag{19}$$

with $\mathcal{S}^*(a;s) \subseteq \mathcal{S}$ the set of states where $a = \arg\max_{\tilde{a}} p(s'|s,\tilde{a})$; when there are ties, we assume that arg max *returns one action based on a given ordering.*

Proof: See Supplementary Materials.

Corollary 1. *For $\ell = 1$, we have*

$$\mathrm{Emp}_1(s) = \underbrace{N_a(s)}_{\textit{mere number of actions}} - \underbrace{\Big(N_a(s) - 1\Big)\Big(1 - \Delta(s)\Big)}_{\textit{regularization regarding action distinction}} \tag{20}$$

where $\Delta(s)$ measures the variability of actions in state s:

$$\Delta(s) := \frac{\sum_{a \in \mathcal{A}(s)} \Delta(a; s)}{N_a(s) - 1} \in [0, 1] \tag{21}$$

with $\Delta(a; s)$ defined in Proposition 1. Additionally, we have $\Delta(s) = 1$ if and only if $p(.|s, a)$ have disjoint supports for different actions; this results in maximum degree-1 empowerment: $\mathrm{Emp}_1(s) = N_a(s)$. On the other hand, we have $\Delta(s) = 0$ if and only if all actions have the exact same $p(.|s, a)$; this results in minimum degree-1 empowerment: $\mathrm{Emp}_1(s) = 1$.

Proof: See Supplementary Materials.

Corollary 2. *If $\mathcal{A}(s) = \{a1, a2\}$, then $\mathrm{Emp}_1(s) = 1 + d_{\mathrm{TV}}\Big[p(.|s, a1), p(.|s, a2)\Big]$, where d_{TV} denotes the total variation distance.*

Proof: See Supplementary Materials.

Another possibility is that the agent feels 'ready' when it is in a state where it has the *highest chance of certainly getting a reward*:

$$\mathbb{P}(V_*(s; G) = 1) = \frac{1}{N_s} \lim_{\ell \to \infty} \sum_{s' \in \mathcal{S}} p^{\ell}_{\max}(s'|s) = \frac{\mathrm{Emp}_{\infty}(s)}{N_s}, \tag{22}$$

where $\mathrm{Emp}_{\infty}(s)$ is the degree-∞ empowerment defined in Eq. 1. This measure counts the total number of certainly reachable states from state s. Hence, being ready in terms of the highest chance of certain rewards resembles maximizing a notion of $\mathrm{C_S}$-Control (Fig. 1C). A third possibility is that the agent feels 'ready' if it is in a state where it has the *lowest chance of NOT getting a reward*:

$$\mathbb{P}(V_*(s; G) > 0) = \frac{1}{N_s} \lim_{\ell \to 0} \sum_{s' \in \mathcal{S}} p^{\ell}_{\max}(s'|s) = \frac{\mathrm{Emp}_0(s)}{N_s}, \tag{23}$$

where $\mathrm{Emp}_0(s)$ is the degree-0 empowerment defined in Eq. 1. This measure counts the total number of possibly (even if extremely unlikely) reachable states from state s; hence, being ready in terms of the lowest chance of NOT getting rewards resembles maximizing a notion of $\mathrm{P_S}$-Control (Fig. 1C).

Extension beyond a to-be-announced goal state. Eq. 17, Eq. 22, and Eq. 23 show how Emp_ℓ relates to different normative objectives, under the assumption that there is only a single rewarding state. Here we show that the correspondence of Emp_ℓ to different normative objectives holds for a much broader class of reward distributions. Specifically, we consider a distribution over state-based reward functions $R_{\mathrm{ext}} : \mathcal{S} \to \mathbb{R}$ with the following constraints (see Supplementary Materials for the precise statement): 1. The distribution over reward function R_{ext} is invariant to permutations of R_{ext} (i.e., there is no *a priori* preference for particular states). 2. There is almost surely only one best state. Theorem 1 below shows that increasing Emp_ℓ increases both upper and lower bounds of different normative objectives.

Theorem 1. *Suppose the assumptions 1-2 hold for the distribution of reward function R_{ext}. We denote* *the maximum reward by the random variable*

$$R^*_{\text{ext}} := \max_{s\in\mathcal{S}} R_{\text{ext}}(s)$$

and the most negative reward value by the random variable

$$R^{-*}_{\text{ext}} := \min_{s\in\mathcal{S}} R_{\text{ext}}(s)\mathbb{1}_{\min_{s\in\mathcal{S}} R_{\text{ext}}(s)<0}.$$

Then, there exist constants c_0, C_0, b_1, c_1, B_1, C_1, and $C_\infty \in \mathbb{R}^+$, dependent solely on the distribution of reward function, such that the following holds.

