# Resource-rational models of human goal pursuit

Ben Prystawski*[,1,2]
benprystawski@gmail.com

Florian Mohnert*[,1]
mail@florian-mohnert.de

Mateo Tošić[1]
mateo.tosic@hotmail.com

Falk Lieder[1]
falk.lieder@tuebingen.mpg.de

November 2, 2020

1: Max Planck Institute for Intelligent Systems, Tübingen, Germany 2: University of Toronto, Toronto, Canada

* These authors contributed equally.

**Keywords**: resource rationality; bounded rationality; goal pursuit; decision making; planning; cognitive modelling

**Word count:** 6,944

**Abstract**

Goal-directed behaviour is a deeply important part of human psychology. People constantly set goals for themselves and pursue them in many domains of life. In this paper, we develop computational models that characterize how humans pursue goals in a complex dynamic environment and test how well they describe human behaviour in an experiment. Our models are motivated by the principle of resource rationality and draw upon psychological insights about people's limited attention and planning capacities. We found that human goal pursuit is qualitatively different and substantially less efficient than optimal goal pursuit. Models of goal pursuit based on the principle of resource rationality captured human behavior better than both a model of optimal goal pursuit and heuristics that are not resource-rational. We conclude that human goal pursuit is jointly shaped by its function, the structure of the environment, and cognitive costs and constraints on human planning and attention. Our findings are an important step toward understanding humans goal pursuit, as cognitive limitations play a crucial role in shaping people's goal-directed behaviour.

# 1 Introduction

Human behavior is fundamentally goal-directed (Carver and Scheier, 2001). People are often unaware of some of their goals and the ways in which they pursue them (Custers and Aarts, 2010). But their behavior is constantly driven by the pursuit of one or more goals nevertheless. The central role that goals occupy in human cognition and behavior raises the question of why people have goals in the first place. This question poses a serious challenge to classical notions of rationality (Lieder and Griffiths, 2020). In fact, from the perspective of rational decision theory (Morgenstern and Von Neumann, 1953), goals appear unnecessary because one should simply choose the series of actions that maximizes one's expected utility in the long run. To cope with the intractability of large decision problems, the mind breaks complex problems down into simpler, more tractable, sub-problems. For instance, the problem of living a good life can be broken down into first quenching one's thirst (a short-term goal) and then thinking about what to do next. Once this goal has been set, deciding what to do becomes much easier. Knowing the goal allows people to focus their limited attention on the most relevant aspects of their environment (e.g., the location of the closest drinking fountain relative to where they are now) and eschew having to plan many steps ahead by using simple heuristics (e.g., walk straight in the direction of the drinking fountain until you encounter an obstacle, then navigate around the obstacle). Goals simplify sequential decision problems because they allow people to use efficient representations and clever heuristics. Although this may appear intuitively clear and there has been some work on heuristics that people use to pursue goals (Newell and Simon, 1972), there is still no formal unified theory of which heuristics and representations people employ to pursue goals efficiently in the light of their bounded cognitive resources.

In particular, it is still unclear which strategies people use to pursue their goals in complex dynamic environments, how they represent those environments depending on their goal, which cognitive resources most strongly constrain their strategies, and how well people perform relative to their cognitive constraints. To address these questions, we propose that the principle that people make optimal use of their finite time and bounded cognitive resources (*resource rationality*; Lieder and Griffiths, 2020), can be used to elucidate how people pursue their goals and why they pursue their goals in this way. We apply this principle to develop a series of resource-rational models that express the implications of limited attention, limited planning, or both on people's representations and heuristics. We test the predictions of these models against the predictions of a model of unboundedly rational goal pursuit and two baseline models in a simulated micro-world paradigm in which participants manage a farm.

The resource-rational approach allowed us to develop computational models that predict human goal pursuit substantially better than either the unboundedly rational model or the baseline models. Our experiment further revealed that the representations and heuristics people use to pursue goals are shaped by both their limited attentional resources and their short planning horizons. Given the central role of goal pursuit in human cognition and behavior these findings are an important step towards understanding how people navigate their lives.

The outline of this article is as follows. We start by summarizing previous approaches to studying how people pursue their goals and our resource-rational modeling paradigm. We then formulate an optimal model of goal pursuit, a series of resource-rational models, and two alternative models. Having formulated these models we then test their predictions in a behavioral experiment and discuss the implications of our findings.

# 2 Background

In this section, we present related work and ideas upon which we draw in developing our models of goal pursuit and designing the task for our experiment.

## 2.1 Theories of human goal pursuit

Various previous theories have construed goal pursuit in terms of feedback control (Carver and Scheier, 2001), planning (Gabaix, 2016; Newell and Simon, 1972; Botvinick and Toussaint, 2012), reinforcement learning (Juechems and Summerfield, 2019), and active inference (Pezzulo et al., 2015, 2018; Botvinick and Toussaint, 2012). Before presenting our own perspective, we briefly review these existing theories, emphasizing their commonalities and differences.

According to feedback control theories, the perceived discrepancy between a system's current state and the goal serves as a feedback signal that drives goal-directed behavior (Wiener, 1948; Miller et al., 1960; Powers, W and Clark, R K and MacFarland, R L, 1960; Carver and Scheier, 2001). This perspective pervades theories in computational models in neuroscience and psychology. The psychological perspective (Carver and Scheier, 2001) views human goal pursuit as a form of feedback control. According to this view, people select their actions so as to reduce the perceived discrepancy between what they perceive to be the case and what should be the case according to the goal(s) they are currently pursuing. According to Carver and Scheier (2001), people's goals are organized in hierarchies. The goals at the top of the hierarchy are abstract long-term goals (e.g., "Make the world a better place."). The goals at the bottom of the hierarchy are very concrete short-term goals (e.g., "Drink a sip of water."). Each goal's subgoals serve to facilitate its attainment or maintenance by making it more actionable. Consistent with this view, the theory of active inference posits that goal-directed behaviour relies on inverting a hierarchical generative model of the relationships between action and perception (Botvinick et al., 2009; Friston, 2010; Pezzulo et al., 2015, 2018).

Goal pursuit has also been studied in the reinforcement learning framework. Supporting this perspective, it has been found that the achievement of subgoals triggers the same dopaminergic reward signals as receiving external rewards (Ribas-Fernandes et al., 2019, 2011; Mas-Herrero et al., 2019). Juechems and Summerfield (2019) have explicitly integrated the cybernetic perspective on goal pursuit into the reinforcement learning framework. Concretely, they present the idea of homeostatically regulated reinforcement learning as a possible explanation for how humans derive value from internal states and external stimuli. In their theory, goals are "cognitive setpoints" and the positive and negative rewards that people experience communicate the perceived reduction and increase in the discrepancy between the current state and the goal state, respectively.

The principles of feedback control and reinforcement learning often lead to simple reactive control laws that specify the system's output as a function of its input. These control laws enable agents to swiftly respond to changes in the environment or their internal state. However, human goal pursuit can also involve planning, wherein a person may deliberate to find a sequence of actions that will likely bring them closer to their goal. Formally, planning can be modelled as search (Newell and Simon, 1972), dynamic programming (Gabaix, 2016), or inference (Botvinick and Toussaint, 2012). Empirical research on planning has generally found that people's capacity for planning is rather limited (Callaway et al., 2018; Newell and Simon, 1972; Gabaix, 2016). For instance, Newell and Simon (1972) ask people to think aloud about how to achieve the goals that they had been asked to achieved in various problem solving tasks. Their data suggested that while people did engage in planning, they were generally unable to plan more than a few steps ahead and to consider more than a few alternatives.

## 2.2 Resource rationality

To characterize humans' goal-directed behavior, we develop computational models which are motivated by the principle of resource rationality (Lieder and Griffiths, 2020), which assumes that the human mind makes optimal use of limited cognitive resources. This theory can be applied to model the way in which people decide in a given environment $E$ by the resource-rational heuristic

$$h^\star = \underset{h \in H_B}{\operatorname{argmax}} \, \mathbb{E}\left[\mathrm{RR(h,E,B)}\right], \tag{1}$$

where $H_B$ is the set of heuristics that the person's brain can implement and the resource-rationality $RR(h, E, B)$ of the heuristic $h$ given the cognitive constraints of the brain $B$ is

$$\mathrm{RR}(h, E, B) = \mathbb{E}_{P(\text{result}|s_0,h,E,B)}\left[u(\text{result})\right] - \mathbb{E}\left[\text{cost}(t_h, \rho) \mid h, s_0, B, E\right], \tag{2}$$

where $u$(result) is the person's subjective utility $u$ of the outcomes (result) of the choices made by the heuristic $h$, $s_0 = (o, b_0)$ comprises the observed information about the initial state of the external world ($o$) and the agent's initial internal state $b_0$, and cost$(t_h, \rho)$ denotes the total opportunity cost of investing the cognitive resources $\rho$ used or blocked by the heuristic $h$ for the duration $t_h$ of its execution. Both the result of applying the heuristic and its execution time depend on the situation in which it is applied. The expected values ($\mathbb{E}$) weigh the utility and cost for each possible situation by their posterior probability given the environment $E$ and the observed characteristics of the current situation ($o$). The brain's computational limitations and uncertainty about the environment limit how effective people's decision strategies can be. In developing resource-rational models of human goal pursuit, we make assumptions about the ways in which human decision-making capacity is limited, then develop models which account for these constraints and reflect optimal behavior subject to them. The principle of resource rationality has been successfully applied to model cognitive processes such as planning (Callaway et al., 2018) as well as explain traditional biases such as the anchoring bias (Lieder et al., 2018b) and availability biases (Lieder et al., 2018a).

The principle of resource-rational applies not only to heuristics but also the representations. One of the first instances of resource-resource rational representations is the idea behind sparse dynamic programming (Gabaix, 2016). Sparse dynamic programming is a method for modelling decision-making which accounts for the cognitive costs of paying attention. Rather than using all available information, the sparse dynamic programming model assumes that people represent the subset of their environment that is most relevant to their decision, then plan the future effects of their action in this subset and choose the best options based on the limited information they attend to. The sparse-max operator is a special case of sparse dynamic programming which does not model planning several steps ahead (Gabaix, 2014).

