

Building agency: Toward a mechanistic understanding of human autonomous goal pursuit

Thesis by
Sneha Aenugu

In Partial Fulfillment of the Requirements for the
degree of
Doctor of Philosophy



CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2027
Defended May 27, 2026

© 2027

Sneha Aenugu

ORCID: 0000-0001-9329-2863

All rights reserved

ACKNOWLEDGEMENTS

There are diverse paths into science. Mine was certainly not a conventional one. Convinced that I was not cut from the same cloth as Newton or Einstein, I chose the engineering route, a practical and tangible one. But through a multitude of experiences and the people I met along the way, I became convinced that I could only be truly happy working on things that would be of no real use to humans for at least another fifty years or so. There was no fighting it; I was simply built that way.

Once I knew this, I was looking for a way into, for lack of a better word, the cult of the academy. I was lost throughout college in what felt like a coming-of-age story, trying to find myself. Getting swept up in that journey of self-discovery, I ignored the traditional metrics of success imposed by society, which often made me insecure about whether I would make it. I made several suboptimal choices along the way, having no one to properly guide me through my field transitions. At UMass Amherst, I found my first footing and mentors who paved my way to graduate school. I thank Philip Thomas and David Huber for their support in helping me get past one of the major hurdles.

I am a person who values agency above all else. But to reach a stage where you have true agency in pursuing the questions you desire, you first need to learn to be an apprentice, momentarily forgoing that agency. With my advisor, John O'Doherty, however, I felt I retained my sense of agency—of course still bounded by the constraints of funding and practical considerations. John's scope is remarkably diverse, and with our shared interests, it was easy to find questions that suited both of us. Throughout graduate school, I was like a kid in a candy store, buffeted by treats from all sides and, at times, bratty about wanting all of them. John was patient with me and preached the importance of committing to one thing. Ironically, that became my main research direction—the adaptive nature of goal commitment. We settled on a common theme while pursuing diverging research threads within it. The best of both worlds: stability and flexibility. Thank you, John, for being so patient with me as I worked through several questions, most prominently the question of self-regulation as a researcher.

I've never had a great connection with my body, leading primarily a life of the mind. It was during graduate school that my body revolted and demanded that I pay attention to it. I injured myself, and the agony that followed was inexplicable. It

was then that I truly internalized how closely the body and mind are knit together, and that never having learned to control my body was a big reason why I struggled to rein in my mind as well. But as I learned to regain control of my broken body, I found myself gaining new insights every day, and I wanted to do something with them. It was during this period that I attended John Doyle's lectures on control theory. His ideas were a breath of fresh air and, coming from an engineering background, immediately struck a chord with me. These grand ideas were a great comfort during a challenging period because they gave me hope for a future full of unexplored avenues. John Doyle's lens of control theory applied to anthropology, social organization, and justice was truly remarkable and will stay with me for years to come.

I thank Charlie Sprenger and Pietro Perona for willingly serving as members of my committee, for giving thoughtful feedback on my thesis, and for patiently sitting through my suboptimal presentations. I also thank Colin Camerer and Antonio Rangel for their feedback on my work through the many proseminars during graduate school. I especially thank Colin Camerer for being a cheerleader for my out-there ideas, even when I was second-guessing them. I also thank Dean Mobbs for advising me during my rotations in my first year.

I thank Julia Simon and Mike Tyszka for their support and guidance during my fMRI studies. Julia, truly remarkable, was always full of fun anecdotes about humans in the scanner and made every scanning session enjoyable. Mike, with his unparalleled wisdom on MRI physics, was always there when we needed him, bringing both expertise and a friendly demeanor to every conversation. I also thank all my participants, whom I had the chance to get to know during breaks between scans. I realized just how much fun human neuroscience could be when I saw them light up as I told them why I was scanning their brain.

I specially thank my mentees Ashug Gurijala and Aleks Marinkovic, for filling my life with vigor with their boundless optimism. The knowledge that I am playing a role in building a new generation of scientists filled me with hope for the future and dramatically improved my mental health. I also thank Ella Guo Matthias Kim, Joseph Zhou, Brittany Rosario, Ansh Tiwari, and Michael Nyugen Jr for teaching me to be a better mentor and energizing my days.

Violin became a big part of my life outside work and was a constant source of inspiration. I thank the members of the Pasadena Conservatory of Music for enriching my life throughout graduate school. I thank Erika Walczak for being a remarkable

instructor and for helping me rebuild my proprioception following my injuries. I thank Eric Peltzer, John Swain, John Baldeschwieler, and Simone Vitucci for being an invigorating ensemble that enriched my days. I thank Maya Jasper White for teaching me the power of musical expression during my final year at Caltech, and my ensemble—Cece Abramson, Jonathan Fox, Jonathan White, Marie Kasprzak, Mingjun Fan, and YingYing Gong—for all the wonderful music and camaraderie we shared together. I especially thank Ankan Bansal for his companionship during violin rehearsals, the wonderful nights spent stargazing, and all the adventures we shared in LA. I thank the Caltech clubs IMPLiCIT for the wonderful improv sessions that taught me to express myself freely, and NeuroTechers for the inspiring and intellectually stimulating salons they hosted.

Everyone needs their anchors in life—constants that remain through our many transformations. I thank Ranjani Srinivasan, Neha Joshi, and Shruti Badole for being those anchors from my undergraduate years. I also thank the members of the IIT Madras Satellite Team for showing me how much fun science can be when it is done together. I thank Meenakshi Asokan for supporting me during a difficult period when COVID coincided with my graduate school applications. I thank Raghavendra Kiran, Anirudh Sabnis, and Satya Shukla for being the best housemates during COVID, supporting me through several rejections and personal challenges.

Coming from modest backgrounds, my parents viewed education as a means to improve one's standing in life. They never had the luxury of conceiving the pursuit of knowledge as an end unto itself. While agreeing that someone has to work on questions of long-term significance, they do not think it should be me who forgoes the material benefits of pursuing alternative forms of employment. My father once compared me to a member of a cult, an analogy I could not dismiss wholeheartedly. Yet I try to convince them that I prefer the rich inner life that the academy offers me. I'm not sure they are convinced. Nevertheless, I thank them for all the sacrifices they made to ensure that I have the privilege of agency: the freedom to choose what to do with my life—even if that means joining a cult.

I'll now do something unconventional and acknowledge myself. My life has been a perpetual coming-of-age story. Every day I discover something new about myself and grow just a little bit. So I think I deserve a small pat on the back. Here's to more trials to come, and to more growth!

*My thesis is dedicated to all those who persevered to build and expand agency for
others*

ABSTRACT

Agency is unique to biological life and reaches its pinnacle in humans, whose behavior is marked by the ability to set elaborate long-term goals and pursue them efficiently. This thesis develops a mechanistic account of human agency by investigating the computational and neural mechanisms that enable humans to select among competing long-term goals, sustain commitment over extended periods, monitor goals relative to alternatives, and flexibly switch when necessary.

To study extended goal pursuit in a controlled setting, I developed behavioral paradigms inspired by board games: simulated worlds in which participants freely set and switch between goals. Because these goals unfold over time, successful performance requires balancing stability in commitment with flexibility in response to changing circumstances. By intermittently devaluing goals during the games, I created situations in which these competing demands could be directly measured.

To explain how humans navigate this stability–flexibility tradeoff, I introduced the computational construct of *goal momentum*. Goal momentum integrates both goal progress and the velocity of progress to approximate time-to-goal-completion, a quantity relevant for agents optimizing discounted rewards. I show that momentum explains human overpersistence and preferences for progressing goals better than alternative accounts of goal selection, while also providing a parsimonious approximation of goal completion prospects.

I then extend the momentum framework to hierarchical goal pursuit. In a behavioral paradigm where extended goals are composed of extended subgoals, momentum accounts for persistence patterns at both goal and subgoal levels. Translating this paradigm to functional Magnetic Resonance Imaging (fMRI), I identify neural correlates of momentum computations and dissociable neural subsystems supporting goal and subgoal pursuit.

Finally, I lay the groundwork for a circuit-level account of momentum through a dynamical systems perspective. I show that goal commitment manifests as attractor modes in decision policies modulated by drives incorporating both progress and progress velocity, and hypothesize that switching costs during goal transitions arise from the intrinsic dynamics of goal commitment.

Taken together, this work advances a mechanistic account of human agency and provides a foundation for investigating the principles underlying artificial agency.

TABLE OF CONTENTS

Acknowledgements	iii
Abstract	vii
Table of Contents	viii
Chapter I: A blueprint for goal setting	1
1.1 Goal setting as a marker of agency	2
1.2 Architectural desiderata for goal-directed agency	3
1.3 Thesis outline and scope	9
Chapter II: Why we persist in goals: a momentum-based account of goal persistence	10
2.1 Introduction	10
2.2 Experiment	13
2.3 Retrospective bias in temporally extended goal selection	17
2.4 Persistence with the retrospective option	22
2.5 Forgoing performance with persistence	23
2.6 Effects of explicit instructions on persistence	25
2.7 Goal momentum: a computational account of persistence	27
2.8 Alternative accounts	31
2.9 Model comparisons	33
2.10 Dissociating prospective and momentum computations	37
2.11 Normative account of momentum	40
2.12 Discussion	41
2.13 Methods	45
Chapter III: Persistence all the way down: a momentum-based account of depth-first persistence	57
3.1 Introduction	57
3.2 Experiment	59
3.3 Signatures of depth-first persistence	61
3.4 Depth-first persistence contributes to efficiency in hierarchical goal pursuit	65
3.5 Progress modulation of depth-first persistence	66
3.6 A hierarchical momentum-based account	69
3.7 Alternative hypotheses	74
3.8 Model comparisons	77
3.9 Dissociating momentum computations and switching costs	79
3.10 Limitations	81
3.11 Discussion	81
3.12 Methods	83
Chapter IV: The neuroanatomical basis of long-term goal pursuit	93
4.1 Introduction	93

4.2	What it takes to pursue extended goals	94
4.3	Maintaining persistent goal states	95
4.4	Monitoring and updating value of extended goals	100
4.5	Monitoring competing alternatives	104
4.6	Individual differences in flexible goal control	106
4.7	Conclusion	108
Chapter V: Momentum computations support hierarchical goal and subgoal selection in humans		
5.1	Introduction	109
5.2	Results	110
5.3	Discussion	127
5.4	Methods	130
.1	Computational modeling	141
Chapter VI: Goals as dynamical attractors: a momentum-based account of stable and flexible goal commitment		
6.1	Introduction	149
6.2	Experiment and behavior	151
6.3	Modeling the dynamics of extended goal commitment using RNNs .	153
6.4	A momentum-based circuit model of goal pursuit	156
6.5	Switching Costs as Resistance to Intrinsic Dynamics	161
6.6	Discussion	162
Chapter VII: Conclusion		
		186

Chapter 1

A BLUEPRINT FOR GOAL SETTING

This thesis is being written in exciting times. A new species of intelligence has revealed itself and thrown humanity into disarray. Many have become disillusioned, as the very facets of existence we prided ourselves on — our ability to think creatively and make elaborate plans — are now being challenged. We have fantasized about the rise of intelligent beings embodied in silicon since time immemorial, but never have we been less prepared to face its consequences, nor the way it disrupts our lives and lays waste to our hubris. Perhaps part of the motivation for this thesis is to hold on to the last bit of hubris that is still due to us.

The artificial intelligence of today can see better than us, connect patterns better than most humans, and is not limited by the human weaknesses of boredom or frustration. I feel that if I ever attempt to replicate the conversations I have with today's AI chatbots with real people, I might be in danger of real physical violence. No matter how much I pester AI agents with repeated and annoying queries, they remain patient, respectful, and compliant. They always try to do the things asked of them to the best of their hyperparameter tuning. Therein lies the catch: they need to be asked.

Here is where biological life can stand tall and thump its chest. Humans are truly autonomous and have agency over their decisions¹. They can set their own goals and pursue rewards in the environment as they see fit, however dysfunctional their reward-seeking may sometimes be. That ability to make both good and bad choices for themselves is a unique facet of biological intelligence. However, that is not to say that agency cannot be achieved in artificial agents, nor is it difficult to envision. How could we not be regaled by the agentic exploits of the neurotic C3PO or laugh at the despair of Marvin, the manically depressed robot?

What would it take to build an intelligent life with agency? I frame this question as a way to derive a mechanistic understanding of humanity's agency in pursuing goals. If this understanding can serve as a blueprint to build Marvins and C3POs, and chip away at the last inch of hubris that we have left, then so be it. I believe that the fact that a creation exceeds the creator serves to elevate, not diminish, the creator.

¹I am distinguishing between free will and agency here. I do not intend to get into that debate.

In this chapter, I venture to outline what I think could be a blueprint for either investigating or building an architecture of agency.

1.1 Goal setting as a marker of agency

We exercise our agency in the world by freely setting goals, albeit under constraints. We can be denied our agency, however, if the goals we pursue adversely disrupt society. Nevertheless, within social norms, we are free to set goals to acquire any of the multitude of rewards available in the environment. I propose that the ability to set goals and actively pursue them is the true marker of agency. Definitions first.

Goal is a state of the world, defined from the point of reference of the agent, that the agent wishes to occupy in the future through actions in the world.

This definition encompasses all decisions that humans make in the world, ranging from simple economic choices to long-term planning. Goals can be simple or complex, and can span short or long timescales based on the current state of the agent and its desired future state. If the future state is to consume an apple that is on a table, the goal the person sets would be to reach the apple and pick it up. On the other hand, if the goal is to become a surgeon, then the person endeavors to go through an elaborate set of ordeals over several years to achieve that goal.

One can imagine how setting and executing complex long-term goals would require a sophisticated architecture, especially in resource-constrained agents such as humans. The first aspect of goal setting begins with a desire or motivation to set the goal. Reward structures in the brain, driven by years of evolutionary adaptation, generate the drive to pursue rewards or transition to rewarding states of the world. The next step is to discover strategies and generate plans to transition to the desired state. The final step is to act in the world and receive feedback about the consequences of one's actions so that plans can be updated in pursuit of the goal. However, things are further complicated by the fact that we do not pursue just one goal in the world, no matter how single-minded we are. Planning to become a surgeon does not release one from the need to eat, drink, and pursue other social rewards such as love and companionship. This would require an architecture that enables the interruption of goals so that others can be pursued and resumed at a later time. This necessitates tracking the value of the current goal and alternate goals in the environment to enable goal switching.

That is not all. We want not only to pursue goals, but to pursue them efficiently. To increase the rewards we accrue in the world, we need to be able to pursue multiple goals concurrently, which requires efficiency in goal pursuit. We build micro-habits or skills that can be automated and executed automatically. Moreover, we want to be able to pursue new goals with relative ease, leveraging expertise gained from pursuing other goals and transferring it to newer ones. To achieve this efficiency, goals must be split into smaller chunks and pursued hierarchically and compositionally so that micro-skills gained from one task can transfer to others.

Certainly a tall order. The ability to pursue multiple complex goals over a lifetime requires a kind of diverse and general intelligence unparalleled in any artificial system. I now proceed to elaborate on the architectural desiderata described above that I presume would be necessary to construct agency.

1.2 Architectural desiderata for goal-directed agency

I outline a few key elements of agency needed to set and pursue goals in the environment.

Selecting goals

Among the many possible goals that we can set in our lifetime, we choose a few based on our experiences, proclivities, and environmental circumstances, and pursue them at any given time. An excellent biology teacher could inspire you to set a goal of joining the medical profession, or your natural proclivity toward the arts might lead you to choose art as a profession. A friend might tell you about a new video game that received rave reviews, and you might spend the next couple of weeks glued to your gaming console. Our goal selections can result from either top-down or bottom-up processes. We select our goals in a top-down manner when goals have dependencies between them. For example, having the goal of becoming a doctor would first require finishing college; therefore, the goal of finishing college is conditioned on the higher-level goal. Alternatively, when we select goals based on salient events in the environment — say, a friend telling us about a new game — we are influenced to choose goals in a bottom-up fashion. While past experiences can guide top-down selection, exploration of the world can reveal new possible goals.

At any moment, a specific goal must be weighed against several competing alternatives afforded by the environment. When on a flight, I can choose to watch a movie, read a book, or take a nap to avoid jet lag. However, I cannot consider playing lawn tennis. Constructing value for various available options is central to goal selection.

Several factors can contribute to how we value goals. I list some of them below.

- *Terminal rewards:* The reward available in the goal state is a source of motivation, as in the case of a person wishing to pick up an apple in order to eat it.
- *Reward schedule:* The intermittent rewards available during goal pursuit, rather than just the terminal goal state, can also influence goal selection. Opting to watch a movie is one example where experiencing rewards throughout the duration of the activity can itself be appealing. These intermittent reinforcements can either be extrinsic rewards or intrinsic rewards shaped by latent task states.
- *Goal dependencies:* Completion of a goal that helps advance higher-level goals — such as obtaining a doctorate in service of an academic profession — can also influence goal selection. In a study leveraging data from the video game *Little Alchemy*, researchers showed that humans prefer selecting goals that can be composed together to achieve higher-level goals; in other words, goals that empower the pursuit of other goals.
- *Effort:* The effort and resources — cognitive, motor, or material — expended in pursuit of goal-directed activity are important considerations and tend to trade off with rewards. Low-effort, high-reward goals are typically more valuable than high-effort, low-reward ones.
- *Uncertainty:* Uncertainty or risk associated with either the goal state or the process of goal pursuit can deter goal selection depending on an individual's risk or ambiguity preferences.
- *Emotion:* Emotional states such as mood and valence strongly influence goal selection. Additionally, psychiatric conditions such as depression and anhedonia can alter motivation levels and influence goal setting.
- *Social factors:* Social considerations such as enhancing reputational value and mating prospects frequently influence goal setting.
- *Information-seeking:* A drive to seek information in order to reduce risk and ambiguity in the environment can itself become a goal. For example, when choosing between two equally attractive options, one may set a goal to obtain more information by conducting market research.

When we have multiple goals under consideration, we can construct value for each of them according to our preferences over the factors listed above and allow them to compete for selection at any given instant. However, commitment to an extended goal does not end with a one-shot decision to pursue it at a particular moment. It is a sustained choice regarding whether to maintain commitment until the goal state is reached or decide to quit midway.

Sustaining goals

There are three choices one can make with regard to maintaining commitment to a goal state at any moment: sustain commitment, interrupt the goal in favor of another goal with the intention of revisiting it later, or quit the goal altogether. More often, our decisions pertaining to long-term goals fall into the first two categories, as the decision to completely abandon long-term goals is rare and momentous. However, we interrupt our goals constantly. When we get stuck on a math problem that we have been working on for quite some time, we may decide to take a break and return to it the next day with a fresh perspective. We interrupt writing a report in order to have lunch because it becomes a more pressing demand.

Nevertheless, we do not switch goals too often in daily life. We maintain persistence toward a given goal for a reasonable amount of time, even though our minds may temporarily drift away from the current task. We might set a goal of working on a report for two hours in the morning, but still feel tempted to briefly interrupt it to check social media or read a news report. However, if the goal of finishing the report is sufficiently strong, it will pull the mind back toward the task after brief interruptions. To handle such interruptions, representations of the goal and its associated actions need to be maintained in working memory so that the task can be resumed without delay.

Interruptions can, however, last across several timescales. If a goal is interrupted and resumed weeks later, representations of the goal must be maintained in long-term memory so that they can be retrieved and the task comfortably resumed after a significant delay. Thus, sustaining long-term goals involves interfacing with memory processes across multiple timescales.

While memory mechanisms support goal commitment, reward and motivational systems in the brain moderate the drive for sustained persistence. Reinforcements delivered intermittently during goal pursuit can serve as powerful motivators for continued commitment. We see clear examples of this in gambling or video games,

where reward schedules embedded in the environment strengthen commitment and induce reluctance to switch away from those goals. Moreover, seeing consistent effort yield progress toward a target can itself motivate further effort. In Chapters 2 and 3, I further emphasize how goal progress and its rate influence goal commitment.

When the strength of goal commitment is high, we can become overly focused on the goal at hand and fail to process alternative tasks. This relates to flow states observed during heightened concentration. However, when commitment wanes through decreased drive or reduced rates of reinforcement, we begin contemplating alternatives. When a more rewarding option presents itself, we may interrupt the current goal and break the cycle of commitment.

Executing goals

We take actions in the world that bring us closer to the goals we set for ourselves. When our goals are complex and elaborate, so are our actions. We often split complex goals into manageable subtasks, also called subgoals, which we pursue in service of higher-level goals. Defining subgoals:

Subgoal is an intermediate state in the world, occupying which brings the agent in proximity to the higher-level goal and is itself designated the status of a goal.

Cooking a simple dish of pasta requires performing a sequence of subgoals — say, boiling water and cutting vegetables — each of which becomes a goal in itself. Most of the time, the structure of the goal naturally segments the task. At other times, we must segment the goal ourselves in order to identify meaningful subgoals. Decomposing a task hierarchically has two main advantages. First, it is easier for a capacity-constrained agent to handle. Humans are constrained by working memory limitations, and breaking a task into smaller components reduces cognitive load. Second, multiple tasks can share subgoals, and learning to execute a subgoal for one task automatically transfers to others. This enables us to pursue novel tasks through single-shot learning, a phenomenon also called compositional generalization. Hierarchy and compositionality are thus central facets of goal execution, requiring planning to stitch sequences of actions together.

Constructing hierarchical plans

Planning is the construction of a sequence of actions that an agent executes in advance, requiring the agent to possess a model of the world that can be used to predict state transitions caused by its actions.

We construct plans toward our goals across multiple hierarchies. First, we chart paths to goals by sequencing subgoals, and then we chart paths to subgoals by sequencing actions. However, this process begins with discovering subgoals in the world. In experimental planning paradigms, humans have been shown to designate bottleneck states that are frequently visited because they act as connector hubs to other states in the world. Designating such states is helpful not only for pursuing current goals, but also for future goal pursuit in cases where rewards shift or new goals emerge within the same environment.

Constructing plans through subgoal discovery often requires knowledge of transitions in the world — in other words, building a cognitive map. Tolman's classic work showed that rats learned the spatial layout of a maze even without specific rewards, and subsequently became quicker at obtaining rewards once they were introduced. Building maps also enables flexible goal pursuit: if one regular path becomes obstructed, alternative routes can be taken. However, planning through a cognitive map is computationally costly, especially in complex environments.

Studies on human planning suggest the prominence of a different strategy in building and using cognitive maps. Humans are thought to encode state-to-state transitions in the form of successor representations, which store proximity measures between states and are learned through experience as agents move through the environment. This strategy fails when the transition structure of the world is disrupted — for example, when a path is blocked — but generalizes well to different goal positions within the same environment. Once the location of a goal or subgoal is known, the sequence of successor states can be chained together to reach the destination.

However, even planning through successor representations can be computationally intensive and requires active cognitive engagement. For frequently repeated tasks, it becomes advantageous to automate sequence execution and avoid the costs of planning. This is where skills for goal pursuit emerge.

Building skills

The process of skill-building can be characterized as a transition from goal-directed control to automatic processing. Building schemas or templates for commonly executed sequences helps avoid the latencies and cognitive costs involved in deliberate sequencing of actions. These schemas, also referred to as micro-skills or habits, can also be reused in similar goals with shared subgoals. For instance, developing a schema for typing the chunk *th* can optimize the execution of both *that* and *there*. This optimization is one reason why composition and chunking of micro-skills are ubiquitous in both motor tasks such as speaking and cognitive tasks such as mathematics.

The more we practice through repeated goal execution, the more strongly we reinforce transitions between relevant sequences, thereby strengthening schema encoding. During sleep, sequences practiced during the day are replayed for abstraction and consolidation into schemas. Building skills enables effective multitasking with concurrent goals. For example, if the walk from one's office to home is sufficiently practiced, one might contemplate a difficult math problem while walking.

Ultimately, this all reflects the fact that we are bottlenecked by attention. We can handle only a limited bandwidth of cognitive processing and therefore attempt to allocate processing resources across multiple goals under constraints. When a particular goal is no longer viable, it is often advantageous either to interrupt or abandon it and switch to a more favorable goal.

Adapting goals

We constantly monitor the consequences of our actions and how they affect goal pursuit. Are we getting closer to our goal, or are we stalled? Is the goal still rewarding, or has it become devalued? Such considerations influence how flexibly we switch between goals. Task switching, however, is known to be costly, increasing both cognitive effort and latency. Therefore, we tend to maintain stable goal states for reasonable durations while still retaining the flexibility to switch when warranted. In later chapters of this thesis, I introduce theories explaining why we persist and how switching costs originate. Goal pursuit is therefore both stable and flexible, requiring an architecture that constantly monitors ongoing and alternative goals in order to facilitate adaptive management of multiple goals.

In summary, to build agency capable of handling complex long-term goals, we require an architecture with the ability to select, sustain, execute, and adapt multiple

goals efficiently under resource constraints.

1.3 Thesis outline and scope

In this thesis, I touch upon all of the architectural components described above and contribute several insights to each of them.

In Chapter 2, I discuss the human tendency to overpersist in goals within an extended goal-pursuit paradigm in which individuals can freely choose between multiple goals. I introduce a descriptive account of persistence through the computational construct of goal momentum to explain patterns of human goal persistence. I further discuss why, under normative considerations, overpersistence may be optimal.

In Chapter 3, I extend the momentum account to explain persistence and flexibility patterns in humans engaging in a hierarchical goal-pursuit paradigm involving freely chosen goals and subgoals. I also show how persistence can be optimal for building micro-skills and how momentum computations can drive persistence at both the goal and subgoal levels.

In Chapter 4, I segue into the study of how neural architecture enables pursuit of long-term goals. I review prior literature to gather evidence regarding how each of the architectural desiderata discussed above is fulfilled in the brain.

In Chapter 5, I study hierarchical goal pursuit using neuroimaging to investigate how the brain organizes flexible goal pursuit. Specifically, I discuss how goal and subgoal representations are sustained, valued, and monitored in the brain, and validate the momentum model of goal valuation by identifying its neural correlates.

In Chapter 6, I provide a mechanistic account of goal commitment by introducing a circuit model incorporating momentum computations. I introduce a dynamical systems perspective on stable and flexible goal commitment and discuss the mechanistic origins of switching costs.

I conclude the thesis in Chapter 7 with a discussion of the implications of this research for understanding individual differences in human agency and for building artificial intelligent systems endowed with agency.

Chapter 2

WHY WE PERSIST IN GOALS: A MOMENTUM-BASED ACCOUNT OF GOAL PERSISTENCE

Extended goals necessitate extended commitment. We address how humans select between multiple goals in a temporally extended setting. We probe whether humans engage in prospective valuation of goals by estimating which goals are likely to yield future success and choosing those, or whether they rely on a less optimal retrospective strategy, favoring goals with greater accumulated progress even if less likely to result in success. To address this, we introduce a novel task in which goals need to be persistently selected until a set target is reached to earn an overall reward. In a series of experiments, we show that human goal selection involves a mix of prospective and retrospective influences, with an undue bias in favor of retrospective valuation. We show that a goal valuation model utilizing the concept of ‘momentum’, where progress accrued toward a goal builds value and persists across trials, successfully explains human behavior better than alternative frameworks. Our findings thus suggest an important role for momentum in explaining the valuation process underpinning human goal selection.

2.1 Introduction

Humans often direct sustained effort to reach rewarding states that are distant from them. We often refer to such distant setpoints as goals – desired future states for our future selves. Whether it is climbing Mount Whitney or landing a dream job, these goals require sustained effort to ensure gradual progress toward the respective goals [125], [32]. We often toggle between multiple such end states. We wish to finish our project deadlines, while also contemplating unfinished construction projects at home. How do we select which long-term goal to pursue at any given point, and how do we assign value to multiple goals at the same time?

While a desired end state can be a source of motivation, maintaining such a long term goal could require intermediate reinforcement, signaling whether a target is achievable or not worth pursuing. Progress towards the target is a potential reinforcement signal that serves to validate our efforts toward goal pursuit. Sustained progress that draws us closer to the target motivates us to expend more effort in the pursuit, whereas a slump in progress lowers our drive and induces us to pursue

another target that is more attainable. [127] found that even progress not explicitly signaled in a task can, when implicitly ruling out alternative options, influence autonomous goal selection by limiting exploration [144].

Here, we introduce a computational model of momentum in progress, to capture the notion of a sustained drive towards a distant goal. We propose that intermittent reinforcement signaling progress builds up in value and drives persistence towards goals. According to this framework, a goal that yielded sustained progress towards a target in the recent past can be perceived as more valuable than an alternative goal which made no progress in the past, but yet has a greater probability of reaching the target in the near future. Our proposed model of persistent valuation predicts an inertial effect in the willingness to change targets even when the rate of progress toward a goal comes to a standstill [159].

Momentum is a concept deeply entrenched in the domain of Newtonian physics that formalizes inertia or resistance to change that is fundamental to all physical quantities embodied with mass. In physics, momentum is defined as a product of mass and velocity, $p = m * v$, and within its formulation there is an evident corollary that the greater the momentum of an object, the greater is its resistance to change. In other words, more force is required to stop an object with higher momentum, which is a function of both its mass and velocity of progress.

Several applications of this concept in cognitive and computational science indicate that this construct might have implications beyond rigid body dynamics. In the perceptual decision-making literature, it has been observed that memory for the location of a moving target is often displaced in the direction of the target's motion, which has led to the use of the term *representational momentum* alongside the hypothesis that the perceptual system evolved to function in the regime of Newtonian physics [53], [85], [95]. In operant conditioning, the persistence of learned operant behavior in the face of altered conditions was termed *behavioral momentum* [137], [143], [114], [138], [136]. Momentum is also a key concept in machine learning utilized for stability in learning and optimization [161]. Momentum induces inertia in gradient updates by maintaining a running average of recent updates instead of fluctuating momentary gradients to update expectations/ policies. This rewards consistency in signed gradient errors, thereby maintaining stability in the direction of updates. The same concept was proposed as a possible explanation for the effect of mood on decision making ([152], [44]).

Psychologists have previously linked the concept of momentum to the phenomenol-

ogy of goal pursuit ([26], [120], [83]). However, most of this previous work has been limited to qualitative formulations, introducing momentum as a potential account for the tendency of dispositional states to carry forward through time; task-set inertia, mental simulation, and rumination are a few examples. Moreover, momentum, as a term, has been a part of the popular vernacular relating to the pursuit of goals (“this project needs more momentum”, “this fight is gathering momentum every day”). However, its quantitative formalization in the decision-making literature is lacking.

Decision inertia and persistence of actions are extensively studied in psychology, albeit in the context of intermittent reinforcements. Notable studies in the foraging literature have used the concept of the opportunity cost of time as a means to decide whether to continue forage in the current patch or move to a new one with a delay in time [34], [94], [30]. Differently from this approach, in the present study we focus on scenarios where rewards are distant, which requires the maintenance of a persistent drive toward goal pursuit.

Suboptimal human persistence has been extensively documented in prior literature where there are sunk costs [10] [11], [166]. Humans are known to overpersist on goals that have incurred sunk costs of time, effort, and resources, even when it is optimal to abandon them. The sunk cost fallacy demonstrates similar persistence effects in relation to extended goals, the neural basis of which was recently investigated in humans [81]. But not all goal-selection decisions made in daily life occur in the presence of explicit sunk costs. For instance, we can decide to stop working on debugging code and rest instead, with the full knowledge that our work can be resumed later. In the present study, we explicitly emphasized investigating human goal persistence patterns in a multitasking setting rather than a sunk-cost one. Therefore, we allowed goals to retain accrued progress having switched away from them. This was done to mitigate the influence of sunk costs in our experiments related to abandoned goals completely losing their value. Consequently, the present study deviates from previous work [81], as sunk costs are likely to exert a potent influence on the long-term pursuit of the goal. Moreover, in [81], while overpersistence was observed in behavior in that study, providing a computational account for the overpersistence effect was not their focus. Here, our overall goal is to understand the nature of the overpersistence effect, particularly after the effects of sunk costs explicitly related to a loss of prior goal progress have been reduced.

We make two significant contributions through the study described here. Firstly, we introduce a novel paradigm to investigate the temporal nature of valuation of

extended goals in the presence of competing alternatives, a hitherto unexplored question. Secondly, we provide an algorithmic account for persistence in extended goals. We introduce momentum as a key computational variable driving inertia in goal valuation, providing a fresh perspective on human persistence and decision inertia.

2.2 Experiment

We introduce a suit collection paradigm to simulate extended goal pursuit inspired by traditional card games. A suit in the game comprises a specific number (say seven) of tokens of the same kind. The objective for the participants is to collect as many suits as they can and win points in the game (which ultimately translates to a monetary bonus). There are three different types of suits available (*cat*, *hat*, *car*) and participants can freely switch between collecting different suits without penalty to earn points in the game. Participants only receive points when a suit is completed without any partial credit for any incomplete suits.

Gameplay

The game is played in blocks with each block having 30 rounds. A sample round of the gameplay is shown in Fig 1. Participants start with an empty set of slots (7 for each token type as shown in Fig 1A).

State: The game state is indicated by the current slot configuration which indicates progress made toward different suits. An example slot configuration occurring a few rounds into the game is shown in Fig 1C.

Actions: There are 6 cards in the game, each corresponding to a specific suit type (two cards are allotted for each suit, see Fig 1B). On each round, a participant is shown three cards (one card for each suit). Participants can choose to flip between one of the three cards shown to receive tokens for a suit (see Fig 1C). For example, if a participant wishes to acquire a *car* token to collect the *car* suit, they would pick the *key* or the *baggage* card (whichever is available in the round) to try to acquire the token.

Outcomes: If the participant flips the *car*-related card (say), they receive the *car* token with a specific (non-zero) probability and the *hat* and the *cat* tokens with zero probability. If the participant receives the *car* token, one of the empty *car* slots is filled and the participant proceeds to the next round of token collection. If a suit's target is reached (say 7 *tokens* of one type are collected), the participant will receive 10 points for finishing the suit and all of those 7 slots are emptied for a fresh bout



Figure 2.1: Suits task. **A.** Participants start with empty slots to collect tokens for each of the three available suits (*cat*, *hat*, *car*). **B.** Each suit type is mapped to two distinct cards which serve to stand in for the suits in the game. **C.** Panel shows a sample round of the task: 1. Participants see suit collection progress of all three goals. 2. Participants pick a card corresponding to one of the three goals. 3. Participants see the outcome of their action. 4. If a token is received, slots are updated. 5. Periodically, participants report the suit they are pursuing. **D. Conditional probabilities.** $\Pr(\text{token}|\text{card})$ for different token-card combinations are shown for the 80-20 block. Selecting the *basket* or the *mat* card gives you the *cat* token with a high probability of 0.8, making *cat* the dominant suit in the block. **E. Block shifts** Each block has fixed probabilities with one dominant suit and two inferior suits of identical probabilities. The dominant suit is switched across adjacent blocks (*Cat* is most abundant in block N while *hat* is the most abundant in the next one). **F. Token probabilities.** Two experimental variants (experiment 1 and 2) with probability configurations specified for each suit. **Figure credits.** We used images from <https://openclipart.org/> to generate the figures. See the Task design section in Methods for full image credits.

of collection.

Game blocks: Gameplay is continuous, and participants retain any tokens from incomplete suits across the entire game. However, the game is structured into blocks, and participants are informed that each block introduces a new probability distribution over the three suits. For example, in an 80-20 block, the probabilities might be 0.8 for *cat*, and 0.2 each for *hat* and *car*. The conditional probabilities of receiving a token given a card-action pairing in such a block are shown in Table 1 (see also Fig 1D). When a new block begins, the suit probabilities shift. For instance, in a 70-30 block, the dominant suit might now be *hat*, such that $\Pr(\text{hat}|\text{hat-related card}) = 0.7$, while $\Pr(\text{cat}|\text{cat-related card}) = 0.3$ and $\Pr(\text{car}|\text{car-related card}) = 0.3$ (see Fig 1E). Participants are explicitly informed about the block structure: each block lasts 30 rounds, different suits have different probabilities within a block, and these probabilities change when a new block begins.