1. *The probability of a certain maximum reward is proportional to degree-∞ empowerment:*

$$\mathbb{P}\big(V_*(s; R_{\text{ext}}) = R^*_{\text{ext}} > 0\big) = C_\infty \text{Emp}_\infty(s). \tag{24}$$

2. *Degree-1 empowerment bounds the expected reward:*

$$b_1 + c_1 \,\text{Emp}_1(s) \le \mathbb{E}[V_*(s; R_{\text{ext}})] \le B_1 + C_1 \,\text{Emp}_1(s) \tag{25}$$

3. *Degree-0 empowerment bounds the probability of possibly receiving a reward higher than the minimum negative reward:*

$$c_0 \,\text{Emp}_0(s) \le \mathbb{P}\big(V_*(s; R_{\text{ext}}) > R^{-*}_{\text{ext}}\big) \le C_0 \,\text{Emp}_0(s) \tag{26}$$

For all bounds, equality holds if $R_{\text{ext}}(s) = \mathbb{1}_{s=G}$ with $G \sim \text{Unif}(\mathcal{S})$.

Proof: See Supplementary Materials.

Remark 1. *All results of Theorem 1 hold if the realized rewards are stochastic, and $R_{\text{ext}} : \mathcal{S} \to \mathbb{R}$ denotes their state-based expectation.*

Corollary 3. *If $R_{\text{ext}}(s) = \mathbb{1}_{s=G} - 1$ with $G \sim \text{Unif}(\mathcal{S})$, then we have $\mathbb{E}[V_*(s; R_{\text{ext}})] = {}^{\text{Emp}_1(s)}/_{N_s} - 1$ for the expected reward, $\mathbb{P}\big(V_*(s; R_{\text{ext}}) = 0\big) = {}^{\text{Emp}_\infty(s)}/_{N_s}$ for the probability of a guaranteed no-punishment, and $\mathbb{P}\big(V_*(s; R_{\text{ext}}) > -1\big) = {}^{\text{Emp}_0(s)}/_{N_s}$ for the probability of possibly no-punishment.*

Proof: See Supplementary Materials.

Emp_ℓ as a function of self-inefficacy and noisy action execution

The notion of self-efficacy[24,26] is primarily concerned with the extent of control an agent feels over executing its desired actions. It is important to note that control in the sense of 'achieving desired outcomes' is distinct from control in the sense of 'executing desired actions', although these are related[24,26]. Here, we show that introducing self-inefficacy in terms of action-execution noise naturally decreases Emp_ℓ. This demonstrates that quantifying controllability via Emp_ℓ is conceptually consistent with these psychological theories[24,26].

Formally, we distinguish between an agent's intended action a in state s and the executed action a_{exc}. We can then model noisy and erroneous action execution by assuming that a_{exc} is most of the time (with probability $1-\epsilon$) equal to the intended action a, but is sometimes (in the case of error, with probability ϵ) sampled from a noise distribution π_{noise} independently of the agent's intended action, i.e.,

$$\pi_{\text{exc}}(a_{\text{exc}}|s, a) := (1-\epsilon)\,\mathbb{1}_{a_{\text{exc}}=a} + \epsilon\,\pi_{\text{noise}}(a_{\text{exc}}|s). \tag{27}$$

Accordingly, if the environment transition probabilities, given the executed action a_{exc}, are described by p, then the agent's perceived transition probabilities, given its intended actions, are described by

$$p_\epsilon(s'|s,a) = (1-\epsilon)\, p(s'|s,a) + \epsilon\, p_{\text{noise}}(s'|s), \tag{28}$$

where $p_{\text{noise}}(s'|s) := \sum_{a'} p(s'|s,a')\pi_{\text{noise}}(a'|s)$ is the next state distribution in the case of a noisy and erroneous action selection.