### 2.3 Simulated Micro-worlds

Simulated micro-worlds (SMWs) are a common paradigm used to study problem solving in complex environments (Brehmer and Dörner, 1993). SMWs are generally designed to model some real-world task, such as running a fictional airline company or being the mayor of a village. A SMW consists of a set of variables which are related to each other and a number of discrete time steps. The person interacting with the SMW can change some variables, but might only be able to influence other variables indirectly. For instance, participants might manipulate the ticket price of the fictional airline company directly but not customer satisfaction. This makes SMWs suitable for studying how people might pursue goals in the real world. So far, SMWs have only rarely been used to study goal pursuit directly (Rohe et al., 2016). Rather, participants were usually just presented with the simulated environment without any explicit goals (Brehmer and Dörner, 1993). In this article, we develop a simulated micro-world paradigm for studying human goal pursuit where people must select inputs to steer a dynamical system toward a target state.

## 3 Models of Human Goal Pursuit

There are three distinct types of questions that one can ask about goal pursuit (Marr, 1982). First, one can ask "What is the purpose of goal pursuit? Which problem does it solve?". Second, one can ask "What are the cognitive strategies and representations that people employ to pursue their goals?". Third, one can ask "Where and how are these processes and representations realized in the brain?" In this article, we focus on the second question. David Marr referred to investigating cognitive systems through the lens of these three questions as the *computational level*, *algorithmic level*, and *implementation level*, respectively (Marr, 1982).

We develop three types of model: a model of optimal goal pursuit, resource-rational models of goal pursuit, and heuristic models of goal pursuit. The optimal model and the resource-rational models require a theory of the function of goal pursuit (i.e., an answer to the first question). We therefore first present our theory of the function of goal pursuit and then introduce the three types of models in turn.

### 3.1 Computational level theory of goal pursuit

In the following, we assume that the function of goal pursuit is to select actions that minimize the discrepancy between the state of the environment ($s$) and a given goal ($g$) by a given deadline ($N$) while minimizing the cost of the efforts that are exerted to achieve it, that is

$$\arg\min_{\pi} C(\pi) \text{ with } C(\pi) = \sqrt{||\mathbf{s}_N - \mathbf{g}||_2^2 + c \cdot \sum_{i=0}^{N-1} \text{cost}(\pi(\mathbf{s_i}))^2}, \tag{3}$$

where the goal pursuit policy $\pi : \mathcal{S} \mapsto \mathcal{A}$ selects actions, such as how much money to invest or many hours to work on given project on a given day, based on the current state, $\text{cost}(\pi(s_i))$ is the cost of those investments, and $c$ controls its relative importance.

Concretely, we modelled goal pursuit in the context of a simulated micro-world (SMW) that models the task of managing a farm (see Figure 1). The dynamics the system we use to model goal pursuit correspond to a system of linear equations (Funke, 1993) with the following transition function:

$$\mathbf{s}_{t+1} = f(\mathbf{s}_t, \mathbf{a}_t) = \mathbf{A}\mathbf{s}_t + \mathbf{B}\mathbf{a}_t, \tag{4}$$

where the vectors $\mathbf{s}_t$ and $\mathbf{a}_t$ contain the current values of the endogenous and exogenous variables, respectively. Matrix $\mathbf{A} \in \mathbb{R}^{D_s \times D_s}$ determines both the effect of endogenous variables both on themselves and each other from time $t$ to $t+1$, while matrix $\mathbf{B} \in \mathbb{R}^{D_s \times D_e}$ determines how the exogenous variables set by the agent at time $t$ affect the endogenous variables at time $t+1$. Participants pursuing a goal must bring the endogenous variables close to a goal state $\mathbf{g} \in \mathbb{R}^{D_s}$. In each round participants could take action to either increase or decrease the levels of various nutrients, herbicides, and pesticides on the fields. We therefore model the cost of the participant's action $\pi(s)$ as the Euclidean norm of the selected changes, that is $\text{cost}(\pi(\mathbf{s})) = ||\pi(\mathbf{s})||_2$. We set the relative importance of these costs to $c$=0.01 to emphasize reaching the goal as relatively more important than keeping inputs small while still approaching the goal state. It is also ideal to distribute the weight of inputs across exogenous variables and time steps as much as possible.

## 3.2 Optimal goal pursuit: the linear-quadratic regulator

We use a linear-quadratic regulator (Kirk, 2004) to model optimal goal pursuit. This model computes the sequence of actions that minimizes the cost function specified in Equation 3. A more detailed description of the linear-quadratic regulator can be found in the Supplementary Material.

## 3.3 Resource-rational models of goal-pursuit

In pursuing a goal in a real or artificial environment, there are often many possible variables and relationships to keep track of. For instance, the goal of learning course material in a class requires keeping track of one's current grade, estimates of confidence in various course concepts, and how those concepts relate to each other in order to plan what to study and how. We might expect human goal pursuit to be sub-optimal in cases where cognitive limitations interfere with the ability to pursue a goal. For instance, limited working memory likely constrains humans' ability to plan multiple steps ahead when solving the Tower of Hanoi problem (Kotovsky et al., 1985). To characterize how resource constraints might shape goal-directed behaviour, we develop resource-rational models of goal pursuit that account for limited attention and limited planning capacity factorially. These models account for limited attention, limited planning, or both. The Supplementary Material contains more detailed descriptions of how these models choose actions.

### 3.3.1 Limited planning: hill-climbing

We use an optimal hill-climbing model to account for limited planning ability. This strategy has been used in the past to model human problem solving (Simon and Newell, 1971). Our model assumes that while people's limited resources do not allow them to look more than a single step into the future, people make rational use of those resources by taking a step that is near-optimal in the short-run in terms of its direction and its length. Concretely, at each time step, it moves in the direction of the negative gradient of its distance to the goal state by just the right amount ($\lambda_{\text{opt}}$) to maximally reduce the distance to the goal after the current step, that is

$$\Delta = \lambda_{\text{opt}} \cdot \nabla_{\mathbf{a}} ||f(\mathbf{s}_t, \mathbf{a}) - \mathbf{g}||_2 \tag{5}$$

where the gradient is evaluated at $a = \mathbf{0}$. To capture that a person's steps might be systematically too small or too large, a free parameter $\lambda$ maps the optimal step onto the chosen action, that is $\mathbf{a}_t = \lambda \cdot \Delta$.

### 3.3.2 Limited attention: sparse LQR

We incorporate attention costs into a model of goal pursuit using a modified version of the *sparse-max* operator that Gabaix (2014) introduced as a psychologically plausible version of maximization. The sparse LQR model assumes that (1) attention is a limited and costly resource and (2) people allocate their limited attention in a near-optimal manner. This model selects an attention vector $\mathbf{m}^\star$ that specifies which variables and which effects are attended to and which are ignored. This attention vector is chosen so as to minimize the weighted sum of the score achieved by planning with this simplified representation and the cognitive cost of attending to it, that is

$$\mathbf{m}^\star = \underset{\mathbf{m}}{\operatorname{argmin}}\, C(\pi_{\mathbf{m}}) + k \cdot \sum_i \mathbf{m}_i \tag{6}$$

where $C$ is the cost function (Equation 3), $\pi_{\mathbf{m}}$ is the plan that the LQR selects when applied to the simplified representation of the environment specified by $\mathbf{m}$, and $k$ is the cost per attended element of the environment.

### 3.3.3 Limited planning and limited attention: sparse hill-climbing

Finally, we use a sparse hill-climbing model to account for planning costs and attention costs simultaneously. This model applies the hill-climbing heuristic (Equation 5) within a simplified mental representation of the environment. This representation is constructed by selecting an attention vector that maximizes the performance attained by applying the hill-climbing heuristic to the resulting representation while minimizing the cost of attention. We developed two versions of this model. In the discrete version, each relationship between variables is either attended to ($\mathbf{m}_i = 1$) or not ($\mathbf{m}_i = 0$). In the continuous version, each relationship between variables is partially attended ($0 \leq \mathbf{m}_i \leq 1$), where 1 represents full attention and smaller numbers lead the model to perceive the relationships to be weaker than they truly are. These models have two free parameters: the step size and the attention cost.

## 3.4 Baseline models

We propose two null models which serve as baselines against which we can compare the fit of our resource-rational models. Null model 1 is based on an intuitive process that people might follow: First, $n$ endogenous variables are selected at random. Then, for each chosen endogenous variable $s$, the agent randomly selects one exogenous variable $a$ that affects $s$. Next, $a$ is set in the direction that brings $s$ toward its target value. Its magnitude is sampled from a uniform distribution between 0 and $b$. $n$ and $b$ are both free parameters of this model.

Null model 2 simply never uses any resources. In each time step, it takes the default action of leaving all the exogenous variables at zero. Therefore, the trajectory of this model is governed entirely by the endogenous transition matrix $\mathbf{A}$. This model has no free parameters, other than the parameters of the observation model.

## 3.5 Observation model

To model how the idealized cognitive processes in each model give rise to concrete goal-directed actions, we model two independent types of noise. *Length noise* describes people taking either larger or smaller steps than the model predicts, while *angular noise* describes people moving in different directions than the model predicts. The details of how we decompose the differences between human and model actions are included in the Supplementary Material.