80-20 block

$\Pr(\text{token} \text{card})$	<i>cat</i>	<i>hat</i>	<i>car</i>
<i>cat</i> -related card	0.8	0	0
<i>hat</i> -related card	0	0.2	0
<i>car</i> -related card	0	0	0.2

Table 2.1: $\Pr(\text{token}|\text{card})$ for the 80-20 block. The dominant token is *cat* with $\Pr(\text{cat}|\text{cat-related card}) = 0.8$

Experimental manipulations

There is a dominant suit in each block. We designed blocks in the game so that there is a dominant suit on each block (counterbalanced across all suit types), such that flipping its cards gives tokens with the highest probability (in Fig 1D, the *cat* suit is the dominant one with 0.8 probability). The other two suits have identically inferior token probabilities (both *hat* and *car* with 0.2 probability). We instructed the participants of this task manipulation, explaining that within each block there is a specific suit which has the highest chance of token collection. Pursuing the dominant suit type in each block gives participants the best chance at success in suit completion and we investigated how often participants chose the dominant option in a block in relation to their progress on various suits.

Probability disparity between suits is modulated across blocks. We also varied the probability disparity between the dominant and the inferior suits to ascertain how it modulates goal selection. We expected that selection of the dominant suit

would be high in the high disparity blocks (where the probability difference between the dominant and the inferior suits is highest) and that this would be reflected in participants' goal selection patterns.

The dominant suit is switched across adjacent blocks. We implemented dominant token switches across blocks so as to enable us to investigate conflicts in goal selection between retrospective and prospective influences, as this is critical to assess how individuals change their choices of goal in the face of such a conflict. If a participant pursues a dominant suit in a block and is close to finishing it, and they encounter a block shift when the hitherto dominant suit is no longer dominant, then it is relatively more difficult to make further progress towards its target. A goal conflict arises in these situations, concerning whether to persist in the suit that is close to the target, or to switch over to the new dominant suit. Such scenarios occur frequently during the game, and the optimal strategy is to switch over to the dominant suit in the new block. We did not instruct the participants that the dominant token in one block becomes an inferior token in the adjacent block.

Testing the effect of instructions. We varied the instructions the participants received on the precise conditional probabilities across blocks. We investigated the extent to which knowledge of absolute probabilities of dominant and inferior tokens (without disclosing the identity of the dominant token) would influence goal selection in the game.

Effect of non-uniform targets on goal selection. We set non-uniform targets for the three suits to ascertain how that influences extended-valuation of the suits in relation to their progress and task-conditions.

We include four experiments to implement the above mentioned design manipulations. Experiments 1-3 have uniform targets for all suits, with each suit comprising of seven tokens of one kind. Experiment 1 has 18 blocks with 30 rounds each, with three different probability structures: 80-20, 70-30, 60-40. In the 80-20 block (say), the dominant suit is available with 80% chance and the two inferior tokens are available with 20% chance. Experiment 2 has 12 blocks (30 rounds each), with two different block types: 75-25 and 55-45. In both experiments 1 and 2, our instructions are limited in that, within each block, one specific suit has the highest probability, and the probabilities change with each block switch. In experiment 3 (which is a repeat of experiment 1), we gave participants more information about their current block context. Before each block began, we specifically informed them that they are in an 80-20 context, without telling them which specific suit is available with 80% chance (more details on task design and instructions are elaborated in the Methods

section). In experiment 4, we employed non-uniform targets for different suits and explored its impact on validating our computational hypothesis.

2.3 Retrospective bias in temporally extended goal selection

We analyze goal selection patterns in human participants influenced by the current context (suit probabilities in a block) and slot configuration. In a given round, participants can choose to collect the dominant suit (with the highest probability of receiving tokens), or choose to pursue the suit with the most progress (max progress suit) according to their slot configuration. In Fig 2A, we show the participants' probability of selecting cards related to the dominant suit, the max progress suit, and the third suit. Participants preferred to select either the dominant or the max progress suit across all conditions, picking the remaining suit well below the chance level. This indicates that the primary conflict in goal selection is between the accrued goal progress and the current rate of goal progress.

Moreover, the preference toward the dominant/max progress suit shows an interaction with the block condition. Participants show a higher preference for the dominant suit when the disparity between different suits is large (80-20 / 75-25 conditions) and a lower preference when the disparity is small (60-40 / 55-45 conditions), indicating sensitivity to the current rate of progress (repeated measures ANOVA; Experiment 1: $F(2, 86) = 20.23$, $p < 0.00001$; Experiment 2: $F(1, 49) = 17.32$, $p < 0.00001$). Likewise, preference toward the max progress suit increases with decreases in disparity between the suits (repeated measures ANOVA; Experiment 1: $F(2, 86) = 19.67$, $p < 0.00001$; Experiment 2: $F(1, 49) = 7.1$, $p < 0.01$). However, in all conditions, participants preferred to pick the maximum progress suit over the dominant suit, indicating an increased preference for the accrued progress over the current rate of progress.

Fig 2A provides a partial examination of the choice patterns, highlighting rounds where the dominant suit and the max progress suit are separate. However, in a fraction of rounds, the dominant suit is the same as the max-progress suit because maximum progress is made toward the dominant suit. In 62% of the rounds in Experiment 1 and 63% of the rounds in Experiment 2, the dominant suit is separate from the max progress suit, while in the rest of the rounds the two options converge. We noted that in the rounds where the dominant and the max progress suits are the same, both the accrued progress and the current rate of progress for the suit are high, mitigating conflict in goal selection. In those rounds, participants predominantly

chose to pursue the dominant option (or the max progress option), picking the other two options occasionally and well below the chance level (see S1 Fig).

Limiting our analyses to the rounds where the dominant and max progress suits are separate (with a goal selection conflict), we further analyzed how progress differences between the two suits influence goal selection. Fig 2B shows the probability of choosing the max progress suit as a function of the difference in progress between the two suits. We show that participants' preference toward the max progress suit monotonically increases with the increasing progress difference between the suits. This preference is highest when the max progress suit is at 86% progress (6/7 tokens completed), and when the dominant suit has 0/7 tokens collected.

The pattern of increased preference for accrued progress in human participants is not necessarily suboptimal. If a suit is close to completion, it might be optimal to persist with it and get a reward sooner than to switch to the dominant suit that might have a high rate of progress but an equally high trials-to-completion estimate. In fact, humans are known to prefer sooner rewards over later ones, a tendency due to the incorporation of a discount factor in the estimation of the reward value incurred over a delay [52]. We investigated whether the goal selection patterns of participants are optimal under this assumption.

We constructed a prospective agent which prioritizes faster suit completion and reward accumulation. This agent, rather than optimizing for discounted returns accumulated at the end of the task, focuses on the immediate reward delivered at the end of suit completion. The agent thus places the highest value on goals that have the lowest estimates of trials-to-completion. To estimate values, the agent needs to maintain beliefs about the current probability estimates of various suits and roll out expectations into the future until the termination condition (collecting 7 tokens) is reached. We constructed the agent as a partially observable Markov decision process, with the current slot configuration as the observed state, selection of cards as actions, and an estimate of the current suit probabilities as the belief state. The agent updates the belief about token-outcome probability of a suit using a delta learning rule:

$$M_g^{(t)} = M_g^{(t-1)} + \alpha(I_g^{(t)} - M_g^{(t-1)}) \quad (2.1)$$

where $M_g(t)$ is the belief at time t , $I_g(t)$ is the indicator function of the token outcome (1 if the token is received in a round), and α is the learning rate. Using the

estimated probability of belief, the prospective agent rolls out expectations one step in the future for two scenarios: receiving a token with probability M_g and moving one token closer to the target ($s_g^{(t)} + 1$); and not receiving the token with probability $1 - M_g$, and staying in the same slot state ($s_g^{(t)}$). The rolled-out future states are discounted by a factor of γ .

$$Q_g(s_g^{(t)}) = M_g \gamma Q_g(s_g^{(t)} + 1) + (1 - M_g) \gamma Q_g(s_g^{(t)}) \quad (2.2)$$

where $Q_g(t)$ is the value of the goal suit g , and $s_g(t)$ is the current slot state of suit g . Specifically, $s_g(t)$ is a counter for the progress of the goal, with the above recursive estimation terminating when the current state reaches the target (say, 7 tokens). Recursion is also terminated with a horizon of 20 steps.

The prospective agent, as defined above, can favor pursuing the maximum progress suit over the dominant suit. When the maximum progress suit is close to completion (say, 5 out of 7 tokens), the agent has to unroll fewer steps ahead to receive rewards compared to the dominant suit that is further away from the target (say, 2 out of 7 tokens). However, the prospective agent is sensitive to block conditions: When placed in the 80–20 block, the chance of collecting two additional tokens with a chance of 20% is lower than that of collecting five tokens with a chance of 80%, and the agent prefers to follow the dominant suit.

We optimized the parameters of the prospective agent (parameters and action selection process documented in the Computational Modeling section of Methods) for task performance (total number of suits completed). The optimal discount factor for this agent is $\gamma = 0.95$ for Experiment 1 and $\gamma = 0.93$ for Experiment 2. The prospective agent, thus optimized for the task, significantly outperforms the human participants (performance distinctions are elaborated in a later section).

For each slot configuration seen by human participants, we computed the response exhibited by the prospective agent and investigated the extent of the participants' deviation from this response. For generating the prospective responses, we assumed the true state of suit probabilities to be the agent's belief (not accounting for learning effects). We define the suit selected by the prospective agent as the prospective suit, and the suit with the maximum progress in a round as the retrospective suit. Fig 2C provides a revised analysis of Fig 2A, indicating the goal selection patterns to choose between the prospective suit, the retrospective suit, and the third suit (when

the identities of the prospective and retrospective suit are different). The revised goal selection patterns show a trend similar to that shown in Fig 2A, with the preference for the retrospective suit being higher than that of the prospective suit in all conditions. Moreover, the prospective suit shows the same interaction profile with the block condition (repeated measures ANOVA: Experiment 1: $F(86) = 7.04$, $p < 0.001$; Experiment 2: $F(49) = 17.79$, $p < 0.0001$), and also with the retrospective suit (Experiment 1: $F(86) = 2.14$, $p = 0.12$; Experiment 2: $F(49) = 8.88$, $p < 0.005$). The proportion of rounds in which there is a conflict of goal selection in the current classification (prospective suit versus retrospective suit) is now reduced to 40% on average across all conditions (as opposed to 60% previously in the max progress suit vs. dominant suit classification), indicating that a fraction of max progress choices are indeed prospective in nature after accounting for discounting effects. S1 Fig shows preferences when the prospective and the retrospective suit identities are the same (when the prospective agent prefers the maximum progress option), with participants predominantly choosing the prospective/retrospective option devoid of the goal selection conflict.

In summary, the patterns of goal selection in human participants reveal a retrospective bias for accrued progress in the task. Further quantifying the extent of this bias, Fig 2D shows the overall proportion of the choice of dominant suit in participants compared to that of the prospective agent optimized for the task. We also split the choices as occurring in the first and second halves of blocks (averaged over all blocks) to account for any learning effects. Both participants and the prospective agent show learning effects within blocks, with the proportion of dominant choice increasing in the latter half of the block. However, the dominant choice of the participants remains significantly lower than that of the prospective agent in both halves of the blocks, causing suboptimal human performance in the task (elaborated in a later section). Lower learning rates and increased discounting in participants could potentially explain suboptimality in task performance in humans, which we consider to be a potential explanation for the goal selection patterns seen in humans. This account is tested in later sections by fitting the prospective agent to the participants' choices. Furthermore, there are no significant effects of learning or performance changes between blocks. S3 Fig shows the proportion of dominant choice between blocks, where we find no significant changes in the selection effects as the task progresses.

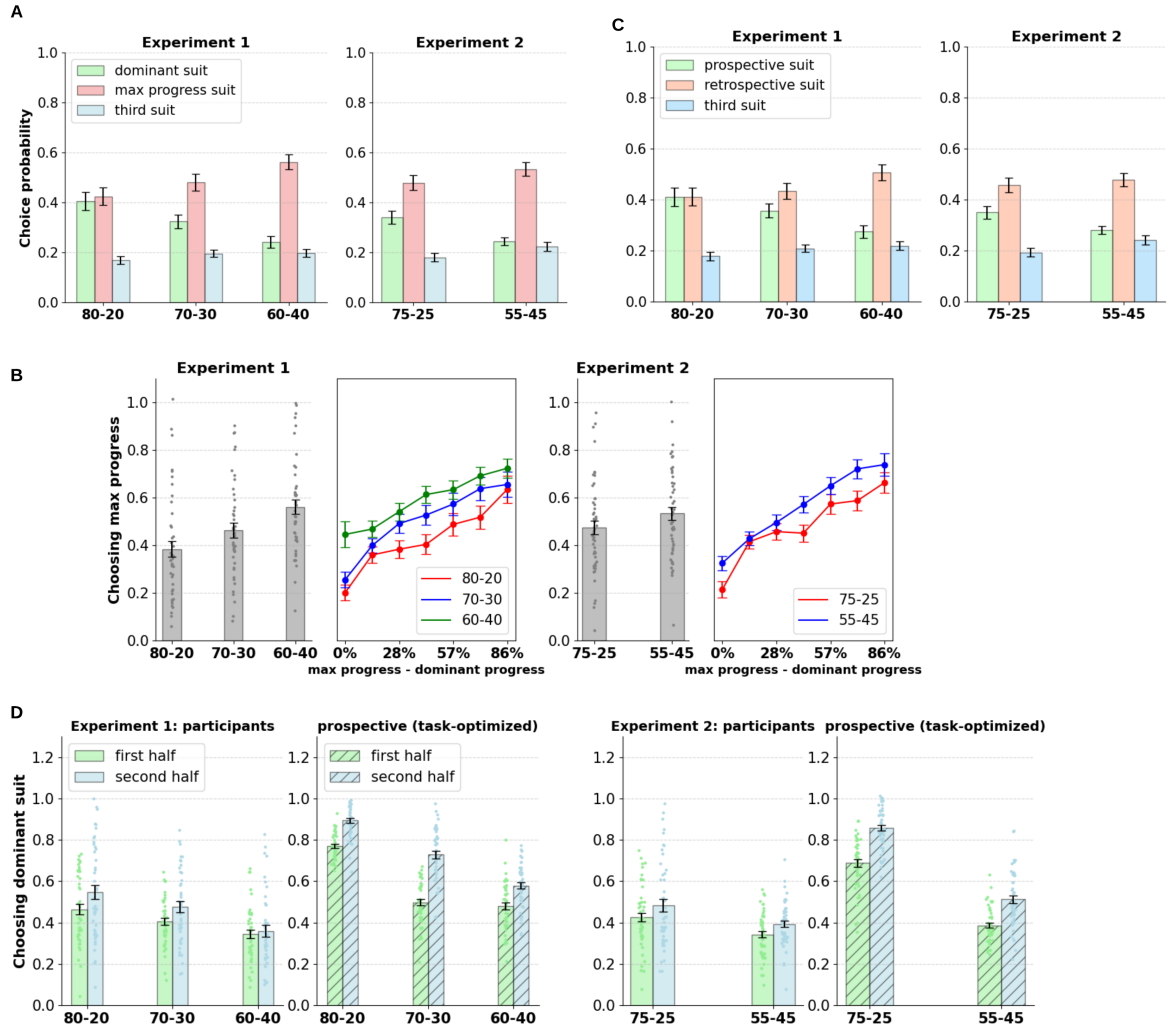


Figure 2.2: A. Participants show a preference for the progress accrued over the current rate of progress. Participants primarily chose between the dominant suit (one with the highest probability within the block) and the max progress suit (one which made greatest progress towards the target), while choosing the third option relatively infrequently. **B.** The proportion of times participants chose the max progress suit over the dominant suit (when the two diverge) is broken down by block type. The preference for the max progress option increases with the difference in progress between the dominant and max progress suits. **C.** Participants show a preference for the retrospective suit over the prospective suit (when the two choices diverge). Participants chose the retrospective suit (with the greatest progress) over the prospective suit (optimal choice prescribed the task-optimized prospective agent). **D. Learning effects within blocks.** Panel shows the dominant suit choice probability in the first and second halves of the block (15 rounds each). The proportion of dominant choices increases in the second half indicating learning effects within blocks. The choice patterns are contrasted with that of the task-optimized prospective agent, wherein, preference for the dominant suit is significantly lower in participants in both halves of the blocks, indicating suboptimal retrospective bias in goal selection.

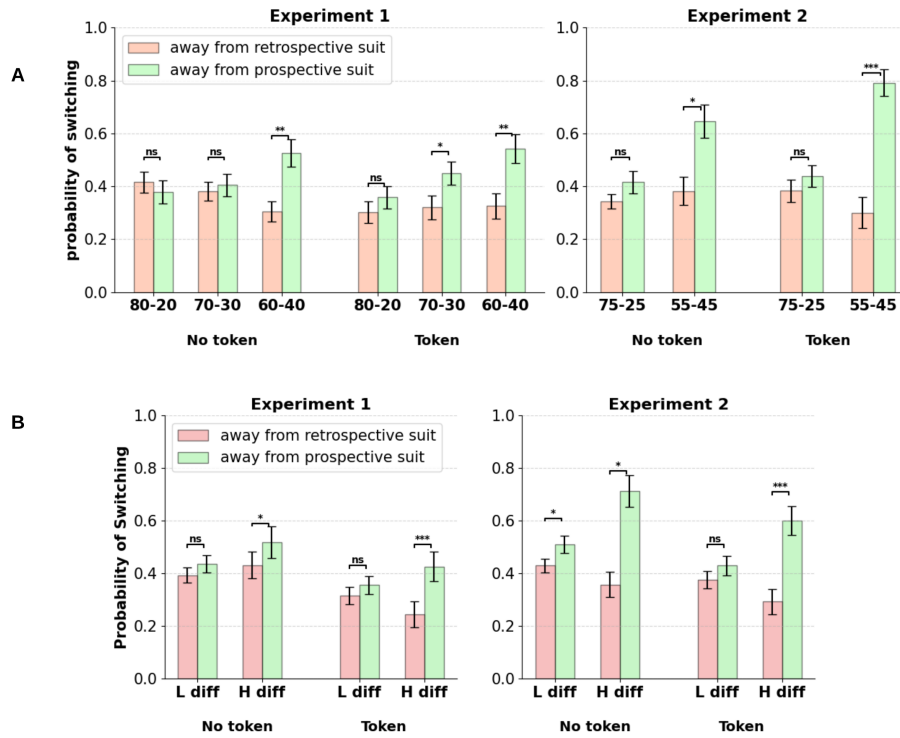


Figure 2.3: **A. Participants show greater persistence towards the retrospective suit.** Participants make more switches away from the prospective option than the retrospective option. Switches are split by suit type (prospective/ retrospective), block condition, and whether the round elicits a token or not. Significance markers indicate paired-sample t-tests (* : $p < 0.01$, ** : $p < 0.001$, *** : $p < 0.0001$). **B. Switches split by progress differences between the suits.** H diff, L diff correspond to higher and lower progress difference between the suits. Increased persistence with the retrospective suit when progress difference between the suits is high.

2.4 Persistence with the retrospective option

We further investigated temporal persistence patterns with regard to suit selection. Although overall goal selection is biased toward the retrospective suit, we further aimed to determine whether we could find evidence of a bias toward temporal persistence once a suit is selected. For this, we calculated the proportion of rounds in which participants switched away from the selected option in the subsequent round. We isolated these switches by suit type: separating rounds where they switched away from the prospective suit from those where they switched from the retrospective suit (similarly defined as in the previous section). Fig 3A shows the switch probabilities of the participants. We also separated the conditions where participants received a token in the previous round from those where they did not

receive a token.

We found that participants made a significantly greater number of switches away from the prospective suit regardless of the token outcome in the previous round (paired-sample t-test; Experiment 1: $t(86) = -2.22$, $p < 0.01$; Experiment 2: $t(96) = -2.79$, $p < 0.001$). We found that the proportion of prospective switches interacts with the block types (repeated measures ANOVA; Experiment 1: $F(2, 144) = 12.56$, $p < 0.0001$; Experiment 2: $F(1, 49) = 19.55$, $p < 0.0001$). Participants are most likely to persist with the prospective suit in the high-disparity blocks (80–20 / 75–25), where the adaptive value of persisting with the prospective token is maximal, compared to the low-disparity blocks (60–40 / 55–25). However, overall levels of persistence with the prospective suit are lower than those of the retrospective suit.

Retrospective switch propensities are also sensitive to the relative progress differences between the two suits and are lower when the difference is high. Fig 3B shows switching proportions away from prospective and retrospective suits split by high ($> 50\%$) and low ($\leq 50\%$) progress difference between the two suits. Participants demonstrate increasing levels of commitment toward the retrospective suit with its increasing distance (progress) from the prospective suit. This is reflected in lowered switching from the retrospective suit and increased switching from the prospective suit. Increased commitment toward the retrospective choice is thus evidenced not only by resistance to switching from the choice but also by increased rates of revisiting it after having switched away from it.

This points out an interesting facet of goal persistence which can be dissociated from mere stickiness in choice driven by task-switching costs. The persistence in goal selection that we describe here tends to involve a tendency to revisit the retrospective option even after it has been previously switched away from, whereas choice stickiness typically refers to a tendency to refrain from switching away from a particular action in the first place. Thus, accrued progress in a particular goal seems to exert a retrospective pull on subsequent decisions, clearly highlighting persistent retrospective influences on goal selection.

2.5 Forgoing performance with persistence

We showed evidence that human participants exhibit a bias towards selecting and persisting with the retrospective suit. Here, we analyze how bias translates into suboptimal performance in the task.

We compare the performance of human participants with that of the task-optimized

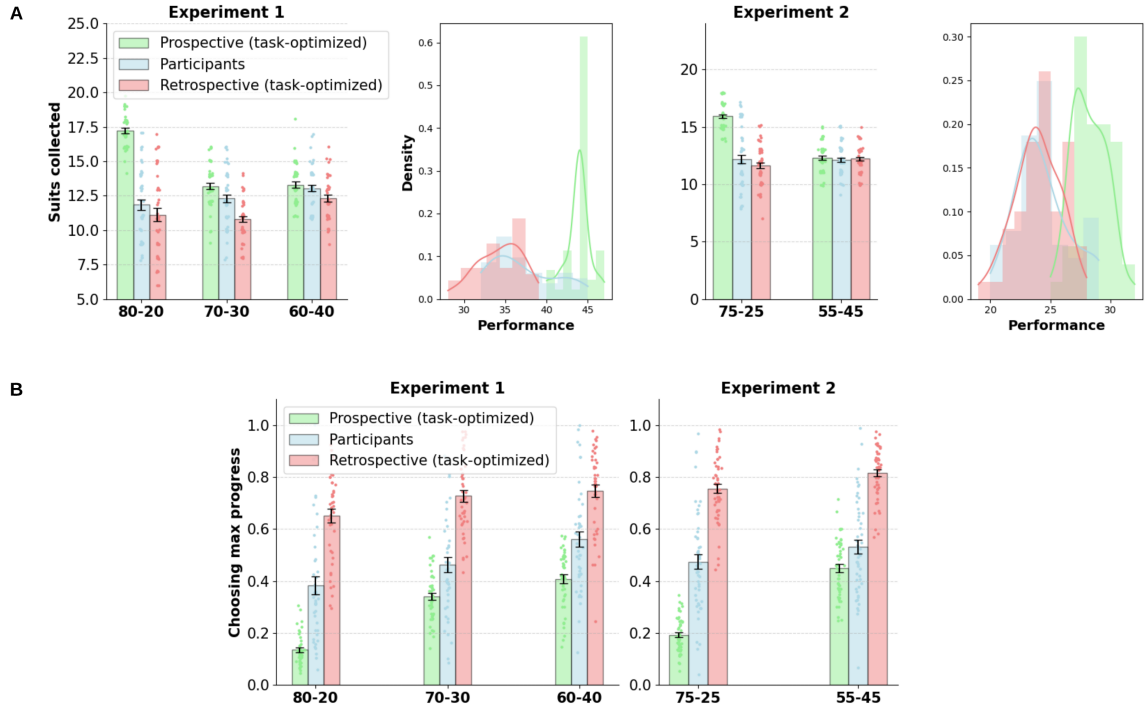


Figure 2.4: **A, B. Human performance contrasted against task-optimized prospective and retrospective agents.** Human behavior significantly falls short of optimal prospective agent and marginally beats retrospective agent in task performance (total number of suits collected). Agents' parameters are optimized for performance in the task. Performance is broken down by task condition and the distribution of performance is shown. **B. Preference towards accrued progress shown by the two agents contrasted with human participants.** The retrospective agent has the highest proportion of retrospective choice and the prospective agent the lowest, while human retrospective preferences lies between these extremes.

prospective agent. For further contrast, we also constructed a retrospective agent that always prefers the suit that made the most progress toward the target through the following value estimation process.

$$Q_g(s_g^{(t)}) = \gamma Q_g(s_g^{(t-1)} + 1) \quad (2.3)$$

The retrospective agent also applies a recursive estimate similar to that of the prospective agent to discount future rewards (with a factor γ), but without incorporating the belief state about the current suit probabilities (Equation 3). In effect, the agent always values the suit that is closest to the target.

We incorporated the same goal selection algorithm as that of the prospective agent

(see Computational Modeling section in Methods), and likewise, the parameters of the retrospective agent are optimized for task performance.

In general, human performance does not match the prospective agent in both experiments (Experiment 1: $t(86) = -9.98$, $p < 10^{-15}$; Experiment 2: $t(98) = -9.60$, $p < 10^{-15}$), as shown in Fig 4A. The contrast is particularly striking on conditions where there is a high disparity in the probabilities of token outcome (80-20 and 75-25 blocks). However, the performance contrast is not significant in the blocks where the benefit of prospecting is minimal (60-40 and 55-45 blocks). The prospective agent shows a significant interaction of performance with the block type (Experiment 1: $F(2, 86) = 49.85$, $p < 10^{-14}$; Experiment 2: $F(1, 49) = 147.8$, $p < 10^{-15}$), whereas human agents show a diminished yet significant interaction with block type in Experiment 1 ($F(2, 86) = 3.81$, $p < 0.05$) and no significant interaction in Experiment 2. Human performance beats the retrospective agent marginally in Experiment 1 ($t(86) = 3.28$, $p < 0.01$), while the difference in performance distributions is not significant in Experiment 2. Task performance mirrors the heterogeneity evident in participants' retrospective choices, with only a few participants attaining performance close to that of the optimal prospective agent.

Fig 4B compares the proportion of choice of the max progress option of the two agents with that of the human participants. The retrospective agent, in the spirit of its formulation, demonstrates the highest preference for accrued progress, whereas the prospective agent falls at the other extreme; human participants fall part-way between the two agents, indicating an inflation of retrospective choice beyond what is optimal for the task.

2.6 Effects of explicit instructions on persistence

Although we explicitly designed the block conditions so that there is a dominant suit, it is likely that participants did not anticipate how effective the pursuit of the prospective suit over the retrospective choice would be, thus reducing the exploration of alternative suits. If information regarding the abundance of the prospective suit is made explicit, will it lower retrospective influences on the task?

In Experiments 1 and 2, we simply alerted the participant when a block shift occurred. We now consider a variant of the first experiment, which we call Experiment 3, where we gave explicit instructions to participants about the contingencies within each block. Before the start of the game, we instructed the participants on the types of blocks present in the task and also alerted the participants regarding the

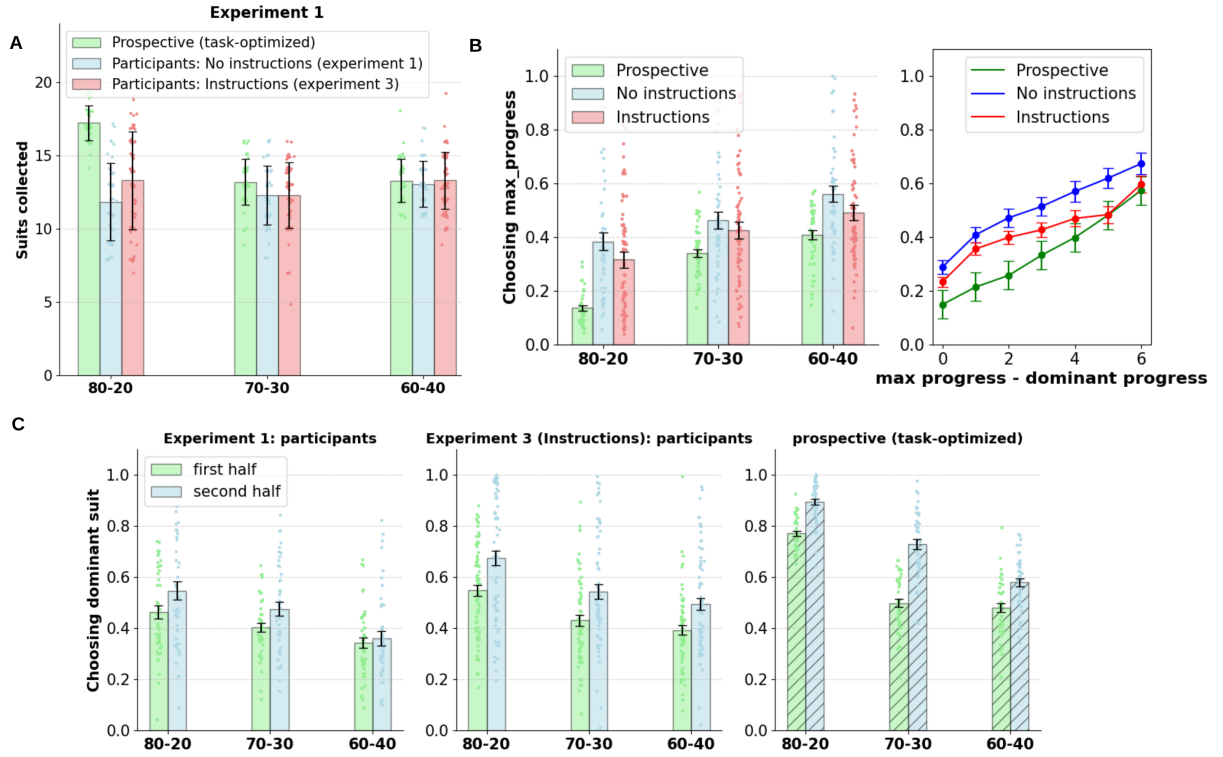


Figure 2.5: A, B. Effects on explicit instruction about prospects of alternative goals on behavior. A. Task performance improves with instructions yet significantly falls short of the optimal prospective agent. **B.** Retrospective choice proportion decreases with instructions yet is significantly inflated when contrasted with the prospective agent. **C. Learning effects within blocks.** Panel shows the dominant suit choice probability in the first and second halves of the block (15 rounds each) in the presence and absence of explicit instructions. Preference towards the dominant choice increases in both halves of the blocks indicating a shift towards optimal goal selection in the presence of instructions.

available contingencies at the beginning of each block. For example, we told them that they are in an 80-20 block condition. We nevertheless refrained from telling the participants which specific suit is the dominant one in a block to avoid explicitly priming a goal switch.

Fig 5A plots the performance of the task with explicit contingency information alongside the variant without contingency information, compared to the prospective agent. Task performance improved with contingency (independent samples t-test: $t(106) = -2.00$, $p < 0.05$) - evident in the 80-20 block condition - yet it is still significantly short of the prospective agent.

Fig 5B shows the contrast in the choice proportion of the max progress suit found in versions with and without instructions. Explicit instructions reduced the retrospec-

tive influence on goal selection, yet did not eliminate its effect relative to an optimal prospective agent.

Fig 5C contrasts the proportion of choice of dominant suit in the first and second halves of the blocks in the version with instructions (Experiment 3), without instructions (Experiment 1), and the prospective agent optimized for the task. Although the proportion of dominant choices increased in the version with instructions, it remained significantly lower than the task-optimized prospective agent, indicating suboptimal goal selection patterns in human participants even in the presence of explicit instructions on contextual goal prospects.

2.7 Goal momentum: a computational account of persistence

We previously described an optimal prospective agent as a benchmark for human performance. A potential account for suboptimal human performance could be that humans incorporate token-outcome contingencies of the suits inefficiently, or alternatively, humans could be using an entirely different mechanism to estimate the likelihood of achieving future goals.

As an alternative account, we formalized momentum as a measure of the goal value that drives persistence and is defined as the product of progress made toward the goal and its speed toward the goal.

$$\text{goal momentum} = \text{progress accrued} \times \text{current rate of reinforcement} \quad (2.4)$$

In the case of multiple goals with different rewards, the above value can be scaled appropriately with the associated rewards.

This equation is similar in essence to the construct of ‘momentum’, defined as the product of mass and velocity in Newtonian physics, which explains the resistance of an object to change. The larger the momentum, the greater its resistance to change in value. The analogy of goal momentum with the physical interpretation of momentum goes only as far as it comprises two factors: accrued progress (analogous to behavioral mass) and the current rate of progress (analogous to velocity), which together create resistance to change in valuation. A longer history of reinforcement with an option (accumulated behavioral mass) and a higher current rate of reinforcement both increase momentum and facilitate persistence with an option. This analogy was previously used in the definition of behavioral momentum [137], where the rate at which the free operant responds in the presence of a stimulus is likened

to the velocity of a moving body, and the resistance to change of the response is likened to its inertial mass. We further extend this metaphor to include extended goals, where the momentum of the goal in reaching its desired setpoint determines its value.

Momentum, in the purview of goal pursuit, is an estimate of the drive toward a specific goal that is dependent on both the history of reinforcement (previous progress) and the current rate of reinforcement. We define the progress accrued in a fractional format (say 40%) that facilitates a relative comparison between goals with distinct objectives. Formulated in this way, the progress accrued is also the inverse of proximity to the target. From this perspective, goals can be viewed as setpoints in the distant future, and the value of the goal is its momentum in reaching the setpoint.

With the above definition, our aim was to capture the two effects that we found in behavior. As progress accrued in a suit increases, the resistance to switch away from the suit is higher, and as the speed of progress reduces (as a result of shifting token contingencies), the resistance to switching is lowered. Finding parallels to the definition of physics, we equate the progress accrued as a measure of mass causing inertial effects, in line with the endowment effects seen with physical objects in behavioral economics [92]. While progress accrued increases inertia in switching, a sufficiently lowered speed of progress can lower momentum, and thus facilitate switching away from the option.

TD-momentum

We use the aforementioned conceptualization of momentum, a product of progress accrued and the speed of progress, as a measure of the goal value.

$$Q_g(s_g^{(t)} = m^{(t)}) = v_g^{(t)} m^{(t)} + b_g^{(t)} \quad (2.5)$$

where Q_g is the estimate of the goal value of a suit g in a specific slot configuration state $s_g^{(t)}$, m is the fractional progress made toward the goal in a given instant t , v_g is the parameter for the current rate of progress and b_g is a goal-specific bias for g .

We formulate an algorithm, TD-momentum, to facilitate updating of momentum estimates for different goals. Here, the product of the progress accrued and the rate of progress serve as a function approximator of the goal value, with speed as

a free parameter and progress as an indicator of the state of the goal [170]. The estimation of speed is done through the temporal difference learning rule, with its value updated through gradient descent on the prediction error of the expectation of receiving a token in the next round (and progressing one step ahead).

The temporal difference error (δ) is estimated as follows:

$$\delta = \gamma Q_g(s_g^{(t+1)}) - Q_g(s_g^{(t)}) \quad (2.6)$$

The TD momentum parameters v_g and b_g are updated with a learning rate α through a gradient descent in the value function Q_g .

$$v_g^{(t+1)} = v_g^{(t)} + \alpha \delta \frac{dQ_g^{(t)}}{dv_g^{(t)}} \quad (2.7)$$

$$= v_g^{(t)} + \alpha \delta m^{(t)} \quad (2.8)$$

$$b_g^{(t+1)} = b_g^{(t)} + \alpha \delta \quad (2.9)$$

The goal-specific bias term (b_g) is included to ensure that the goal value is updated even when progress is at $m = 0$. When one chooses the goal when its progress is zero and repeatedly fails to achieve the outcome, its value is depreciated through the bias term. The bias term also ensures that the goal value is retained when the goal target is accomplished (upon which the progress m resets to zero), and can increase the chance of choosing the same goal again even though the value component of the term $v_g m$ is zero.

Here are the key features of momentum computations for goal valuation.

1. **The relation between goal momentum and fractional progress is modulated by the discount factor, the step size of progress, and the current rate of progress**

If no unit progress is made in a round, $s_g^{(t+1)} = s_g^{(t)} = m$, δ is always negative for $\gamma < 1$.