This formulation connects the execution of desired actions to the achievability of desired outcomes. Specifically, our formalism closely matches the original conceptualization of self-efficacy in psychology[26], and the equivalents of our noise probability ϵ have indeed been used previously as operationalizations of self-inefficacy in experimental studies[18,80]. Theorem 2 below shows that, for all ℓ, the perceived degree-ℓ empowerment of state s indeed decreases by increasing the probability ϵ of the execution noise. Theorem 2 additionally shows that Emp_ℓ accounts for the causation-prediction distinction proposed by ref.[14] as a necessary property of formal measures of the sense of control.

Theorem 2. *Suppose the action execution is noisy as in Eq. 27. Then, the perceived degree-ℓ empowerment of state s, derived based on $p_\epsilon(s'|s,a)$ in Eq. 28 and denoted by $\text{Emp}_{\ell,\epsilon}(s)$, is a decreasing function of ϵ, i.e.,*

$$\frac{\partial \text{Emp}_{\ell,\epsilon}(s)}{\partial \epsilon} \leq 0 \tag{29}$$

for all $s \in \mathcal{S}$, $\epsilon \in [0,1]$, $\ell \geq 0$, π_{noise}, and p.

Proof: See Supplementary Materials.

Existing measures of the sense of control

In this section, we review different existing measures of the sense of control in Fig. 1C. We also demonstrate how the various measures relate to different conceptualizations.

Mere number of actions. The simplest existing measure is the mere number $N_a(s)$ of actions available in state s (e.g., see ref.[1]). This serves as a crude measure of A-Control. In the toy example of Fig. 1A, an agent's sense of control based on $N_a(s)$ follows the ranking in Fig. 1B2. However, the main issue with $N_a(s)$ is that it does not account for the degree to which the actions are distinct. For example, duplicating the action to the painting class in c_2 would increase its 'controllability' according to $N_a(s)$. Two more appropriate alternatives are the (1-step) Klyubin-empowerment[13] $\text{Emp}_{\text{Kly}}(s)$ and the maximum outcome entropy[11] $H_{\max}(s)$; we discuss these alternatives next.

Klyubin-empowerment. The (1-step) Klyubin-empowerment $\text{Emp}_{\text{Kly}}(s)$ is defined as the maximum mutual information between the next state S' and the current action A, conditioned on the current state $S = s$:

$$\begin{aligned}\text{Emp}_{\text{Kly}}(s) :&= \max_{\pi_s} \underbrace{I_\pi(S', A|S=s)}_{\text{transfer entropy}^{14}} \\ &= \max_{\pi_s} \Big(\underbrace{H_\pi[A|S=s]}_{\text{action diversity}} - \underbrace{H_\pi[A|S', S=s]}_{\text{regularization regarding action distinction}} \Big),\end{aligned} \tag{30}$$

where I_π and H_π denote mutual information and entropy, respectively, under policy π; we note that, in the 1-step case considered here, Eq. 30 is equivalent to other formulations and generalizations of the Klyubin-empowerment[15,64]. In the reformulation of mutual information (the second equality in Eq. 30), the first term $H_\pi[A|S=s]$ quantifies the action diversity at state s and is a proxy of the total number of actions at state s; e.g., its maximum value is $\log N_a(s)$. On the other hand, $H_\pi[A|S', S=s]$ quantifies the ambiguity of action A in light of S and S' (i.e., how accurately the agent can infer its past action based on the transition). Hence, $e^{\mathrm{Emp}_{\mathrm{Kly}}}(s) \in [1, N_a(s)]$ can be seen as a regularized version of $N_a(s)$ and a measure of A-Control, similarly to $\mathrm{Emp}_1(s)$ (Corollary 1).

In the toy example of Fig. 1A, an agent's sense of control based on $\mathrm{Emp}_{\mathrm{Kly}}(s)$ follows the ranking in Fig. 1B2. The mutual information $I_\pi(S', A|S = s)$ is also known as (conditional) transfer entropy, which is another proposed measure of the sense of control[14]. It is straightforward to show that, for any choice of π, an agent's sense of control based on transfer entropy also follows the ranking in Fig. 1B2. Hence, conceptually, $\mathrm{Emp}_{\mathrm{Kly}}$ (and transfer entropy) quantifies a notion of action-distinction and hence A-Control.

However, action distinction is the only quality of a state that $\mathrm{Emp}_{\mathrm{Kly}}$ quantifies. Specifically, $\mathrm{Emp}_{\mathrm{Kly}}$ does not consider the outcome distribution of different actions *as long as they are distinct.* Proposition 2 below makes this statement precise, connects $\mathrm{Emp}_{\mathrm{Kly}}$ to the Emp_1, and showcases that $\mathrm{Emp}_{\mathrm{Kly}}$ is blind to certainty-possibility dichotomy.