## 3.6 Qualitative predictions of goal pursuit models

Our resource-rational models make two qualitative predictions about how humans pursue goals. First, the limited-attention models represent only a subset of the relationships between variables, leaving some exogenous variables disconnected from the environment. Our models predict that people will usually leave some the variables unchanged. In contrast, the optimal model always manipulates all of the exogenous

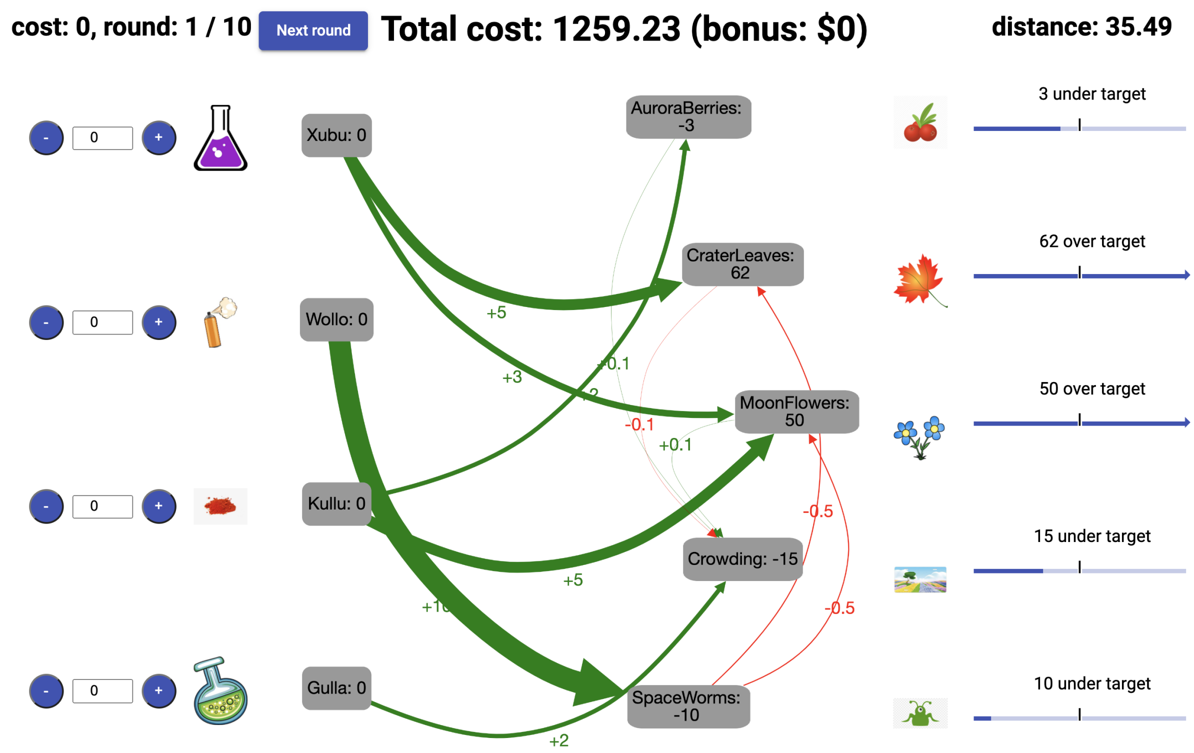



Figure 1: Screenshot of the simulated micro-world task shown to participants in Experiment 1.

variables. Second, our limited planning models predict that the magnitude of inputs will vary significantly over time. The limited planning models always minimize the immediate costs of their input in the current step only. They therefore take larger steps when they are farther from the goal and smaller steps when they are closer to the goal. This variability contrasts with the optimal model, which balances its inputs roughly evenly across rounds. Importantly, these predictions are inherent to the general structure of our models and do not depend on the specific settings of their free parameters.

# 4 Experiment: Which model best explains human goal pursuit?

We had two goals for our experiment. First, we wanted to test which of our models best explains human goal pursuit and which parameter values best fit the data. Second, we wanted to test whether a model of unboundedly-optimal goal pursuit ever explained humans' actions better than bounded models. To achieve these goals, we designed an experiment in which participants were asked to pursue a goal in a SMW. Each action they took was recorded along with the total cost they achieved. Each model was then fit to each participant's data individually and Bayesian model selection (Stephan et al., 2009) was applied to identify the model that best explains the data from each individual participant.

### 4.1 Methods

We recruited 180 participants via *Positly*, an online participant recruitment service which interfaces with Amazon Mechanical Turk. 111 of the participants passed the attention checks and did not report a lack of understanding of the task in the experiment feedback, so their data was used in subsequent analysis. On average, participants took 34.6 minutes to complete the experiment. We paid them a base payment of $2.70 if they watched the video instructions and completed (either passed or failed) the quiz. If they passed the quiz and completed the full experiment, they received another $1.50 in addition to a bonus of up to $1 based on the score they achieved. We believe the rate of passing the attention check was relatively low due to the complexity of the task.

Before attempting the task, participants watched three instructional videos explaining the dynamics of the SMW, the objective function, and how to manipulate the endogenous variables. The task was presented as managing a farm on an alien planet. Participants were tasked to wisely invest costly "resources" to efficiently achieve a target within ten rounds. The target was defined in terms of a combination of five "farming measures". Causal relationships between inputs and variables were represented as weighted edges. A screenshot of the experiment can be seen in Figure 1.

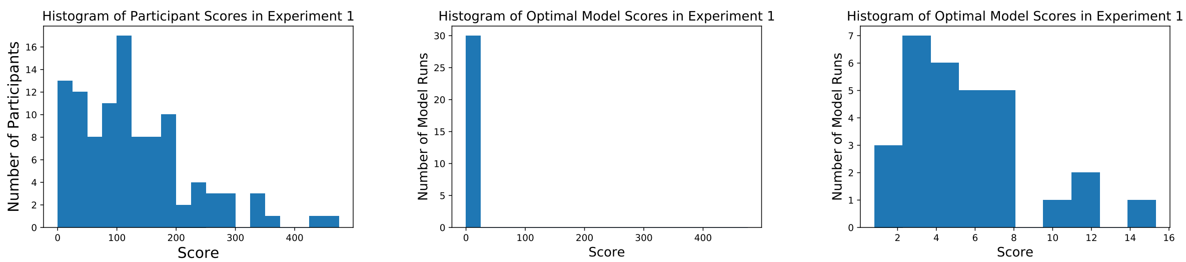
In between videos, participants were given the opportunity to try the functionality for four rounds. Participants then took a quiz consisting of two questions which tested their understanding of the task and completed a training season consisting of 6 rounds. If they achieved a goal distance of at 30 or less during the training round, they could move on to the final task, where they were assigned a situation and asked to pursue a goal over 10 rounds.

Participants were randomly assigned to one of 30 situations. A situation in this experiment is defined by the initial values of the endogenous variables. We selected situations for the experiment by first generating 3000 situations whose endogenous values were each sampled uniformly from the interval $[-250, 250]$. From these, we selected 30 situations using two criteria reflecting the two goals of the experiment. To discover which model describes human goal-directed actions best, we designed an *informative* condition which maximizes the probability of finding strong evidence in favour of one model over the others. To test whether our models of goal pursuit can accurately predict which situations humans do well on and which situations they find hard, we designed an *easy* condition, where the situations where the bounded models performed best relative to the optimal model were selected. The goal of distinguishing between models is ultimately more important, so we chose 20 informative situations and 10 easy situations.

We applied a Bayesian optimal experimental design approach (Chaloner and Verdinelli, 1995) to create 20 situations that maximize the probability that the model generating the data can be correctly inferred

| Model | Expected model prob. | Exceedance prob. |
|---|---|---|
| optimal | 0.01 | 0.00 |
| limited attention | 0.27 | 0.11 |
| limited planning | 0.36 | 0.89 |
| limited attention and planning (discrete) | 0.17 | 0.00 |
| limited attention and planning (continuous) | 0.12 | 0.00 |
| null model 1 | 0.05 | 0.00 |
| null model 2 | 0.02 | 0.00 |

Table 1: Results of mixed-effects Bayesian model selection applied to the data from the experiment. Expected model probabilities denote the expected proportion of participants who are best explained by the model. Exceedance probabilities denote the probability that the proportion of people whose data is best explained by a given model is larger than for any other model.




(a) Scores for the human participants. Each bin contains a score range of 25.

(b) Scores for the optimal model. Each bin contains a score range of 25.

(c) Scores for the optimal model. Each bin contains a score range of 2.

Figure 2: Histograms of the scores achieved by participants in the experiment (left), the optimal model (centre), and the optimal model on a finer scale (right). 6 of the 111 participants were excluded from the human histogram due to their scores being above 500.

from the resulting state-action sequences (*informativeness*). To achieve this we generated simulated noisy data from the LQR, sparse hill-climbing model, and null model 2, then fit each model to the simulated data and computed its posterior probability. We chose the parameters for generating the data by fitting the models to data from a small pilot experiment. We then chose the situations which maximize the sum of the posterior probability of each model given the data it generated. The remaining 10 situations were in the *easy* condition; they were the situations where the sparse hill-climbing model achieved a score closest to the optimal model. Of the included participants, 37 (33%) were assigned to the easy condition and 74 (67%) were in the informative condition.

## 4.2 Results

We find four main results from this experiment. First, human goal pursuit is highly sub-optimal. Second, resource-rational models capture deviations from optimality much better than simpler alternative models. Third, considering limitations on both attention and planning is important to capture human goal pursuit. Finally, people differ in which of these resource constraints best explains how their goal pursuit deviates from optimal goal pursuit.

Human goal pursuit was far from optimal in this experiment. Figure 2 shows a histogram of the scores achieved by participants. They were mostly clustered around 100, with a few participants who did much worse. The median score achieved by humans was 116.0. In contrast, the LQR achieved a median score of 4.7, outperforming the median human participant by a factor of over 24, which is highly significant according to a Kruskal-Wallis H test ($H(111) = 180$, $p < 0.0001$).

The resource-rational models achieved scores much closer to the human median, as shown in Table 2. To compare the fit of each model to the participant data more precisely, we fit the free parameters and noise parameters of all the models to each participant's data individually using Bayesian optimization (Snoek et al., 2012) as implemented by Nogueira (2014) with 500 initial evaluation points and 500 iterations. We then performed mixed-effects Bayesian model selection at the group level (Stephan et al., 2009) to estimate the proportion of people whose behavior is best explained by each model (expected model probability $\hat{w}$) and the probability that it best explains a larger proportion of the population than any other model (exceedance probability $\phi$). To do so, we used Akaike's Information Criterion (AIC) (Akaike, 1998) as an estimate of each model's marginal likelihood at the participant level. Table 1 summarizes the results of this analysis. The LQR was least compatible with our the data by far. It did not fit any participant's data best it and appears to be appropriate for less than 1% of the population ($\hat{w} < .01$). The four resource-rational models, which accounted for limited attention, limited planning, or both, explained the experimental data better than both the optimal model and the baseline models which do not rely on the theory of resource-rationality. According to family-level Bayesian model selection (Penny et al., 2010), more than 97% of all people appear to be best described by one of our resource-rational models, indicating very strong evidence in favour of those models ($\phi > 0.9999$). We also include the parameter estimates that best fit human data in the Supplementary Material.