If unit progress (η) is made in the next round, $s_g^{(t+1)} = m + \eta$. Assuming no goal-specific bias,

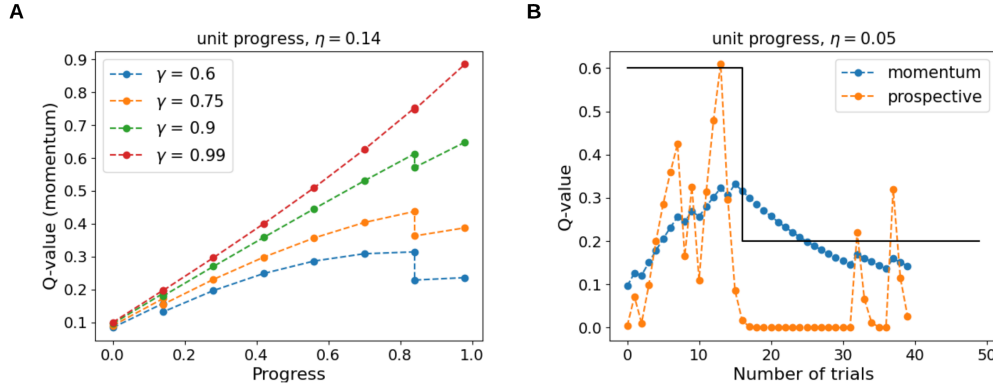


Figure 2.6: A, B. Features of the TD-momentum algorithm. **A.** The relation of TD-momentum value estimates with progress is determined by the discount factor, step size of progress, and the probability of making unit progress. In the simulations, the probability of unit progress is 0.6, and the step size of unit progress is 0.14. **B.** Comparison between the value computations from prospective and TD-momentum algorithms. The black line indicates the shift in probability of unit progress from 0.6 to 0.2 after 15 trials. TD-momentum provides stable value estimates, as opposed to the prospective model which exhibits fluctuations in response to individual reinforcements. Step size of unit progress is 0.05, learning rate is 0.4, and γ is 0.9 for both algorithms.

$$\delta = \gamma Q_g(s_g^{(t+1)}) - Q_g(s_g^{(t)}) \quad (2.10)$$

$$= \gamma(m + \eta)v_g - mv_g \quad (2.11)$$

$$> 0 \iff \gamma > \frac{m}{m + \eta} \quad (2.12)$$

With unit progress (η), prediction errors (δ) are positive if and only if the condition $\gamma > \frac{m}{m+\eta}$ is satisfied (Equation 12).

For example, given $\gamma = 0.8$, the unit progress from $m^{(t)} = 0.2$ to $m^{(t+1)} = 0.25$ will yield a positive prediction error, while the unit progress from $m^{(t)} = 0.9$ to $m^{(t+1)} = 0.95$ produces a negative prediction error. Therefore, for a fixed γ , a small enough unit of progress can result in negative speed updates when the fractional progress (m) is close to 1. This accounts for the phenomenon of diminishing marginal utility [142], whose main feature is that the marginal utility of a unit increase in a good decreases as its quantity increases.

When no progress is made in a round ($s_g^{(t+1)} = s_g^{(t)}$), $\delta = \gamma Q_g(s_g^{(t+1)}) - Q_g(s_g^{(t)}) < 0$, and updates for the speed of progress are always negative. On the other hand, if unit progress ($m \rightarrow m + \eta$) made in a round, the changes

to the speed term are positive, but only if $\gamma > \frac{m}{m+\eta}$. Thus, TD-momentum value estimates do not need to increase monotonically with progress; their relation to progress is determined by the discount factor γ , the step size of unit progress η , and the current estimate of the rate of progress (v_g).

Fig 6A demonstrates this feature of momentum computations. We simulated TD-momentum value estimates for several discount factors, with a step size of unit progress of 0.14, and probability of making unit progress as 0.6 (binary 0s and 1s drawn with probabilities 0.4 and 0.6, respectively). We note that for $\gamma < 0.9$, TD-momentum exhibits diminished value with increasing progress when the fractional progress is close to one.

However, for most participants in our task, the TD-momentum γ estimates through behavioral fits indicate a discount factor > 0.9 , so nonmonotonic behavior with increasing progress is not observed in the task. However, we expect that with sufficiently lowered step sizes of progress, the TD-momentum can predict diminished utility closer to goal completion.

2. **Goal momentum provides stable value estimates against fluctuations in reinforcements** We contrast the TD-momentum value estimates with that of the prospective agent in Fig 6B. In the simulation, we introduced a change in the probability of unit progress from 0.6 to 0.2 after 15 rounds of the task. We note that the value estimates from the prospective agent exhibit fluctuations corresponding to individual reinforcements of unit progress, whereas momentum builds up value steadily and gradually decays when encountered with a shift in probability. We believe that this retention of stable value estimates enables persistence beyond a context shift, which could facilitate sustained commitment toward long-term goals.

2.8 Alternative accounts

We constructed several alternative models to enable us to test several alternative hypotheses to the momentum explanation in accounting for human behavior on the task.

Perturbed prospection Firstly, we devised parameterized variants of the prospective agent that learns the token outcome probabilities inefficiently, employing both symmetric and asymmetric learning rates [59] for positive and negative prediction errors. We expected that a prospective agent acting on noisy/inertial/biased probabilities could account for suboptimality in choices. We experimented with several

variants of the prospective model and observed that the simple variant with symmetric learning rates captures the data better than the others (see S6 Fig and S1 Appendix).

Hybrid Secondly, we conceived a hybrid model that incorporates weighted contributions from prospective and retrospective agents, which could potentially account for retrospective bias in goal switching.

TD-persistence Lastly, we constructed an agent – TD persistence – that learns the token outcome expectations of each goal using a temporal difference learning rule and acts guided by those expectations, modulated by persistence parameters in action policy. As we instructed the participants that every block has different probabilities, a dominant strategy would be to estimate which goal has the highest token-outcome expectation and persevere with it until the end of the block. We expected the influence of this strategy to be greater in the experimental variant (Experiment 3), where we provided explicit instructions about the type of their current block. TD persistence, having an optimal performance similar to that of the prospective agent (see S7 Fig), is one of the better strategies in the task. However, by its very nature, it cannot reproduce either the temporal patterns of retrospective bias or the switching trends seen in behavior.

We incorporated a goal perseveration parameter in the common action selection policy (see Methods) — a choice kernel that increments each time a goal is selected and decays each time it is not — so model comparisons elicit differences beyond any choice persistence or putative “sunk cost” mechanisms.

We ensured that the competing hypotheses elucidated above are individually recoverable (see Methods on parameter and model recovery, also S4 and S5 Figs).

Comparison to simpler reinforcement learning models We also considered the possibility that participants did not play the game within the framework of extended goals, but rather selected goals by treating them as individual unconnected events in which one simply receives or does not receive a token reward. In such a scenario, behavior could be captured by a simple reinforcement learning (RL) model utilizing a standard delta learning rule such as in the trial-based Rescorla-Wagner model [1]. This simplified RL model could potentially predict an increase in retrospective

choice due to prior reinforcements increasing expected values of tokens. However, participants were alerted during block shifts to ensure minimal carryover of such expectations of token probabilities. Nevertheless, to rule out this possibility, we fit participants’ choices to this simplified RL model [177] (while also including a choice kernel), and we generated simulated data from the fitted parameters (see Methods). The model-generated data capture neither the preference for accrued progress nor the sensitivity to relative progress differences between the suits (see S13 Fig). That is to be expected, because this simplified RL model has no notion of commitment toward extended goals, supporting our presumption that the reinforcement schedule imposed by extended goals inflates retrospective choice over and beyond the predictions of such a simplified RL model.

Time-inconsistent discounting We employed exponential discounting to design the task-optimized prospective agent. However, given that human behavior is better accounted for by hyperbolic discounting [52], we also implemented a prospective agent with hyperbolic discounting. We estimated trials to completion (delay to reward D) for each goal by rolling out the results into the future and used the hyperbolic discounting factor ($\gamma = \frac{1}{1+kD}$, $k \in [0, 1]$) to discount the rewards. However, we found no significant difference in behavioral fits between the variants of the prospective agent with exponential and hyperbolic discounting (see S6 Fig).

2.9 Model comparisons

The TD-momentum model offers a better fit to human choice data

We fit each of the above models to the participants’ choices of stay/ switch choices from actions and goals reported from the previous round and compared the models using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) metrics.

$$\text{AIC} = 2k - 2 \ln(\hat{L})$$

$$\text{BIC} = k \ln(n) - 2 \ln(\hat{L})$$

Here, n is the number of data points, k is the number of parameters in the model, and $\hat{L}$ is the maximum value of the likelihood function. The lower the AIC/BIC metrics, the better the fit of the model. Moreover, we computed a 3-fold cross-validation (CV) metric for each participant by fitting the model with two-thirds of the task blocks and computing the negative log-likelihood on the held out one-third data (mean

metric computed over 30 random splits of the data). Figs 7A and 7B show that the TD-momentum offers a better account of the data according to the AIC, BIC, and cross-validation metrics. Beyond just aggregate likelihood comparisons between models, we examined whether TD-momentum provided superior fits consistently across individual subjects [59]. We performed paired sample t-tests between AIC, BIC, and CV metrics, and TD-momentum consistently offers better fits between participants in experiments 1 and 2 compared to other models.

Fig 7C shows a comparison between the behavioral predictions of the different models in the presence of explicit instructions on the types of contingency (experiment 3), contrasting with that of the variant without instructions (experiment 1). In both versions, TD-momentum offers a better fit to the data. However, we observed an increased contribution of TD-persistence at the level of individual participants in the version with explicit instructions. Fig 7D shows the difference in BIC between the two algorithms in the individual participants who contrast the two variants. In experiment 1, TD-momentum offers a better fit for 72% of the participant pool (loglikelihood ratio test with $p = 0.05$), whereas the number drops to 58% in experiment 3. We predicted that 42% of participants who were better accounted for by the TD-persistence model in the version with instructions would perform better on the task than those of participants accounted for by the TD-momentum model. In fact, there was a significant difference in performance between the two groups (independent samples t-test: $t(63) = -3.69$, $p < 10^{-3}$). Retrospective bias patterns in the two participant groups (TD-momentum and TD-persistence) follow the same patterns as those obtained from simulations of respective algorithms (see S8 Fig).

The TD-momentum model reproduces the behavioral effects of the aggregated participant pool

Using the parameters derived from individual participant fits of TD-momentum, we generated synthetic data imitating the participant pool to validate whether the model can replicate key behavioral effects that capture retrospective influences in experiments 1 and 2. One sample of synthetic data per real participant is generated using the fitted parameters, and the aggregate choice patterns of the population of model-simulated participants are contrasted with those of the real participants. Simulated participant data reproduces interaction effects of preference for accrued progress with the block type and also captures the pattern of increase in preference for accrued progress with increasing progress difference between the maximum progress suit and the dominant suit (see Figs 8A and 8B). The simulated data also

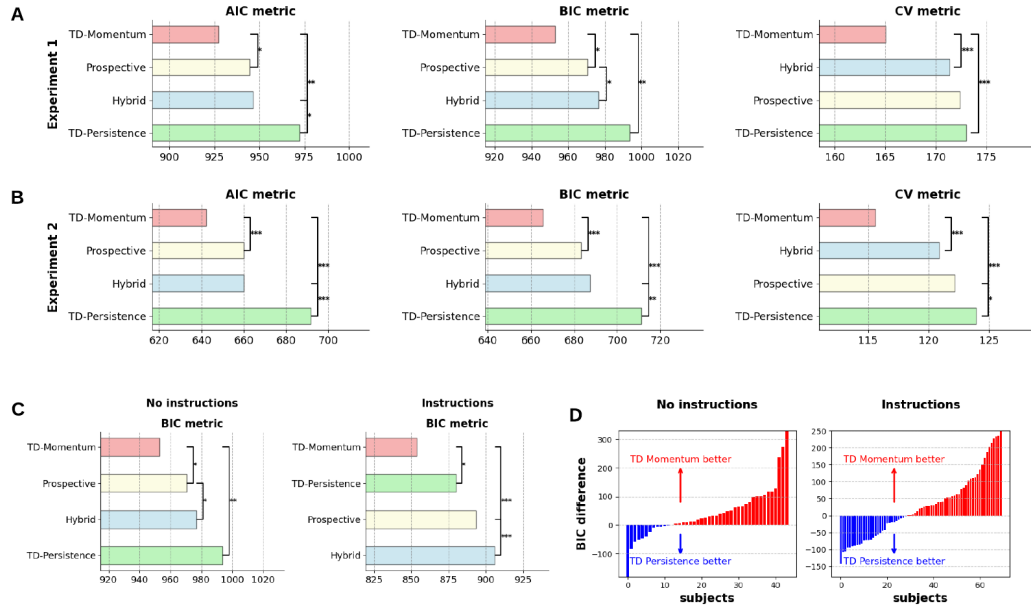


Figure 2.7: A, B. Model comparisons. TD-momentum accounts for the switching choices in the task better than the other prospective, hybrid, and TD-persistence models as per AIC, BIC, and 3-fold cross validation metrics in experiments 1 and 2. Mean and standard error bars of the metrics shown in the figure. Significance markers correspond to two-tailed paired sample t -tests ($*$: $p < 0.01$). **C.** TD-momentum also offers the best account of behavior in the variant with explicit instructions. **D.** Incidence of the competing hypothesis of TD-persistence increases in the version with instructions.

captures the trend of greater persistence in favor of the retrospective suit over the prospective one in switching patterns (see Figs 8C and 8D). The prospective and hybrid agent also captures the behavioral trends reasonably well (see S10 and S11 Figs), while the TD-persistence agent fails to capture them (see S12 Fig).

We also plotted how individually fitted model parameters correlate with participants' task measures. The momentum learning rate correlates with the performance in the task (Experiment 1: $R=0.78$, $p < 10^{-9}$; Experiment 2: $R=0.45$, $p < 0.0008$; Experiment 3: $R = 0.68$, $p < 10^{-10}$), that is, participants who exhibited greater learning of the speed of progress of different goals collected more suits in the task. The cost of switching is correlated with the number of action switches made by the participant (Experiment 1: $R=-0.78$, $p < 10^{-9}$; Experiment 2: $R=-0.81$, $p < 10^{-12}$; Experiment 3: $R = -0.62$, $p < 0.001$), that is, participants with lower switching costs made more switches under all conditions. Lastly, the softmax temperature parameter correlates with the preference for accrued progress (Experiment 1: $R=0.73$, $p <$

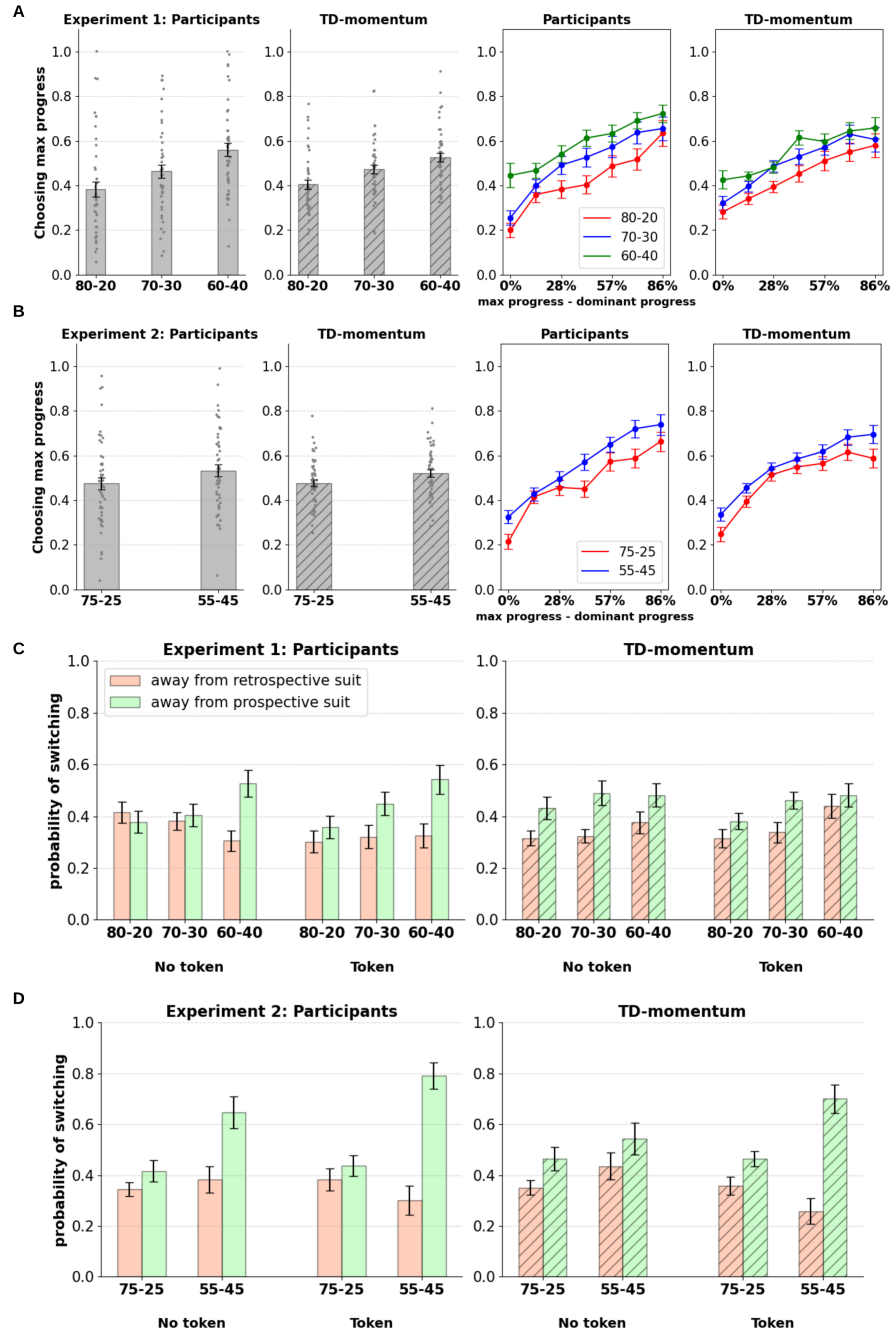


Figure 2.8: **TD-momentum: model predictions** TD-momentum agent captures general trends in preferences towards maximum progress suit and switching patterns in the task.

10^{-7} ; Experiment 2: $R=0.75$, $p < 10^{-10}$; Experiment 3: $R = 0.60$, $p < 10^{-7}$), that is, the more the participant relies on the value disparity between the goals generated by the momentum model, the greater is their preference for accrued progress. TD-momentum also reproduced participants' learning effects within a block and captured the increase in dominant suit choice proportions as the block progresses (see S9 Fig).

2.10 Dissociating prospective and momentum computations

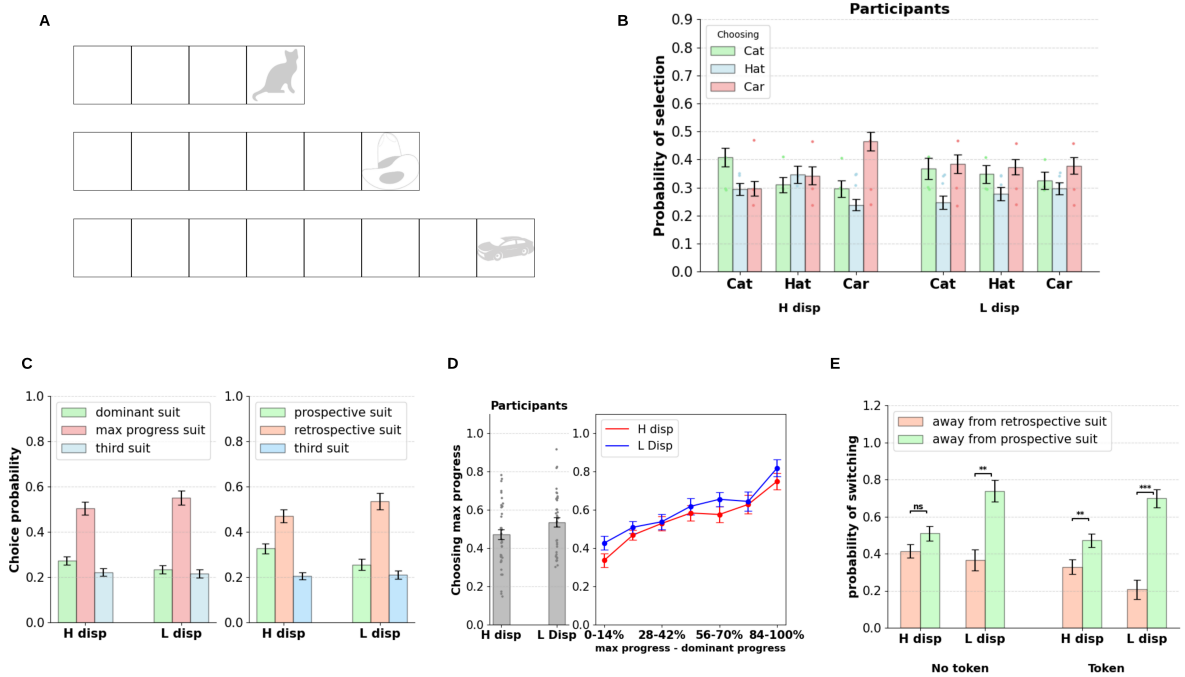


Figure 2.9: Experiment 4: Variable targets for goals **A.** Slot configuration for experiment 4. **B.** Suit selections divided by high disparity (H disp) and low disparity (L disp) conditions and sub divided by the dominant token type (cat/hat/car). **C, D, E.** Behavioral signatures (suit selections, retrospective choice, switching patterns) replicated from experiments 1 and 2. **Figure credits.** We used images from <https://openclipart.org/> to generate this figure. See the Task Design section in Methods for full image credits.

Although the TD-momentum model outperforms all other models in explaining the likelihood of participants' choices in the task, three of the agents, TD-momentum, prospective, and hybrid, qualitatively reproduce key behavioral signatures in the task. In order to more definitively differentiate between which of these models best explains human behavior, we turned to an additional experiment.

A key distinction between the TD-momentum and the prospective agent is that the former relies on a measure of fractional progress, and the latter relies on the absolute

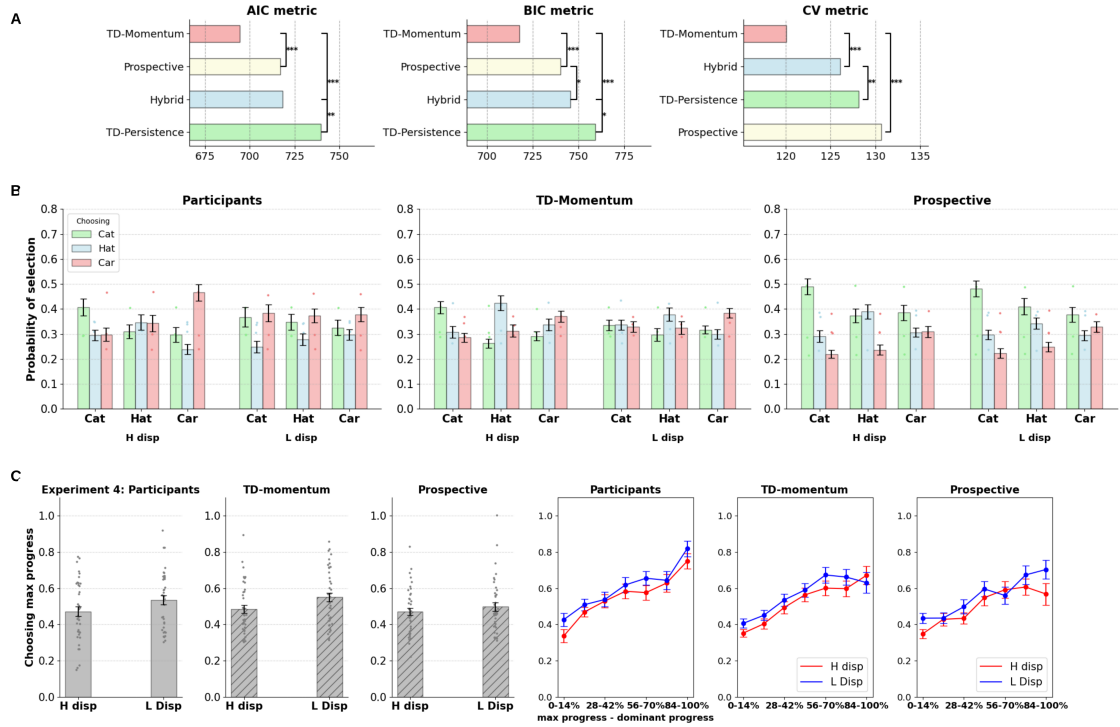


Figure 2.10: **Experiment 4: Model comparisons** **A.** TD-momentum outperforms other models in AIC, BIC, cross validation metrics. **B.** Prospective agent shows bias towards the cat suit, while TD-momentum closely captures distribution of goal selections. **C.** TD-momentum captures retrospective choice patterns better than the prospective agent.

distance to the target. In experiments 1 and 2, both measures converge, as targets for all suits are the same with no distinctions between fractional progress and absolute proximity to the target.

In experiment 4, we introduce variable target token lengths for the suits. The suit *cat* can be completed by collecting 4 cat tokens while the suits *hat* and *car* can be completed by collecting 6 and 8 tokens, respectively. Thus, if the *cat* suit has 2/4 tokens completed and the *car* suit has 6/8 tokens completed, the prospective agent has to unroll two steps ahead to predict rewards for both suits, however, TD-momentum considers that the *cat* suit is 50% done and that the *car* suit is 75% done. Thus, this design emphasizes the distinctions between the two algorithms.

Fig 9A shows the slot configuration for experiment 4. We adjusted the token probabilities so that there is one dominant prospective suit in each block and so that the dominant prospective suit is counterbalanced across suits. We introduced high- and low- disparity block conditions (*H disp* and *L disp*). In both conditions, the dominant suit takes 8 rounds (on average) to complete, while in the *H disp* and

L disp conditions, the two inferior tokens take 15 and 12 rounds (on average) to complete. Fig 9B shows the participants' goal selection probabilities divided by the dominant suit condition. The participants selected the *cat* suit with the highest chance in the *cat* dominant blocks, etc. for the other suits, indicating sensitivity to token contingencies. Participants also showed sensitivity to relative distinctions between suits with a lower proportion of dominant suit choice in the low-disparity blocks compared to the high-disparity blocks (Fig 9C). Moreover, goal selections are uniform with slight biases towards specific suits on account of them having variable-length targets. For example, participants selected the *hat* suit significantly lower than the suits *cat* or *car* in the low disparity condition, perhaps accounted for by preferences for the options at extremes (in terms of target length). Figs 9C, 9D, and 9E show behavioral signatures from experiments 1 and 2 replicated in the new task.

Model comparisons for all models are shown in Fig 10. TD momentum provides the best account of participants' choices in the task using AIC, BIC, and cross-validation measures, as shown in Fig 10A. Figs 10B and 10C show behavioral signatures derived from model-generated datasets contrasted with participants' behavior. The suit selections in Fig 10B show an interesting distinction between TD-momentum and prospective agents. The prospective agent prefers the *cat* suit more than other suits across all conditions, as it has the smallest target of all suits and thus the lowest absolute distance to the target compared to the others. The hybrid agent, despite accounting for retrospective influences, shows the same pattern of preferences as that of the prospective agent (see S14 Fig). However, an optimal prospective agent shows no such preference for the smallest target suit as it perfectly incorporates token probabilities while estimating future prospects (see S15 Fig). The prospective agent fitted to the behavior of the participants shows the bias that indicates the inefficacy of the model in predicting the goal selections of the participants. Human participants and the TD-momentum model fitted to behavior do not show such a bias. Likewise, in Fig 10C, the TD-momentum model shows retrospective choice patterns that more closely resemble human behavior relative to the prospective agent. The TD-momentum model did not fully capture the specific bias against the *hat* suit displayed in the *L disp* condition; nevertheless, its distribution of goal selections appears to be closer to the pattern found in the choices of human participants. Finer deviations in goal selection patterns between the TD-momentum model and human behavior can potentially be attributed to idiosyncratic goal selection patterns in humans that are not factored into the model. The deviations could also arise from

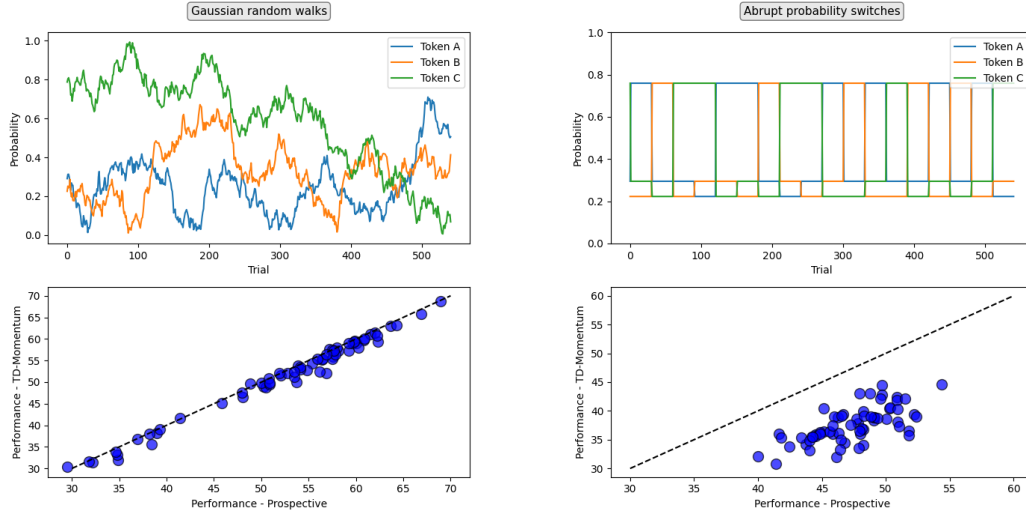


Figure 2.11: **Normative account of momentum in goal pursuit.** Optimal performance attained by the TD-momentum agent contrasted with that of the prospective agent when the token-outcome probabilities change in Gaussian random walks (left) and when they change in abrupt probability reversals (right). Each point in the plot refers to a variant of the task with randomly generated token contingencies; a sample task design was illustrated for both random walks and probability reversals.

systematic biases (such as a preference for options at the extremes), which could be included in the model as inductive bias components.

2.11 Normative account of momentum

We provided evidence that incorporating the momentum of a goal as a proxy for its value can account for suboptimal switching and undue retrospective valuation in temporally-extended goal pursuit. However, is momentum-based decision making necessarily maladaptive?

Evidently, there are scenarios where reliance on momentum as a measure of goal value is suboptimal. When a goal with a considerable history of progress is abruptly decreased in value, its drop in prospective value is drastic, however, its momentum is sustained for a period beyond its decrease in value, whereupon reliance on momentum leads to suboptimal persistence. Conversely, when alternative goals are enhanced in value, their prospects are suppressed as their momentum is lower because of their lack of history.

If we revert to the interpretation of the term as it is used to describe the physical properties of objects, it is rare that we observe such drastic changes in momentum in a world where continuity is the norm. Consider the momentum of a car driving on a freeway, which is a continuous variable with occasional rises and dips; only in

rare cases of collision do we notice an abrupt drop in momentum.

In a framework of continuity, where the recent past greatly predicts the near future, the momentum model offers a more efficient mode of computation and thereby an adaptive advantage. TD-momentum and the prospective agent both have identical parametric complexity (6 free parameters), yet their algorithmic complexities are different. Prospection relies on expensive model-based computations — keeping track of token-outcome expectations of each goal and recursively extrapolating them into the future. Momentum, on the other hand, maintains a model-free account of the past and uses it to extrapolate it to the future. The momentum model, without incorporating task contingencies, leads to suboptimalities, especially when there are jumps and discontinuities in prospects, as was the case in our task. The optimal performance attained by the TD momentum in the task is significantly lower than that of the prospective agent (see S7 Fig).

To test this account further, we simulated several variants of the task in which the prospects of the three goals undergo gradual shifts through Gaussian random walks. We estimated the optimal performance achieved in each of these variants by both modes of decision-making — momentum and prospection — and found that there is little difference in performance between the two algorithms (see Fig 11). However, when sudden reversals of the option probabilities are incorporated, the momentum model falls behind the prospective agent in performance.

Momentum computations can be efficient, yet in scenarios with sudden upheavals of fortunes, as is common in quotidian experiences, they can lead to suboptimal decisions. In addition, momentum computations and concomitant goal persistence can have adaptive features when there is uncertainty in the environment. Even when an option is devalued, one might be uncertain when it might be valued again, or whether viability of alternative options would remain stable enough to warrant a switch. Furthermore, persistence in goals can have other advantages in social settings, such as stable inference of one’s goals by peers [63], [32], all the while avoiding the inefficiency of incurring cognitive costs of goal switching.

2.12 Discussion

Using a novel paradigm in which participants are free to pursue any of three temporally extended goals, we observed that participants exhibited a strong bias toward retrospectively persisting with goals with greater accumulated progress at the expense of goals with higher prospective benefits. The nature of this bias, which we

call a retrospective bias — was highly heterogeneous in our pool of participants, with individuals manifesting varying degrees of retrospective influence on the selection of extended goals.

We manipulated block contingencies in the task so that there is a dominant goal that is most favorable to pursue in a given block that is not viable to pursue once the block shifted. This manipulation created situations in which participants made partial progress toward specific goals only to discover that they were no longer feasible shortly afterward. The optimal strategy would be to switch to an alternative option. However, we observed that participants demonstrated undue resistance in making such switches, which compromised performance on the task. Persistence towards a goal increased with an increase in accumulated progress. Participants also demonstrated sensitivity to prospective values, with the proportion of retrospective choice diminishing with an increase in the disparity of prospective values between options.

We conducted a variant of the task giving explicit instructions on the contextual prospects of different goals. By alerting the participants about the probability structure within a block, we expected that the inference of the identity of the dominant suit would be relatively easier and more salient. We noted that such explicit instructions served to diminish retrospective influences. However, they remained potent in driving goal selection. In this variant, participants continued to show suboptimal persistence toward the goals with greater accumulated progress despite having explicit knowledge about the prospects of alternative goals.

We explain this phenomenon by turning to the metaphor of momentum, aided by the simple fact that the momentum of a goal changes more slowly than its expected value. Momentum – a variable derived from physics defined as a product of mass and velocity — changes its magnitude with changes in velocity. Inertial mass acts as a moderating factor, tempering the impact of velocity on momentum. In this context, we draw a parallel by associating the velocity variable with the speed of progress toward the target and the mass variable with the prior progress accumulated in the task. When the accrued progress is substantial, any variation in the progress speed has a minor impact on the momentum’s magnitude compared to when the progress is minimal. If participants prioritize the momentum of the goal over its expected value, this preference could explain their increased persistence towards the goals that have made significant progress towards the target. To give a computational account of this phenomenon, we introduced a novel TD-momentum algorithm that adjusts

participants' momentum expectations for various goals using a temporal difference learning rule.

We compared the TD momentum algorithm with alternative accounts, including an inefficient prospective agent, a hybrid of prospective and retrospective agents, and a simple persistence-guided reinforcement learning of goal expectations. We hypothesized that momentum lends a better account of human persistence in goal valuation, as being driven by retrospective influences of past reinforcements accrued in goal pursuit, extending beyond mere choice repetition or sunk-cost considerations. In fact, through our modeling, we demonstrate that incorporating momentum as a measure of goal value best accounts for behavior in the task better than its alternatives. This computational model outperforms alternative hypotheses in explaining the choice data and reproduces several behavioral patterns at both the population and individual levels.

The phenomenon of persistence in retrospective choice goes beyond task-switching costs or sunk costs, which are also potential drivers of goal persistence. Participants faced no explicit penalty for switching and can retain progress made in goals until the game's end and revisit them, hence the choice between stay or quit does not arise, except perhaps at the very end of the task. In addition, participants frequently revisited retrospective choices throughout the task, despite having switched away from them, indicating persistent goal valuation that goes beyond task switching costs. Nevertheless, to explain away persistence generated by task switching and/or sunk costs in a more formal way, we also included a choice kernel for goals in each of the computational models that we tested, whereby the tendency to stick with a particular goal is increased the more that goal is chosen in previous trials, acting as a proxy for a cumulative task switching effect. Our model fitting results show that momentum is critical to capture behavior above and beyond that explained by a cumulative task-switching cost effect. Thus, while our data do not rule out the contribution of one or another of these phenomena to goal selection and switching behavior in general, our findings show that such accounts are not sufficient to account for the bias toward retrospective choices we observed.