Proposition 2. *Suppose $\mathcal{A}(s)$ consists of $M(s)$ partitions $\{\mathcal{A}_1(s), \ldots, \mathcal{A}_{M(s)}(s)\}$, where: 1. All actions within each partition are identical, i.e., $p(.|s, a) = p(.|s, a')$ if a and a' are in the same partition. 2. Actions in different partitions lead to distinct sets of next states, i.e., $p(.|s, a)$ and $p(.|s, a')$ have disjoint supports on $\mathcal{S}$ if a and a' belong to different partitions. Then, we have $\mathrm{Emp}_1(s) = e^{\mathrm{Emp}_{\mathrm{Kly}}(s)} = M(s)$, independently of the exact transition probabilities.*

Proof: See Supplementary Materials.

Maximum outcome entropy. The (1-step) maximum entropy $H_{\max}(s)$ quantifies the diversity of the potential next states that an agent can reach:

$$H_{\max}(s) := \max_{\pi_s} H_\pi[S'|S = s], \tag{31}$$

where $H_\pi[S'|S = s]$ denotes the Shannon entropy of S' under policy π (Fig. 1C). In the toy example of Fig. 1A, an agent's sense of control based on $H_{\max}(s)$ follows the ranking in Fig. 1B3. Accordingly, $H_{\max}(s)$ can be seen as a measure of $\mathrm{C_S}$-Control, but it is important to note that $H_{\max}(s)$ ignores any notion of certainty of outcomes or distinctiveness of actions.

Sensing control in deterministic states

A deterministic state is uniquely characterized by the transition function $T(., s) : \mathcal{A}(s) \to \mathcal{S}$ that determines the next state s' given the previous state-action pair s, a, i.e.,

$$p(s'|s, a) = \mathbb{1}_{s'=T(s,a)}. \tag{32}$$

We call two deterministic actions 'distinct' regarding their outcome if they take the agent to different next states, i.e., if $T(s, a) \neq T(s, a')$; otherwise, the two actions are indistinguishable regarding their outcome. As a result, in a deterministic state, the number of distinct actions is equal to the number of reachable next states expressed as $|\{T(s, a) : a \in \mathcal{A}(s)\}|$. Proposition 3 below shows that, in deterministic states, different measures of the sense of control are indistinguishable from each other. Notably, they all assign the highest value of the sense of control to central hubs[13,60,70].

Proposition 3. *Suppose $s \in \mathcal{S}$ is a 'deterministic' state with transition function T as in Eq. 32. Then, we have $\mathrm{Emp}_\ell(s) = e^{\mathrm{Emp}_{\mathrm{Kly}}(s)} = e^{H_{\max}(s)} = |\{T(s, a) : a \in \mathcal{A}(s)\}|$ for all $\ell \geq 0$.*

Proof: See Supplementary Materials.

Acknowledgement

We thank Sophia Becker, Franziska Brändle, Johanni Brea, Wulfram Gerstner, Lucas Gruaz, Kai Sandbrink, and Elena Zamfir for insightful discussions on related topics. We also thank Susanne Haridi, Milena Rmus, Tianyuan Teng, Andrew Webb, Kristin Witte, and many other members of the Dayan Lab and the Institute for Human-Centered AI for their help with the experiment design and data collection. This research was supported by the Max Planck Society (PD), the Humboldt Foundation (PD), and the European Research Council (ES).

Competing Interests statement

The authors declare no competing interests.

Code and data availability

All code and data needed to reproduce the results reported in this manuscript will be made publicly available after publication acceptance.

References

1. Leotti, L. A., Iyengar, S. S. & Ochsner, K. N. Born to choose: The origins and value of the need for control. *Trends in Cognitive Sciences* **14**, 457–463 (2010).

2. Moscarello, J. M. & Hartley, C. A. Agency and the calibration of motivated behavior. *Trends in Cognitive Sciences* **21**, 725–735 (2017).

3. Wang, K. S., Yang, Y.-Y. & Delgado, M. R. How perception of control shapes decision making. *Current Opinion in Behavioral Sciences* **41**, 85–91 (2021). Value based decision-making.