The resource-rational models made three qualitative predictions, two of which were confirmed by the experimental data. First, as predicted by the models with a limited planning horizon, the magnitude of humans' inputs varied across rounds much more than the LQR's inputs (14.68 vs. 1.10, $H(141) = 70.1$, $p < 0.0001$). The discrete sparse hill-climbing model captures this variability much better than the LQR model (see Figure 3). A notable exception is that in some environments the discrete sparse hill-climbing model sets all inputs to zero in all rounds, leading to a standard deviation of 0. Second, as predicted by the resource-rational models with bounded attention,people manipulated significantly fewer variables at a time than the would have been optimal (2 vs. 4, $H(141) = 84.2$, $p < 0.0001$). As Figure 4 shows, the continuous sparse hill-climbing model captures the number of variables manipulated by people much more accurately than the LQR. Plots of input norm standard deviation and number of variables manipulated for all models

| Model | Median Score | Median # Vars Manipulated | Input Norm Stdev |
|---|---|---|---|
| optimal | 4.7 ( 4.0) | 4 (0) | 1.10 ( 0.88) |
| limited attention | 94.4 ( 96.1) | 1 (2) | 9.33 (13.40) |
| limited planning | 126.7 (104.5) | 4 (0) | 2.63 ( 3.02) |
| limited both (discrete) | 164.6 (167.6) | 0 (1) | 9.28 (10.83) |
| limited both (continuous) | 263.0 (299.6) | 2 (1) | 7.95 (25.45) |
| null model 1 | 459.9 (488.3) | 1 (0) | 0.48 ( 1.10) |
| null model 2 | 514.8 (541.6) | 0 (0) | 0.00 ( 0.00) |

Table 2: Comparison of the scores and numbers of variables manipulated by each model. These values were computed by running each model on each situation 60 times for each participant whose data that model fit best, using the set of parameters that best fit that participant. We then computed medians across all of those runs. Inter-quartile ranges appear in parentheses next to numbers. Scores vary both due to noise and due to the models being run on different situations.

are available in the Supplementary Material. Third, the prediction that participants in the easy condition would do better than those in the hard condition (median scores of 137.9 vs. 169.8 (H(120)=6.28, $p = 0.012$) according to the sparse hill climbing model) was not confirmed in the experiment. The median scores were 117.3 in the easy condition and 116.4 in the informative condition. A Kruskal-Wallis H test did not reveal a significant difference between median scores (H(111)=0.0006, p=0.98).

As Table 1 shows, individual participants varied notably in which model explained their actions best. The expected proportion of people described best by the limited attention model ($\hat{w}$) was 27%, compared with 36% for the limited planning model and 17% and 12% for the discrete and continuous sparse hill-climbing model. Null models 1 and 2 explained an expected 4.6% and 2.3% of participants best, respectively. This indicates that different people might allocate their cognitive resources differently. Participants best described by the hill-climbing model might allocate their resources to representing the problem in full detail instead of planning, while those best described by the limited attention model might represent only a subset of the system, but plan several steps ahead. Participants whose actions were best explained by the limited attention model achieved a better median score (112.0) than participants whose actions were best explained by the hill-climbing model (161.6; H(72)=4.69, p=0.030). This suggests that the most effective allocation of people's limited cognitive resources might be to create a simplified mental model that makes it tractable to plan several steps ahead. Table 2 shows the median score of each model, along with qualitative aspects of models' performance. Overall, models with a limited planning horizon were best for a larger proportion of people than models that do not (67% vs. 33%, $\phi = 0.999$) and models accounting for bounded attention best explained the data from a larger proportion of participants than models that do not (57% vs. 43%, $\phi = 0.90$). This suggests that bounded attention and limits on planning should both be considered in modeling human goal pursuit, even though their relative importance may differ from person to person.

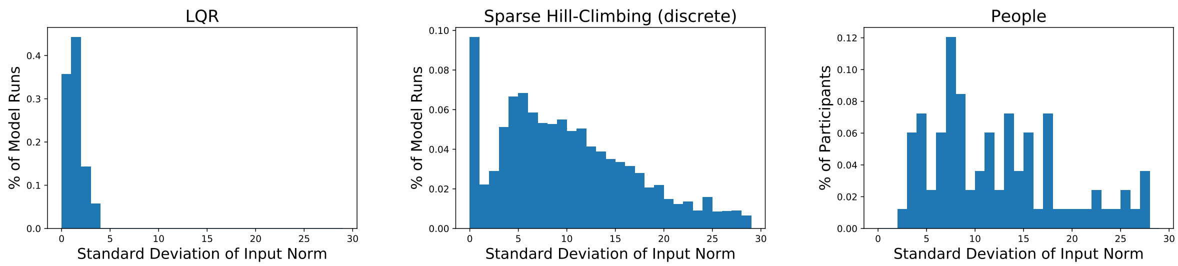



Figure 3: Comparison of the standard deviation of input norms (i.e. how much the magnitude of inputs varied between rounds on a particular task) between the optimal model (LQR), the sparse LQR, and humans.

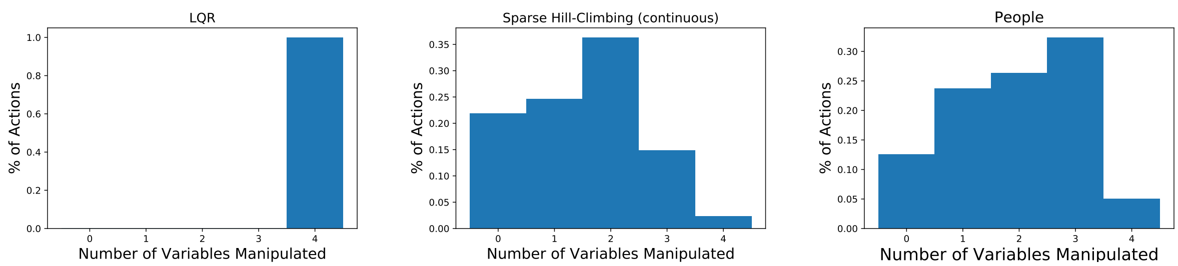



Figure 4: Comparison of the number of variables manipulated (i.e. the number of non-zero input variables) between the optimal model (LQR), the continuous sparse hill-climbing model, and humans.

# 5 Discussion

In this paper, we introduced resource-rational models of human goal pursuit which incorporate planning and attention costs to pursue a goal within a simulated micro-world. We then tested these models in a behavioural experiment. We found that the resource-rational models explained participants' actions better than both the optimal model and our two baseline models. One interesting aspect of our resource-rational models of goal pursuit under limited attentional resources is that they adapt which information they use to the structure of the environment and the goal. The success of our resource-rational models therefore suggests that people did not act optimally, but made rational use of their finite cognitive resources.

The LQR was an extremely poor fit to the human data, which suggests that bounded planning and attention abilities are significant constraints on human goal pursuit. One possible explanation for the high degree of individual variation in which model fit best is that people have a finite pool of cognitive resources which they can allocate between planning and attention. The difference between the scores achieved by participants best described by the limited attention model and those best described by the limited planning model might reflect differences in the relative cognitive costs of paying attention and planning for different participants. Alternatively, some participants might simply be better at allocating limited resources than others.

## 5.1 Relationship to other models and frameworks

Our experimental task was motivated by theories of feedback control in psychology, where goal pursuit is framed as an attempt to bring multiple variables close to a goal state (Carver and Scheier, 2001). We have expanded upon these ideas by developing computational models of how resource constraints can affect macroscopic goal pursuit. While optimal control theory accurately describes how people achieve small-scale goals such as in motor control (Christopoulos and Schrater, 2011), we have shown that the same model does not provide a plausible account of goal pursuit in a larger-scale simulated micro-world task.

Several of our models draw heavily on the work of Gabaix (2016) on sparse dynamic programming as an account of decision making and planning under resource constraints. However, while sparse dynamic programming only accounts for attention constraints, we combine sparse attention with planning constraints in our sparse hill-climbing model. This enables us to understand more closely how different cognitive constraints interact with each other and model some humans' behavior more accurately.

Quantum probability theory has previously been applied to explain deviations from rationality, such as people making decisions which seem to violate the It has traditionally been applied to decision-making under uncertainty and errors in probabilistic judgement (i.e. Busemeyer et al., 2011). The task we gave participants was deterministic, so the differences between classical and quantum probability theory likely do not play a role in our experiment. In future work, quantum probability theory could be applied to model how people represent their uncertainty about the state and the dynamics of the environment. Quantum probability theory is consistent with the general principle of resource-rationality in that it can be seen as an account of how people use limited representational resources to perform complex probabilistic computations.

The Probabilistic Language of Thought framework posits that mental representations rely on both structure and probabilistic uncertainty. One idea from this framework that might be relevant to modeling how people manipulated our task is that of clustering. It has been proposed that people form meaningful categories by identifying clusters of related objects (Kemp et al., 2012). This idea could be incorporated into future models of goal pursuit. Those models might postulate that people's mental representations cluster variables or the relationships between variables into categories based on features like magnitude and direction. Such a model might plan over a mental representation where the weight of each edge has been replaced by the mean edge weight of its cluster's centroid. This fits into the theory of resource rationality as an alternative formulation of how people optimally allocate limited attention, since it reduces the number of unique relationships between variables that have to be considered. The Probabilistic Language of Thought framework also emphasizes that representations are situated in the larger context of the mind's causal models (Krynski and Tenenbaum, 2007). This could be used to model how people's representations of elements of our experimental task were influenced by both the narrow context of their overall causal model of the task and the broader context of being given a problem-solving task for an online experiment.