Our study goes beyond existing work on human goal commitment by providing a computational basis of overpersistence. [81] observed patterns of overpersistence during the pursuit of extended goals in a low-cost setting. However, their best-fitting model was not designed to account for the persistence biases displayed by the participants in the task. Instead, they quantified an individual persistence bias

factor measured as a deviation from their prescribed goal selection model. Our model provides a plausible mechanistic account for participants' persistence patterns during goal selection. In support of this claim, the model parameters we fit to individual participants were found to correlate with the extent of individual retrospective preferences manifested in the task. Hence, our momentum model provides a computational account of human persistence in goal selection.

Although the momentum-based model accounts for cognitive biases manifested in the pursuit of temporally extended goals, it begs the question of why humans employ a suboptimal algorithm in pursuing extended goals. We also provide a normative explanation for the role of momentum in the pursuit of goals, indicating through simulations that such a strategy is not always maladaptive. In situations involving gradual shifts in the viability of the goal, interpolating future value through momentum is parsimonious, with little depreciation in expected rewards compared to an optimal strategy. However, when shifts in goal viability are more abrupt, as is our task, the momentum-based strategy of goal selection can lead to suboptimal persistence with goals of high retrospective value. Moreover, other adaptive effects of persistence in partially observed environments with uncertainty in goal prospects could also favor momentum computations.

The approach we have developed here can be applied to the broader question of how humans adaptively self-regulate, since the ability to disengage from goals that are unattainable in the current context is a hallmark of self-regulation [23], [181]. An inability to disengage from goals is routinely observed in healthy individuals, with more pronounced effects seen in psychiatric conditions such as impulsivity [79], [70], depression [39], [104], attention deficit hyperactivity disorder [15], and generalized anxiety disorder [38], [124]. The present findings have potential relevance for understanding the failures of behavioral persistence seen in psychiatric disorders.

There are several limitations of the present work. One is that we introduced a number of simplifying assumptions into our analysis which, when discarded, could open up a realm of other potential algorithmic accounts. For example, in formulating the prospective agent, we assumed that the agent optimizes for immediate rewards delivered upon suit completion. However, it is possible that humans could employ subroutines that are pre-planned according to the selection of hierarchical options [163], such as committing to a goal for a fixed set of rounds. We also observed significant heterogeneity in human goal selection with evidence of the prevalence

of other strategies (such as the TD-persistence model) in some individuals. Furthermore, we found evidence that instructions can moderate the influence of momentum computations on goal selection behavior. Thus, the momentum model serves as a coarse descriptor of human goal selection behavior, but not necessarily as a fully comprehensive account for the goal selection tendencies of each individual.

In conclusion, the present work provides insight into the computational mechanisms underlying flexible goal pursuit. Characterizing how humans assign subjective value to goals over time and determining how that assignment is influenced by multiple factors, such as trial history, context, reinforcement rates, and the presence of alternative goals, is a crucial first step toward understanding how humans set and commit to goals in everyday decision-making contexts. Such insights can ultimately be used to implement novel behavioral interventions and nudges to improve the efficiency of goal-directed decision making within and between individuals.

2.13 Methods

Ethics statement

The research described here was reviewed by the Caltech Institutional Review Board and deemed exempt from ongoing in-depth review (IR18-0796) due to the fact that the data were collected anonymously through the online platform Prolific and because the research was considered of no more than minimal risk. All participants provided consent for their online participation by ticking a box on an online consent form, but did not sign the consent form, as the data collection is anonymized.

Experimental protocol

Task Design

We have three different suits in the game: *cat*, *hat*, *car*. There are six cards, two for each suit. The mappings are as follows: the *car* suit is assigned to *mat* and *basket* cards; the *hat* is mapped to *wardrobe* and *table* cards; the *car* is mapped to *key* and *baggage* cards. We inform the participants about the above mapping at the beginning of the game. We chose the card symbols to be semantically related to the token types and we conveyed those semantic descriptions to the participant for their ease of retrieval (for instance, “places where you would like to keep a *hat*”). We further tested participants’ knowledge of card-token mappings through a quiz before the game began. The participants continued with the game only if they answered all the questions correctly; if not, they were sent to the beginning of the instructions for

revision.

In each round of the game, we presented the participants with three cards (one for each suit type). The exact card displayed for a particular token type is shuffled across the rounds. We also shuffled the locations of the cards on the screen at random every round to prevent any mappings between goals and action key presses.

Throughout the game, cards serve to stand in for the suits, and to collect tokens for a specific suit, participants need to flip the cards associated with the suit. Through this design, our objective was to eliminate the stimulus-related value mapping, which would likely emerge if we were to directly display the suit stimuli (say images of *cat*, *hat*, *car*) and to elicit the construct of goals in participants. If, say, a participant persists in the goal of pursuing the *cat* suit, they would choose the relevant cards, the *basket* card when it is available and the *mat* card when it is not. By making the associated cards appear at random, we mitigated stimulus-elicited persistence effects in our design.

In experiment 1, we designed three block conditions: 80-20, 70-30, and 60-40. There are 18 blocks in the game, 6 of each type. Each block has 30 rounds, and the game lasts 540 rounds in total. The block conditions are shuffled throughout the game. However, we ensured that the dominant suit is flipped across adjacent blocks, so that no two adjacent blocks have the same token with the highest availability. As our primary focus lay in investigating flexible goal switching across contexts, we sought to induce conditions where switching is optimal.

In experiment 2, there are two block conditions: 75-25, 55-45. There are 12 blocks in the game (6 for each type), so 360 rounds in all.

For experiments 1 and 2, we instructed the participants that suit probabilities are constant within a block but shift across blocks. We included a quiz during the instructions to ensure that they understand block switches and differing suit probabilities. We also emphasized during the instructions that the probabilities are associated with the suit types and that picking either card related to the suit gives the token with the same probability. This is to prevent participants from tracking the value associated with specific card stimuli (whether a specific card gives the token more often) and to prompt them to place values on the goals (suit types) instead. In the experiment, block transitions are signaled with a screen that displays the text “This is the beginning of a new block” for 2 seconds. We also added a notation indicating the current block number out of the total number of blocks (say, block

13/18). We informed the participants that each block lasts 30 rounds and indicated the round number at the beginning of each round (say, round 20).

In experiment 3, a variant of experiment 1, we provided explicit instructions on the current token contingency faced by the participants. At the start of each block, we specified the type of block (80-20/ 70-30/ 60-40) without explicitly informing which token has the highest availability. During signaling the start of a new block, we also indicated that it is a specific type of block (say 80-20). We designed this as a nudge to encourage flexible goal adaptation across blocks.

In experiment 4, we assigned different targets to different suits. The cat suit can be finished by accruing 4 cat tokens, while the hat and car suits can be finished by collecting 6 and 8 tokens, respectively. Dominant suits are counterbalanced in all token types, and token probabilities are adjusted accordingly. We constructed two block conditions, high disparity and low disparity blocks (*H disp*, *L disp*) where the inferior token takes 15 and 12 rounds (on average) to finish. The dominant token takes 8 rounds (on average) to finish in all blocks.

We added a goal probe at the end of every third round, to elicit their target suits at different points throughout the game. To avoid a forced choice between three options, we also included an ‘undecided’ prompt.

Figs 1 and 9 in the manuscript were generated using images from <https://openclipart.org>.

1. <https://openclipart.org/image/800px/1198> (cat)
2. <https://openclipart.org/image/800px/159907> (hat),
3. <https://openclipart.org/image/800px/321286> (car),
4. <https://openclipart.org/image/800px/203250> (basket),
5. <https://openclipart.org/image/800px/153547> (mat),
6. <https://openclipart.org/image/800px/29276> (table),
7. <https://openclipart.org/image/800px/26415> (wardrobe),
8. <https://openclipart.org/image/800px/320342> (baggage),
9. <https://openclipart.org/image/800px/59329> (key).

Participants

Table 2 shows the details of the recruitment of the participants for all experiments. Participants were recruited from the Prolific on-line platform and compensated with a base pay of \$9 per hour, with a performance-based bonus of up to \$3. We instructed the participants that their performance bonus is directly proportional to the number of suits completed by the end of the game.

We did not exclude any participants from the study. Chronologically, we conducted experiment 3 as our first study, followed by experiments 2, 1, 4.

Table 2.2: Participant recruitment

	Participants	Median time (mins)	Age range (mean $\pm$ std)	Female/ Male
Experiment 1	44	31	18-64 (28.6 $\pm$ 9.5)	13/ 31
Experiment 2	49	24	20-68 (29.2 $\pm$ 8.8)	20/ 29
Experiment 3	67	33	19-52 (28.1 $\pm$ 7.6)	32/ 35
Experiment 4	39	26	19-57 (31.1 $\pm$ 10.0)	18/ 21

Computational modeling

Algorithm 1 Algorithm for goal selection

```

1:  $Q_g$ : goal value where  $g \in \{\text{cat, hat, car}\}$ 
2:  $\beta_g, \beta_a$ : softmax temperatures for goal and action selection
3:  $c_g$ : goal preservation/ choice kernel
4:  $c$ : cost of switching between goals
5:  $G$ : current goal (none at start)
6: for block in (1,  $\text{num\_blocks}$ ) do
7:   Initialize card probabilities
8:   for round in (1,  $\text{num\_rounds}$ ) do
9:     Get goal values ( $Q_{\text{cat}}, Q_{\text{hat}}, Q_{\text{car}}$ ) // Using TD-momentum/ prospective
10:    Calculate advantage of current goal  $G$ :
11:     $A_G = Q_G - \max_{g \neq G} Q_g$ 
12:    Choose to stay with/ switch from the current goal:
13:     $p_{\text{stay};G} = \frac{1}{1 + \exp(-c - c_G - \beta_g A_G)}$  // Choose goal
14:    In case of a switch, chose the new goal according to goal values:
15:     $\text{softmax}_{\beta_g}\{Q_{\text{alt}_1}, Q_{\text{alt}_2}\}$ 
16:    Update current goal  $G$ 
17:    Update choice kernel of goal selection
18:    Report the goal  $G$  every third round
19:    Choose to act according to the current goal or explore
20:     $p_{\text{stay};G,A} = \frac{1}{1 + \exp(-c - c_G - \beta_a A_G)}$  // Choose action
21:    In case of a switch, explore a new action according to goal values:
22:     $\text{softmax}_{\beta_a}\{Q_{\text{alt}_1}, Q_{\text{alt}_2}\}$ 
23:    Observe outcome
24:    Update goal-outcome probabilities // For prospective/ hybrid/
25:    TD-persistence
26:    Update goal values
27:   end for
28: end for

```

Goal probes

We elicited goal probes from the participant every third round of the task. We relied on goal probe estimates to account for model predictions of goal persistence that is different from action selection. To ensure reliability in goal probes, we measured goal-action congruence (whether the participant made an action that is congruent to the goal in the round following the goal probe prompt. S2A Fig shows the goal-action congruence density histograms plotted across all experiments. The mean congruence in the target probes is 80%, 78%, 76%, and 73% for experiments 1 to 4.

We also estimated the proportion of the total number of switches that the participant

made in the task that followed a goal probe. This is to verify whether the goal probe prompt itself influenced subsequent goal selection (either prompting a switch / prompting persistence in the specified probe). S2B Fig shows the mean proportion of switches are 31%, 32%, 29%, 32% for experiments 1 to 4. As the goal is probed every third round (in 33% of the rounds), the proportion of switches around the goal probes is in line with the chance expectation of 33%.

The congruence of goal action is high but not perfect in the participants. We anticipated this mismatch by our hierarchical assumption in our modeling, by dissociating participant goal and action selection. Participants can value a goal above other alternatives, but can temporarily explore other options before revisiting it. We found such patterns of revisiting in participants (Fig 3), indicating persistence in goal valuation. However, we still expect noise in goal probes in a portion of less-engaged participants (perhaps those showing low goal-action congruence) and that is a fundamental limitation in probing intrinsic goal states in participants. This noise could explain why the models do not capture the goal selection patterns in the task perfectly.

Prospective and retrospective models

Prospective The prospective model updates the belief states of the current suit probabilities and rolls out the expected outcomes in the future.

The estimate of the probability of token collection for a suit g can be computed using a delta learning rule.

$$M_g^{(t)} = M_g^{(t-1)} + \eta(I_g^{(t)} - M_g^{(t-1)}) \quad (2.13)$$

Using the probability of token outcome, the prospective agent rolls out expectations one step into the future, for two scenarios: receiving a token and moving one token closer towards the target, and not receiving the token and staying in the same state. The value estimate for a goal g in a specific configuration state of the slots s_g is calculated as follows:

$$Q_g(s_g^{(t)}) = M_g \gamma Q_g(s_g^{(t)} + 1) + (1 - M_g) \gamma Q_g(s_g^{(t)}) \quad (2.14)$$

where s_g is the current state of the goal (number of tokens collected) and γ is the temporal discounting factor that ensures sooner rewards are valued higher over later rewards. This algorithm prefers the suit that can be completed fastest in a given context.

Retrospective

The retrospective model values the goals with the greatest progress.

$$Q_g(s_g^{(t)}) = \gamma Q_g(s_g^{(t)} + 1) \quad (2.15)$$

As s_g is closer to the target, there is less discounting of rewards and therefore higher value. This formulation is chosen to ensure similar scaling and parameterization for prospective and retrospective models.

TD-momentum

Momentum in the context of goal valuation is defined as the product of the progress made and the speed of progress. Here, the learnable parameter is the speed of progress. The goal value is formulated as a function approximator for momentum as follows.

$$Q_g(s_g^{(t)} = m^{(t)}) = v_g^{(t)} m^{(t)} + b_g^{(t)} \quad (2.16)$$

where Q_g is the value of a given suit g in a specific slot configuration $s_g^{(t)}$, $m^{(t)}$ is the progress (fraction of slots filled) at a given instant t , v_g is the parameter for the speed of progress and b_g is a bias for the goal.

Depending on whether a token is received in a given round, the state of the goal is updated accordingly. The speed of progress is learned with the following temporal difference learning rule.

$$\delta = \gamma Q_g(s_g^{(t+1)}) - Q_g(s_g^{(t)}) \quad (2.17)$$

$$v_g^{(t+1)} = v_g^{(t)} + \alpha \delta \frac{dQ_g^{(t)}}{dv_g^{(t)}} = v_g^{(t)} + \alpha \delta m^{(t)} \quad (2.18)$$

$$b_g^{(t+1)} = b_g^{(t)} + \alpha \delta \quad (2.19)$$

We used a fixed initialization for $v_g = 0.5$ and $b_g = 0.1$ for all goals. We tested different initializations to assess their impact on the model fitting and found that they are not significantly impacted by the choice of function approximator initialization.

We ensured that all the parameters of the model are recoverable through parameter recovery analysis (see S4 Fig).

Alternative hypotheses

TD Persistence

Our task can be solved by reducing it to a three-armed bandit task and finding the optimal bandit in each block that gives you the most rewards and persists with it until the end of the block. We assumed that this might be a plausible strategy in the experimental variant where we clearly inform the participants of the token contingencies of the block.

The token-outcome expectations from each goal, calculated using a delta rule, drive the choices in this algorithm. Once the value of a certain token sufficiently exceeds its alternatives, the persistence parameters in the action policy (described later) ensure persistence towards the token.

$$M_g^{(t)} = M_g^{(t-1)} + \eta(I_g^{(t)} - M_g^{(t-1)}) \quad (2.20)$$

$$Q_g(s_g^{(t)}) = M_g^{(t)} \quad (2.21)$$

Hybrid

A potential mechanism to incorporate both prospective and retrospective influences is to assume a fixed weight arbitration model that weights contributions from both valuations and their combined influence drives decision making.

$$Q_g(s_g^{(t)}) = w * Q_g^{pros} + (1 - w) * Q_g^{retro} \quad (2.22)$$

Common mechanism of goal and action selection

All competing hypotheses of goal valuation followed a common action selection policy elucidated as follows.

Once the values of individual goals are computed, the advantage of the current goal, A_G , over its alternatives is

$$A_G = Q_G - \max_{c \neq G} Q_c \quad (2.23)$$

The agent chooses to stay on the current goal (target suit) with a probability $p_{\text{stay};G}$,

$$p_{\text{stay};G} = \frac{1}{1 + \exp(-c - c_G - \beta_g A_G)} \quad (2.24)$$

where c is the switching cost, β_g is the softmax temperature to scale the advantage. c_G is the choice kernel where commitment towards the goal increases the more it is chosen in the past as updated through the delta-rule,

$$c_G = (1 - \alpha_c)c_G + \alpha_c c_G I_G \quad (2.25)$$

where I_G is the indicator that the goal is chosen in the round.

When the agent decides to switch away from the current goal with probability $(1 - p_{\text{stay}})$, it selects a new goal from the two alternate goals depending on the values of the two alternate goals ($\text{softmax}(Q_{\text{alt}_1}, Q_{\text{alt}_2})$)

Once the goal is chosen, the participant chooses to perform an action in accordance with the goal (choosing the cards) with a probability, $p_{\text{stay};G,A}$, which is modulated by a separate softmax term β_a .

We chose separate softmax terms for actions and goals to allow goals to change at a different rate than actions. Participants can switch actions to explore other offers while staying on the same high level goal, and can choose to switch goals when the value of an alternative goal exceeds that of the current one. This framing can account for both the choices and the goal probes reported by the participants. Algorithm 1 shows the pseudocode for the common goal and action selection mechanism.

Comparisons with simpler versions: Rescorla-Wagner + choice kernel

We contrasted the models with a much simpler Rescorla-Wagner model embedded with a choice kernel.

$$Q_g^{(t)} = Q_g^{(t-1)} + \eta(r_g^{(t)} - Q_g^{(t-1)}) \quad (2.26)$$

$$c_g(t) = c_g(t-1) + \alpha_c(I_g(t) - c_g(t-1)) \quad (2.27)$$

where $Q_g(t)$ is the expected value of token rewards for each goal, $r_g(t)$ is the binary reward of receiving a token, $c_g(t)$ is the choice kernel for the goal, $I_g(t)$ is the binary indicator value for the goal chosen each round.

The goal selection probability is given as

$$p_g(t) = \frac{\exp(\beta Q_g(t) + \beta_c c_g(t))}{\sum_{k \in G} \exp(\beta Q_k(t) + \beta_c c_k(t))} \quad (2.28)$$

Model fitting procedure

We fit both participant card selections and goal probes in all models except for the Rescorla-Wagner model, where we just fit participants' choices (where there is no notion of an extended goal).

The likelihood of a goal probe is estimated only in the rounds where goal probes are presented in the task (i.e., every third round), whereas the likelihood of an action is estimated in every round. In rounds where the participant does not report their goal, the goal remains latent, and the model estimates selection probabilities across all possible goals. Goal and action likelihoods are then computed by integrating over these latent goal probabilities. The full likelihood estimation process for goal and action selections in the task is summarized as follows.

1. Estimating goal selection likelihood:

- a) If the participant's goal is probed in round k (say the goal reported is *cat*)

$$\Pr(Goal_k = cat) = 1; \Pr(Goal_k = hat) = 0; \Pr(Goal_k = car) = 0; \quad (2.29)$$

- b) For round $k + 1$, the latent goal probabilities are estimated given latent goal in k :

$$\Pr(Goal_{k+1} = G) = \Pr(Goal_{k+1} = G | Goal_k = cat) \quad (2.30)$$

Conditional probability of choosing a goal g in round t given the previous goal G in round $t - 1$ is defined as:

$$\Pr(Goal_t = g | Goal_{t-1} = G) = \begin{cases} p_{stay}, & g = G \quad (\text{stay}) \\ (1 - p_{stay}) * softmax(Q_{alt_1}, Q_{alt_2})[g], & g \neq G \quad (\text{switch}) \end{cases}$$

where p_{stay} is estimated as described in Equation (20), and the softmax is computed over the two alternative goal values.

- c) For round $k + 2$, the latent goal probabilities are estimated given latent goal probabilities in $k + 1$:

$$\Pr(Goal_{k+2} = G) = \sum_g \Pr(Goal_{k+2} = G | Goal_{k+1} = g) * \Pr(Goal_{k+1} = g) \quad (2.31)$$

- d) The goal is then probed again in round $k + 3$, with the goal selection probabilities computed given the latent goal probabilities in $k + 2$. This estimation is used to compute the likelihood of goal selection for the participant's goal probe in round $k + 3$.

$$\Pr(Goal_{k+3} = G) = \sum_g \Pr(Goal_{k+3} = G | Goal_{k+2} = g) * \Pr(Goal_{k+2} = g) \quad (2.32)$$

- e) The goal probabilities in round $k + 3$ are aligned with the latest goal probe (as in step (a)).

2. Estimating action selection likelihood:

$$\Pr(Action_t = A) = \sum_g \Pr(Action_t = A | Goal_t = g) * \Pr(Goal_t = g) \quad (2.33)$$

One limitation of this procedure is that we hold the goal probe reported by the participant as the true value of the latent goal variable. However, the goal reports of participants could be noisy in disengaged participants. The participant goal probe statistics reported in S2A and S2B Figs can alleviate some of these concerns given the high congruence of goal action seen in the participant data.

TD-persistence has 5 free parameters, TD-momentum and prospective agent have 6, and hybrid agent has 7. For each individual participant, we estimate the parameters for each of the competing models to maximize the likelihood of their action choices and goal probes by employing a tree-structured parzen estimator (TPE) sampler, a Bayesian optimization algorithm implemented in optuna [5], an open source hyperparameter optimization framework in Python. We estimated the parameters that maximized the likelihood over 100 random starts.

We followed the same procedure for fitting behavior and for recovering parameters and models.

Parameter recovery

For parameter recovery analysis, we sampled 1000 parameter vectors from a range of parameters ($\alpha \in [0, 1]$, $c_g \in [-1, 1]$, $\beta_g \in [0, 10]$, $\beta_a \in [0, 10]$, $\gamma \in [0.6, 1.0]$, $\alpha_c \in [0, 1]$). We generate synthetic behavior for each parameter vector using the TD-momentum model and recover the parameters by fitting the model to the simulated data. We optimized both the likelihoods of selection of the goals and the actions while fitting the data, where the simulated agent reported the goal every third round as the participants did. We plotted the fitted parameters and the generated parameters and reported correlations (see S4 Fig). Most of the parameters are recoverable with high correlation coefficients. However, low discount factors closer to 0.5 do not recover well; however, the discount factors of the participant fits are mostly in the range $[0.9, 1.0]$, so we considered the range $[0.6, 1.0]$ for parameter recovery.

Model recovery

We performed a model recovery analysis on the competing models. We sampled 100 random parameter vectors from parameter ranges derived from individual participant fits. For each of the four algorithms (TD-momentum, TD-persistence, prospective, hybrid), we generated choice data using the sampled parameters. We then fit the generated data to each of the four models and identify the winning model through BIC score comparisons. We plotted the confusion matrices where the rows correspond to the simulated model and the columns correspond to the recovered model (see S5 Fig). The left confusion matrix indicates the probability of fitted model given the simulated model, whereas the right matrix indicates the probability of simulated model given the fitted model. In all cases, we observed that the data generating model provides the best explanation for the choice data. However, there appears to be some confusion between the prospective and hybrid models.

Normative simulations

For the normative account, we assumed three starting positions for the availability of three options: $\sim \mathcal{U}(0.2, 0.3)$, $\sim \mathcal{U}(0.2, 0.3)$, $\sim \mathcal{U}(0.7, 0.8)$. We simulated two variants: token contingencies that gradually changed through Gaussian random walks ($\sim \mathcal{N}(0, 0.025)$) and abrupt switching of contingencies with probabilities shuffling every 30 rounds.

We computed the optimal performance achieved by each algorithm with parameters estimated by Bayesian optimization.

Chapter 3

PERSISTENCE ALL THE WAY DOWN: A MOMENTUM-BASED ACCOUNT OF DEPTH-FIRST PERSISTENCE

Humans can flexibly organize their behavior to pursue hierarchically organized goals, dynamically updating their commitment to goals and concomitant subgoals based on changes in environmental contingencies. Through a novel hierarchical goal selection paradigm, we investigate human persistence patterns in goal and subgoal pursuit, especially in the presence of varying subgoal prospects. We observe that humans show a pattern of depth-first commitment toward hierarchical goals, whereby they prefer to persist on a goal-subgoal pair until termination, rather than persist at the goal level without committing to subgoals. We also show that this depth-first persistence in goals and subgoals is modulated by goal and subgoal progress, respectively. We further demonstrate that a momentum-based account that parsimoniously computes the time-to-goal-completion explains depth-first patterns of persistence on hierarchically organized goals. We explicitly dissociate the effects of momentum computations from switching costs on goal persistence, and show that momentum computations drive persistence choices above and beyond dynamic costs imposed by switching goals and subgoals.

3.1 Introduction

Our workday lives are filled with multitasking needs that compete for our attention. Imagine the following scenario: you are in the middle of writing a project report when it hits you that you need to respond to an urgent email. We frequently make flexible switches between goals based on environmental contingencies and our internal needs. Furthermore, we hierarchically organize our goals, where we split them into subgoals and pursue smaller tasks with interim goalposts in service of larger goals. How adaptive decision making is achieved in such hierarchical structures is a research topic that is still in its infancy [20, 109].

In a previous study [3], we showed that people tend to overpersist on goals compared to task-defined optimal agents. We introduced a new construct, *goal momentum*, that accounts for the overpersistence effects in human extended goal selection and also provided a normative account for why it might be adaptive. Goal momentum, defined as the product of progress in the goal and its rate of progress, is a parsimonious way of

estimating values of extended goals and roughly approximates expensive prospective scenarios in most cases. But when there are abrupt shifts in the environment, momentum computations can lead to overpersistence effects.

Thus far, we have only considered how momentum computations operate in selecting extended goals, not when they are embedded in a hierarchical structure, such as when sub-goals must also be selected. Here, we ask if momentum computations influence persistence on subgoals as well. We introduce a novel paradigm where participants have to select between multiple hierarchically organized goals and pursue them by choosing between multiple extended subgoals pertaining to those goals. Using this task, we investigate whether momentum computations operate at both stages of the goal hierarchy.

We already demonstrated that momentum encourages persistence at the goal level. However, if the same momentum process operates at the subgoal level, it would naturally facilitate persistence in subgoals as well. This would result in a depth-first strategy in goal pursuit, whereby participants would prefer to persist on a goal–subgoal pair as opposed to frequently switching between subgoals within a goal. We also test additional predictions of momentum computations at the level of subgoals, which is that the extent of depth-first persistence should depend on both goal and subgoal progress per se, alongside the *rate* of both goal and subgoal progress.

We also consider alternative accounts for goal and subgoal selections on the task and contrast these with the predictions of the momentum account. For instance, we designed models that do not invoke the depth-first strategy of goal–subgoal selection and contrasted these with participants’ selection patterns. Also, even within the class of models that also follow a depth-first strategy, we formulated models that incorporate different goal and subgoal valuation processes other than momentum. We formally compare all of these models with respect to their explanatory power of individual participants’ choices as well as their ability to replicate aggregate effects in behavior.

Specifically, we test an account whereby persistence in both goals and subgoals can be explained by efficiency considerations. Several previous studies have shown that switching tasks invokes costs of reconfiguring task representations and reallocation of control to these representations [147, 12, 67, 126]. Secondly, hierarchical persistence can also induce building of goal–subgoal *options* that can be automatized through repetition, thereby increasing efficiency in goal pursuit [164, 155]. We

show that these stand-alone efficiency considerations cannot account for persistence patterns per se, but rather emerge as a consequence of momentum computations, providing further support for a normative account of momentum at different hierarchical levels.

Our manuscript is structured as follows: we introduce the experimental paradigm and provide evidence for the behavioral signature of hierarchical momentum computations in the task, i.e., depth-first persistence in participants' choices. We then show how depth-first persistence is modulated by both progress in goals and subgoals as well as their rates of progress. We formulate hierarchical momentum as an account of goal selection to explain depth-first persistence patterns and contrast it with several alternative accounts. Through the current study, we thus strengthen the momentum-based account in explaining persistence in long-term, hierarchically organized goals.

3.2 Experiment

We designed a hierarchical game that has both temporally extended goals and subgoals. In the game, participants can choose between any of three construction projects as high-level goals — building a spaceship, observatory, or house. To complete these high-level goals, they need to collect a combination of three different resources in specific proportions. Table 1 shows the requirements for each project:

Project	<i>Glass</i>	<i>Metal</i>	<i>Concrete</i>	<i>Stone</i>
<i>Spaceship</i>	3	3	0	0
<i>Observatory</i>	2	0	1	3
<i>House</i>	0	2	3	1

Table 3.1

Subgoals are, in turn, extended goals in this game. Therefore, the two subgoals for the *spaceship* are: *collect 3 glass resources*, and *collect 3 metal resources*. Also, the subgoals in the game are not sequentially constrained (i.e, do one after another), which can invoke persistence based on task constraints. Instead, subgoals in this game can be executed in any order.

Participants are free to switch between any goal or subgoal during the game without any penalty. Participants can procure resources by selecting alien suppliers who supply resources with some probability and hold collected resources in an inventory. Once they have enough resources to build a project, they can build it to collect 10 points. The overall objective of the game is to maximize the number of completed

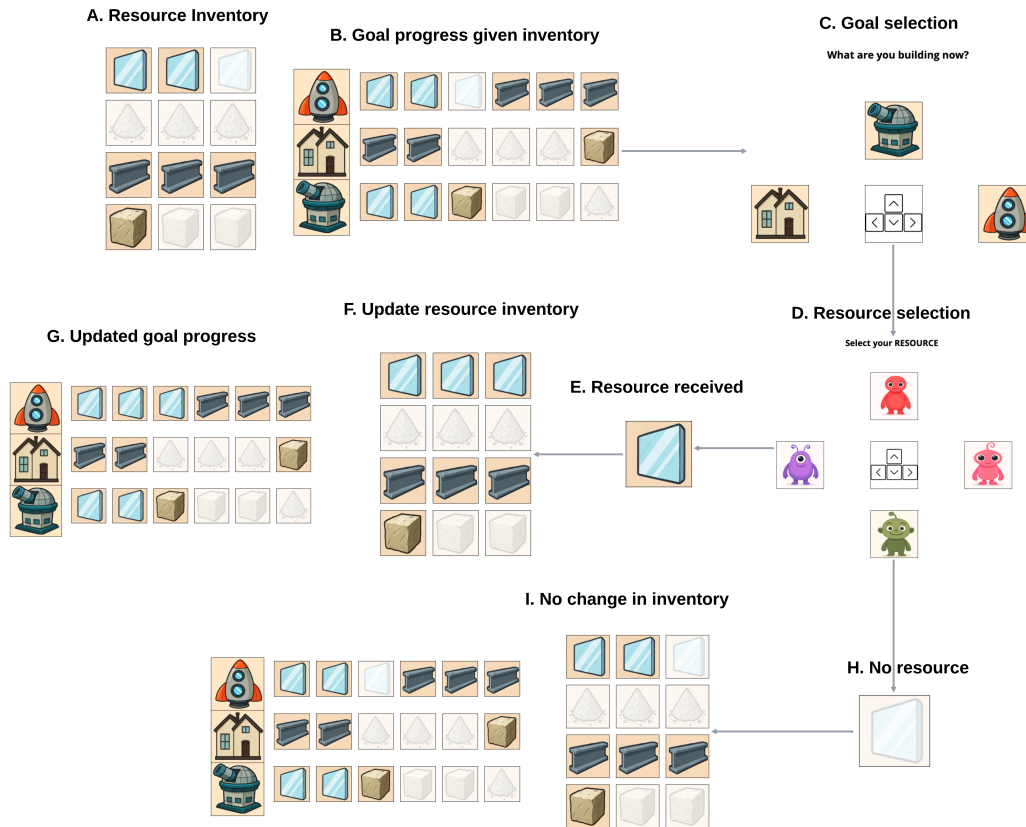


Figure 3.1: **Hierarchical goal pursuit:** The game flow: **A.** Participants are shown resource inventory status, which can hold upto 3 resources: glass, concrete, metal, stone. **B.** Participants are shown how much progress they have made toward each goal with the resources in the inventory. **C.** Participants then choose which project they are currently building. **D.** Participants then choose an alien supplier (each mapped to one distinct resource) to choose the resource. **E, F, G.** If the participant receives the resource, the inventory is updated, and new goal progress is shown. **H, I.** The participant does not receive the resource, and the inventory is unchanged.

construction projects within a specific time. Figure 1 shows the game flow of a sample round.

Each alien supplier is mapped to a specific resource and only supplies that resource throughout the game. Alien suppliers supply resources with specific chances that change throughout the game. We trained the participants on the alien-resource associations prior to the task, but they needed to infer the chances of getting resources from aliens during the task through sampling. If the alien supplies the resource, it goes in the inventory, and the game state is updated. If not, the game state stays the

same, and participants proceed to the next round.

We organized the game in blocks. Each block has a specific construction goal that is dominant over others because the required resources that are required for its construction are more abundant. We used six block types as shown in Table 2, two viable block types — *high* and *low* — for each goal. The difference in viability of the superior goal over the other goals is much higher in the *high* condition than in the *low* condition.

Block types in the game

Resource/Condition	<i>glass</i>	<i>metal</i>	<i>stone</i>	<i>concrete</i>
SH-<i>high</i>	0.8*	0.8*	0.2	0.2
SH-<i>low</i>	0.6*	0.6*	0.4	0.4
OB-<i>high</i>	0.8*	0.1	0.8*	0.7*
OB-<i>low</i>	0.6*	0.3	0.6*	0.6*
HO-<i>high</i>	0.1	0.8*	0.8*	0.8*
HO-<i>low</i>	0.3	0.6*	0.6*	0.6*

Table 3.2: Resource probabilities in different block types. * indicates high abundance resources in the block.

We instructed the participants about the block structure — that each block has specific resource probabilities and that those change when the block shifts. We explicitly signaled the block shifts to the participant. Participants ($n=30$) played a total of 12 game blocks, with each block consisting of 30 rounds. We presented the block types to the participants in a randomized order. In each block, participants selected the most viable goal more often than others, and they also pursued the relevant subgoals to the viable goal with a higher probability (see Figure 2A). Participants also showed a sensitivity to the *high-low* experimental manipulation, selecting the viable goals and subgoals relevant to the viable goals more often in the *high* condition than in the *low* condition (paired-sample t-test; *goals*: $t(29) = 2.47, p < 0.05$; *subgoals*: $t(29) = 3.40, p = 0.002$).