4. Goddu, M. K. & Gopnik, A. The development of human causal learning and reasoning. *Nature Reviews Psychology* **3**, 319–339 (2024).

5. Nussenbaum, K. & Hartley, C. A. Understanding the development of reward learning through the lens of meta-learning. *Nature Reviews Psychology* **3**, 424–438 (2024).

6. Eysenbach, B., Gupta, A., Ibarz, J. & Levine, S. Diversity is all you need: Learning skills without a reward function. In *International Conference on Learning Representations (ICLR)* (2019).

7. Choi, J., Sharma, A., Lee, H., Levine, S. & Gu, S. S. Variational empowerment as representation learning for goal-conditioned reinforcement learning. In *International Conference on Machine Learning (ICML)*, 1953–1963 (PMLR, 2021).

8. Leibfried, F., Pascual-Díaz, S. & Grau-Moya, J. A unified bellman optimality principle combining reward maximization and empowerment. In Wallach, H. *et al.* (eds.) *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 32 (Curran Associates, Inc., 2019).

9. Sontag, E. D. *Mathematical Control Theory: Deterministic Finite Dimensional Systems* (Springer New York, NY, 2013).

10. Ogata, K. *Modern Control Engineering* (Prentice Hall, 2010), 5th edn.

11. Huys, Q. J. & Dayan, P. A Bayesian formulation of behavioral control. *Cognition* **113**, 314–328 (2009).

12. Cover, T. M. *Elements of Information Theory* (John Wiley & Sons, 1999).

13. Klyubin, A., Polani, D. & Nehaniv, C. Empowerment: a universal agent-centric measure of control. In *2005 IEEE Congress on Evolutionary Computation*, vol. 1, 128–135 Vol.1 (2005).

14. Ligneul, R. Prediction or causation? towards a redefinition of task controllability. *Trends in Cognitive Sciences* **25**, 431–433 (2021).

15. Abel, D. *et al.* Plasticity as the mirror of empowerment. *arXiv preprint arXiv:2505.10361* (2025).

16. Mancinelli, F., Roiser, J. & Dayan, P. Internality and the internalisation of failure: Evidence from a novel task. *PLoS Computational Biology* **17**, e1009134 (2021).

17. Dorfman, H. M. & Gershman, S. J. Controllability governs the balance between pavlovian and instrumental action selection. *Nature Communications* **10**, 5826 (2019).

18. Sandbrink, K., Hunt, L. & Summerfield, C. Understanding human meta-control and its pathologies using deep neural networks. *PsyArXiv* (2024).

19. Grahek, I., Frömer, R., Prater Fahey, M. & Shenhav, A. Learning when effort matters: neural dynamics underlying updating and adaptation to changes in performance efficacy. *Cerebral Cortex* **33**, 2395–2411 (2022).

20. Ligneul, R., Mainen, Z. F., Ly, V. & Cools, R. Stress-sensitive inference of task controllability. *Nature Human Behaviour* **6**, 812–822 (2022).

21. Frömer, R., Lin, H., Dean Wolf, C., Inzlicht, M. & Shenhav, A. Expectations of reward and efficacy guide cognitive control allocation. *Nature Communications* **12**, 1030 (2021).

22. Stolz, D. S., Müller-Pinzler, L., Krach, S. & Paulus, F. M. Internal control beliefs shape positive affect and associated neural dynamics during outcome valuation. *Nature Communications* **11**, 1230 (2020).

23. Limbachia, C. *et al.* Controllability over stressor decreases responses in key threat-related brain areas. *Communications Biology* **4**, 42 (2021).

24. Ajzen, I. Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *Journal of Applied Social Psychology* **32**, 665–683 (2002).

25. Seligman, M. E. & Maier, S. F. Failure to escape traumatic shock. *Journal of Experimental Psychology* **74**, 1–9 (1967).

26. Bandura, A. Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review* **84**, 191–215 (1977).

27. Rotter, J. B. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied* **80**, 1–28 (1966).

28. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* (MIT press, 2018).

29. Puterman, M. L. *Markov Decision Processes: Discrete Stochastic Dynamic Programming* (John Wiley & Sons, 1994).

30. Bertsekas, D. *Dynamic Programming and Optimal Control: Volume I* (Athena Scientific, 2012).

31. Twitter, Inc. Twemoji: Twitter emoji (2025). URL https://github.com/twitter/twemoji.

32. Levenson, H. Activism and powerful others: Distinctions within the concept of internal-external control. *Journal of Personality Assessment* **38**, 377–383 (1974).