Our resource-rational modelling framework (Lieder and Griffiths, 2020) is consistent with the adaptive heuristics framework (Payne et al., 1988; Gigerenzer and Todd, 1999; Simon, 1956; Hertwig et al., 2019) which postulates that the mind relies on heuristics that are fast, frugal, and adapted to the structure of the environment. The goal pursuit strategies entailed by our resource-rational models are heuristics because they are much simpler than the optimal solution method of the LQR. They are fast in that they only perform a limited amount of planning. They are frugal in that they attend to only a small subset of all causal relationships in the environment. And they are adaptive in that they are optimized for the structure of the environment. The latter is accomplished by selecting the representation that leads to highest performance in the given environment. While research on fast-and-frugal heuristics typically focuses on one-shot decisions, we use resource-rational analysis to have modelled the strategies people use to solve the sequential decision problems entailed by goal pursuit. Future work should therefore strive to extract and characterize the goal-pursuit heuristics that our resource-rational models entail for different environments as it is currently being done for risky choice (Lieder et al., 2017; Gul et al., 2018; Krueger et al., 2021).

While our model can capture goal pursuit in continuous domains, most goal-based agent models are rely on symbolic search-based methods that only work in discrete environments (Russell and Norvig, 2002). In a recent instance of this approach, Correa et al. (2020) used breadth-first search and A* search as models of goal pursuit in a grid-world environment. These algorithms have been used extensively, but are difficult to scale up to continuous environments. They computed optimal subgoals using a differentiable form of value iteration which can compute subgoals efficiently, but which does not easily extend to continuous state spaces. Callaway et al. (2018) gave participants a task which required them to balance costly exploration of a state space with taking a path through that space and used meta-level Markov decision processes to model uncertainty about the true rewards attained at different states. The authors found that the resource-rational model for this task described participants' behavior the best, which aligns with our finding that the resource-rational models fit humans' actions better than the optimal model and baseline models in our task.

## 5.2 Limitations

While the task used for our experiment was designed to be a more realistic environment to study goal pursuit than many discrete puzzles used in past research, it is still much simpler than the real-world environments in which human goal pursuit takes place. For instance, people often have incomplete information in real-world goal pursuit. In our task, we showed participants the exact values of all relevant variables and the complete dynamics of the system in which they pursued the goal. However, in many real world environments the effects of one's actions are not perfectly predictable, even with unbounded attention and planning abilities. In trying to find a good restaurant to eat dinner at, one must make decisions without perfect knowledge of the quality of the food or how crowded the restaurant is. This additional layer of uncertainty is not represented in our deterministic experiment but it likely plays an important goal in real-world goal pursuit.

While our models of goal pursuit did not set any subgoals in the experimental situations, it is possible that some participants implicitly set their own subgoals in completing this task (Newell and Simon, 1972). People might have intuitively decomposed the task into smaller tasks and solved those one-by-one, while our models treat the entire goal as one task. Furthermore, while our models either looked one step ahead or 10 steps ahead, people may often use an intermediate planning horizon (Keramati et al., 2016).

### 5.3 Directions for future work

Future work should improve upon our models of goal pursuit by adding adaptive planning horizons. It is possible that people are somewhere between these two extremes, either planning only a few steps ahead or decomposing the task into smaller sub-tasks. Drawing upon other models of planning to develop a model that plans a variable number of steps ahead could be a fruitful direction of development. For example, Callaway et al. (2018) developed a resource-rational model of human planning that balances the benefits of learning about possible future states against the costs of tracing the effects of actions far into the future.

Furthermore, developing models that can account for incomplete information would be a crucial step to developing models that can describe goal pursuit in a wider variety of situations. Stochastic optimal control (Todorov, 2005) is an extension of optimal control theory that deals with these additional challenges. Normative models of planning under uncertainty have been developed in organizational an financial decision-making (i.e. Friedman and Segev, 1976) and the method of stochastic programming has been used to find optimal solutions to such problems (Huang and Ahmed, 2010).

Future work should also investigate how the amount of effort that people invest in planning how to pursue their goals depends on their self-efficacy and the predictability and controllability of their environment (Lieder et al., 2013).

## 6 Conclusion

Through our experiment and modelling, we have shown that human goal pursuit is far from optimal. Extending the notion of rational goal pursuit with assumptions about humans' cognitive constraints was necessary to accurately predict people's goal-directed actions, even in a simplified environment with a goal that is much simpler to pursue than real-world goals. Our results suggest that these deviations can, at least partly, be understood as a consequence of people's limited attentional resources and limited planning horizon. Our resource-rational models not only out-performed the optimal model and baseline models in terms of exceedance probability, they also captured qualitative aspects of goal pursuit such as the number of variables participants manipulated and the variability in the magnitude of actions.

These findings illustrate that extending standard notions of rationality into the principle of resource rationality can be immensely beneficial in modelling cognitive functions such as goal pursuit. Taking cognitive constraints into account allowed our resource rational models to predict human goal pursuit significantly better than both the unboundedly rational model and baseline models. This finding has important implications for understanding how people pursue goals. This is an important step forward because goal pursuit is a central organizing principle of all aspects of psychology and human behavior, ranging from motor control to decision-making, problem solving, work, social interaction, and pursuing one's dreams (Carver and Scheier, 2001).

Overall, the success of our resource-rational models of goal pursuit further strengthens the view that incorporating cognitive constraints into rational models of cognition is a promising approach to explaining human behavior.

## References


Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike*, pages 199–213. Springer.

Botvinick, M. and Toussaint, M. (2012). Planning as inference. *Trends in cognitive sciences*, 16(10):485–488.

Botvinick, M. M., Niv, Y., and Barto, A. C. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*, 113(3):262–280.

Brehmer, B. and Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior*, 9(2-3):171–184.

Busemeyer, J. R., Pothos, E. M., Franco, R., and Trueblood, J. S. (2011). A quantum theoretical explanation for probability judgment errors. *Psychological review*, 118(2):193.

Callaway, F., Lieder, F., Das, P., Gul, S., Krueger, P. M., and Griffiths, T. L. (2018). A resource-rational analysis of human planning. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society.*

Carver, C. S. and Scheier, M. F. (2001). *On the self-regulation of behavior*. Cambridge University Press.

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, pages 273–304.

Christopoulos, V. N. and Schrater, P. R. (2011). An optimal feedback control framework for grasping objects with position uncertainty. *Neural computation*, 23(10):2511–2536.

Correa, C. G., Ho, M. K., Callaway, F., and Griffiths, T. L. (2020). Resource-rational task decomposition to minimize planning costs. In *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*.

Custers, R. and Aarts, H. (2010). The unconscious will: How the pursuit of goals operates outside of conscious awareness. *Science*, 329(5987):47–50.

Friedman, Y. and Segev, E. (1976). Horizons for strategic planning. *Long Range Planning*, 9(5):84 – 89.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, 11(2):127–138.

Funke, J. (1993). Microworlds based on linear equation systems: A new approach to complex problem solving and experimental results. In *Advances in psychology*, volume 101, pages 313–330. Elsevier.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, 129(4):1661–1710.

Gabaix, X. (2016). Behavioral macroeconomics via sparse dynamic programming. Technical report, National Bureau of Economic Research.

Gigerenzer, G. and Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. In *Simple heuristics that make us smart*, pages 3–34. Oxford University Press.

Gul, S., Krueger, P. M., Callaway, F., Griffiths, T. L., and Lieder, F. (2018). Discovering rational heuristics for risky choice. In *The 14th biannual conference of the German Society for Cognitive Science, GK*.

Hertwig, R., Pleskac, T. J., and Pachur, T. (2019). *Taming uncertainty*. MIT Press.

Huang, K. and Ahmed, S. (2010). A stochastic programming approach for planning horizons of infinite horizon capacity planning problems. *European Journal of Operational Research*, 200(1):74 – 84.

Juechems, K. and Summerfield, C. (2019). Where does value come from? *Trends in cognitive sciences*.

Kemp, C., Shafto, P., and Tenenbaum, J. B. (2012). An integrated account of generalization across objects and features. *Cognitive Psychology*, 64(1-2):35–73.

Keramati, M., Smittenaar, P., Dolan, R. J., and Dayan, P. (2016). Adaptive integration of habits into depth-limited planning defines a habitual-goal–directed spectrum. *Proceedings of the National Academy of Sciences*, 113(45):12868–12873.

Kirk, D. E. (2004). *Optimal control theory: an introduction*. Courier Corporation.

Kotovsky, K., Hayes, J., and Simon, H. (1985). Why are some problems hard? evidence from tower of hanoi. *Cognitive Psychology*, 17(2):248 – 294.

Krueger, P., Callaway, F., Gul, S., Lieder, F., and Griffiths, T. (2021). Discovering rational heuristics for risky choice. in preparation.

Krynski, T. R. and Tenenbaum, J. B. (2007). The role of causality in judgment under uncertainty. *Journal of Experimental Psychology: General*, 136(3):430.

Lieder, F., Goodman, N. D., and Huys, Q. J. M. (2013). Controllability and resource-rational planning. *Cosyne abstracts*, 2013.

Lieder, F. and Griffiths, T. L. (2020). Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 43:1–85.

Lieder, F., Griffiths, T. L., and Hsu, M. (2018a). Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. *Psychological review*, 125(1):1.

Lieder, F., Griffiths, T. L., Huys, Q. J., and Goodman, N. D. (2018b). The anchoring bias reflects rational use of cognitive resources. *Psychonomic bulletin & review*, 25(1):322–349.

Lieder, F., Krueger, P. M., and Griffiths, T. (2017). An automatic method for discovering rational heuristics for risky choice. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. MIT press.

Mas-Herrero, E., Sescousse, G., Cools, R., and Marco-Pallarés, J. (2019). The contribution of striatal pseudo-reward prediction errors to value-based decision-making. *NeuroImage*, 193:67–74.

Miller, G. A., Galanter, E., and Pribram, K. H. (1960). Plans and the structure of behavior.