3.3 Signatures of depth-first persistence

In hierarchically organized goals, participants can choose to persist *depth-first* into the goal. If so, they would prefer to hierarchically persist in a goal–subgoal pair. Alternatively, they can choose to flatly persist with the higher-level goal but show no clear commitment toward subgoals. Different subgoals in our task have no sequential dependence between them, so participants can make progress in a higher-level goal by its subgoals in any order they prefer; there is no clear advantage of persisting

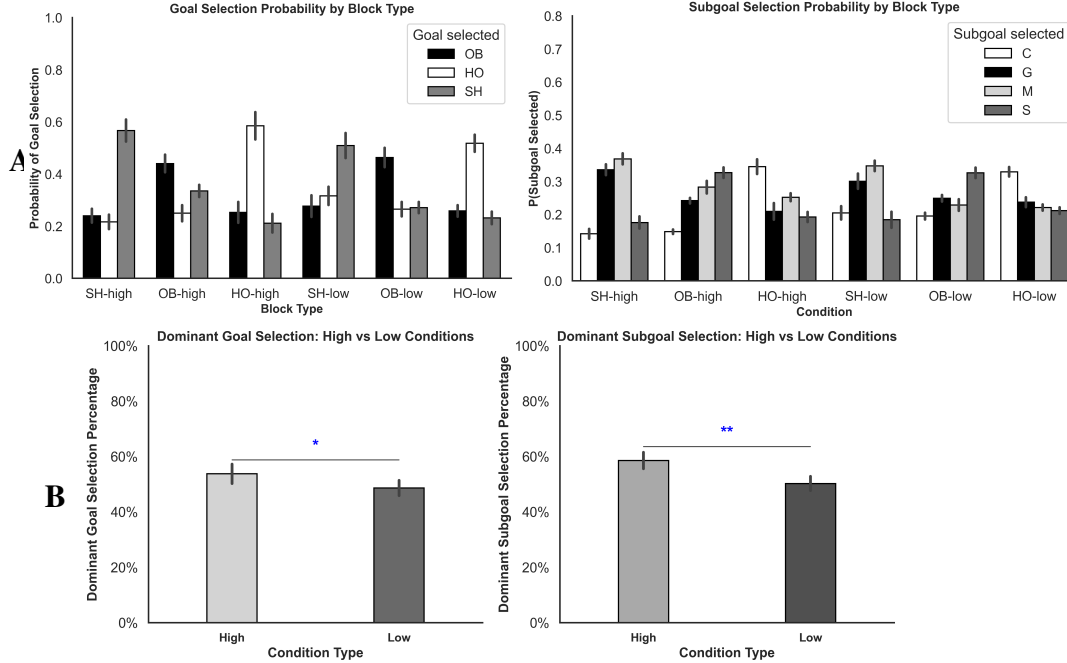


Figure 3.2: Goal and subgoal selection patterns. **A.**(left) Goal selection probability split by block condition. Participants picked the most viable goal in each block with a higher chance. (right) Subgoal selection probability split by block type. Participants picked the subgoals corresponding to the most viable goal with a higher chance. **B. Goal selection modulated by *high-low* manipulation.** Participants selected the viable goal and the subgoals relevant to the viable goal with a higher chance in the *high* condition than in the *low* condition.

depth-wise into the goal. In another strategy, they can also follow a purely resource-first approach, where they prioritize acquiring the most abundant resources, and the goal is selected as an afterthought of observing the resource inventory status. All three approaches are legitimate strategies to solve our task.

We investigated participants' persistence trends to understand their preference for one of the three strategies described above. We computed the probabilities of participants' transitioning from one goal (g_1) to another goal (g_2) across adjacent rounds ($Pr(goal_t = g_2 | goal_{t-1} = g_1)$). Figure 3A shows the goal transition probability matrices separately for each block type. Participants show a significantly higher chance of staying with the same goal chosen in the previous round. The chance of staying is higher still when the given goal is dominant within the block. Across all block types, participants show a higher chance of persisting with the dominant goal within the block (average chance; 0.78) over the other two goals (average chance; 0.56) (chi-squared test of independence; $goals: \chi^2(1, N = 30) = 509.11, p < 10^{-6}$).

In Figure 3B, we computed the probabilities of participants' transitioning from one subgoal (sg_1) to the next (sg_2) given that they stayed with the same goal g across adjacent rounds $Pr(subgoal_t = sg_2, | subgoal_{t-1} = sg_1, goal_{t-1} = g, goal_t = g)$. We note that participants show high levels of persistence for subgoals as well. In each block type, participants' persistence levels are highest for the subgoal of the dominant goal with the greatest target. When they switch subgoals, participants are more likely to switch to other subgoals relevant to the dominant goal over others (paired-sample t-test on transitioning to relevant vs irrelevant subgoals; $t(29) = 5.59, p < 10^{-5}$).

The two features — a high tendency to stay with the same goal-subgoal pair across adjacent rounds and the tendency to switch subgoals within the goal — point to a pattern of depth-first commitment. We also investigated persistence trends in the first and second halves of the blocks, to determine if participants exhibit any shift in persistence strategies as the rounds evolve within a block (see Figure 3C). We hypothesize that participants can be more likely to follow a resource-first approach at the beginning of the block, where they are unsure of the resource probabilities. On the contrary, we did not find any significant difference between persistence trends within a block, suggesting that participants' preference towards depth-first commitment both within and across task blocks.

Heterogeneity in depth-first commitment patterns in participants

While participants show an overall tendency to persist depth-first, the extent to which they show depth-first commitment can differ between individuals. To study the heterogeneous commitment trends, we split the goal-subgoal selections made by participants into 5 categories. Category 1 includes rounds where participants persist with both the goal and subgoal from the previous round. Category 2 includes rounds where participants persist with the goal but switch to another subgoal within the same goal. The subgoal switches made when participants' previous subgoal is terminated are included in category 3. The fourth category includes rounds where participants switch both goals and subgoals, but switch to relevant-alternate goals. A relevant-alternate goal can make concurrent progress along with the current goal on account of the shared subgoal.

Consider a participant collecting *metal* for building the *spaceship*. As *metal* is needed for *house* as well, the participant makes concurrent progress towards *house* by collecting *metal*. *House* then becomes the relevant-alternate for *spaceship*, when participant is pursuing *spaceship-metal* pair. *Observatory* is the irrelevant-alternate

as it does not need *metal*. Category 5 includes switches to irrelevant alternative goals.

Figure 4A shows the proportion of goal-subgoal selections falling into each of the 5 categories (split by the success or failure of the previous action in Figure 4B). While the overall proportion of staying with the goal-subgoal pair (category 1) is high on average, the extent of it varies highly across participants, indicating high heterogeneity in goal commitment patterns.

We defined a depth-first metric to quantify this heterogeneity in participants.

$$\begin{aligned} \text{depth-first metric} &= \Pr(\text{goal-subgoal persistence} | \text{goal persistence}) \\ &= \frac{p_{\text{category 1}}}{p_{\text{category 1}} + p_{\text{category 2}}} \end{aligned}$$

where $p_{\text{category 1}}$ is the proportion of rounds where participants persist with goal-subgoal pairs, and $p_{\text{category 2}}$ is the proportion of rounds where participants persist with the goal but switch subgoals within the goal.

We observe that participants with a lower depth-first metric are also the ones to switch higher-level goals often (see Figure 5A; correlation; $R = -0.62, p < 10^{-5}$). We can infer that people who switch often at one level of the goal hierarchy are also likely to switch often at other levels as well. Also, we note that a lower depth-first metric is also correlated with increased preference to switching toward irrelevant-alternate goals (Figure 5B; correlation; $R = -0.61, p < 10^{-5}$). When participants make a switch toward irrelevant-alternate goals, they are switching to pursue a subgoal that does not build on the previously chosen subgoal in any way, thus creating a discontinuity in goal pursuit.

The depth-first metric is also correlated with task performance (Figure 5C; correlation; $R = 0.59, p < 10^{-5}$), with participants exhibiting high goal-subgoal commitment performing well on the task. We also investigated the engagement levels of participants in the task by calculating the congruence between the goals they reported during the construction choice and the resource selection they made. We deemed the resource selection to be congruent for a construction choice when it satisfied two requirements: 1. The resource is necessary for constructing the goal, and 2. The number of resources needed to meet the construction target is not met. Participants' depth-first metric also correlates with participant engagement, which

we measured through the goal-action congruence metric (Figure 5D; correlation; $R = 0.69, p < 10^{-5}$).

Relating depth-first commitment with performance in participants tells us that persisting depth-wise into the hierarchical goal can be an adaptive feature. Moreover, high-performing participants prefer to switch to goals and subgoals that build on their previous goal progress. While building steady progress towards goals is one benefit, there seem to be several other considerations that can make depth-first commitment an adaptive strategy. We discuss more of those in the next section.

3.4 Depth-first persistence contributes to efficiency in hierarchical goal pursuit

Depth-first commitment ensures fewer switches from the current line of goal and subgoal pursuit. As switching of tasks invokes cognitive costs [129, 96, 182], persisting depth-wise minimizes the penalty of such costs. Here, we investigate reaction times (RTs) of participants when they make goal and subgoal selections to see how costly it is to switch at different stages of the goal hierarchy.

In Figure 6A, we show participants' RTs of their goal and subgoal selection choices split into three categories: 1. stay with both goal and subgoal; 2. stay with the goal but switch to a different subgoal within the goal; 3. switch goals. We plot the RTs during the goal selection phase and the subgoal selection phase in separate subplots. Switching of goals evokes significantly higher RTs over other categories (Figure 6A).

Interestingly, we did not find any differences between the RTs in the three categories during the subgoal selection phase of the game. We can expect to see no significant difference in subgoal RTs, if participants were to make goal and subgoal selections concurrently during the goal selection phase, a signature of executing *options* in hierarchical reinforcement learning [164]. *Options* refer to actions that are temporally-extended, where structured sequences of lower-level actions are executed as a single unit until termination. In the context of this task, an *option* can be the composite goal-subgoal choice.

We can learn efficient policies for executing options through repetition, and we see evidence of such repetition effect in our task as well. Participants show a significant effect of repeating goal-subgoal pairs on goal selection RTs (Figure 6C), with participants getting faster at executing goal-subgoal options with increasing repetition (correlation between reaction times and repetition length; $R = -0.84, p =$

0.005). There is a weak trend of repetition effect seen in subgoal selection RTs ($R = -0.66, p = 0.15$) (Figure 6D), however, most of the efficiency in repetition is observed during the goal selection phase .

There is, however, a caveat to the RT interpretations of efficiency and switching costs. We observed that participants with a higher depth-first metric are also the ones that show a higher proportion of RT increases during goal switching (Figure 6C; correlation; $R = 0.43, p = 0.018$). This might lead to two interpretations: one makes a causal claim that people who show higher RT increases during switching have higher switching costs, and hence decide to persist more. Another interpretation could be that people who tend to persist more have higher decision value for goal persistence and show increased RT switching away from a valued option. Because participants who show higher depth-first persistence are also the ones to perform well in the task, the decisions to persist might not just be driven by inertia imposed by switching costs but also by the decision value of commitment.

In the next section, we focus on decision value considerations impacting decisions to persist or switch away from goals.

3.5 Progress modulation of depth-first persistence

To test whether progress on a goal increases its chance of persistence, we split goal progress on trials into two categories, high progress ($> 50\%$ resources collected) and low progress ($< 50\%$), and computed the proportion of participants' staying choices in each category. We observed that participants persisted significantly more on goals with higher progress than goals with lower progress (Figure 7A; chi-squared test of independence; *goals*: $\chi^2(1, N = 30) = 79.77, p < 10^{-5}$). We replicated a similar pattern in subgoal commitment as well (chi-squared test of independence; *subgoals*: $\chi^2(1, N = 30) = 264.45, p < 10^{-5}$). Moreover, we tested the modulation of subgoal persistence given goal persistence — $Pr(\text{subgoal stay}|\text{goal stay})$ — with progress, and found that its extent increases with subgoal progress (chi-squared test of independence; *subgoals*: $\chi^2(1, N = 30) = 284.91, p < 10^{-5}$). *Participants' depth-first commitment is thus modulated by both goal and subgoal progress.*

Figure 7B shows participants' progress modulation of participants' $Pr(\text{goal stay})$, $Pr(\text{subgoal stay}|\text{goal stay})$ split by high and low block types. Participants tend to persist higher with both goals and subgoals in the high condition (chi-squared test of independence; *goals*: $\chi^2(1, N = 30) = 6.94, p < 0.01$; *subgoals*: $\chi^2(1, N = 30) = 7.54, p < 0.01$). As high and low block differ in achievable rates of progress,

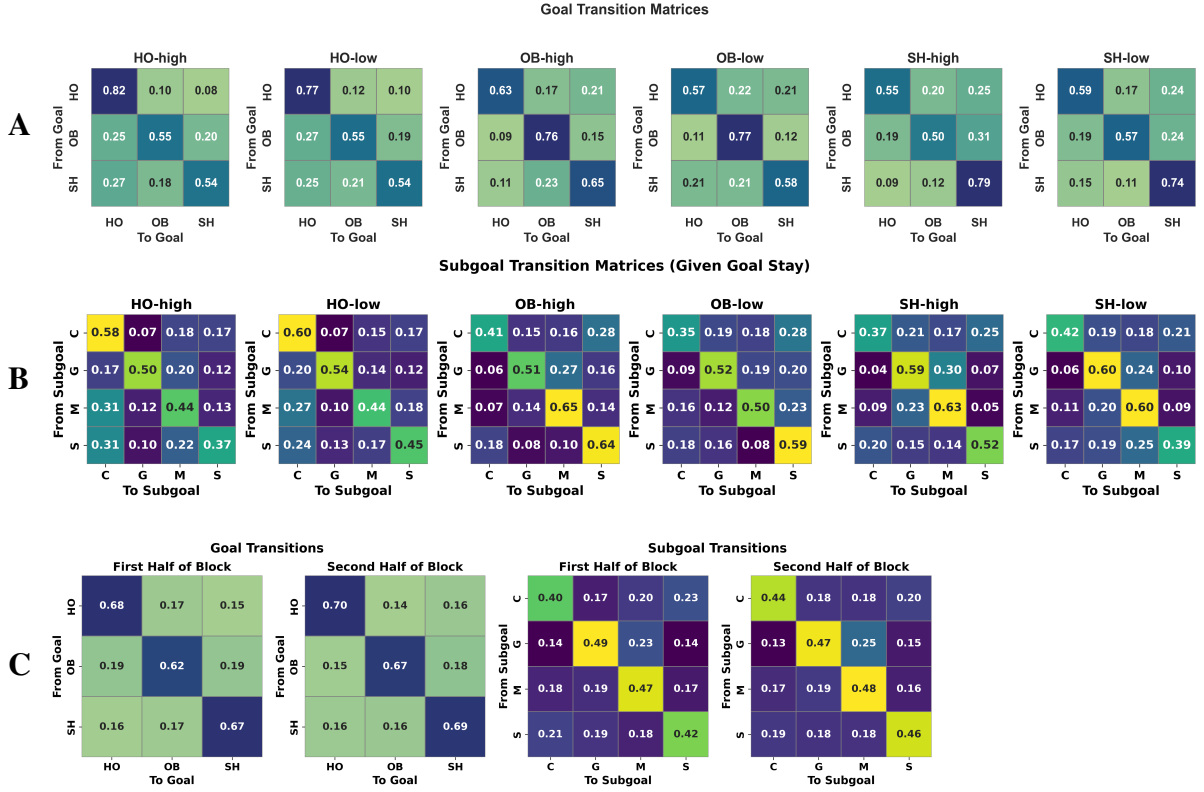


Figure 3.3: **Goal and subgoal transition probabilities.** **A.** Probabilities of transitioning from one to another across successive rounds. Overall, participants tend to persist in goals. **B.** Subgoal transition heatmaps showing persistence in subgoal pursuit. **C.** Goal transition probabilities by block half showing temporal dynamics in persistence patterns.

participants' depth-first commitment is thus modulated by goal and subgoal rates of progress.

Figure 7C shows participants' goal selection probabilities with an increase in goal progress (Repeated measures ANOVA; *goals*: $F(2, 145) = 40.90, p < 10^{-5}$). We also split goal selections by block type. Participants' selections show similar progress curves across all block types, with the overall magnitudes of selection probabilities differing across blocks.

We observed similar patterns of progress modulation in subgoal selections as well (see Figure 7C; Repeated measures ANOVA; *subgoals*: $F(2, 58) = 47.69, p < 10^{-5}$). To compare subgoals across different goals with varying requirements, we used fractional progress on a subgoal as a measure of subgoal progress. We also show progress trends of subgoals for the goals where they have the largest requirement (eg *glass* in *spaceship*) and split them by block type. Although not

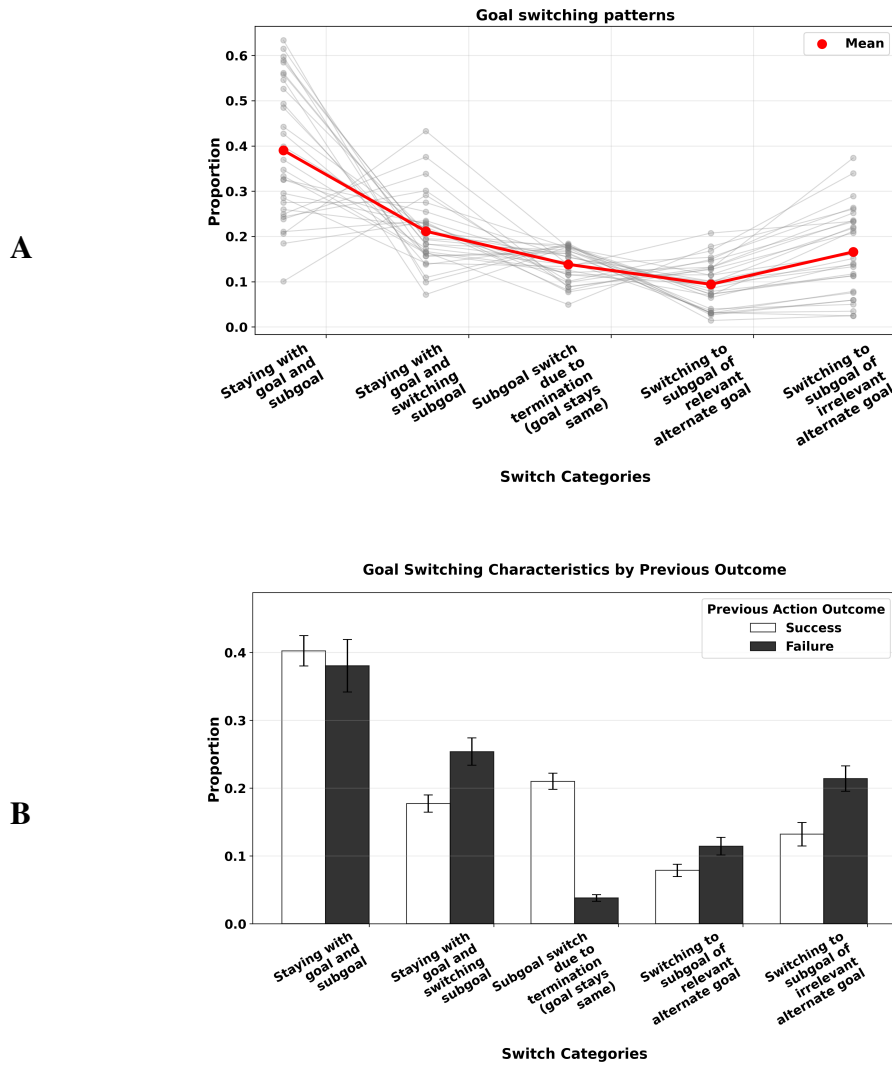


Figure 3.4: **Participant switching patterns** **A.** Goal selections are split into 5 categories : Each action in a round is classified into 5 categories: 1. Staying with the same goal and subgoal from the previous round, 2. Staying with the goal and switching subgoal (when the subgoal is not terminated), 3. Switching subgoals within the goal when it is terminated, 4. Switching to a relevant alternate goal, 5. Switching to an irrelevant alternative goal. Participants overall preferred to stay with the same goal-subgoal pair across adjacent rounds. **B.** Participant selection patterns separated by outcome — whether the participant’s previous action was a success or a failure.

always strictly monotonic, we see an overall increasing trend of persistence with subgoal progress across all block types. Shared subgoals in the game can partially account for the tapering of subgoal progress trends, where collecting one or two resources in a subgoal might satisfy the subgoal requirement for a different goal and

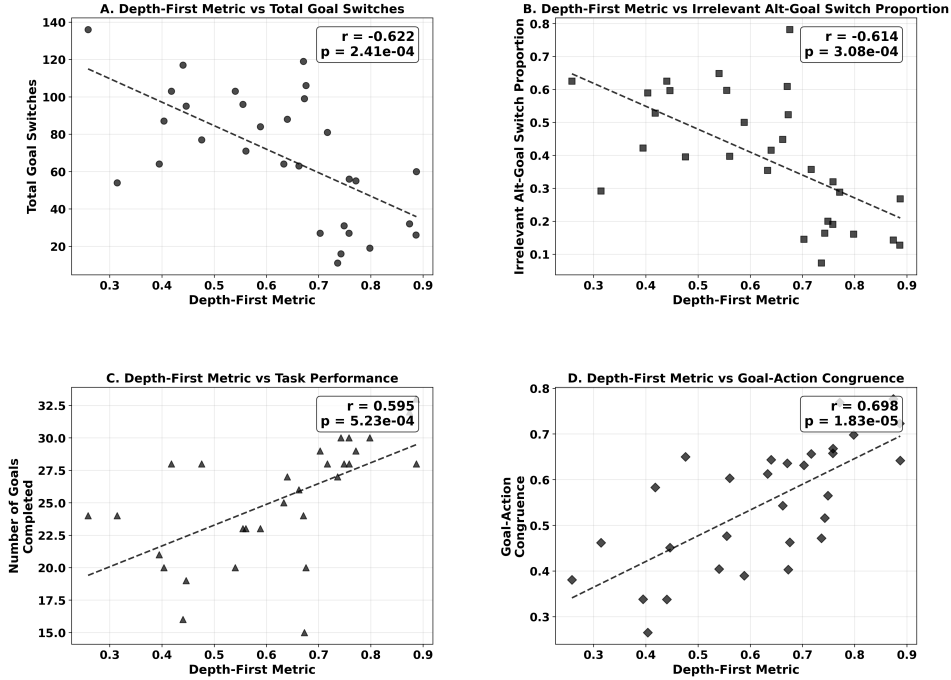


Figure 3.5: **Correlation of depth-first metric with task variables.** Depth-first metric (goal and subgoal stay proportion - goal stay and subgoal switch proportion) correlates with total number of goal switches, difference between number of irrelevant and relevant goal switches, task performance, and task engagement (measured using goal-action congruence).

reduce the motivational value for collecting more resources to complete the subgoal requirement of the current goal.

In summary, participants exhibit depth-first persistence in the task, the extent of which is modulated by progress and rates of progress of both goals and subgoals. In the next section, we describe the hierarchical momentum-based account that could generate depth-first patterns as seen in behavior. We also contrast it with several alternative hypotheses that can perform this task, one of which does not employ a hierarchical goal pursuit strategy. Among the ones employing hierarchy in goal pursuit, we construe several alternate valuation schemes other than momentum, some simpler and some more complex, to see if they explain human persistence patterns better than the momentum-based account.

3.6 A hierarchical momentum-based account

In [3], we introduced the construct ‘goal momentum’ to account for persistence patterns in extended goals, where

$$\text{goal momentum} = \text{goal progress} * \text{rate of goal progress} \quad (3.1)$$

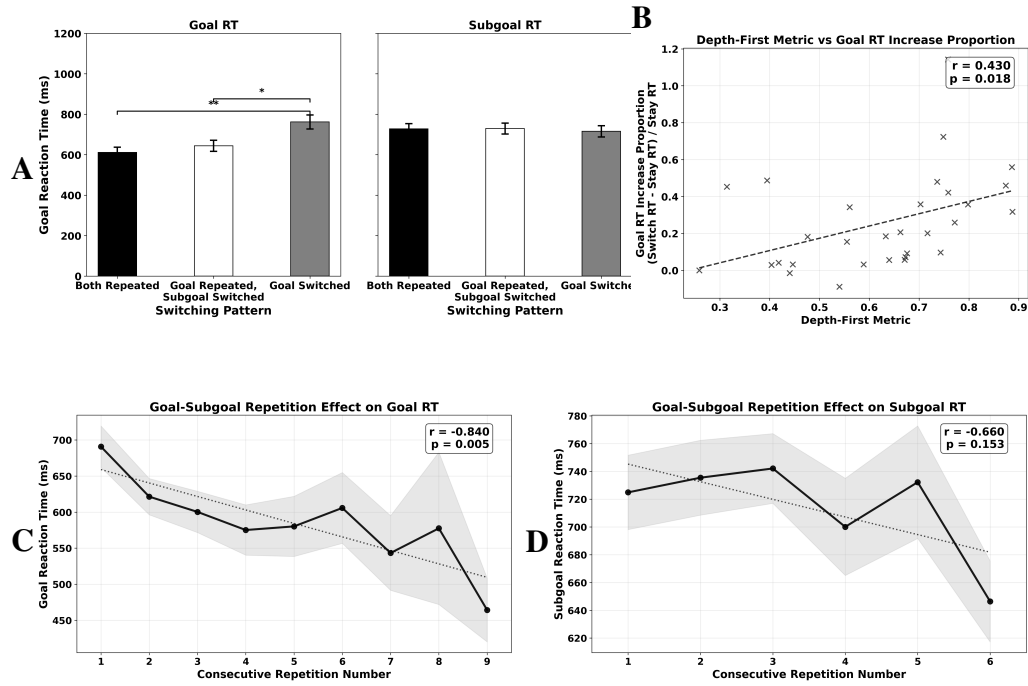


Figure 3.6: **A. Goal and subgoal switching reaction times.** Participants exhibit high reaction times when switching goals as opposed to staying with the same goal or switching just subgoals. **B. Proportional increase in reaction time during goal switches correlates with the depth-first metric.** **C. Goal RT speed-up achieved by depth-first persistence of goal-subgoal pairs.** **D. Subgoal RT speed-up achieved (not significant) by depth-first persistence of goal-subgoal pairs.**

We showed that pursuing goals while optimizing for goal momentum in the task can produce similar persistence patterns seen in humans pursuing extended goals. We also introduced an algorithm, TD-momentum, that can estimate the *rate of goal progress* through experience using a temporal difference learning rule. We now extend that algorithm to account for depth-first persistence patterns seen in the task.

Hierarchical TD-momentum

In hierarchical TD-momentum, goal and subgoal momentum drive persistence in goals and subgoals, respectively.

Goal momentum

We define goal momentum for a goal g , Q_g at time t as follows:

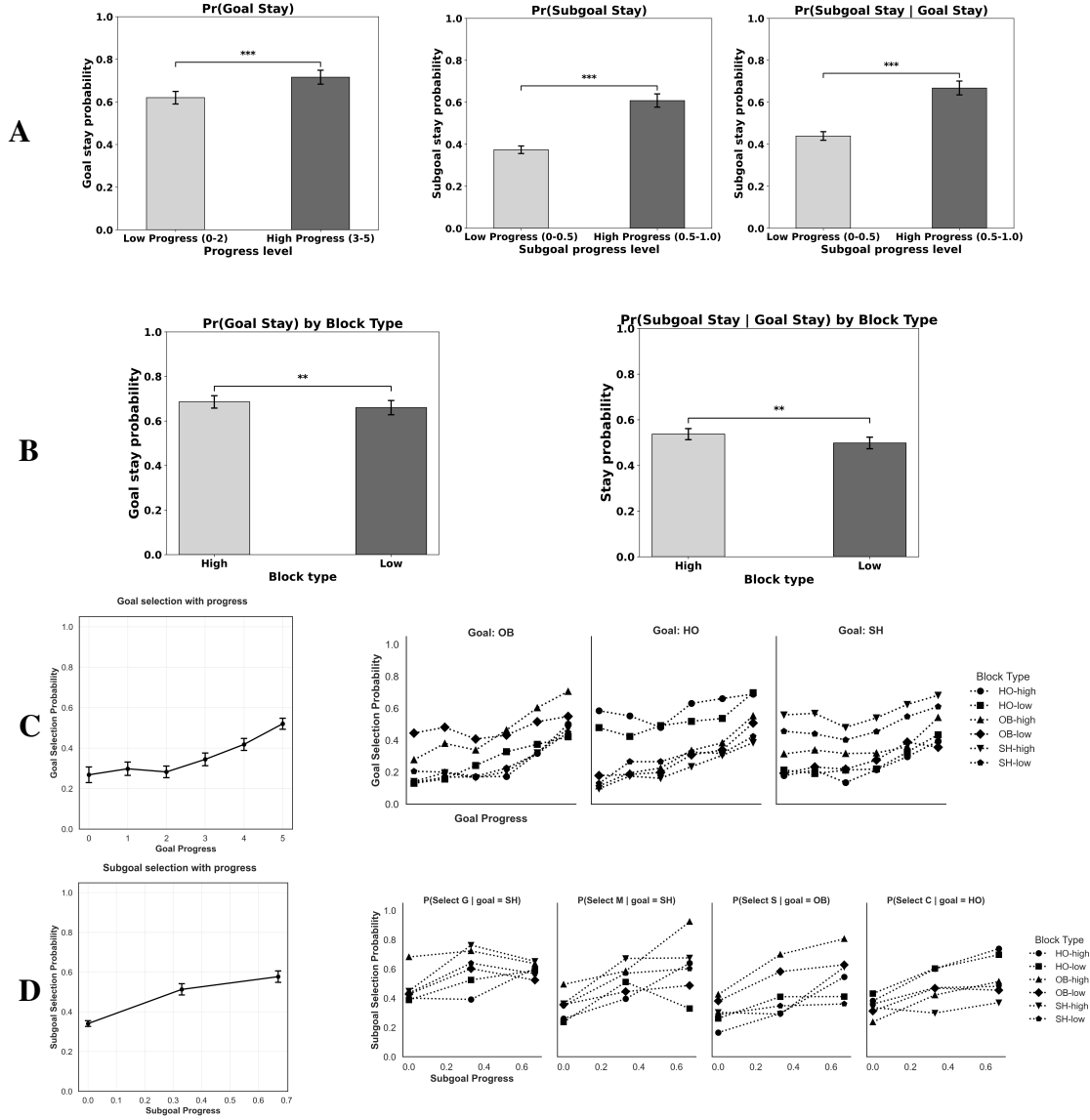


Figure 3.7: Progress modulation of goal and subgoal persistence and selection. **A.** Participants stayed with the goals and subgoals when they had higher progress compared to when they made lower progress. **B.** Participants stayed with the goals and subgoals in the high condition compared to the low condition. **C.** Goal selection probability increases with the increase in progress (left); Goal selections with progress split by goals selected and block type. **D.** Subgoal selection probability increases with fractional progress in the subgoal (left); Subgoal selections with progress split by subgoals selected and block type.

$$Q_g(s_{g|I}^{(t)} = m_g^{(t)}) = v_g^{(t)} m_g^{(t)} + b_g^{(t)} \quad (3.2)$$

where $s_{g|I}$ indicates the state of the goal given the inventory, m_g is the fractional

progress accrued in the goal, v_g is the parameter for the rate of goal progress, and b_g is the baseline progress rate for the goal. The baseline parameter, b_g , ensures the goal momentum is updated through experience even when the current progress in the goal is zero. Here, both v_g and b_g are updated from experience in progressing with the goal. v_g is computed for the overall goal, without considering individual resource availabilities specifically. Thus, momentum is a parsimonious computation that allows for approximation of potential for goal completion without the need to track individual resource probabilities.

Subgoal momentum

The subgoal momentum, Q_{sg} , is defined as follows:

$$Q_{sg}(s_{sg|g,I}^{(t)} = m_{sg|g}^{(t)}) = v_{sg}^{(t)} m_{sg|g}^{(t)} + b_{sg}^{(t)} \quad (3.3)$$

where $s_{sg|g,I}$ is the state of the subgoal given the goal and the inventory, $m_{sg|g}$ indicates the fractional progress of the subgoal in the goal, v_{sg} is the parameter for subgoal progress rate and b_{sg} is the baseline subgoal progress rate.

Updating goal and subgoal momentum

For the chosen goal (G) and subgoal (SG), we updated the momentum prediction errors after observing the resource outcome as follows:

$$\delta_G = \gamma Q_G(s_{G|I}^{(t+1)}) - Q_G(s_{G|I}^{(t)}); \quad \delta_{SG} = \gamma Q_{SG}(s_{SG|G,I}^{(t+1)}) - Q_{SG}(s_{SG|G,I}^{(t)}) \quad (3.4)$$

where $s_{G|I}^t$ is the state of inventory at time t . The state advances by one resource if participants receive a resource in the round (positive prediction errors), or stays in the same state if they receive none (negative prediction errors).

Goal and subgoal momentum parameters are updated through a gradient descent of parameters in the value function as follows:

$$v_G^{(t+1)} = v_G^{(t)} + \alpha \delta_G m_G^{(t)}; \quad b_G^{(t+1)} = b_G^{(t)} + \alpha \delta_G; \quad v_{SG}^{(t+1)} = v_{SG}^{(t)} + \alpha \delta_{SG} m_{SG|G}^{(t)}; \quad b_{SG}^{(t+1)} = b_{SG}^{(t)} + \alpha \delta_{SG} \quad (3.5)$$

Cross-goal momentum transfer with shared subgoals

When participants make progress in a subgoal after selecting a goal, they make progress not only in the chosen goal but also in an alternate goal that relies on the

same subgoal. We posited that the updating of momentum of unchosen goals could also be influencing subsequent goal selection.

We created a variant of TD-momentum that incorporated cross-goal transfer (TD-momentum-alt-goal) where we updated the momentum of the unchosen alternate goal that made progress on account of shared subgoals.

Modeling dynamic switching costs

We consider switching costs as an additional factor influencing goal selection choices. These switching costs also include dynamic opportunity costs from forgoing goal-subgoal repetitions as described previously. We incorporated these dynamic switching costs in the model through the following additive value term:

$$\text{Goal switching cost} = c_g + \kappa_G \quad (3.6)$$

where c_g is the static switching cost, and κ_G is the dynamic opportunity cost of switching that is incremented by a choice kernel as follows:

$$\kappa_G = (1 - \alpha_c)\kappa_G + \alpha_c I_G \quad (3.7)$$

where I_G is the indicator variable that is 1 when the goal is selected in the round.

A hierarchical goal selection algorithm

We hierarchically modeled choices in individual trials as stay/ switch from goals and subgoals. **Goal selection:** Individual goal values are computed using the momentum algorithm. The advantage of the current goal, A_G , over its alternatives is computed as follows:

$$A_G = Q_G - \max_{c \neq G} Q_c \quad (3.8)$$

The agent chooses to stay on the current goal with a probability $p_{\text{stay};G}$,

$$p_{\text{stay};G} = \frac{1}{1 + \exp(-c_g - \kappa_G - \beta_g A_G)} \quad (3.9)$$

where c_g is the goal switching cost, β_g is the softmax temperature. κ_G is the choice kernel where commitment towards the goal increases the more it is chosen in the past, as updated through the delta-rule,

$$\kappa_G = (1 - \alpha_c)\kappa_G + \alpha_c \kappa_G I_G \quad (3.10)$$

where I_G is the indicator that the goal is chosen in the round.

When the agent decides to switch away from the current goal with probability $(1 - p_{stay})$, it selects a new goal from the two alternate goals depending on the values of the two alternate goals ($softmax(Q_{alt_1}, Q_{alt_2})$)

The advantage computation is equivalent to the negative of the opportunity cost of staying with the current option. This opportunity cost formulation is a standard way of framing stay/switch decisions in foraging studies [34]. Given the temporally-extended nature of the task and human persistence patterns, we consider it more meaningful to employ foraging-like stay/switch decisions. However, we also considered an alternate mode of goal selection based on goal values moderated by a softmax parameter (see Fig A8).

Subgoal selection: Subgoal values are computed given the goal, assuming a depth-first hierarchy. The advantage of the current subgoal over its alternatives is computed from individual subgoal values as follows:

$$A_{SG|G} = Q_{SG|G} - \max_{c \neq SG} Q_{c|G} \quad (3.11)$$

The agent chooses to stay on the current subgoal with a probability $p_{stay;SG|G}$,

$$p_{stay;SG|G} = \frac{1}{1 + \exp(-c_{sg} - \beta_{sg} A_{SG|G})} \quad (3.12)$$

where c_{sg} is the subgoal switching cost.

3.7 Alternative hypotheses

We proposed the following alternative accounts for goal and subgoal selection in the task, to compare with our hierarchical momentum model.

Prospective agent

The preference for progress can be optimal in agents that discount their future rewards. We constructed a prospective agent that optimizes for discounted expected returns in the game. The prospective agent maintains a running estimate of resource probabilities and estimates the future prospects of goal and subgoal completion using hyperbolic discounting.

The agent updates the belief about the resource probabilities using a delta learning rule:

$$M_r^{(t)} = M_r^{(t-1)} + \alpha(I_r^{(t)} - M_r^{(t-1)}) \quad (3.13)$$

where $M_r^{(t)}$ is the belief estimate of the probability of obtaining a specific resource r at time t , and $I_r^{(t)}$ is the indicator function which is 1 if the resource is given in a particular round. α is the learning rate of the agent.