33. Carleton, R. N., Norton, M. P. J. & Asmundson, G. J. Fearing the unknown: A short version of the intolerance of uncertainty scale. *Journal of Anxiety Disorders* **21**, 105–117 (2007).

34. Bradley, R. A. & Terry, M. E. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika* **39**, 324–345 (1952).

35. Agresti, A. *Categorical Data Analysis* (John Wiley & Sons, 2013).

36. Rigoux, L., Stephan, K. E., Friston, K. J. & Daunizeau, J. Bayesian model selection for group studies—revisited. *NeuroImage* **84**, 971–985 (2014).

37. Wilson, R. C. & Collins, A. G. Ten simple rules for the computational modeling of behavioral data. *eLife* **8**, e49547 (2019).

38. Efron, B. & Hastie, T. *Computer age statistical inference* (Cambridge University Press, 2016).

39. Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J. & Friston, K. J. Bayesian model selection for group studies. *NeuroImage* **46**, 1004–1017 (2009).

40. MacKay, D. J. *Information theory, inference and learning algorithms* (Cambridge university press, 2003).

41. Hastie, T., Tibshirani, R., Friedman, J. H. & Friedman, J. H. *The elements of statistical learning: data mining, inference, and prediction*, vol. 2 (Springer, 2009).

42. Kahneman, D. & Tversky, A. Prospect theory: An analysis of decision under risk. *Econometrica* **47**, 263–291 (1979).

43. Pachur, T., Schulte-Mecklenbeck, M., Murphy, R. O. & Hertwig, R. Prospect theory reflects selective allocation of attention. *Journal of Experimental Psychology: General* **147**, 147–169 (2018).

44. Peterson, J. C., Bourgin, D. D., Agrawal, M., Reichman, D. & Griffiths, T. L. Using large-scale experiments and machine learning to discover theories of human decision-making. *Science* **372**, 1209–1214 (2021).

45. Baratta, M. V., Seligman, M. E. & Maier, S. F. From helplessness to controllability: toward a neuroscience of resilience. *Frontiers in Psychiatry* **14**, 1170417 (2023).

46. Maier, S. F. & Seligman, M. E. Learned helplessness at fifty: Insights from neuroscience. *Psychological Review* **123**, 349–367 (2016).

47. Beck, B., Di Costa, S. & Haggard, P. Having control over the external world increases the implicit sense of agency. *Cognition* **162**, 54–60 (2017).

48. Giersiepen, M., Schütz-Bosbach, S. & Kaiser, J. Freedom of choice boosts midfrontal theta power during affective feedback processing of goal-directed actions. *Biological Psychology* **183**, 108659 (2023).

49. Deci, E. L. & Ryan, R. M. The support of autonomy and the control of behavior. *Journal of Personality and Social Psychology* **53**, 1024 (1987).

50. Gregor, K., Rezende, D. J. & Wierstra, D. Variational intrinsic control. *arXiv preprint arXiv:1611.07507* (2016).

51. Bharadhwaj, H., Babaeizadeh, M., Erhan, D. & Levine, S. Information prioritization through empowerment in visual model-based RL. In *International Conference on Learning Representations (ICLR)* (2022).

52. Pathak, D., Agrawal, P., Efros, A. A. & Darrell, T. Curiosity-driven exploration by self-supervised prediction. In *Proceedings of the 34th International Conference on Machine Learning (ICML) - Volume 70*, ICML'17, 2778–2787 (JMLR.org, 2017).

53. Shelhamer, E., Mahmoudieh, P., Argus, M. & Darrell, T. Loss is its own reward: Self-supervision for reinforcement learning. *arXiv preprint arXiv:1612.07307* (2016).

54. Salge, C., Glackin, C. & Polani, D. Empowerment–an introduction. *Guided Self-Organization: Inception* 67–114 (2014).

55. Volpi, N. C. & Polani, D. Goal-directed empowerment: combining intrinsic motivation and task-oriented behavior. *IEEE Transactions on Cognitive and Developmental Systems* **15**, 361–372 (2020).

56. Becker-Ehmck, P., Karl, M., Peters, J. & van der Smagt, P. Exploration via empowerment gain: Combining novelty, surprise and learning progress. In *ICML 2021 Workshop on Unsupervised Reinforcement Learning* (2021).