Morgenstern, O. and Von Neumann, J. (1953). *Theory of games and economic behavior*. Princeton university press.

Newell, A. and Simon, H. A. (1972). *Human problem solving*, volume 104. Prentice-Hall Englewood Cliffs, NJ.

Nogueira, F. (2014). Bayesian Optimization: Open source constrained global optimization tool for Python.

Payne, J. W., Bettman, J. R., and Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3):534.

Penny, W. D., Stephan, K. E., Daunizeau, J., Rosa, M. J., Friston, K. J., Schofield, T. M., and Leff, A. P. (2010). Comparing families of dynamic causal models. *PLOS Computational Biology*, 6(3):1–14.

Pezzulo, G., Rigoli, F., and Friston, K. (2015). Active inference, homeostatic regulation and adaptive behavioural control. *Progress in neurobiology*, 134:17–35.

Pezzulo, G., Rigoli, F., and Friston, K. J. (2018). Hierarchical active inference: A theory of motivated control. *Trends in cognitive sciences*, 22(4):294–306.

Powers, W and Clark, R K and MacFarland, R L (1960). A general feedback theory of human behavior. *General Systems-Yearbook of the Society for General Systems Research*, pages 63–83.

Ribas-Fernandes, J. J., Shahnazian, D., Holroyd, C. B., and Botvinick, M. M. (2019). Subgoal-and goal-related reward prediction errors in medial prefrontal cortex. *Journal of cognitive neuroscience*, 31(1):8–23.

Ribas-Fernandes, J. J., Solway, A., Diuk, C., McGuire, J. T., Barto, A. G., Niv, Y., and Botvinick, M. M. (2011). A neural signature of hierarchical reinforcement learning. *Neuron*, 71(2):370–379.

Rohe, M. S., Funke, J., Storch, M., and Weber, J. (2016). Can motto-goals outperform learning and performance goals? influence of goal setting on performance and affect in a complex problem solving task. *Journal of Dynamic Decision Making*.

Russell, S. and Norvig, P. (2002). *Artificial intelligence: a modern approach*. Pearson.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological review*, 63(2):129.

Simon, H. A. and Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist*, 26(2):145.

Snoek, J., Larochelle, H., and Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, pages 2951–2959.

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., and Friston, K. J. (2009). Bayesian model selection for group studies. *Neuroimage*, 46(4):1004–1017.

Todorov, E. (2005). Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system. *Neural Computation*, 17(5):1084–1108.

Wiener, N. (1948). *Cybernetics or Control and Communication in the Animal and the Machine*. MIT Press.

# Supplementary material for resource-rational models of human goal pursuit

Ben Prystawski[*,1,2] Florian Mohnert[*,1] Mateo Tošić[1] Falk Lieder[1]

November 2, 2020

1: Max Planck Institute for Intelligent Systems, Tübingen, Germany 2: University of Toronto, Toronto, Canada

* These authors contributed equally.

## 1 Mathematical Descriptions of Models of Goal Pursuit

### 1.1 Optimal goal pursuit: the linear-quadratic regulator

The optimal model for our task, which originates in optimal control theory (Kirk, 2004), finds the sequence of actions which minimizes a quadratic cost function of both the states and inputs in a linear dynamical system. Optimal control theory is a branch of mathematics focused on optimizing objective functions in dynamical systems over time and has previously been applied to modelling human motor control (e.g. Harris and Wolpert, 1998; Christopoulos and Schrater, 2011; Todorov, 2005). It is plausible that it also describes behavior on more macroscopic tasks such as goal pursuit. Since our cost function is the square root of a quadratic and minimizing the square root of a non-negative function is equivalent to minimizing that function, we can use the LQR to minimize our cost function. It uses backward induction to compute this solution, and thus relies on long-term planning along with full attention to the system's dynamics.

Given some discrete linear dynamical system, where $\mathbf{s}_{t+1} = \mathbf{A}\mathbf{s}_t + \mathbf{B}\mathbf{a}_t$, and time horizon $N$, the LQR computes the action sequence $\mathbf{a}$ that minimizes the cost function

$$C(\mathbf{a}) = \mathbf{s}_N^T \mathbf{Q}_f \mathbf{s}_N + \sum_{i=0}^{N-1} (\mathbf{s}_i^T \mathbf{Q} \mathbf{s}_i + \mathbf{a}_i^T \mathbf{R} \mathbf{a}_i), \tag{1}$$

where $\mathbf{s}_N^T$ is the transposed version of the vector $\mathbf{s}_N$ reporting the values of all state variables in the final time step ($N$), $\mathbf{Q}_f$ is the cost matrix for the final endogenous state, $\mathbf{a_i}$ is the action taken at time step $i$, $s_i$ is the endogenous state at time step $i$, $\mathbf{Q}$ is the cost matrix that determines the penalty for endogenous states being away from $\mathbf{0}$, and $R$ is the cost matrix for actions. If we choose $\mathbf{Q}_f = \mathbf{I}$, $\mathbf{T} = c \cdot \mathbf{I}$, and let $\mathbf{Q}$ be the zero matrix, this cost function becomes the squared cost function in the task for humans.

## 1.2 Resource-rational models of goal-pursuit

### 1.2.1 Limited planning: hill-climbing

We use a hill-climbing model to account for limited planning ability. This strategy has been used in the past to model human problem solving (Simon and Newell, 1971). The hill-climbing model plans only one step ahead at each round. At each time step, it moves in the direction of the negative gradient of its distance to the goal state. This model has one free parameter, which is the step size. This determines what multiple of the optimal step size the model takes.

To determine how large of a step the model takes, we can compute the multiple of the gradient $\lambda_{\text{opt}}$ which minimizes the cost (Equation 1) that would be achieved if the task ended in the next round. However, people might not always take the optimal step size. They might take smaller steps than the "optimal" size for the hill-climbing model because they know that they can get closer to the goal in subsequent time steps even though they do not explicitly plan actions for those time steps. Therefore, we introduce one free parameter to this model: the step size $\lambda$. The step size determines which multiple of the optimal step size the agent takes. Formally, the hill-climbing model selects an action $\mathbf{a}_t$ via the following equation:

$$\mathbf{a}_t = -\lambda \cdot \lambda_{\text{opt}} \cdot \nabla_{\mathbf{a}_t} \left( ||f(\mathbf{s}_t, \mathbf{a}_t) - \mathbf{g}||_2 \right), \tag{2}$$

where $\lambda_{\text{opt}}$ is the optimal step size, $\lambda$ is a free parameter, and the gradient is evaluated at $\mathbf{a_t} = \mathbf{0}$.

### 1.2.2 Limited attention: sparse LQR

As mentioned in the main paper, we model attention costs by jointly minimizing the cost resulting from planning in the simplified representation of the environment defined by attention vector $\mathbf{m}$ and the cost of paying attention the elements of the environment represented by $\mathbf{m}$. Here, we describe the details of how this is accomplished in our goal pursuit environment.

First, we define an *edge* in the SMW as a causal relationship between two variables. Each green or red arrow in Figure 1 in the main text is an edge. These correspond to the non-zero elements of either transition matrix $\mathbf{A}$ or $\mathbf{B}$, except for the diagonal of ones in $\mathbf{A}$. Each edge determines how one variable influences one other variable from one round to the next. An agent with limited attention might only focus on a subset of the edges, based on how important those edges are in deciding how much of each resource to use. Attention could be discrete, where each edge is either fully attended to or fully ignored, or continuous, where edges can be attended to partially. We can represent the subset of edges which a sparse-max model attends to with the attention vector $\mathbf{m}$. The discrete-attention version uses an attention vector $\mathbf{m} \in \{0,1\}^{|E|}$, where $|E|$ is the set of edges of the SMW, while the continuous-attention version uses attention vectors $\mathbf{m} \in [0,1]^{|E|}$. We only present the discrete-attention version of the sparse LQR. The continuous-attention version is much more complicated as it would require computing derivatives of entire action sequences which the LQR plans via backward induction.

This model selects the optimal attention vector $\mathbf{m}$ which minimizes the cost it achieves by planning in the reduced version of the microworld plus a cost of attention which scales with the number of edges attended to. The fewer edges the model pays attention to, the worse its ability to plan and minimize the cost function in Equation 1. In minimizing the weighted sum of its cost function and cognitive cost of attention, the model balances the cognitive cost of paying attention with the improvement resulting from additional attention. Formally, this means the model selects the attention vector $\mathbf{m}$ that minimizes the following function:

$$\mathbf{m} = \underset{\mathbf{m}\in\{0,1\}^{|E|}}{\operatorname{argmin}} \sqrt{||\mathbf{s}_t - \mathbf{g}||^2 + c\sum_{i=0}^{t-1}||\mathbf{a}_i||^2} + k \cdot \sum_{i=1}^{|E|}\mathbf{m}_i \tag{3}$$

Where $k$ is the only free parameter of the model, representing the cost of attention. When $k = 0$, the sparse LQR behaves equivalently to the standard LQR. As k increases, the model attends to fewer edges.

#### 1.2.3 Limited planning and limited attention: sparse hill-climbing

The sparse hill-climbing model combines the attention constraints of the sparse LQR with the planning constraints of the hill-climbing model. Furthermore, it can have either discrete or continuous attention.

The discrete version of the model chooses the optimal attention vector $\mathbf{m} \in \{0,1\}^{|E|}$ which minimizes the sum of its distance to the goal in the next round after taking a step in the attention-reduced SMW and the cost of paying attention to the edges used to determine that step. It then uses the hill-climbing action in the simplified representation to take a step in the real SMW.

The optimal discrete attention vector is defined as follows:

$$\mathbf{m} = \underset{\mathbf{m}\in\{0,1\}^{|E|}}{\operatorname{argmin}} ||\mathbf{f}(\mathbf{s}_t, \mathbf{a}_\mathbf{m}) - \mathbf{g}||_2 + k \cdot \sum_{i=1}^{|E|}\mathbf{m}_i, \tag{4}$$

where $\mathbf{a}_\mathbf{m}$ is the hill-climbing action taken when only the edges that the model attends to are taken into account.