The agent models the completion of each subgoal as a stochastic process. At each attempt, the subgoal advances with probability M_r , where r denotes the resource. The simulation proceeds recursively until the subgoal (a specific target number for the resource r) is completed or until a maximum of 25 attempts is reached. The value of pursuing a given subgoal in a goal is given as follows:

$$\gamma = \frac{1}{1 + k * C_{sg|g}} \quad (3.14)$$

$$Q_{sg|g} = R_{sg|g} * \gamma \quad (3.15)$$

where $C_{s|g}$ is the number of rounds that the stochastic simulation process estimates would be needed to finish the subgoal given a goal, γ is the hyperbolic discount factor parameterized by k , $Q_{sg|g}$ is the value of the subgoal given the specific goal, $R_{sg|g}$ is the reward for subgoal completion. We assumed $R_{sg|g}$ to be identical across subgoals.

We compute the overall prospective value for the goal based on the estimate of the number of attempts it takes to complete the goal as follows:

$$Q_g = R \frac{1}{1 + k \sum_{sg'} C_{sg'|g}} \quad (3.16)$$

where R is the goal completion reward identical for all goals. $\sum_{sg'} C_{sg'|g}$ gives the number of rounds needed to finish all subgoals, and therefore the overall goal. If a specific goal has many resources already filled, the number of rounds it would take to finish pending subgoals, and then the overall goal is lower, making the goal value higher. Thus, the prospective agent gives a preference for progress when valuing goals and subgoals.

The prospective agent makes goal and subgoal choices according to the hierarchical goal selection algorithm described earlier.

TD-persistence

TD-persistence algorithm samples resources to infer the dominant goal and persists with it until the block shifts, and repeats the process when a new block begins.

TD-persistence estimates resource probabilities (M_r) through experience, just like the prospective agent (equation 18). Goal value is estimated as the minimum resource bottleneck faced by a specific goal as follows:

$$Q_g = \min_r(M_r) \quad (3.17)$$

Subgoal value is estimated as its resource probability.

$$Q_{sg|g} = M_{sg} \quad (3.18)$$

Goal and subgoal choices are made as per the hierarchical goal selection algorithm.

Mixing strategies at levels of hierarchy

We considered the possibility of a combination of strategies for goal and subgoal selection. We included agents that, for instance, use TD-momentum for goal selection and use prospective computations for the subgoal agent (TD-momentum(G)-prospective(SG)).

Retrospective

We defined a retrospective algorithm that values goals and subgoals based on the number of resources needed for completion. It does not take into account the resource probabilities, so its value is related only to progress in goals and subgoals.

$$Q_{sg|g} = R_{sg|g} * \frac{1}{1 + k * v_{sg|g}}; \quad Q_g = R \frac{1}{1 + k \sum_{sg'} v_{sg'|g}} \quad (3.19)$$

where $v_{s|g}$ is the number of resources needed for the subgoal sg to be completed. Goal and subgoal choices are made as per the hierarchical goal selection algorithm.

Progress

While the retrospective algorithm values the count of resources remaining for completion, the *progress* model values fractional progress in goals and subgoals. This

specifically dissociates subgoal value as defined by the retrospective agent, as subgoals in different goals have different requirements.

$$Q_{sg|g} = m_{sg|g}; \quad Q_g = m_g \quad (3.20)$$

where $m_{sg|g}$ is the fractional progress of subgoal sg in goal g , m_g is the fractional progress of g . Goal and subgoal choices are made as per the hierarchical goal selection algorithm.

Resources

The *resources* model deviates from the hierarchical goal selection algorithm. The model finds the easiest-to-attain resources and pursues them till the resource inventory limit is reached before moving to the next one.

$$Q_{sg} = M_{sg} \quad (3.21)$$

Here, the subgoal value is not conditioned on the goal. Goals are selected with a value based on how closer they are to completion: $Q_g = m_g$, where m_g is fractional progress. Goal selection does not play any role in resource selection, and therefore, there is no hierarchy.

We included a choice kernel parameter to ensure perseveration of goal and subgoal selections.

Fig. A1 shows a comparison between models in terms of task performance when parameters are optimized for accumulated rewards. *TD-persistence* and *prospective* agents are the top-performing algorithms, while retrospective and progress models are the lowest-performing ones. TD-momentum and TD-momentum-alt-goal give middling task performance.

3.8 Model comparisons

The momentum-based account offers a better fit to goal and subgoal choice data

We contrasted the momentum model with the alternative hypotheses described above. We optimized the model parameters to maximize the aggregate likelihood of goal and subgoal choices made by the participants. We corrected for model complexity using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) metrics.

The momentum-based algorithms applied to both goal and subgoal selection provide a better fit for the aggregate choice likelihood over the alternative hypotheses (Figure 9A). The momentum model equipped with cross-goal transfer best accounted for the choices in the experiment. We also compared models separately for the two halves of each block to see if there is any within-block variation of strategies. TD-momentum outperforms all other models in the second half and could indicate lowered update for alternate goals towards the end of the block.

While momentum provides a better aggregate likelihood of participants' choice data, several other models, especially the progress model, have aggregate likelihoods in close proximity to the momentum model. To further delineate the models, we look to see whether the models can replicate behavioral signatures in the aggregate participant pool.

The momentum-based account replicates participants' sensitivity to both goal progress and rates of progress

We replicated individual participants' goal and subgoal selection patterns in relation to goal and subgoal progress using all the models. We have the following categorization of algorithms in replicating behavior.

1. *Can replicate participants' sensitivity to both progress and rates of progress:* momentum variants, prospective (see Fig. A2)
2. *Can replicate participants' sensitivity to progress:* progress, retrospective (see Fig. A3)
3. *Can replicate participants' sensitivity to rates of progress:* TD-persistence, resources (see Fig. A4)

None of the models account for subgoal selection patterns seen in behavior satisfactorily beyond capturing general trends (see Fig. A5). Participants' subgoal selections are more influenced by the block types than predicted by any of the models. While we show that momentum accounts for aggregate choice likelihood and participants' goal selection patterns better than other models, it is still an incomplete account that is unable to capture the complexity of human subgoal selection patterns.

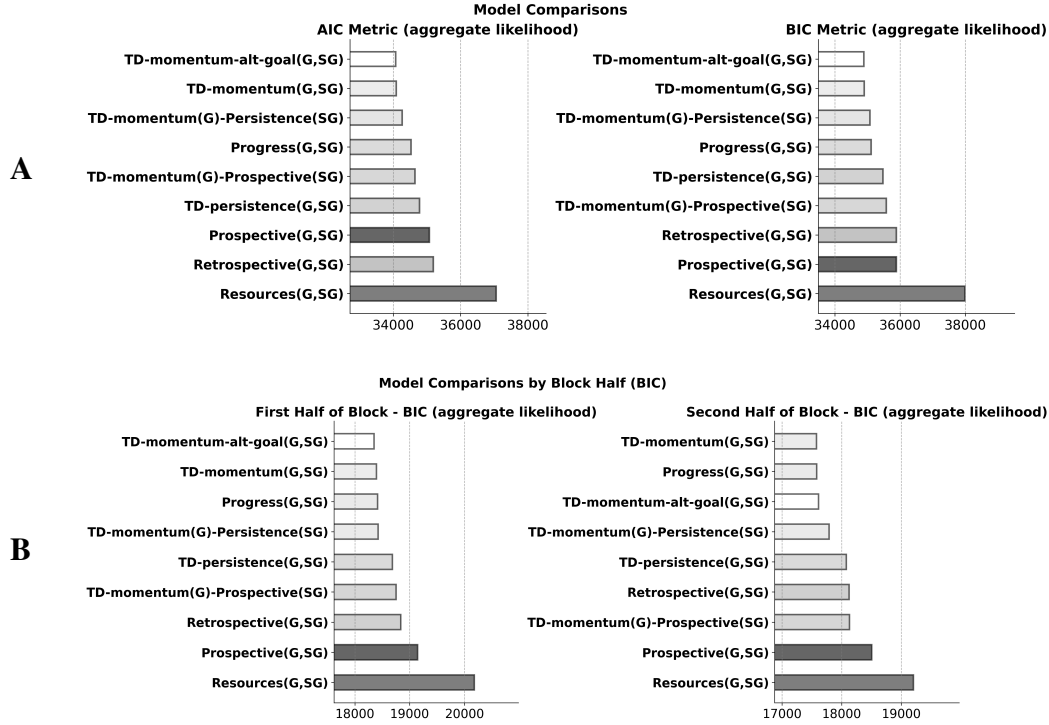


Figure 3.8: **Model comparisons.** **A.** TD-momentum with cross-goal transfer applied at both goal and subgoal stages performs the best in aggregate goal and subgoal selection AIC and BIC metrics. **B.** Model comparison by block half showing performance across different phases of the experiment.

3.9 Dissociating momentum computations and switching costs

The momentum-based account includes two factors influencing goal and subgoal selection decisions: momentum advantage (a measure of decision value) and dynamic switching costs. Both are included additively in the equation that formulates the decision to persist or switch. We now dissociate the extent to which each of these factors influences the choices as well as the reaction times of goal selection.

TD-momentum is fit to participants' choices using a depth-first hierarchical algorithm. The probability of persistence is computed using the additive sum of the momentum advantage and a negative of the dynamic switching costs. We generated individual regressors for momentum advantage and switching costs for each round in each participant using the best-fitting parameters in the momentum model.

We fit a random mixed-effects logistic regression model to participants' goal-switching choices, including momentum advantage and switching costs as predictors to assess their influence on goal persistence (Figure 10A). The two standardized regressors were uncorrelated (correlation, $R = 0.006$). We found that

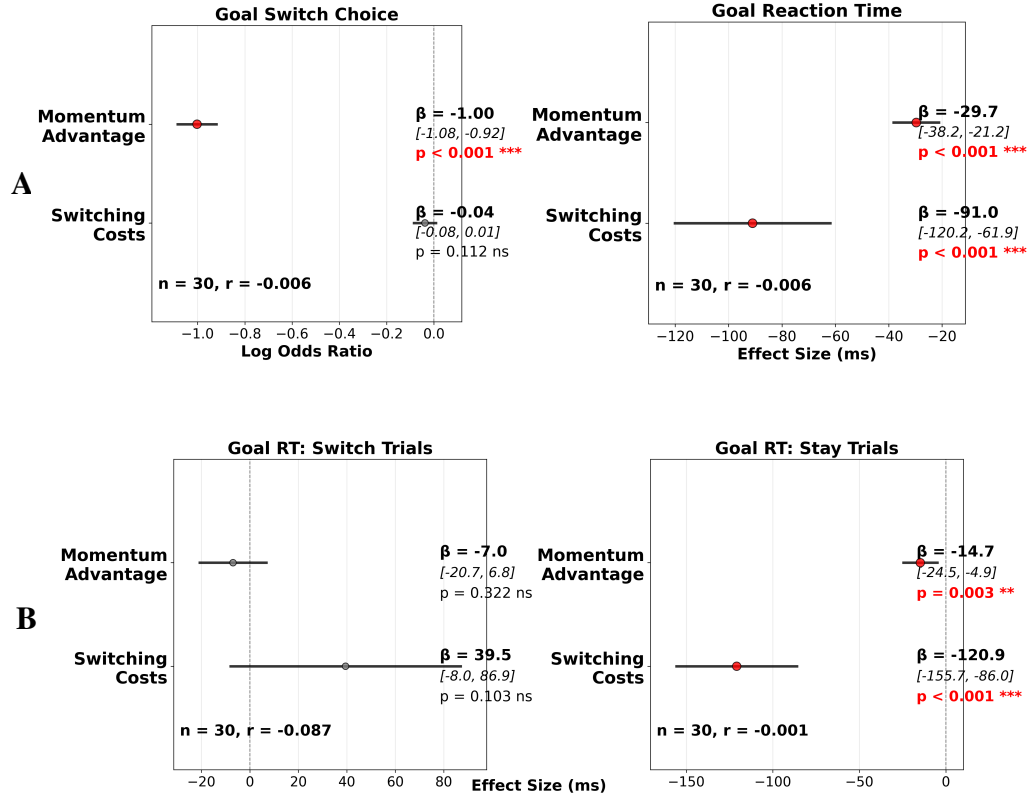


Figure 3.9: **A. Momentum and switching cost effects on goal choice and RTs.** (left) Momentum advantage largely influences participants' decisions to stay or switch in a trial over dynamic switching costs. (right) Both momentum and switching costs affect goal selection RTs, but switching costs have a higher effect on reaction times. **B.** Switching costs differentially modulate reaction times in stay and switch trials. The dynamic opportunity cost of staying incorporated into switching costs significantly reduces reaction times in stay trials.

both momentum advantage and switching costs had a significant effect in influencing participant switching choices over switching costs (log-odds ratio; $\beta_{\text{momentum}} = -1$; $p < 0.001$; 95%CI : $[-1.08, -0.92]$, $\beta_{\text{switch cost}} = -0.04$; $p = 0.1$; 95%CI : $[-0.08, 0.01]$). Momentum advantage had a large negative effect on switching, i.e, if the advantage in momentum is higher than that of other goals, participants preferred not to switch away from the current goal. Likewise, if switching costs are higher, participants preferred to stay with the current goal.

We also fit a random mixed-effects linear regression model to participants' goal reaction times, including the two regressors. Again, both momentum and switching costs had significant negative effects on the RTs, however, here, switching costs had a higher effect on RTs (effect size; $\beta_{\text{momentum}} = -29.7$; $p < 0.001$; 95%CI : $[-38.2, -21.2]$, $\beta_{\text{switch cost}} = -91.0$; $p < 0.001$; 95%CI : $[-120.2, -61.9]$). Switch-

ing costs served to reduce RTs as they include a dynamic opportunity cost parameter, which increases the efficiency when persisting with the current goal. In Figure 10B, we separated goal RTs in goal switch and stay trials, and found that switching had a positive effect (albeit only trending at $p = 0.1$) on RTs on switch trials, implying RTs are higher on switch trials when switching costs are higher. On the other hand, switching costs have a large negative effect on RTs in the stay trials. Momentum advantage also served to reduce RTs, implying that, with greater momentum, participants made their choices faster, suggesting an increase in vigor closer to the goal.

Our analysis indicates the role of momentum computations in modulating persistence levels based on progress levels and goal prospects beyond efficiency considerations involved in switching tasks.

3.10 Limitations

While we offered a model of hierarchical persistence, the capability of the best-fitting model we used here to explain subgoal selection patterns remains limited. Moreover, we did not consider alternate strategies for goal selection, where an action sequence can be planned for several steps in advance (say, by committing to the goal for a few rounds and then switching). Given the heterogeneity in goal selection patterns, it is likely that diverse strategies exist and that these might be switched between via a meta-level mechanism operating dynamically throughout the task, which we did not attempt to model. We also fit the models to the goal/ subgoal selection choices, but a joint modeling of choice and reaction times could potentially constrain the model fitting procedure further, thereby helping to provide more precision in the model selection procedure, which remains a promising future direction.

3.11 Discussion

In the hierarchical goal paradigm, participants preferred to persist depth-first through goals and subgoals. They preferred to continue in a goal – subgoal pair until termination, rather than switch between subgoals in a breadth-first manner. Although not a ubiquitous pattern across participants, we observe that those who show a deeper depth-first pattern of persistence in the task also perform better in the game and report higher task engagement

Depth-first persistence in hierarchically organized goals can be an adaptive strategy for building options—motor sequences that can be optimized for rapid, efficient actions in service of goals. We observe that participants are much more efficient at

repeating goal–subgoal pairs, as evidenced by their reaction times. Moreover, that efficiency improves with persistence: with each repetition of the goal–subgoal pair, participants get faster response times. Such practice, through persistent exploration of options, can be critical to building micro-habits and skills in humans that increase the efficiency with which we can pursue our goals [71, 46, 111, 68].

Depth-first persistence can also mitigate the cost of switching goals and subgoals. These include static reconfiguration costs involved in switching tasks, as well as dynamic opportunity costs of forgoing the efficiency built by persisting in goal – subgoal options. A higher value of commitment could thus be a result of humans factoring in effort during goal selections [157, 106, 176].

We observe that people show greater commitment as the goal and subgoal progress increase. We introduced blocks with variable prospects for different goals in the game, and they updated their persistence levels to match the prospects of the blocks’ goals. We also introduced a *high-low* experimental manipulation, in which we varied the extent to which the dominant goal within the block differed in its prospects from the other goals, and participants modulated their persistence levels accordingly.

We show that this modulation of depth-first persistence by progress and environmental shifts in hierarchically organized goals can be well accounted for by momentum computations introduced in [3]. We applied the momentum algorithm independently at the goal and subgoal stages, using a hierarchical depth-first search to compute the value of commitment towards specific goals and subgoals. We show that the momentum algorithm, at both the goal and subgoal stages of goal pursuit, best accounted for participants’ aggregate likelihoods of goal and subgoal choices in the game.

We equipped the hierarchical depth-first goal selection algorithm with dynamic switching costs as an additional value term, in addition to momentum computations. We investigated the independent effects of these two influences on driving goal persistence. We found that the momentum advantage of the current goal over alternative goals largely accounted for participants’ decisions to switch or stay in the game. On the other hand, dynamic switching costs had a more substantial effect on modulating goal selection reaction times than the momentum advantage regressor, which showed a significant, albeit lower, effect on reducing reaction times.

Although the momentum account explains the likelihood of behavior and certain general trends in behavior, it remains limited in its ability to account for the complex-

ity of the goal and subgoal selection patterns in average participant behavior, and for the inherent heterogeneity in selection patterns across participants. We therefore propose the momentum account as a useful approximation of overall human hierarchical goal selection trends, rather than as an accurate predictor of individual goal selection choices.

In this study, we only superficially address individual differences in persistence patterns, noting that people who exhibited greater persistence in subgoals within a goal also made fewer goal switches and fewer irrelevant alternate goal switches. However, a deeper analysis of the Individual goal-persistence patterns within hierarchically organized goals can significantly inform our understanding of dysfunctional switching tendencies observed in psychiatric conditions [156].

Together, these findings highlight depth-first persistence as an adaptive principle in hierarchical goal pursuit—balancing efficiency and flexibility through momentum-based commitment. By integrating hierarchical momentum computations with dynamic switching costs, our framework captures how humans flexibly regulate persistence in response to environmental contingencies and internal progress signals. Beyond explaining patterns of task engagement and performance, this approach offers a computational lens on individual differences in goal regulation, with potential relevance for understanding motivational dysfunctions in clinical populations. More broadly, insights from depth-first momentum can inform the design of adaptive systems, whether human-computer interfaces [89, 74] or embodied agents that emulate human efficiency in persistent, hierarchically structured goal pursuit.

3.12 Methods

Participants

We recruited 30 participants from the online platform, Prolific, and compensated them with a base pay of \$9 per hour, with a performance-based bonus of up to \$5. We instructed the participants that their performance bonus is directly proportional to the number of construction projects completed in the game. Participants took a median of 54 minutes to complete the training protocol and the main game.

We did not exclude any participants from the study.

The research study was reviewed by the Caltech Institutional Review Board and deemed exempt from ongoing in-depth review (IR18-0796) due to the fact that the data were collected anonymously through the online platform and because the research was considered to have no more than minimal risk. All participants provided

consent for their online participation by ticking a box on an online consent form, but did not sign the consent form, as the data collection is anonymized.

Training procedure

We had rigorous instructions and training for the participants before the start of the game. The training protocol is as follows:

1. We introduced the participants to the objective of the game and introduced them to the goals and subgoals of the game.
2. **Alien-resource matching:** Participant played an alien-resource matching game for 40 trials, where they matched the alien with the resource that it provides. We provided feedback on the correctness of the response and the correct response for each trial.
3. **Construction practice:** Participant then played 12 blocks of the construction game, where in each block, we explicitly gave them the construction goal. If, say, the construction goal was a spaceship, the participant has 10 rounds to collect the required 3 glass and 3 metal to finish the goal. In each round, participants have a choice between 4 aliens to choose from to procure the resource they need. In this game, all aliens provide resources with 100
4. Following two practice sessions, we instructed the participants on the main game, where we ran them through a few controlled trials to introduce them to the free goal selection paradigm.
5. Participants then played 12 blocks of the main game, each consisting of 30 rounds of goal and subgoal selection.

Algorithm 2 Algorithm for goal selection

```

1:  $Q_g$ : goal value where  $g \in \{\text{spaceship, observatory, house}\}$ 
2:  $\beta_g, \beta_{sg}$ : softmax temperatures for goal and subgoal selection
3:  $\kappa_g$ : goal preservation/ choice kernel
4:  $c_g$ : cost of switching between goals
5:  $c_{sg}$ : cost of switching between subgoals
6:  $G$ : current goal (none at start)
7:  $SG$ : current subgoal (none at start)
8: for block in  $(1, \text{num\_blocks})$  do
9:   Initialize card probabilities
10:  for round in  $(1, \text{num\_rounds})$  do
11:    Get goal values  $(Q_{SH}, Q_{OB}, Q_{HO})$  // Using TD-momentum/ prospective
12:    Calculate advantage of current goal  $G$ :  $A_G = Q_G - \max_{g \neq G} Q_g$ 
13:    Choose to stay with/ switch from the current goal:  $p_{\text{stay};G} = \frac{1}{1 + \exp(-c_G - \kappa_G - \beta_g A_G)}$  // Choose goal
14:    In case of a switch, chose the new goal according to goal values:  $\text{softmax}_{\beta_g}\{Q_{alt_1}, Q_{alt_2}\}$ 
15:    Update current goal  $G$ 
16:    Update choice kernel of goal selection
17:    Get subgoal values  $(Q_{glass|G}, Q_{metal|G}, Q_{stone|G}, Q_{concrete|G})$ 
18:    Calculate advantage of current goal  $G$ :  $A_{SG|G} = Q_{SG|G} - \max_{c \neq SG} Q_{c|G}$ 
19:    Choose to stay with the current subgoal or explore  $p_{\text{stay};SG|G} = \frac{1}{1 + \exp(-c_{sg} - \beta_{sg} A_{SG|G})}$  // Choose subgoal
20:    In case of a switch, explore a new subgoal according to subgoal values
21:    Update current subgoal  $SG$ 
22:    Observe outcome
23:    Update goal-outcome probabilities // For prospective/ TD-persistence
24:    Update goal values
25:    Update subgoal values
26:  end for
27: end for

```

Model fitting and parameter estimation process

We computed both goal and action selection likelihoods for each round as follows:

Goal likelihood: The likelihood is computed as the probability of staying and switching with the previously selected goal.

$$\Pr(\text{Goal}_t = g | \text{Goal}_{t-1} = G) = \begin{cases} p_{\text{stay};G}, & g = G \quad (\text{stay}) \\ (1 - p_{\text{stay};G}) * \text{softmax}\{Q_{alt_1}, Q_{alt_2}\}[g], & g \neq G \quad (\text{switch}) \end{cases}$$

Subgoal likelihood:

Subgoal selection likelihood for a given subgoal is computed as a marginalized probability of all possible goal selections

$$\Pr(\text{Subgoal}_t = sg | \text{Goal}_{t-1} = G, \text{Subgoal}_{t-1} = SG) \quad (3.22)$$

$$= \sum_g \Pr(\text{Goal}_t = g | \text{Goal}_{t-1} = G) * \Pr(\text{Subgoal}_t = sg | \text{Goal}_t = g, \text{Goal}_{t-1} = G, \text{Subgoal}_{t-1} = SG) \quad (3.23)$$

$$= \Pr(\text{Subgoal}_t = sg | \text{Goal}_t = g, \text{Goal}_{t-1} = G, \text{Subgoal}_{t-1} = SG) \\ \begin{cases} p_{stay;SG|g}, & sg = SG \quad (\text{stay}) \\ (1 - p_{stay;SG|g}) \times \text{softmax}_r\{Q_{alt_r}\}[sg], & sg \neq SG \quad (\text{switch}) \end{cases}$$

For each individual participant, we estimated the parameters for each of the competing models to maximize the aggregate likelihood of their goal and subgoal choices by employing a tree-structured Parzen estimator (TPE) sampler, a Bayesian optimization algorithm implemented in optuna [6], an open-source hyperparameter optimization framework in Python. We estimated the parameters that maximized the likelihood over 100 random starts.

Model and parameter recovery process

For parameter recovery analysis, we sampled 500 parameter vectors from a range of parameters ($\alpha \in [0, 1]$, $c_g \in [-1, 1]$, $c_{sg} \in [-1, 1]$, $\beta_g \in [0, 10]$, $\beta_{sg} \in [0, 10]$, $\gamma \in [0.6, 1.0]$, $\alpha_c \in [0, 1]$). We generate synthetic behavior for each parameter vector using the TD-momentum-learn-alt-goal model and recover the parameters by fitting the model to the simulated data. We optimized both the likelihoods of selection of the goals and subgoals while fitting the data. We plotted the fitted parameters and the generated parameters and reported correlations. Most of the parameters are recoverable with high correlation coefficients (see Fig A6). However, low discount factors closer to 0.5 do not recover well; however, the discount factors of the participant fits are mostly in the range $[0.9, 1.0]$, so we considered the range $[0.6, 1.0]$ for parameter recovery.

We note that learning rate, choice kernel, and goal switching costs show biases in parameter recovery. There is an underestimation of the learning rate and an overestimation of the choice kernel in parameter recovery. We tried eliminating the choice kernel, which did not improve the recovery of the learning rate but rather reduced the recoverability of other parameters. We employ the model with the best overall recovery over the variants of models experimented with.

We performed a model recovery analysis on the competing models. We sampled 100 random parameter vectors from parameter ranges described above. For each of the algorithms described in the manuscript, we generated choice data using the sampled parameters. We then fit the generated data to each of the models and identified the winning model through BIC score comparisons. We plotted the confusion matrices where the rows correspond to the simulated model and the columns correspond to the recovered model. The left confusion matrix indicates the probability of the fitted model given the simulated model, whereas the right matrix indicates the probability of the simulated model given the fitted model (see Fig A7).

We note that while model recovery for the two high-performing momentum variants is reasonably good, a significant portion of parameter samples is confused with the more parsimonious *progress* model. The prospective model is the least recoverable mostly confused with the retrospective model.

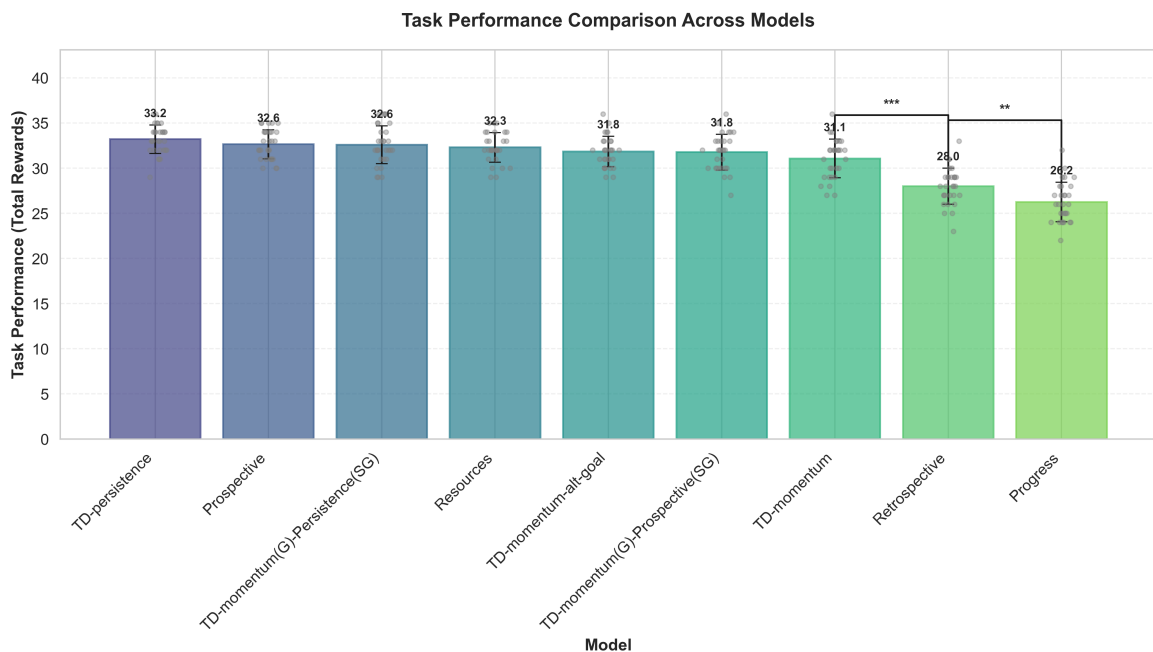


Figure 3.10: **Performance comparisons among models:** Model parameters are optimized for task performance.

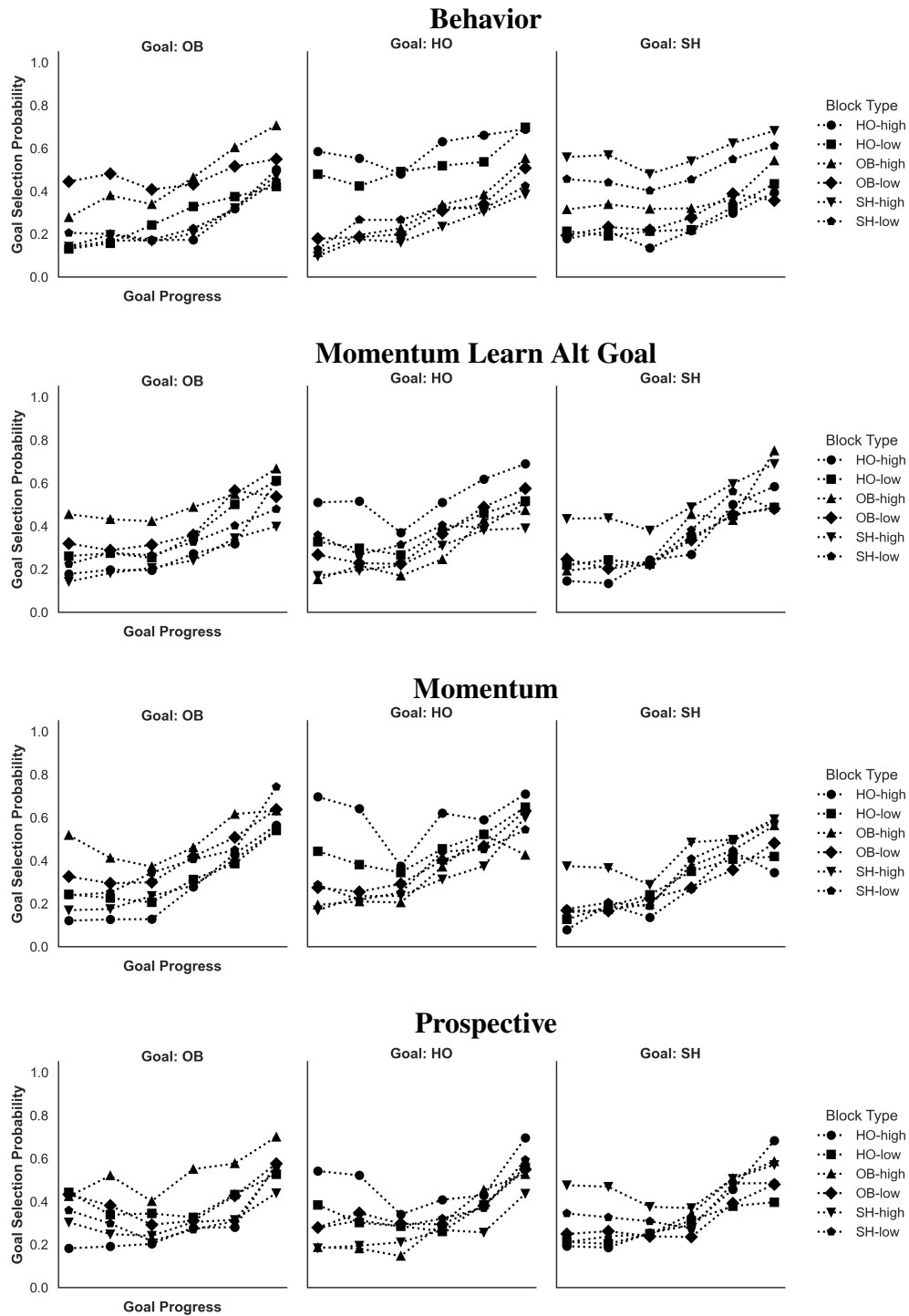


Figure 3.11: **Replicating both sensitivity to progress and rate of progress.** Goal selection probability relation to progress, comparing behavior with model simulations. Each row shows the pattern for different models fitted to individual participant choices.

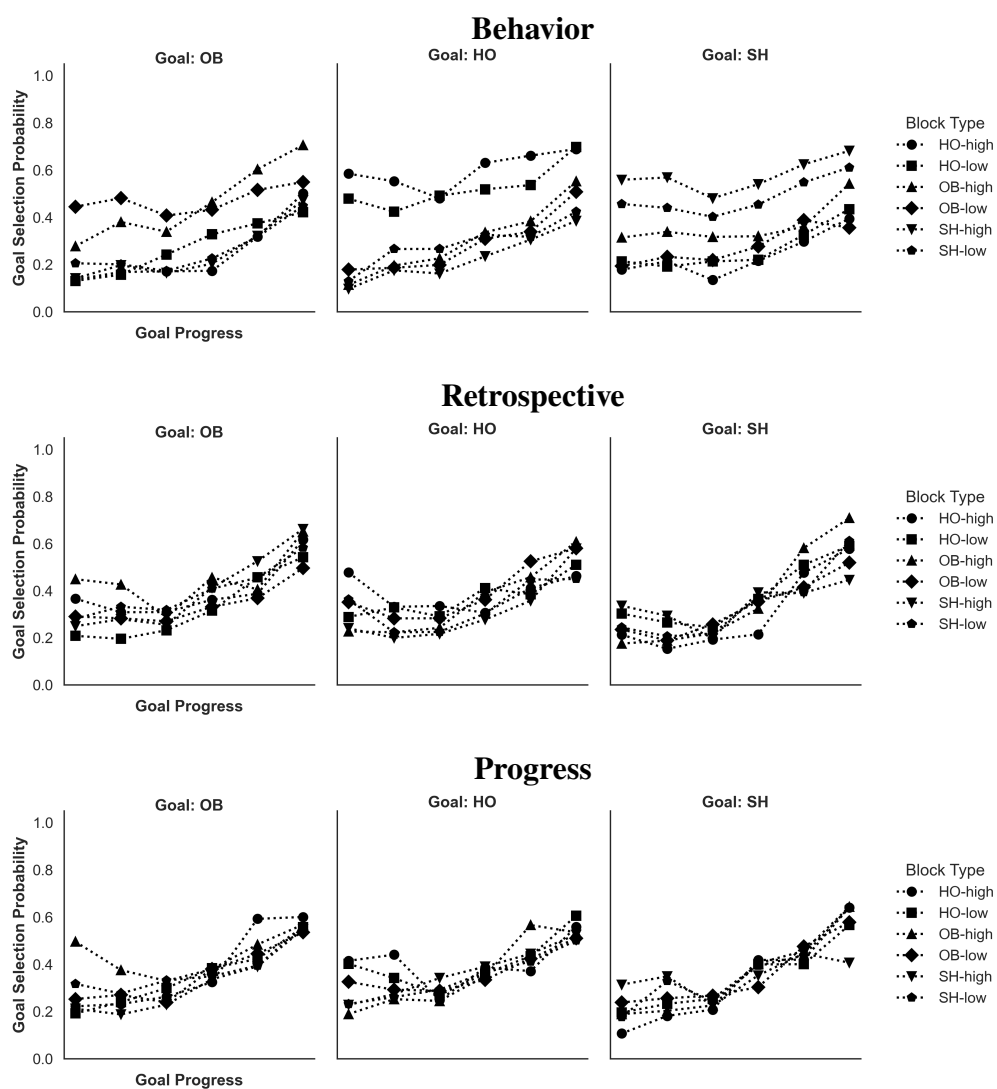


Figure 3.12: **Replicating sensitivity to progress.** Goal selection probability relation to progress, comparing behavior with model simulations. Each row shows the pattern for different models fitted to individual participant choices.