57. Cao, H. *et al.* Towards empowerment gain through causal structure learning in model-based reinforcement learning. In *The Thirteenth International Conference on Learning Representations (ICLR)* (2025).

58. Jung, T., Polani, D. & Stone, P. Empowerment for continuous agent—environment systems. *Adaptive Behavior* **19**, 16–39 (2011).

59. Dai, S., Xu, W., Hofmann, A. & Williams, B. An empowerment-based solution to robotic manipulation tasks with sparse rewards. *Autonomous Robots* **47**, 617–633 (2023).

60. Gruaz, L., Modirshanechi, A., Becker, S. & Brea, J. Merits of curiosity: a simulation study. *PsyArXiv (to appear in Open Mind)* (2024).

61. Singh, S., Lewis, R. L., Barto, A. G. & Sorg, J. Intrinsically motivated reinforcement learning: An evolutionary perspective. *IEEE Transactions on Autonomous Mental Development* **2**, 70–82 (2010).

62. Aubret, A., Matignon, L. & Hassas, S. A survey on intrinsic motivation in reinforcement learning. *arXiv preprint arXiv:1908.06976* (2019).

63. Modirshanechi, A., Kondrakiewicz, K., Gerstner, W. & Haesler, S. Curiosity-driven exploration: foundations in neuroscience and computational modeling. *Trends in Neurosciences* **46**, 1054–1066 (2023).

64. Capdepuy, P. *Informational principles of perception-action loops and collective behaviours*. Ph.D. thesis, University of Hertfordshire (2011).

65. Anvari, F., Billinger, S., Analytis, P. P., Franco, V. R. & Marchiori, D. Testing the convergent validity, domain generality, and temporal stability of selected measures of people's tendency to explore. *Nature Communications* **15**, 7721 (2024).

66. Witte, K., Thalmann, M. & Schulz, E. How should we measure exploration? *PsyArXiv* (2024).

67. Vrizzi, S., Najar, A., Lemogne, C., Palminteri, S. & Lebreton, M. Behavioral, computational and self-reported measures of reward and punishment sensitivity as predictors of mental health characteristics. *Nature Mental Health* (2025).

68. Jach, H. K. *et al.* Individual differences in information demand have a low dimensional structure predicted by some curiosity traits. *Proceedings of the National Academy of Sciences (PNAS)* **121**, e2415236121 (2024).

69. Lidayan, A. *et al.* Intrinsically-motivated humans and agents in open-world exploration. *arXiv preprint arXiv:2503.23631* (2025).

70. Brändle, F., Stocks, L. J., Tenenbaum, J. B., Gershman, S. J. & Schulz, E. Empowerment contributes to exploration behaviour in a creative video game. *Nature Human Behaviour* (2023).

71. Langer, E. J. The illusion of control. *Journal of Personality and Social Psychology* **32**, 311 (1975).

72. Zorowitz, S., Solis, J., Niv, Y. & Bennett, D. Inattentive responding can induce spurious associations between task behaviour and symptom measures. *Nature Human Behaviour* **7**, 1667–1681 (2023).

73. Daw, N. Trial-by-trial data analysis using computational models. *Decision making, affect, and learning: Attention and performance XXIII* **23** (2011).

74. Barber, D. *Bayesian Reasoning and Machine Learning* (Cambridge University Press, 2012).

75. Ge, H., Xu, K. & Ghahramani, Z. Turing: a language for flexible probabilistic inference. In *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain*, 1682–1690 (2018).

76. Modirshanechi, A., Lin, W.-H., Xu, H. A., Herzog, M. H. & Gerstner, W. Even if suboptimal, novelty drives human exploration. *bioRxiv* (2025).

77. Kass, R. E. & Raftery, A. E. Bayes factors. *Journal of the american statistical association* **90**, 773–795 (1995).

78. Bishop, C. M. *Pattern Recognition and Machine Learning* (Springer New York, NY, 2006).

79. Bates, S., Hastie, T. & Tibshirani, R. Cross-validation: What does it estimate and how well does it do it? *Journal of the American Statistical Association* **119**, 1434–1445 (2024).

80. Sandbrink, K. & Summerfield, C. Modelling cognitive flexibility with deep neural networks. *Current Opinion in Behavioral Sciences* **57**, 101361 (2024).