In the continuous version, the choice of $\mathbf{m}$ is performed as follows,

$$\mathbf{m} = \underset{\mathbf{m}\in[0,1]^{|E|}}{\operatorname{argmin}} \;\; \frac{1}{2}\sum_{e_i \in E}(1-\mathbf{m}_i)^2\Lambda_i + k \cdot \sum_{i=1}^{|E|}\mathbf{m}_i \tag{5}$$

Where $\mathbf{m}_i$ is the attention paid to edge $e_i$, $c$ is the cost of attention, and $\Lambda_i$ is the cost of inattention for edge $e_i$. The cost of inattention is a proxy for the loss of goal pursuit effectiveness resulting from imperfect attention based on the first term of the Taylor expansion. It is defined as $-|e_i|\frac{\partial \mathbf{a}_t}{\partial m_i}\frac{\partial^2 C}{\partial \mathbf{a}^2}\frac{\partial \mathbf{a}_t}{\partial m_i}$, with $u$ denoting the utility function that the model of human goal pursuit is optimizing and $a_t$ denoting the action taken by the model. (Gabaix, 2014).

An attention vector $\mathbf{m}$ defines a new reduced SMW where the weight of each edge $w_{e_i}$ is replaced with $w_{e_i}\mathbf{m}_i$. This reduces the perceived weight of edges that the model pays limited attention to.

After choosing attention vector $\mathbf{m}$, both the discrete and continuous sparse hill-climbing models choose the hill-climbing action defined in Equation 2 in the simplified representation of the SMW defined by $\mathbf{m}$.

Both sparse hill-climbing models have two free parameters: the step size $\lambda$ and the attention cost $k$.

### 1.3 Observation Model

To model how the idealized cognitive processes in each model give rise to concrete goal-directed actions, we model two independent types of noise. *Length noise* describes people taking either larger or smaller steps than the model predicts, while *angular noise* describes people moving in different directions than the model predicts. To measure this noise, we first convert each $D_s$-dimensional endogenous state at time $t$ to spherical coordinates, where we have one radial coordinate $\phi_{t,1}$ and $D_s - 1$ angular coordinates $\phi_{t,2}, \ldots \phi_{t,D_s}$. We then model the error in the radial component as following an exponential distribution with parameter $\gamma$. This is length noise. The errors in the angular components can be modelled as following a von Mises distribution, which is essentially a Normal distribution over a circle. This von Mises distribution is centered at 0 and has concentration parameter $\kappa$, which is analogous to $\frac{1}{\sigma^2}$ in a Normal distribution. We refer to this type of noise as *angular noise*.

Therefore, the likelihood of a participant's data $\mathcal{D}$ under model $\mathcal{M}$ over $T$ time steps with noise parameters $\gamma, \kappa$ is

$$p(\mathcal{D}|\mathcal{M}, \gamma, \kappa) = \prod_{t=1}^{T} p_{ex}(\Delta_{t,1}|\gamma) \cdot \Big[\prod_{j=2}^{|\phi_t|} p_{vm}(\Delta_{t,j}|0, \kappa)\Big] \tag{6}$$

Here, $T$ denotes the number of time-steps in the participant data, $p_{ex}(\cdot|\gamma)$ denotes the likelihood under an exponential distribution with parameter $\gamma$ and $p_{vm}(\cdot|0, \kappa)$ denotes the likelihood under a Von Mises

| Model | Expected Model Prob. | Exceedance Probability |
|---|---|---|
| optimal | 0.0088 | 0 |
| resource-rational | 0.93 | > 0.9999 |
| null models | 0.0580 | 0 |

Table 1: Family-level Bayesian model selection results, where the optimal model, all resource-rational models, and the null models are each families.

| Model | Expected Model Prob. | Exceedance Prob. |
|---|---|---|
| no attention constraints | 0.43 | 0.10 |
| attention constraints | 0.57 | 0.90 |

Table 2: Family-level Bayesian model selection results where the families are models with and without limited attention.

distribution centred at 0 with concentration parameter $\kappa$.

## 2 Family-level Bayesian model comparison

In addition to the Bayesian model selection included in the main paper, we perform family-level Bayesian model selection for each constraint individually in order to assess the importance of each type of constraint (Penny et al., 2010). we group together models into families based on whether or not they incorporate each type of constraint. This means we compare models with no attention constraints (optimal, hill-climbing, and null models) against those with attention constraints (sparse LQR, sparse hill-climbing) and models with no planning constraints (optimal, sparse LQR, null model 1, null model 2). Results of these comparisons are reported in Table 2 and Table 3. In both cases, the models that incorporate constraints have a higher family-level model probability (i.e. the expected probability that any given participant is best explained by that model). The models with attention constraints have an expected model probability of 0.57 while the models with planning constraints have an expected proability of 0.67. The exceedance probability, which is the probability that that family describes participants best overall, is 0.90 for the version with attention constraints and 0.999 for the version with planning constraints.

We also report family-level model selection results where we simply group the discrete and continuous-attention sparse hill-climbing models into a family and leave the other models in families of their own in Table 4. Finally, we report the group-level AIC and number of participants best described by each model in

| Model | Expected Model Prob. | Exceedance Prob. |
|---|---|---|
| no planning constraints | 0.33 | 0.001 |
| planning constraints | 0.67 | 0.999 |

Table 3: Family-level Bayesian model selection results where the families are models with and without limited planning.

| Model | Expected Model Prob. | Exceedance Prob. |
|---|---|---|
| optimal | 0.0085 | 0 |
| limited attention | 0.2822 | 0.11 |
| limited planning | 0.3784 | 0.75 |
| limited attention and planning | 0.28 | 0.14 |
| null model 1 | 0.046 | 0 |
| null model 2 | 0.025 | 0 |

Table 4: Family-level Bayesian model selection results, where the discrete and continuous versions of the limited attention and planning models are collapsed into a single family. All other families consist of just one model.

| Model | # participants best fit | AIC | Expected model prob. | Exceedance prob. |
|---|---|---|---|---|
| optimal | 0 | 27,600 | 0.0085 | 0 |
| limited attention | 35 | 18,763 | 0.27 | 0.11 |
| limited planning | 33 | 18,042 | 0.36 | 0.89 |
| limited attention and planning (discrete) | 19 | 18,434 | 0.17 | 0.0008 |
| limited attention and planning (continuous) | 14 | 18,595 | 0.12 | 0 |
| null model 1 | 5 | 21,101 | 0.046 | 0 |
| null model 2 | 5 | 18,966 | 0.023 | 0 |

Table 5: Results of mixed-effects Bayesian model selection applied to the data from the experiment. Expected model probabilities denote the expected proportion of participants who are best explained by the model. Exceedance probabilities denote the probability that the proportion of people whose data is best explained by a given model is larger than for any other model. We also include the number of participants best fit and group-level AIC, which is the sum of the AICs achieved by each participant.

| | Best-fitting parameters | | | |
|---|---|---|---|---|
| Model | Length Noise | Angle Noise | Attention Cost | Step Size |
| limited attention | 0.050 | 4.01 | 116.5 | |
| limited planning | 0.065 | 6.10 | | 0.468 |
| limited both, discrete | 0.105 | 3.67 | 17.4 | 0.568 |
| limited both, continuous | 0.031 | 5.22 | 11.0 | 0.295 |

Table 6: Average parameter estimates for each resource-rational model, taken across participants best explained by that model. Higher noise parameters indicate less noisy actions.

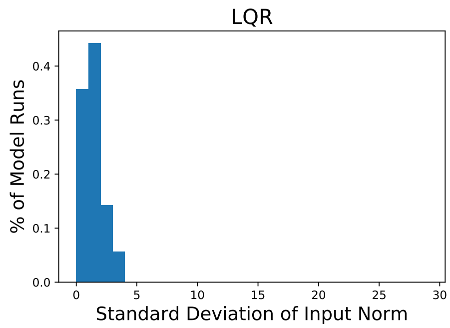
| | Best-fitting parameters | | | |
|---|---|---|---|---|
| Model | Length Noise | Angle Noise | $n$ | $b$ |
| optimal | 0.0093 | 1.46 | | |
| null model 1 | 0.077 | 5.89 | 1 | 5.06 |
| null model 2 | 0.039 | 7.21 | | |

Table 7: Average parameter estimates for each non-resource-rational model, taken across participants best explained by that model. For the optimal model, the best-fitting parameters were averaged across all participants because nobody was best explained by the optimal model. Higher noise parameters indicate less noisy actions.

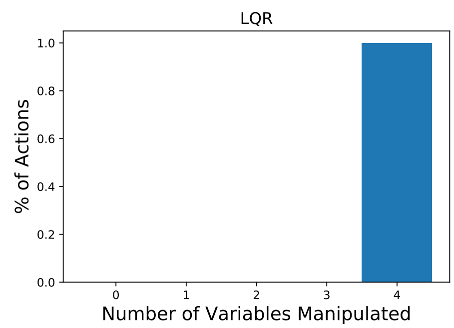
Table 5.

## 3 Interpretation of Model Parameters

The average parameter values which optimize the model fit to each participant's actions are shown in Table 6 (resource-rational models) and Table 7 (other models). Higher angle and length noise parameters correspond to less noise. This means that the discrete limited attention and limited planning model seems to capture the sizes of steps people took best, with the highest average length noise parameter of 0.105. The limited planning model captured the direction of people's movement well with the highest angular noise parameter among the resource rational models (6.10). Null model 2 had the highest angular noise parameter of all (7.21), indicating that it was good at capturing the the direction of movement for people who did not manipulate the exogenous variables much. Step sizes tended to be close to half of the optimal step size, with 0.468 for the limited planning model, and 0.568 for the discrete limited attention and limited planning model. The step size was somewhat lower for the continuous limited attention and limited planning model than for the other two models with step sizes. Finally, note that the attention cost was an order of magnitude higher for the limited attention models compared to the limited attention and limited planning models (116.5 vs. 17.4 and 11.0). This is because the limited attention model plans all 10 actions in one simplified representation of the environment, meaning the attention cost has to balance out the benefits to planning all 10 actions with additional information. In contrast, the limited attention and limited planning models create new simplified representations on each time step, so the attention cost only needs to balance out the benefit for one round.