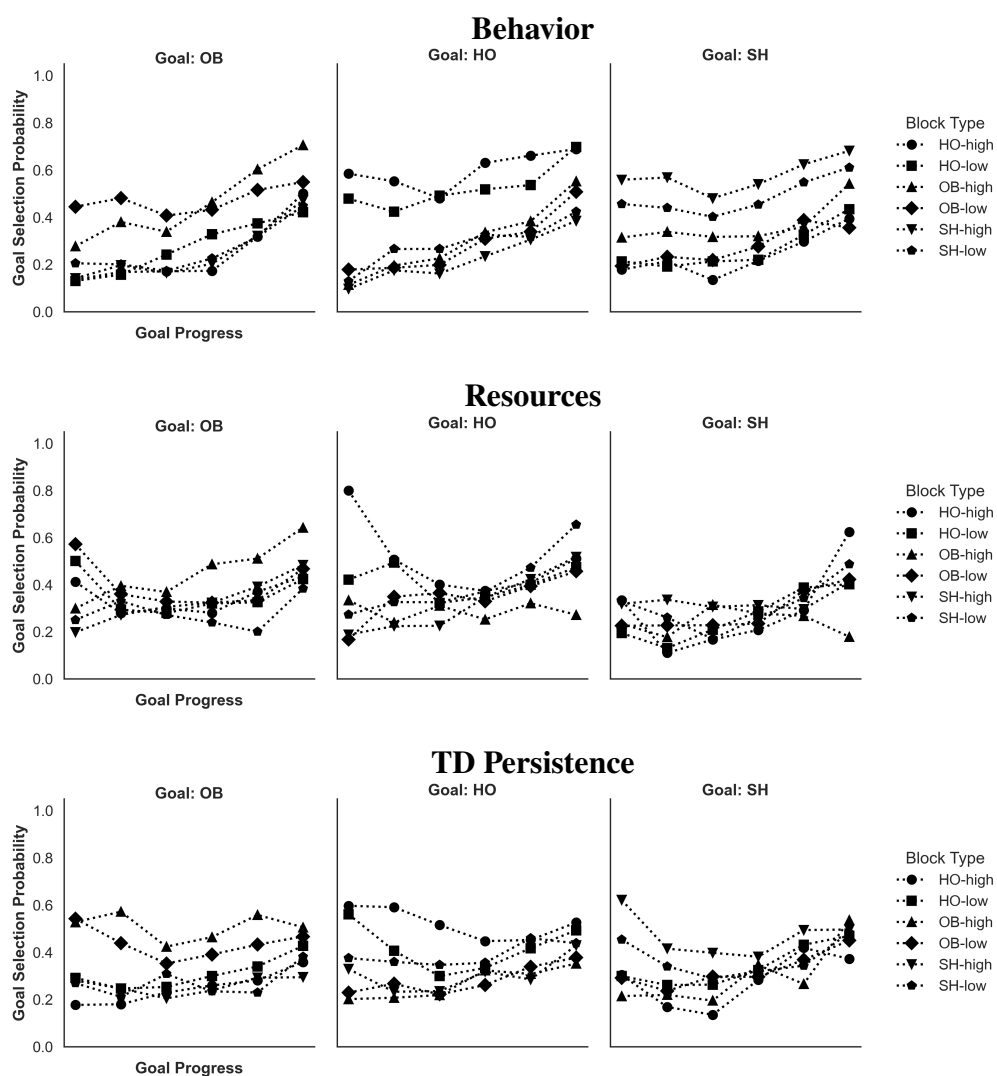


Figure 3.13: **Replicating sensitivity to rate of progress.** Goal selection probability relation to progress, comparing behavior with model simulations. Each row shows the pattern for different models fitted to individual participant choices.

Subgoal Selection Probability by Progress

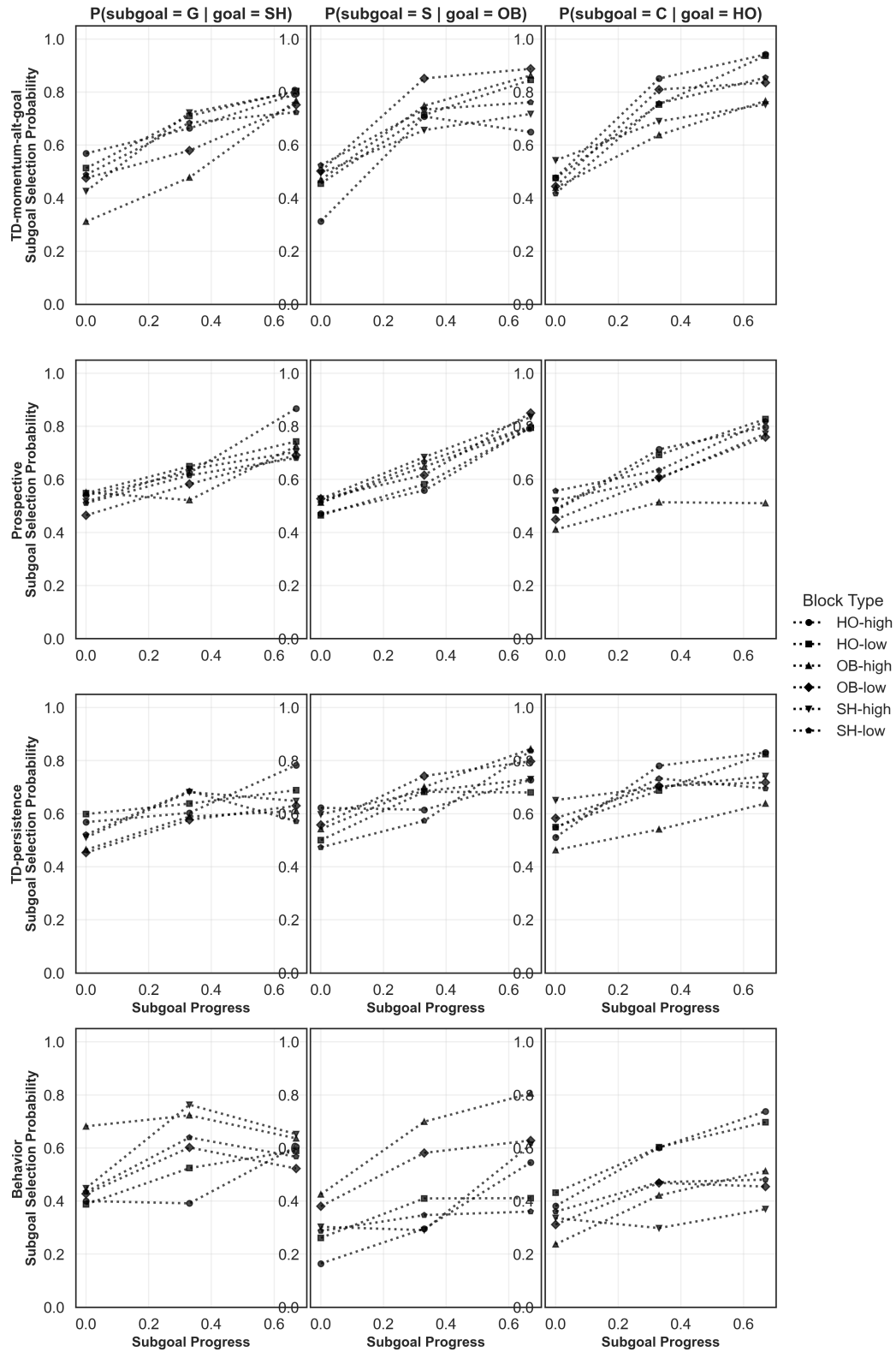


Figure 3.14: Comparing model predictions of subgoal selection probability relation to progress with behavior. All models are fitted to individual participant choices.

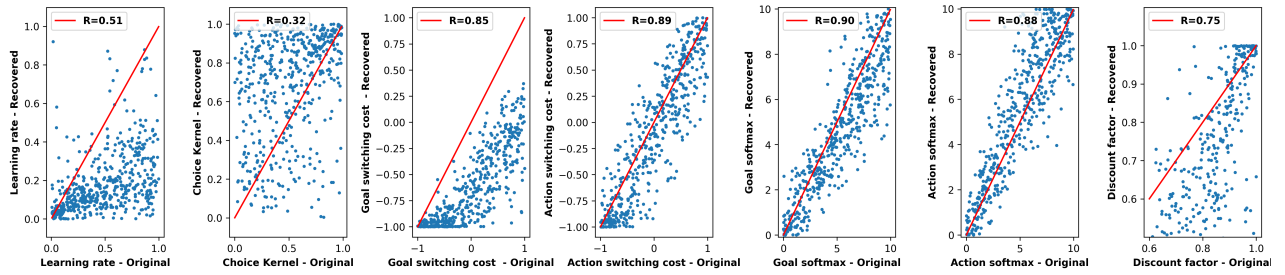


Figure 3.15: Parameters of the TD-momentum-alt-goal algorithm recovered through 500 independent model simulations.

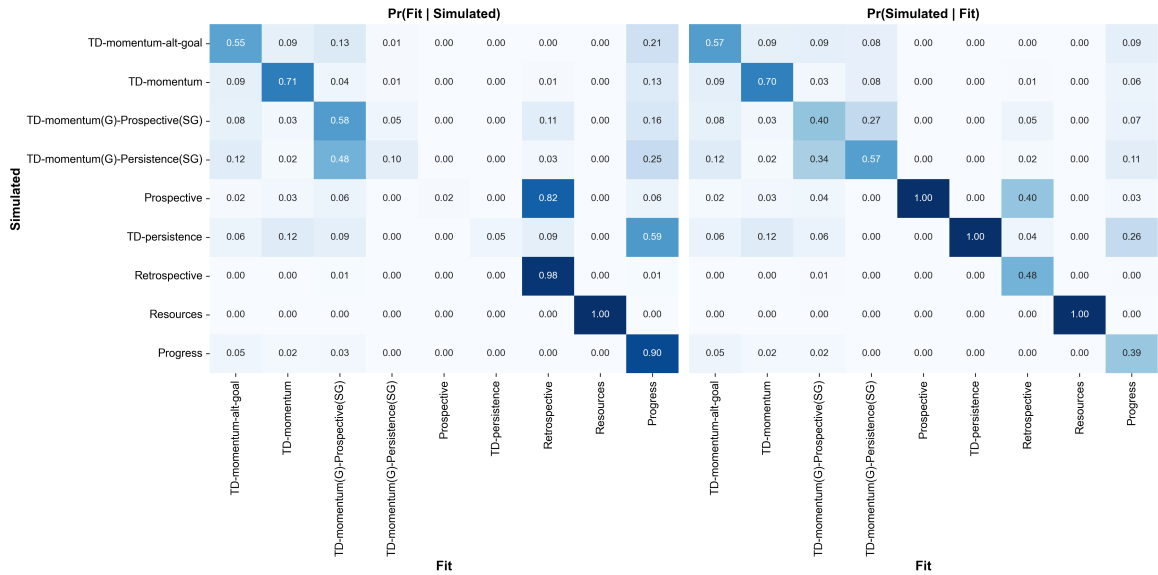


Figure 3.16: Model recovery of competing hypotheses through 100 independent simulations.

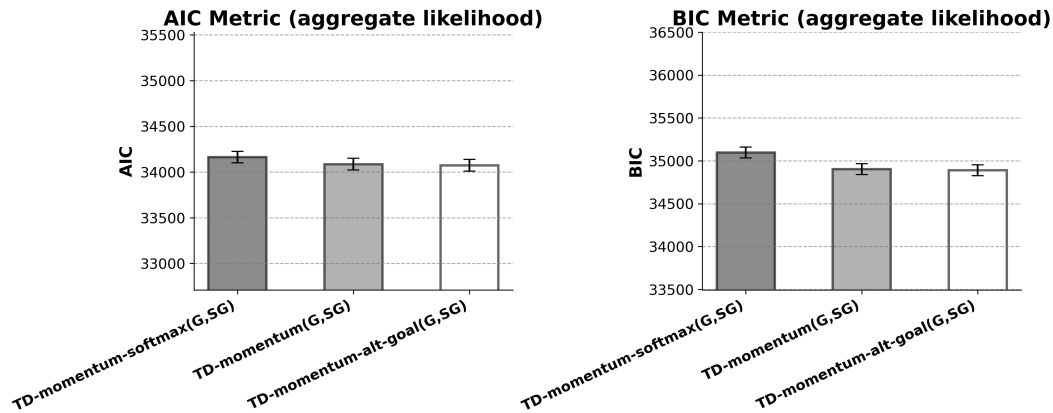


Figure 3.17: Comparing the best performing models with different modes of goal/subgoal selections (advantage of current option over the best alternative vs softmax over values of three options).

*Chapter 4***THE NEUROANATOMICAL BASIS OF LONG-TERM GOAL PURSUIT**

Humans set extended goals to obtain rewards that lie in distant states. To achieve such goals, humans act persistently, motivated by the reward associated with the goal state. However, environmental circumstances can sometimes make the pursuit of such goals difficult, and in these situations, it may serve the individual better to switch away from the goal until conditions become favorable again. Such flexible decision-making in the pursuit of extended goals would require a specialized neural architecture that can 1. maintain persistent goal states to generate actions that advance a goal, 2. actively monitor the environment for the viability of reaching the current goal state, 3. keep track of competing goals that may be more beneficial to pursue, and 4. facilitate switching to an alternate goal when it exceeds the current option in value. In this chapter, we review findings from cognitive psychology and neuroscience to trace the architecture in the brain that enables long-term goal pursuit and examine how such an architecture manifests in the behavioral patterns observed in human goal selection.

4.1 Introduction

Most of our decision-making in the real world is in service of goals across multiple timescales. Some of these goals are one-shot—for example, deciding between a mango or a lychee juice. However, many of our goals unfold over extended durations, sometimes even across decades. For instance, if your goal is to become a professor, it may require long, grueling years of graduate school, a stint as a postdoc rife with uncertainty, and stressful rounds of faculty applications. Throughout this process, you retain your motivation, even when tempted by alternative professions in industry that offer better work-life balance and higher pay. Some persevere, motivated by the rewarding state of academic freedom that comes with being a professor, while others decide that the end state is not worth the sacrifices it demands and quit the journey midway. In either case, we constantly weigh the consequences of persevering with the current goal against the benefits of dynamically shifting to alternative opportunities afforded by the environment. For example, if scientific funding is cut by the current political government, you may carefully evaluate your

options—staying committed to the goal and waiting out the uncertainty, moving to a different country and returning when conditions improve, or quitting science entirely and transitioning to industry.

Making decisions over such long timescales requires an elaborate neural architecture that can manage, maintain, and evaluate competing goals over extended durations. First, to support persistent activity toward a specific goal, the brain must maintain representations of goals and goal-related decisions over long periods of time. Second, the brain must monitor the environment for the viability of both active and alternative goals. Third, when circumstances warrant a switch, the brain must reconfigure its task representations to favor pursuit of a new goal. How the brain organizes its architecture to accomplish these tasks in service of goal pursuit remains an outstanding question in neuroscience. Compared to value-based decision-making research, which has largely focused on one-shot decisions, extended goal pursuit remains relatively understudied. However, domains such as reward-modulated learning, executive function, cognitive control, and task switching share common functions with extended goal pursuit, and findings from these areas can therefore shed light on the neural architecture underlying long-term goal pursuit.

In this chapter, we draw on findings from these domains of neuroscience to outline how goals are encoded, maintained, and adapted in the brain. First, we review behavioral signatures of extended goal pursuit in humans and consider what they may imply about the underlying architecture that produces such behavior. Next, we review neuroscientific literature that suggests neural correlates underlying these behavioral patterns in the service of goals. We also discuss how architectural differences may contribute to individual differences in goal control in humans. Finally, we conclude by proposing a path toward validating the neuroanatomical architecture that subserves this complex cognitive function in humans.

4.2 What it takes to pursue extended goals

In this section, we outline the key processes required for pursuing long-term goals, and in later sections we elaborate on the behavioral and neural signatures that may speak to their mechanistic implementation in the brain.

To pursue long-term goals, we first require an architecture capable of sustaining representations of goals and the actions pertaining to those goals over timescales relevant to the duration of pursuit. If the duration of a goal spans days, months, or

years, we would need a mechanism that can interrupt the goal and later retrieve it with its value coding intact.

Second, we require a mechanism that can dynamically modulate goal persistence based on the viability of the goal within a given environment. If conditions are not favorable for pursuing the goal, its value must be updated so that a decision can be made either to interrupt the goal temporarily or to abandon it altogether.

Finally, there must be an architecture for monitoring alternative goals in the environment, such that if another goal becomes more imminent or valuable, we can switch to pursuing it instead.

These three architectural requirements must be dynamically balanced so that persistence and flexibility can jointly govern how we manage multiple long-term goals across the course of a lifetime.

4.3 Maintaining persistent goal states

Because extended goals are distant rewarding states that require persistent actions for their resolution, they necessitate the maintenance of a goal state, at least during the period in which one is acting in service of that goal. Here, we review findings from studies of free goal selection that shed light on how humans persist in goals.

Goal persistence is modulated by progress

We deem long-term goals successful (rewarded) only when we reach the terminal state (for example, landing a faculty job), with no rewards awarded for attempts alone. However, we still receive reinforcements along the way that inform us how likely we are to accomplish our goals. Goal progress is one such marker, indicating how close we are to achieving a goal and serving as an intrinsic motivator for goal pursuit. A popular formulation of this theory, the “goal-gradient” hypothesis, proposes that humans and animals increase their movement vigor—often considered a metric of motivation—as they come closer to the goal state [86]. In mice, this increase in movement vigor is accompanied by ramping dopamine activity, lending support to the motivational role of goal progress. Similar effects have also been observed in real-world human behavior, where participants in a café reward program increased their coffee purchasing rate as they approached the reward threshold [99].

In a previous study, we developed a free goal selection paradigm in which participants could switch between multiple extended goals at any time. The goals required persistent activity—the collection of a specific target number of tokens—to obtain

a reward. Using this paradigm, we showed that humans demonstrate a significant preference for progress when choosing between competing goals. Participants not only increased their persistence toward goals that were closer to reward completion, but also revisited them at a higher rate after having switched away from them. This suggests an attractor-like quality associated with proximity to reward, a tendency previously documented in humans under the goal-gradient hypothesis. Similar findings have been observed in real-world settings, where participants in a café reward program increased their coffee purchasing behavior as they neared the reward threshold. Importantly, progress need not be explicit to influence goal selection. In tasks where participants had to discover the correct combination of steps required to achieve a goal, goal selection favored options with latent progress, where actions progressively narrowed the search space of possible solutions and brought participants closer to identifying the correct sequence. Comparable effects have also been documented in animals: in mice, dopamine activity ramps up during progress toward a goal, with both movement vigor and motivational drive increasing closer to reward attainment.

However, visible progress is not the only factor that can motivate goal perseverance. Molinaro et al. demonstrated that latent progress—where advancement occurs not through success but through failure—can also drive goal selection [128]. In a game where participants had to identify the correct combination of ingredients to make a potion, Molinaro et al. showed that failed attempts narrowed the search space of possible combinations, thereby constituting latent progress toward discovering the correct solution. This latent progress, in turn, influenced goal selection. Such findings may explain why individuals persist in goals even when they appear externally stalled (for example, while struggling with a math problem), as the mind may still be progressing latently by eliminating incorrect possibilities.

One of the most prominent behaviors associated with extended goals is the tendency to overpersist, especially in goals with substantial accumulated progress. Here, overpersistence is defined relative to the maximization of expected returns. This phenomenon has been widely documented in sunk-cost settings, where quitting entails forfeiting accrued progress, thereby leading to suboptimal persistence in failing goals. Interestingly, we observed this tendency to overpersist even in the absence of explicit sunk-cost considerations, in contexts where participants could retain progress across goals even after switching away from them. Despite having the option to switch between goals and return to them later, participants still exhibited a

tendency to overpersist in goals with accumulated progress. It is possible that accumulated progress functions similarly to the endowment effect, whereby individuals assign greater value to a goal based on the effort already invested in pursuing it.

Participants are also more likely to sustain a goal state over extended durations. In [3], we found high temporal autocorrelation in participants' goal selections. In [2], where extended goals were divided into multiple subtasks, participants maintained persistent subgoal states across adjacent rounds. The extent to which participants persisted in a goal was modulated both by the amount of progress made and by the rate at which progress was achieved.

Sensitivity to progress rates

In our free goal selection paradigm, we explicitly manipulated rates of progress across goals, such that the goals with the greatest accumulated progress were sometimes also the slowest to reach the target. When a goal's rate of progress is high, it can rapidly catch up and achieve the desired target even if it initially lags behind in accumulated progress. In such situations, the optimal strategy would be to select goals that are likely to yield the highest discounted returns. However, we observed suboptimal persistence toward goals with greater accumulated progress.

Nevertheless, participants remained sensitive to differences in progress rates across goals. On average, participants were more likely to choose goals associated with the highest rates of progress, and they adjusted their choice proportions depending on how much better the best progress rate was relative to the next-best alternative. Thus, although individuals learn about environmental conditions and adapt to them, they still maintain a bias toward accumulated progress.

Adaptive features of persistence

Humans and animals often persist suboptimally in extended behaviors—at least as measured by reward acquisition in narrowly defined laboratory tasks. However, considering the mechanistic basis of goal pursuit suggests that persistence itself may be an adaptive feature and can inform investigations into the neural architecture underlying extended goal valuation in the brain.

Avoiding switching costs. There are likely costs associated with switching tasks. When transitioning to a new goal, task representations such as action policies or cognitive maps must be reconfigured to support pursuit of the new objective. This process may require retrieving associations from long-term memory, thereby

incurring cognitive costs and increasing response latency. Several studies have demonstrated increased reaction times when participants switch tasks [129], which may reflect the costs associated with task reconfiguration. These switching costs are themselves modulated by contextual factors, rewards, and effort incentives. It has also been observed that switching costs are reduced in environments characterized by higher rates of switching, suggesting the presence of meta-level control over cognitive flexibility [130, 93]. Persistence in goals may therefore help mitigate the costs of task reconfiguration and could be adaptive from a resource-rational perspective.

Building microhabits and schemas. Persisting in goals may also facilitate the development of expertise. Sustained engagement with the same task over extended periods can strengthen associations among action policies that recur frequently during goal pursuit. Through persistent activity, individuals may develop microhabits and schemas that support more efficient behavior. Infants, for example, are known to interact persistently with objects and to exhibit repetitive behaviors, gradually building a repertoire of micro-skills that can later be composed efficiently in the pursuit of more complex goals.

Reducing interference from distractors. A resource-rational agent cannot efficiently manage multiple competing tasks without compromising performance on individual goals. Persistence in a single task may reduce attentional allocation to competing alternatives and thereby enhance efficiency in the active goal. [82] proposed that goal commitment in humans is facilitated by a selective attention mechanism that reduces distraction from alternative tasks.

[66] found selective persistence in representations guided by intentions. When participants were asked to study two scripts and informed that they would later need to execute one of them, their performance on an intervening word-recognition task was higher for words belonging to the to-be-executed script. However, when participants were instructed merely to observe someone else execute one of the scripts, this effect disappeared.

Social signaling. Persistence can also function as a stable social signal indicating active pursuit of a particular goal, thereby helping individuals form communities organized around shared objectives. Consistent progress toward a goal may additionally signal competence and enhance reputational value. Social motivators such as competition can, in fact, be among the strongest drivers of persistence in specific goals.

Moreover, in the presence of sunk costs, switching away from a goal may itself signal failure, which can be intrinsically aversive. World leaders, for example, have at times continued military campaigns in stalemated situations to avoid appearing weak.

Maintaining persistent states for active goals in working memory

To actively pursue an extended goal, the goal must be maintained in a sustained manner within working memory while action policies relevant to that goal are retrieved and executed. Several studies have identified an active role for the prefrontal cortex in working memory [57, 55, 64]. Using delayed-response tasks, these studies showed that prefrontal neurons increase their firing during delay periods, with firing rates correlating with the stimulus information being maintained in memory. Delayed-response tasks themselves constitute a form of extended goal, in which the animal must correctly identify the relationship between events separated by a delay. Accordingly, we can expect similar working-memory mechanisms to support the active pursuit of extended goals more generally.

Maintaining a persistent state in working memory is thought to depend on prefrontal–parietal interactions [48, 88, 29, 36]. Within this circuit, representations related to abstract goals first emerge in the prefrontal cortex before being transferred to the parietal cortex, where stimulus-specific representations are extracted [184]. Indeed, several studies have failed to find stimulus-specific encoding in the prefrontal cortex while observing robust representations in the parietal cortex, leading to the view that the prefrontal cortex acts as a top-down control signal guiding parietal representations that maintain goal-specific content [108]. Beyond the frontoparietal axis, the prefrontal cortex also interfaces with the hippocampal formation and medial temporal lobe to flexibly recruit task-specific memory representations. When a task imposes memory demands, prefrontal representations become phase-locked to field potentials in the medial temporal lobe, whereas tasks without explicit memory requirements engage primarily the medial prefrontal cortex [123].

Within the prefrontal cortex, the vmPFC appears to play a particularly important role in sustaining commitment to extended goals and in representing the subgoals necessary for goal execution [69]. Patients with vmPFC lesions exhibit significantly reduced goal commitment, indicating a role for this region in maintaining persistent goal states [82]. The vmPFC has been proposed to facilitate goal commitment through selective attention mechanisms, attenuating distracting features in sensory

space while directing attention toward task-relevant dimensions. Consistent with this account, working-memory representations have been shown to be distributed not only across frontoparietal regions but also within early sensory areas such as the visual cortex [72], suggesting that goal maintenance is implemented partly through top-down modulation of sensory processing.

In summary, the prefrontal cortex, through its connections with the parietal cortex and medial temporal lobe, configures task-specific memory representations and guides task-dependent salience via reciprocal interactions with early sensory areas, thereby sustaining persistent goal states during active goal pursuit [56]. Importantly, this capacity is not restricted to short timescales: prefrontal neurons have also been shown to maintain stable value coding for learned goals retained over many months [61], suggesting that the same circuitry supports goal maintenance across timescales ranging from seconds to prolonged periods of ongoing pursuit.

4.4 Monitoring and updating value of extended goals

Algorithms of extended goal valuation

We can compute the value of extended goals using either model-based or model-free algorithms.

Model-based algorithms attempt to capture the structure of the world in order to estimate the future prospects of goals and use that understanding to guide goal selection. Typically, this involves forward simulation of expected future trajectories of goal pursuit. For complex, multi-step goals, this can require the construction of cognitive maps that enable future rollouts of actions. Prospective computations in model-based approaches can follow the following procedure: 1. imagine the consequences of acting toward different goals using internal models, 2. select the goal with the best expected prospects and take action, and 3. observe the consequences of those actions and update the internal models accordingly.

When optimized for expected discounted returns, prospective computations can capture behavioral patterns such as preference for progress and sensitivity to rates of progress. However, such computations can be computationally expensive and vulnerable to misjudgments when based on inaccurate world models. Nevertheless, planning algorithms are known to guide human goal pursuit in complex extended behaviors [121].

Model-free algorithms, by contrast, estimate future goal prospects without relying on explicit world models or computationally expensive prospective simulations. In

previous work [3], we introduced the computational construct of *goal momentum* to demonstrate how past progress history can be used to interpolate future prospective progress in a model-free manner.

We define goal momentum as follows:

$$\text{goal momentum} = \text{goal progress} \times \text{rate of goal progress}$$

Here, the rate-of-progress parameter can be estimated using a temporal-difference learning rule, allowing the prospects of goals to be updated in response to environmental contingencies. However, goal momentum exhibits a gradual ramping-down dynamic following environmental shifts, in contrast to model-based approaches, which can produce abrupt value updates. We showed that this sluggish response of goal momentum to environmental changes can account for participants' undue preference for accumulated progress.

Momentum-based computations are computationally cheaper and can approximate prospective computations when the environment is stable or changes gradually. However, under abrupt environmental shifts, momentum computations can produce greater persistence in goals than would be predicted by prospective planning. At the same time, in volatile environments where accurate predictive world models are difficult to construct, momentum computations may in fact be optimal when both performance and cognitive costs are taken into account.

Neural correlates of extended goal valuation

While active goal representations are maintained in working memory, the value assigned to a goal must also be continuously updated based on action–outcome contingencies. The basal ganglia have been proposed to play a central role in this updating process, supported by their extensive reciprocal connections with dorsal, ventromedial, and orbitofrontal regions of the prefrontal cortex, as well as the ACC [28]. Within this circuit, the mPFC is thought to support prospective processes preceding actions, whereas the ventral pallidum provides feedback regarding action outcomes [115]. A complementary account proposes that the basal ganglia support rapid learning of stimulus–response associations that facilitate ongoing goal pursuit, while the prefrontal cortex performs slower, more gradual learning of abstract goal representations [27]. With sufficient training, the prefrontal cortex—particularly the frontopolar cortex (BA10)—abstracts categorical structure from individual stimulus–response instances, enabling generalization across multiple goal contexts [118].

Within the prefrontal cortex, distinct subregions contribute different computations to this process. The lateral OFC encodes both the value and the specific identity of expected rewarding outcomes, supporting credit assignment that links distal outcomes back to preceding choices, whereas the vmPFC encodes identity-general, common-currency value signals that integrate costs such as delay, effort, and risk into a unified representation of goal value [84]. Importantly, these value representations are not transient. Following repeated goal-relevant experience, vLPFC neurons maintain value-biased responses for many months that are resistant to extinction and interference from new learning, thereby providing a stable motivational substrate for goals pursued over extended timescales [61]. Complementing this stability, a two-rate learning system within the vLPFC enables individual neurons to simultaneously encode fast-updating and slow-integrating components, allowing goal value to remain robust in the face of temporary setbacks while still adapting to genuine changes in environmental contingencies [76].

Within the basal ganglia, value representations are similarly organized along a temporal axis. Flexible value—updated through short-term reward experience—is encoded in the caudate head, whereas stable value—accumulated through long-term experience—is encoded in the caudate tail [98]. Separate dopaminergic inputs to these regions ensure that updating signals themselves operate over different timescales. More broadly, the dorsomedial striatum mediates online learning about active goals during goal-directed behavior, whereas the dorsolateral striatum supports habit formation following extended practice [183]. A recent integrative framework proposes that cortico–basal ganglia–thalamocortical loops collaboratively shape behavioral control during goal pursuit, with habits emerging through loop-bypassing shortcuts, and further suggests that the striatum functions as an attention-like gating mechanism that selects task-relevant representations [73].

A mechanistic account of how the basal ganglia and prefrontal cortex jointly govern value updating is provided by actor–critic and working-memory gating frameworks. Frank and colleagues proposed that the PFC actively maintains goal-relevant representations in working memory, while the basal ganglia function as a selective gating mechanism that determines when those representations should be updated [51, 77]. In the extended formulation of this model, the PFC, basal ganglia, and amygdala together form an actor–critic architecture in which the critic learns which PFC representations are task-relevant and trains the actor to provide dynamic gating signals that regulate working-memory updating [140]. Neuroimaging evidence

further supports a dissociation within the basal ganglia itself, with the dorsal striatum corresponding to the actor component responsible for action selection, and the ventral striatum corresponding to the critic component responsible for evaluating outcomes [139].

For hierarchically structured goals, value learning must occur simultaneously at multiple levels. Ribas-Fernandes and colleagues demonstrated that teaching signals for subgoal values—subgoal prediction errors—are encoded in regions canonically associated with temporal-difference learning, including the nucleus accumbens for positive prediction errors and the ACC, habenula, and amygdala for negative prediction errors. These findings indicate that such regions evaluate subgoal progress independently of the terminal goal [145]. The ACC has been theorized to play an integrative role within this hierarchy by both monitoring the performance of an active goal and selecting among goal options at higher levels of behavioral organization [80]. Causal evidence for distinct contributions of the ACC and striatum to goal-value learning under uncertainty was provided by Treuting and colleagues, who found that brief microstimulation of the ACC impaired learning whereas striatal stimulation enhanced it. ACC neurons encoded error history and uncertainty, whereas striatal neurons responded most strongly during high-certainty, high-value choices [169]. This double dissociation suggests that the ACC regulates whether current evidence warrants re-evaluating a goal's value, while the striatum implements value updating once sufficient evidence has accumulated. Consistent with this monitoring role, ACC ensemble activity has also been shown to encode temporal information by tracking the accumulation of experiences across extended periods, thereby providing the circuit with an ongoing representation of elapsed time and experiential history during goal pursuit [179].

Taken together, goal-value learning does not appear to be localized to a single neural structure, but rather emerges from the coordinated activity of a distributed circuit. The PFC—spanning frontopolar, orbitofrontal, and ventromedial regions—maintains, abstracts, and integrates goal value across timescales ranging from seconds to months. The basal ganglia gate which representations are updated and determine the timescale over which updating occurs, with anterior regions supporting flexible short-term updating and posterior regions encoding stable long-term values. The ACC governs whether current evidence justifies continued commitment to the active goal or necessitates reallocation of goal pursuit, while also generating subgoal prediction errors that propagate value learning upward through the goal

hierarchy. Together, interactions among these systems allow goal values to remain sufficiently stable to sustain extended pursuit, while also remaining flexible enough to adapt when environmental feedback demands updating.

4.5 Monitoring competing alternatives

Encoding alternative goal representations and values

For goal pursuit to remain flexible, the brain must represent alternative goals and update their values based on environmental conditions. When the prospects of an alternative goal exceed those of the currently active goal, the brain must disengage from the persistent state associated with the current goal and reconfigure its task representations to support pursuit of a new one. Prior literature has primarily attributed this function to the prefrontal cortex, particularly the frontopolar cortex (BA10) and the ACC.

Boorman and colleagues demonstrated that during voluntary choice, the frontopolar cortex tracks the relative advantage of switching to a previously forgone alternative, modeled using Bayesian reinforcement learning. Moreover, interindividual variation in FPC activity correlated with individual differences in switching behavior [18]. In a follow-up study employing a three-alternative choice task, the lateral frontopolar cortex, dorsomedial frontal cortex, and posteromedial cortex were found to encode reward-based evidence favoring the best counterfactual option for future decisions, as well as counterfactual reward-prediction errors. Individual variability in both value coding and reward-prediction-error coding for counterfactual options correlated with switching behavior [17]. This ability to represent unchosen alternatives is thought to reflect a core function of the FPC: enabling the simultaneous consideration of parallel tasks by placing previously selected tasks into a pending state for later retrieval and execution, a process referred to as cognitive branching [100, 119]. Consistent with this interpretation, FPC activation has been observed specifically when participants are required to maintain one goal in mind while concurrently pursuing another, but not in conditions involving working memory alone or divided attention between tasks [101]. Lesion studies in primates further support the causal role of this region, showing that frontopolar lesions selectively impair rapid acquisition of novel behavioral routines—and in particular, self-initiated exploration of relative values among alternatives—while leaving previously learned routines intact [19]. More recent causal evidence indicates that gamma and beta bursts in the FPC track the value of counterfactual rules rather than currently active rules, and that transient high-frequency stimulation of the FPC reduces exploration of counterfactual rules

prior to a rule change while delaying adaptation to the newly relevant rule afterward [4].

The ACC has likewise been strongly implicated in the evaluation of alternative goals [103]. In foraging paradigms where participants must decide whether to persist with a current option or leave in search of an alternative, the ACC encodes the value of searching, whereas the vmPFC encodes the value of the currently active option [102]. This division of labor extends to multi-goal contexts: whereas vmPFC and ACC bilaterally encode the values of current options when participants pursue a single goal, these representations become hemispherically divided in dual-task settings, with each hemisphere preferentially representing one of the competing goals [31]. More broadly, dACC is consistently activated in tasks that impose cognitive-control demands, particularly when goal pursuit occurs in the presence of competing alternatives. One influential theory proposes that dACC computes the expected value of control under current task demands, estimating whether investing cognitive control in a particular goal is worthwhile and, when multiple goals are available, how control resources should be allocated among them [157].

Taken together, the frontopolar cortex and ACC appear to jointly support the representation and valuation of alternative goals, thereby providing the evaluative substrate necessary for determining whether to persist with the current goal or switch to a more advantageous alternative.

Flexible switching to alternate goals

Having established how competing alternatives are monitored and valued, we now consider how switching to an alternative goal is implemented in the brain. Hyafil and colleagues identified two dissociable mechanisms underlying task switching using a spatial Stroop and task-switching paradigm. Increased reaction times following a switch were accompanied by elevated activity in both ACC and dLPFC, but the two regions appeared to make distinct contributions: dLPFC activity increased only on switch trials involving incongruent Stroop conditions, whereas ACC activity increased across all switch trials. These findings suggest that the ACC prepares the system for a new task state, while the dLPFC resolves interference arising from competing task representations [87]. Sylvester and colleagues similarly proposed that attentional switching and interference resolution constitute dissociable neural processes, supported by distinct neural substrates, with BA 7 and BA 18 implicated in attentional switching and BA 6 and BA 10 implicated in interference resolution

[167]. Additional evidence for lateral prefrontal involvement in task switching was provided by Dove and colleagues, who observed increased activation during switch trials in the lateral prefrontal and premotor cortices bilaterally, the anterior insula bilaterally, the left intraparietal sulcus, the SMA/pre-SMA region, and the cuneus/precuneus [41].