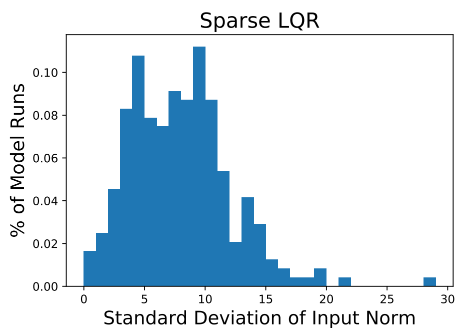



(a) Standard deviation of Input Norms

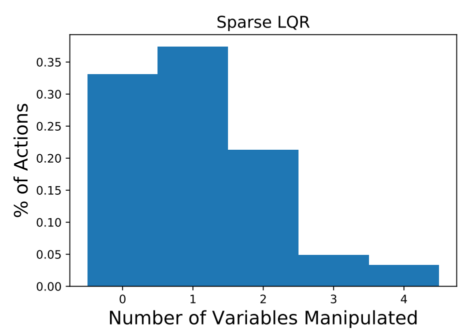



(b) Number of variables manipulated

Figure 1: Standard deviation of input norms and number of variables manipulated for the optimal model (LQR).




(a) Standard deviation of Input Norms




(b) Number of variables manipulated

Figure 2: Standard deviation of input norms and number of variables manipulated for the limited attention model (sparse LQR).

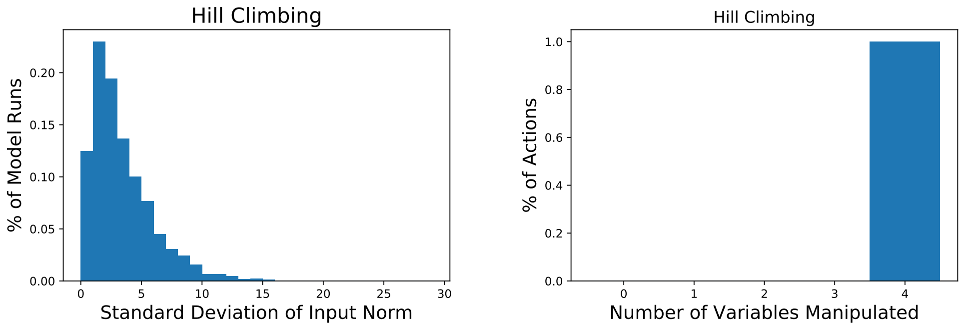
# 4 Qualitative aspects of goal pursuit models

Figures 1-8 show histograms of two qualitative aspects of goal pursuit: the standard deviation of the norms of the input vectors and the number of variables manipulated. To generate this data, each model was run ten times on each situation for every participant best fit by that model. The parameters that best fit that participant's data were used in those runs of the model, with the exception of the noise parameters. we used a length noise of parameter of 0.1 and an angular noise parameter of 40 for all models and participants in order to capture the model's behaviour more accurately.

# References

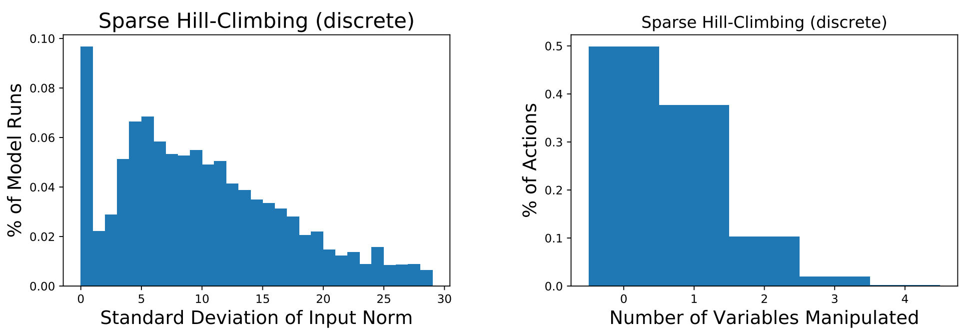
Christopoulos, V. N. and Schrater, P. R. (2011). An optimal feedback control framework for grasping objects with position uncertainty. *Neural computation*, 23(10):2511–2536.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*,



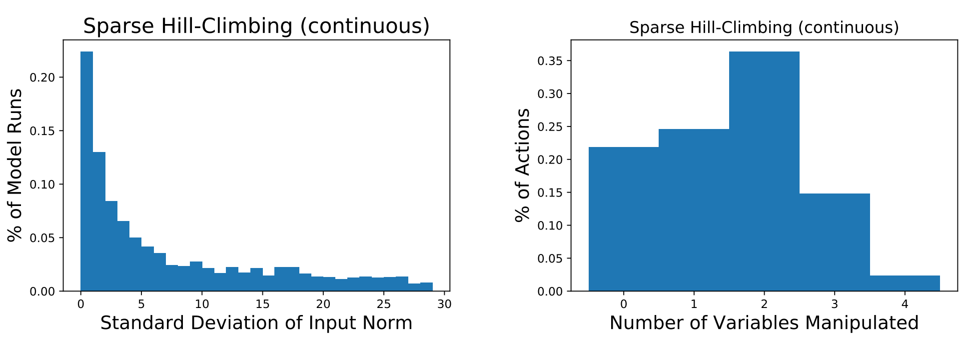

(a) Standard deviation of Input Norms (b) Number of variables manipulated

Figure 3: Standard deviation of input norms and number of variables manipulated for the limited planning model (hill-climbing).



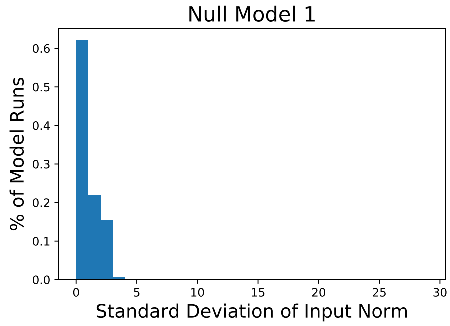

(a) Standard deviation of Input Norms (b) Number of variables manipulated

Figure 4: Standard deviation of input norms and number of variables manipulated for the discrete limited attention and limited planning model (discrete sparse hill-climbing).

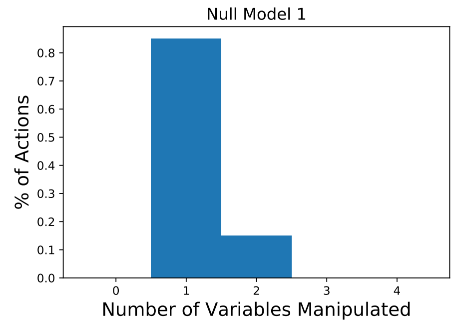



(a) Standard deviation of Input Norms (b) Number of variables manipulated

Figure 5: Standard deviation of input norms and number of variables manipulated for the continuous limited attention and limited planning model (continuous sparse hill-climbing).

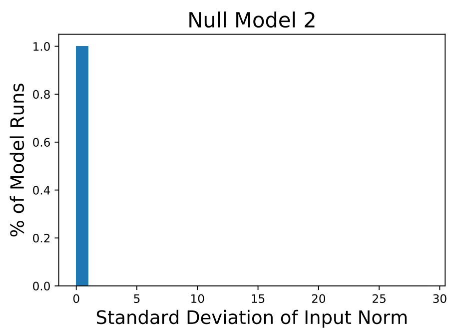



(a) Standard deviation of Input Norms

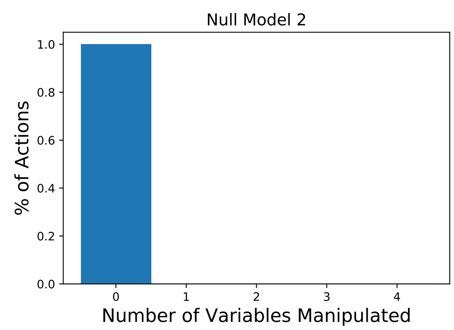



(b) Number of variables manipulated

Figure 6: Standard deviation of input norms and number of variables manipulated for null model 1.

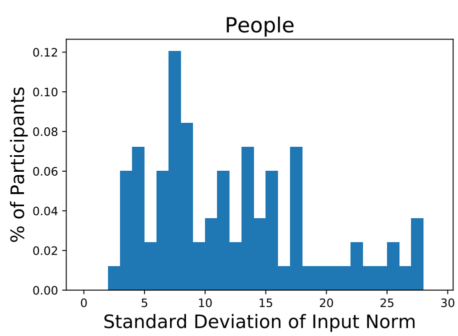



(a) Standard deviation of Input Norms

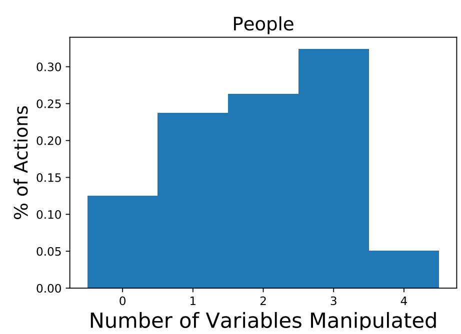



(b) Number of variables manipulated

Figure 7: Standard deviation of input norms and number of variables manipulated for null model 2.




(a) Standard deviation of Input Norms




(b) Number of variables manipulated

Figure 8: Standard deviation of input norms and number of variables manipulated for humans.

129(4):1661–1710.

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

Kirk, D. E. (2004). *Optimal control theory: an introduction.* Courier Corporation.

Penny, W. D., Stephan, K. E., Daunizeau, J., Rosa, M. J., Friston, K. J., Schofield, T. M., and Leff, A. P. (2010). Comparing families of dynamic causal models. *PLOS Computational Biology*, 6(3):1–14.

Simon, H. A. and Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist*, 26(2):145.

Todorov, E. (2005). Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system. *Neural Computation*, 17(5):1084–1108.