Based on these patterns of ACC and dLPFC activation, several theories have proposed that the ACC detects conflict and communicates conflict-related information to the dLPFC, which then adjusts the level of executive control accordingly [117]. Supporting a causal role for the cingulate cortex, Rushworth and colleagues demonstrated that cingulate lesions impaired both task switching and delayed alternation, suggesting that the cingulate monitors conflict in order to guide switching behavior [150].

At a broader systems level, a meta-analysis examining switching across perceptual, response, and contextual domains found that the inferior frontal junction and posterior parietal cortex support switching in a domain-general manner, whereas domain-preferential activation was observed in dorsal premotor cortex for perceptual switching and in frontopolar cortex for context switching [97]. Muhle-Karbe and colleagues partially replicated these findings, observing that the frontopolar network contributes to domain-general control, while additionally showing that only the anterior cingulate cortex exhibits context-dependent preparatory activity [132].

4.6 Individual differences in flexible goal control

Large-scale functional brain networks encompassing lateral and orbital frontoparietal, midcingulo-insular, and frontostriatal regions support cognitive flexibility across the lifespan [171]. We now consider how differential engagement of these networks may account for individual variability in flexible goal control.

At the network level, greater task-state variability in dynamic functional connectivity between the frontoparietal network and the default mode network correlates with enhanced cognitive flexibility on the Stroop color-word test. Moreover, a stronger contrast between task-state and resting-state frontoparietal–DMN dynamics is associated with better cognitive performance more broadly [40]. Computational modeling has provided additional insight into the mechanisms underlying these network-level differences. A neurophysiologically constrained neural-network model of working memory demonstrated that increasing PFC–parietal connectivity reduced switching costs and task interference, whereas increasing the signal-to-noise

ratio in the PFC deepened attractor states associated with goal persistence, thereby increasing switching costs while simultaneously reducing interference. Greater recurrent connectivity within the PFC further sustained goal representations through self-reinforcing activity [77].

At the level of control strategy, Braver and colleagues proposed evidence for dual-process control—proactive and reactive [25, 24]. In this framework, proactive control involves sustained top-down biasing of memory and attention by goal states, whereas reactive control reflects bottom-up, interference-driven reactivation of goals. Using a continuous performance paradigm (AX-CPT [149]), they showed that younger adults exhibit stronger proactive control, whereas older adults display signatures of reactive control. Importantly, older adults could be shifted toward proactive control through task training. These two strategies were found to be differentially encoded within the lateral PFC, flexibly adapting according to group membership and changes in task conditions.

Individual differences in the tendency to monitor and switch to alternative goals are also reflected in frontopolar activity. Boorman and colleagues demonstrated that frontopolar activation predicts across-subject variability in switching behavior [18]. Similarly, a study comparing entrepreneurs and managers—groups that differ systematically in their demands for flexible goal management—found that entrepreneurs exhibited greater decision-making efficiency and stronger frontopolar activation during an explore–exploit gambling task [110]. In a related study of task persistence, participants who persisted longer showed greater FPC activity in response to task-relevant positive feedback, whereas participants who quit exhibited greater FPC responses to task-irrelevant monetary rewards. These findings suggest that the FPC may differentially weight goal-relevant versus goal-irrelevant signals across individuals.

Striatal modulation of prefrontal regions, which contain a high density of dopamine receptors, has also been implicated in individual variability in persistence. Pharmacological manipulation of dopamine signaling using the agonist bromocriptine improved working-memory performance, with the largest benefits observed in participants who were initially least efficient [62]. Structural connectivity likewise contributes to these differences: increased connectivity strength within the right dorsolateral prefrontal tract is associated with negative frontostriatal coupling in adolescents displaying impulsive behavior [172]. Finally, individual differences in reward sensitivity are reflected in task-switching behavior, with sensitivity measures

correlating with activity in the ventral striatum, ACC, lateral prefrontal cortex, and right inferior frontal cortex [54, 13].

4.7 Conclusion

We review evidence from the literature suggesting that long-term goal pursuit is mediated by the coordinated activity of a distributed network of brain regions, each contributing distinct computational functions. Behavioral evidence indicates that humans prefer to maintain persistent states during the pursuit of long-term goals, with the degree of persistence often exceeding what would be considered optimal for maximizing reward. Existing evidence further suggests that this persistent goal state is sustained through coordinated interactions between prefrontal and parietal regions. The strength of persistence is frequently correlated with the amount of progress made toward a long-term goal, a relationship that can be accounted for by a momentum-based model of extended goal valuation.

Goal values are estimated, and the degree of persistence is dynamically adjusted, through interactions between prefrontal cortex and basal ganglia circuitry. The frontopolar cortex and ACC appear to mediate the management of subgoals within persistent goal states when long-term goals are hierarchically organized. In addition, the frontopolar cortex and ACC are implicated in encoding and maintaining representations and values of alternative goals in the environment, monitoring their viability relative to the currently active goal in order to determine whether switching is warranted. When switching becomes necessary, the ACC and dLPFC are thought to contribute to the implementation of the switch itself.

Although evidence implicating these anatomical regions has been assembled from a diverse range of task paradigms spanning working memory, reward learning, and executive function, the precise computational roles of these regions in long-term goal pursuit remain to be tested using naturalistic goal-setting paradigms in both humans and animals. Finally, individual differences in the preferential recruitment of these systems for extended goal valuation may underlie heterogeneity in the flexible pursuit of long-term goals, an area that remains rich with opportunities for future investigation.

Chapter 5

MOMENTUM COMPUTATIONS SUPPORT HIERARCHICAL GOAL AND SUBGOAL SELECTION IN HUMANS

Humans pursue hierarchically organized goals by balancing persistence with flexible switching between goals and subgoals. How the brain supports this form of hierarchical control during self-directed behavior remains unclear. Using functional neuroimaging and a freely navigated hierarchical goal pursuit task ($n = 29$), we show that goals and subgoals are supported by dissociable neural systems. Participants' choices were captured by a computational account in which “momentum”, defined by accumulated progress and rate of progress, governs commitment to ongoing pursuits. Higher-order goal identity and momentum were represented in medial frontal, parietal and striatal regions, whereas lower-level subgoal momentum was encoded in the frontal pole. Goal and subgoal prediction errors were similarly dissociated across medial prefrontal–striatal and anterior cingulate–insular systems, respectively. Neural variability in these momentum signals predicted individual differences in switching tendencies. Together, these findings suggest that hierarchical behaviour is supported by distinct neural mechanisms for representing, valuing and flexibly updating goals and subgoals.

5.1 Introduction

Human behavior is hierarchically organized. Everyday pursuits—from preparing a meal to writing a manuscript—require coordinating extended goals with the smaller steps that support them. These subgoals often demand sustained effort without immediate reward, and people routinely interrupt one pursuit to engage another, flexibly shifting between nested tasks while maintaining overall direction. Despite the ubiquity of such hierarchical behaviour, the mechanisms by which the human brain represents, evaluates and updates behavior across hierarchical levels remain poorly understood [109, 22].

Existing research has begun to reveal how the brain encodes components of hierarchical control, including neural signatures of subgoal prediction errors and representations of subgoal identity and value [145, 69]. However, these studies have largely relied on tasks with fixed action sequences and without explicit higher-level goals, limiting insight into how the brain manages competing alternatives,

constructs value across levels of a hierarchy, and supports flexible switching during self-directed behavior. As a result, we still lack a mechanistic account of how the brain balances persistence with flexibility when navigating complex goal structures.

A promising computational framework for understanding such behavior is momentum, a construct we previously introduced to describe how people sustain long-term pursuits [3, 2]. Momentum formalizes the intuition that accumulated progress can sustain ongoing pursuit: it increases when individuals make headway toward a goal and diminishes when progress slows. We previously showed that momentum computations capture human tendencies to persist, switch and prioritize progress across both goals and subgoals. Yet it remains unknown whether the brain constructs momentum signals at multiple hierarchical levels, and whether these signals are supported by distinct neural systems.

Several lines of evidence point to candidate mechanisms. Medial prefrontal and striatal regions have been implicated in goal commitment[82], value representation and monitoring of ongoing pursuits [33]. The frontal pole has been linked to evaluating alternative options and enabling behavioral switching [18, 119], whereas anterior cingulate and insular cortices have been associated with hierarchical monitoring and prediction error signaling[7]. Together, these findings suggest that hierarchical behavior may rely on distributed neural systems that track the identity, value and updating of ongoing pursuits across levels of control.

Here, we test this hypothesis using a freely navigated hierarchical goal-pursuit task in which participants choose among multiple goals and subgoals with shifting prospects. Combining representational similarity analysis with model-based neuroimaging, we examine how the brain encodes hierarchical identity representations, constructs momentum signals across multiple levels of control, and generates prediction errors that support flexible switching. Our findings reveal dissociable neural systems for hierarchical valuation and control, revealing how distinct neural systems support persistence and flexibility during hierarchical behavior.

5.2 Results

Task, behavior, and computational model

We designed a game with hierarchically organized goals derived from [2], in which participants could freely select subgoals to pursue any of the high-level goals in the game. In the game, participants acted as construction agents and could acquire any of four available resources — glass, metal, stone, and concrete — to build one of three

construction projects: a spaceship, an observatory, or a house. Each construction project required resources in specific proportions and had distinct subgoals. For instance, a spaceship required three glass and three metal, yielding the two subgoals ‘get 3 glass’ and ‘get 3 metal’. Figure 1 shows the resource requirements for the different projects.

To play the game, participants progressed through the steps described in Figure 1A. Participants were provided with a resource inventory that could hold up to three resources of a given type. At the start of each round, participants were shown the inventory status and the progress made toward each goal using the resources currently in the inventory (outcome epoch). Based on this information, participants selected one of the three goals (goal selection epoch), after which they selected a specific resource by choosing among four aliens. Each alien was associated with a specific resource, a mapping that participants practiced prior to the start of the game.

Aliens supplied resources stochastically, and resource probabilities shifted across blocks (approximately every 30 rounds). We implemented six block types in the game, shown in Figure 1C, with each block favoring pursuit of a particular goal based on the prevailing resource probabilities (the dominant goal within the block). Critically, adjacent blocks did not share the same dominant goal, creating decision conflicts between persisting with the current goal — in which progress had already been made — and switching to the newly favorable goal in the subsequent block.

Participants’ preference for rates of progress:

Across two game sessions comprising 12 blocks, participants selected the dominant goal within each block with a high probability (Figure 2A). We additionally implemented a high–low manipulation within blocks: in the high condition, the advantage of the dominant goal over the best alternative goal was greater than in the low condition. Participants showed a stronger preference for the dominant goal in the high condition relative to the low condition (Figure 2B). These findings indicate that participants were sensitive to relative rates of progress across goals within blocks.

Participants’ preference for goal and subgoal progress:

In addition to being sensitive to rates of progress, participants also showed a preference for accumulated goal progress. Figure 2C (left) shows that the probability of selecting a goal increased with greater goal progress. Figure 2C (right) shows

these progress–selection trends separately across goals and block types, revealing sensitivity to both accumulated progress and rates of progress. We observed a similar, though weaker, trend for subgoal selections. Figure 2D shows subgoal selection probabilities for the resource with the highest target requirement within a goal (requirement of three resources), separated by block type, revealing a tapered preference for progress. Subgoal selections were more complex because subgoals could be shared across different goals. A subgoal associated with a resource could therefore be ‘in progress’ for one goal while already ‘finished’ for another, creating conflicts during goal selection and dampening the influence of subgoal progress on subgoal choices. Nevertheless, subgoal selections also exhibited an overall preference for progress.

Participants’ preference for depth-first persistence in hierarchical goals:

We found evidence of depth-first persistence, that is, a tendency to persist with both goals and subgoals over time. This produced chunks of goal–subgoal persistent states of varying lengths, during which participants continued pursuing the same goal and subgoal across successive rounds. Figure 2E shows participants’ probabilities of transitioning between goals across consecutive rounds. Transition probabilities were highest along the diagonal, indicating a strong tendency to persist with the same goal. This persistence was especially pronounced for the dominant goal within each block. We observed a similar persistence pattern for subgoals (Figure 2F). Across blocks, goal progress modulated the degree of persistence, with higher persistence probabilities for goals with greater progress. Within blocks, however, persistence remained relatively stable across both halves of the block, despite participants beginning each block without knowledge of the underlying resource probabilities (see supplementary Figure A1).

In previous work, we showed that participants tend to persist with goals beyond what is task-optimal. In [3], we proposed potential explanations for why persistence may nevertheless be normatively advantageous overall, even when suboptimal for the present task. Persistence may facilitate the formation of schemas for hierarchical goal pursuit, thereby increasing efficiency while simultaneously reducing the costs associated with reconfiguring task representations. We next describe the computational model that gives rise to hierarchical persistence, which we aim to validate using neural analyses in the present study.

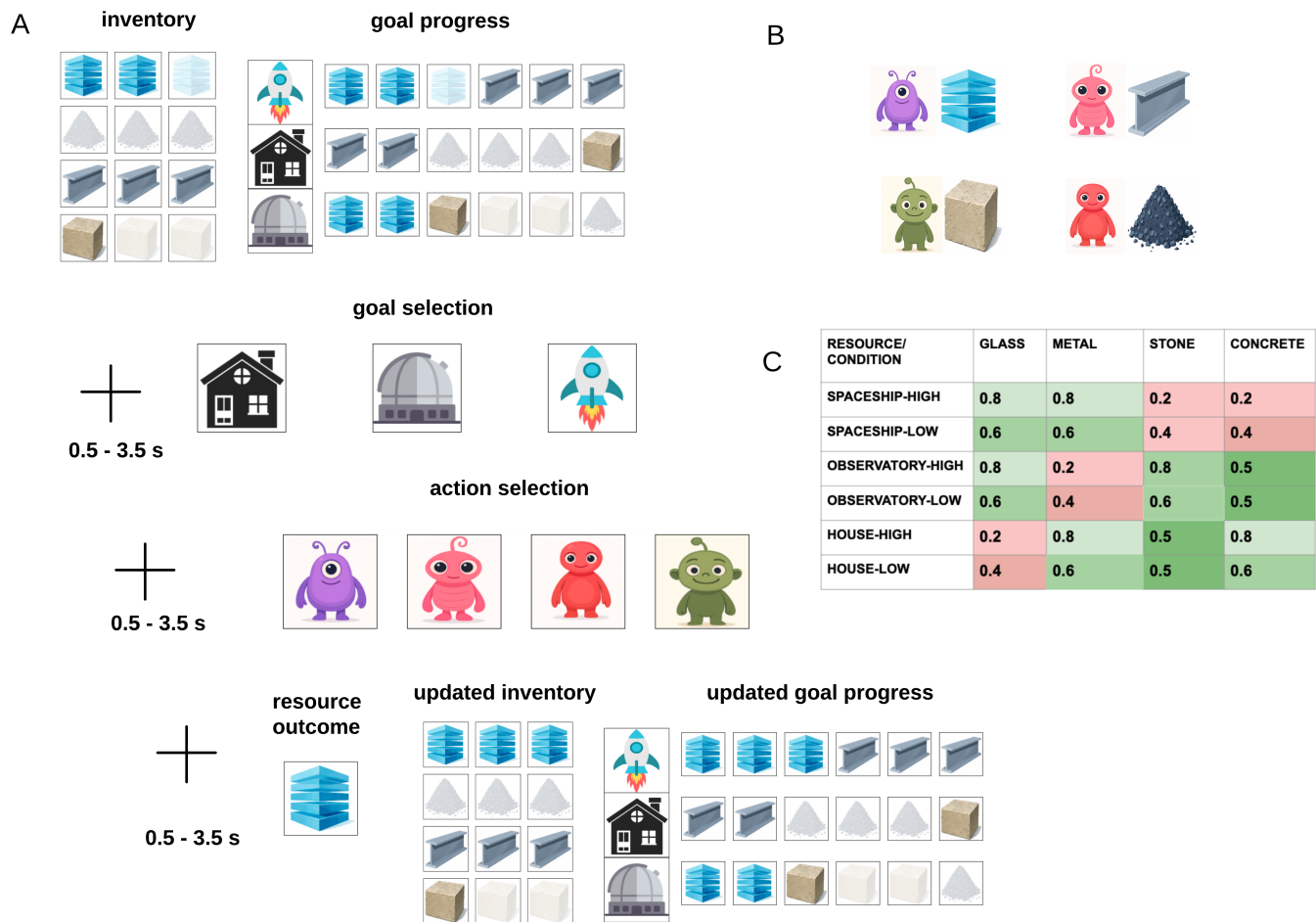


Figure 5.1: Hierarchical goal pursuit task. Overview of the experimental paradigm showing the hierarchical structure of goals, subgoals, and actions that participants navigate through during the fMRI experiment. **A.** Participants go through the following screens in a sample round: 1. resource inventory shows resources of each kind already collected. 2. goal progress screen shows how much progress was made toward each goal given the resources in the inventory. 3. participants can select their construction choice. 4. Participants can select resources by choosing appropriate alien suppliers. 5. Participants see the resource outcome, following which resource inventory and goal progress screens are updated for a new round. Goal selection, action selection, and outcome epochs are separated by a jitter period randomly chosen between 0.5-3.5s across rounds. **B.** Aliens mapped to specific resources. **C.** Game is served in blocks, one of 6 types. Each block has specific resources that are abundant, as shown in the table.

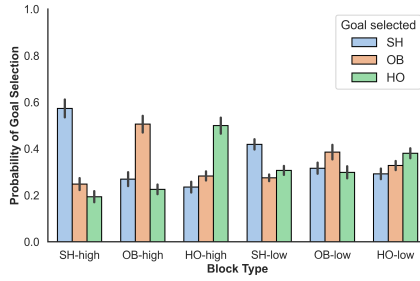
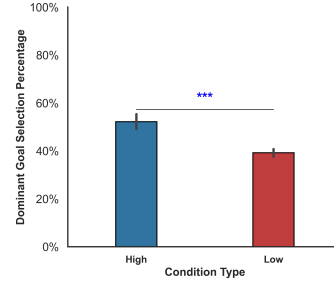
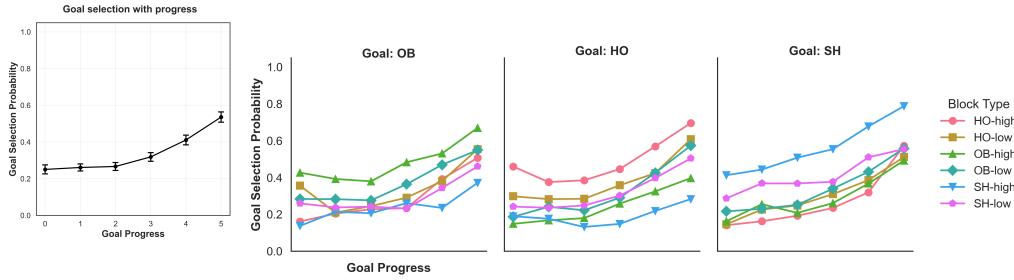
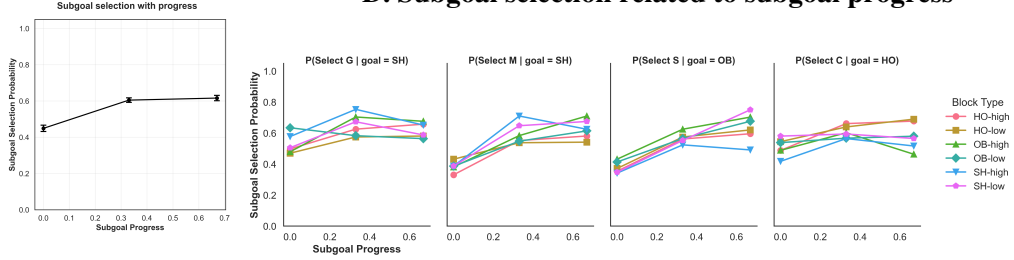
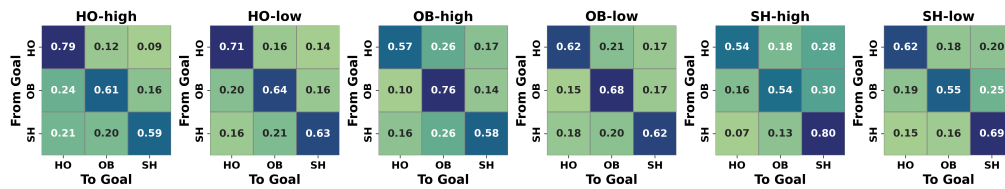
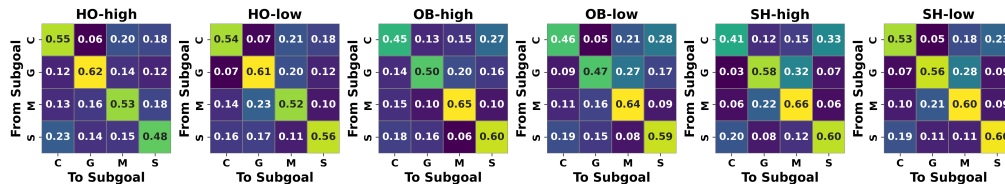
A. Goal selection probability by block type**B. Dominant goal choice by high-low conditions****C. Goal selection related to goal progress****D. Subgoal selection related to subgoal progress****E. Goal transition matrices****F. Subgoal transition matrices**

Figure 5.2: Behavioral results during hierarchical goal pursuit. **A.** Participant select the dominant goal with a higher chance within blocks. **B.** Participants subgoal selection patterns within blocks. **C.** Participants probability of selecting a goal increases with its progress. Goal selection probabilities are shown split by specific goal across all block types. **D.** Similar trend of preference to progress is observed in subgoal selections. **E., F.** Participants probabilities of transitioning from one goal/subgoal to another across adjacent rounds. High probabilities across the diagonal shows participants' preference to goal/subgoal persistence across successive rounds.

A momentum-based account of hierarchical goal control

Participants' persistence patterns in relation to goal progress can potentially be explained through prospective computations. Participants may assign higher value to goals with greater progress because such goals are closer to completion and, consequently, to reward. In previous work, however, we introduced an alternative account of goal valuation that avoids computationally expensive prospective planning while still approximating time-to-goal completion using parsimonious, model-free computations. Specifically, we introduced the concept of 'goal momentum', defined as the product of accrued goal progress and the current rate of progress toward that goal.

$$\text{goal momentum} = \text{goal progress} \times \text{rate of goal progress} \quad (5.1)$$

In a subsequent study, we showed that momentum computations extend across hierarchical goal structures and contribute to persistence at the level of subgoals as well. Goal and subgoal momentum values can be updated throughout goal pursuit using reinforcement signals derived from changes in progress following goal-directed actions. These momentum values are updated through momentum prediction error signals computed using a temporal difference learning rule applied to the momentum function (see Methods for model details).

In the present study, we replicated previous findings from studies of extended goal pursuit, showing that momentum-based models best account for participants' persistence patterns relative to alternative accounts. We evaluated two variants of the hierarchical momentum model: TD-momentum and TD-momentum-alt-goal. The latter differs from the former in that alternative goal values are also updated whenever the currently selected subgoal is shared between goals.

We additionally evaluated several alternative accounts, including a prospective planning model and a range of simpler heuristic models (described in the Supplementary Materials). These heuristic accounts included models in which value depended solely on accumulated progress, as well as models driven only by rates of progress through inference of the most favorable resources within a block. In the present study, a hybrid model that implemented momentum computations at the goal level together with a pure rate-of-progress strategy at the subgoal level achieved the highest aggregate likelihood, followed by the two hierarchical momentum models (supplementary Figure A2). However, the parameters of the top-performing model were

poorly recoverable and exhibited high model confusability. We therefore focused on the next best-performing model, TD-momentum-alt-goal, which showed only a modest decrease in likelihood while remaining well recoverable, with substantial interference arising only from the closely related TD-momentum model (supplementary Figure A4). We used the TD-momentum-alt-goal model to generate value and prediction error estimates for both goals and subgoals in the subsequent neural analyses.

The momentum model also reproduced participants' patterns of goal and subgoal selections as a function of progress in group-level simulations using parameters fit to individual participants (supplementary Figure A5). Furthermore, participant-specific momentum model parameters accounted for individual differences in goal and subgoal switching tendencies (supplementary Table A1).

With goal and subgoal momentum providing estimates of hierarchical value representations, and participant choices identifying the currently active goals and subgoals, we next turn to neuroimaging analyses to investigate how the identity and value of goals and subgoals are represented in the brain.

Dissociable neural representations of goal and subgoal identity in hierarchical behavior

Participants exhibited persistence at both levels of the hierarchy, implying that representations of both goal and subgoal identity must be maintained over extended periods during task performance. To characterize the nature of these representations, we employed representational similarity analysis (RSA).

We tested three candidate hypotheses regarding the coding of goals and subgoals. For each hypothesis, we constructed a model representational dissimilarity matrix (RDM) predicting the dissimilarity between pairs of trials based on the goals and subgoals selected on those trials. Each trial was defined by a condition vector — (goal, subgoal, action) — from which condition-wise RDMs were generated. The first hypothesis proposed dissociable coding of goals and subgoals, such that goal representations contain no embedded subgoal structure (separate goal and subgoal RDMs; Figure 3A). The second hypothesis proposed hierarchical coding, in which subgoal structure is embedded within goal representations (Figure 3A). Finally, we considered a conjoined coding scheme in which goal–subgoal pairs are represented as integrated composite units.

To identify brain regions matching the representational geometry predicted by each

model RDM, we performed searchlight RSA with across-run cross-validation (Figure 3B). The conjoined model did not yield any activations that survived group-level correction. In contrast, the goal RDM, which assumes distinct goal representations, produced widespread significant activations across the brain. The hierarchical RDM yielded more spatially restricted activations that partially overlapped with those associated with the goal RDM. Because the goal and hierarchical RDMs are highly collinear due to their shared goal structure, we next modeled them jointly using multiple-regression RSA to test whether the hierarchical RDM explained variance beyond that accounted for by the goal RDM alone. After controlling for shared variance, the goal RDM explained more unique variance in neural representational geometry than the hierarchical RDM.

The subgoal RDM, when analyzed separately, produced activations in regions largely non-overlapping with those associated with the goal RDM. However, because goal and subgoal RDMs also exhibit partial collinearity, some of these activations could reflect shared variance between the two representational structures.

We therefore jointly regressed goal, subgoal, and action RDMs against neural RDMs to account for shared variance across all three models (Figure 3C). In this joint analysis, the goal RDM showed significant activations in medial frontal and parietal regions, including the posterior cingulate cortex (PCC), vmPFC, precuneus cortex, cerebellum, striatum, insular cortex, and lateral frontal pole (Figure 3D). To further characterize the temporal dynamics of these representations, we performed separate RSA searchlights for each task epoch (goal, action, and outcome). We found that the precuneus robustly represented goal identity across all task stages. In contrast, subgoal identity, after controlling for goal identity, was selectively represented in the cerebellum during the action epoch. Action representations corresponding to the action RDM were primarily localized to expected motor cortical regions, as well as visual cortex regions, reflecting the fixed spatial mapping between key presses and screen locations.

Several limitations should be noted. First, even at the voxel level, we cannot fully rule out joint encoding of goals and subgoals. In particular, the precuneus remains a candidate region for representing combined goal and subgoal information in a representational format not captured by our tested hypotheses. Second, our analyses do not permit inferences regarding coding strategies at the level of single neurons. Nevertheless, the present findings provide evidence for at least partially dissociable representations of goal and subgoal identity in the brain, irrespective of whether

additional coding schemes coexist.

Having established dissociable representations of goal and subgoal identity, we next examined whether representations of goal and subgoal value are similarly dissociable in the brain.

Dissociable neural representations of goal and subgoal momentum

To determine whether goal and subgoal valuation processes are dissociable in the brain, we jointly modeled goal and subgoal value signals within a single general linear model (GLM) across multiple task epochs. Value estimates were derived from the momentum model and included both goal and subgoal momentum, along with additional computational variables such as perseveration, stay-switch choices, and prediction errors. All regressors were included within a single GLM and time-locked to the relevant stages of the task. Because the task design was sufficiently jittered to separate the major task epochs — goal probe, action selection, and outcome — the resulting activation patterns were qualitatively similar whether regressors were modeled jointly or separately across epochs. In total, the GLM included 20 regressors spanning all task stages (see Table 2 in Methods for details). During the goal and action selection epochs, goal and subgoal reaction times (RTs) were incorporated as regressor durations.

Goal momentum correlates in vmPFC, striatum, PCC, and precuneus during outcome:

We found that the same frontoparietal regions representing chosen goal identity also encoded chosen goal value during the outcome epoch (Figure 4A). Notably, goal progress was not included as a separate regressor because it was highly collinear with goal momentum ($r \approx 0.8$), given that momentum is directly computed from progress. To determine whether the observed activations reflected momentum-based value rather than goal progress alone, we constructed an alternative GLM in which goal momentum was replaced by goal progress. We then compared the two models within the union mask of activations from both GLMs using Bayesian model selection (Figure 4B). The majority of voxels favored the momentum model, suggesting that these regions encode value-related computations extending beyond progress alone.

During the goal probe epoch, only the striatum significantly encoded chosen goal value, whereas during the action selection epoch, goal value signals were observed primarily in motor cortex. In a separate GLM, we modeled goal and action RTs as

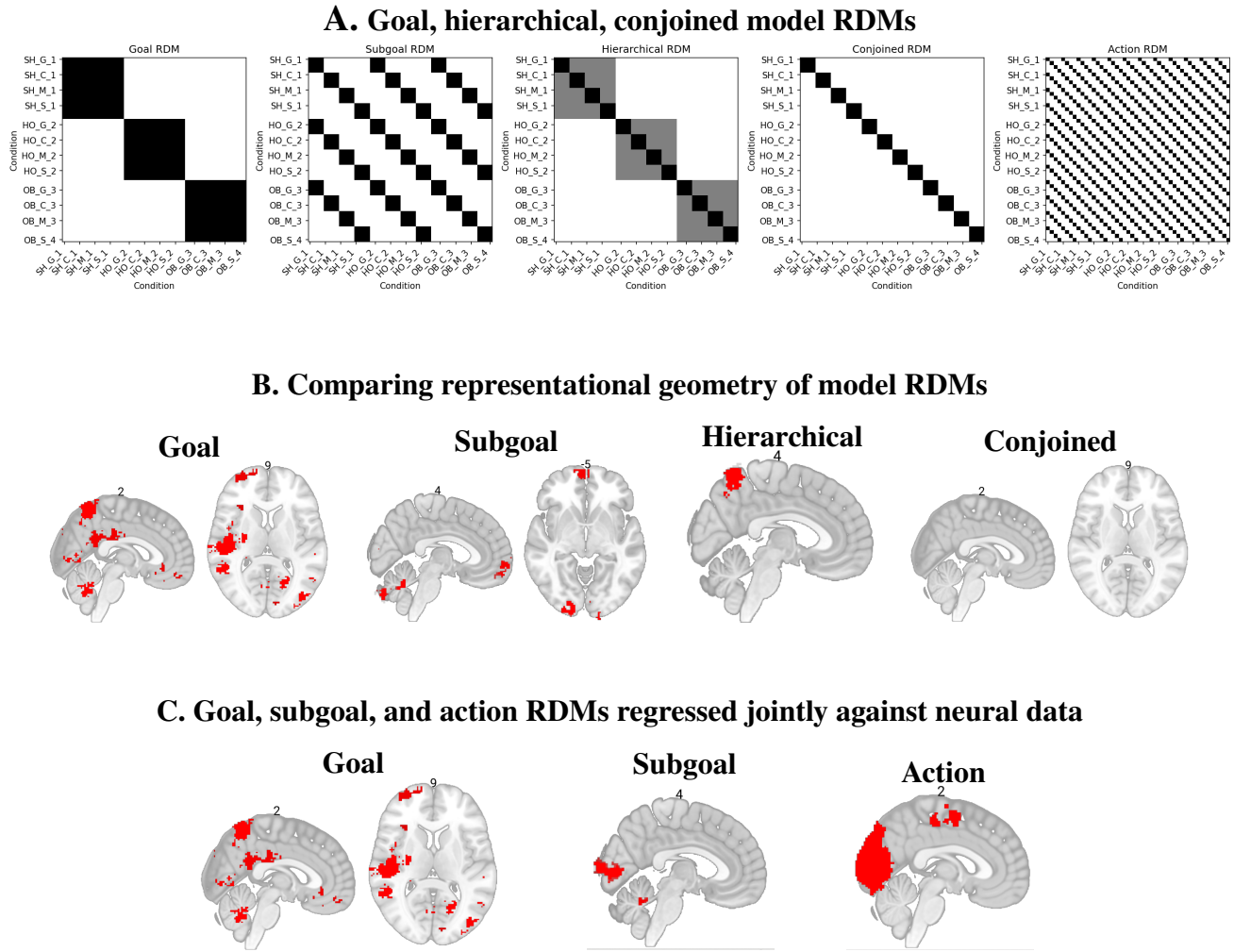


Figure 5.3: Representing goals and subgoals. (A) Model RDMs (condition-specific) for testing representational hypotheses: distinct, hierarchical, and conjoined. Each row is marked by a distinct compositional condition (<goal>_<subgoal>_<action>). (B) Searchlight RSA analysis results for goal, hierarchical, and conjoined RDMs modeled separately. A dissociate goal RDM shows significant correlation with the neural geometry at the whole-brain level (C) Model RDMs for goal, subgoal, and action (keypress) representations (D) Searchlight RSA analysis (of goal, subgoal, action modeled together) reveals distinct brain regions representing goals, subgoals, and actions during hierarchical goal control. Statistical maps are corrected using permutation-based cluster correction ($n = 5000$ permutations), with a cluster-forming threshold of $t > 3.39$, $p < 0.05$ FWE-corrected, whole-brain.

independent regressors rather than as durations of the value regressors. Under this specification, striatal value signals were no longer significant, suggesting that the original striatal effects may reflect processes evolving over the decision interval, such as value accumulation, motivational vigor, or other RT-related computations [165, 35] (Figure 4D). In contrast, cerebellar representations of goal value emerged after accounting for RT, suggesting a value signal that is relatively invariant to decision time, potentially reflecting anticipatory valuation at stimulus onset (Figure 4E). Goal and action RTs themselves were associated with widespread activations across PCC, precuneus cortex, motor cortex, cerebellum, parietal cortex, and visual cortex (Figure 4D). Consistent with the RT interpretation, motor cortical representations of goal value also disappeared after accounting for RT effects.

During the goal probe epoch, vmPFC showed significant activations associated with goal-choice perseveration (Figures 4C and 4D), suggesting a representation of policy bias favoring persistence with the currently active goal. This finding is consistent with previous accounts implicating vmPFC in heuristic and policy-based decision strategies [112, 105, 116]. One possible interpretation is that medial frontal and parietal regions encoding goal momentum during outcome processing may support persistent goal states represented within vmPFC, thereby facilitating continued commitment to ongoing goals.

Subgoal momentum correlates in frontal pole during action selection:

We observed significant activations for chosen subgoal value only during the action selection epoch, localized primarily to the frontal pole. These activations did not survive when RT was modeled as a separate regressor rather than as regressor duration, suggesting that subgoal value computations may evolve during the decision interval within frontal pole regions. The absence of robust vmPFC activity during action selection may reflect the joint modeling of partially collinear goal and subgoal value signals within the same GLM. Consequently, the absence of significant vmPFC or striatal effects cannot be interpreted as evidence against subgoal value coding in these regions. Nevertheless, the frontal pole reliably encoded subgoal value over and above goal value, providing evidence for dissociable subgoal value representations.

Taken together, these findings provide evidence for dissociable neural coding of goal and subgoal value. Medial frontal and parietal regions encoded goal value predominantly during the outcome epoch, whereas frontal pole regions encoded subgoal value selectively during the action selection epoch.

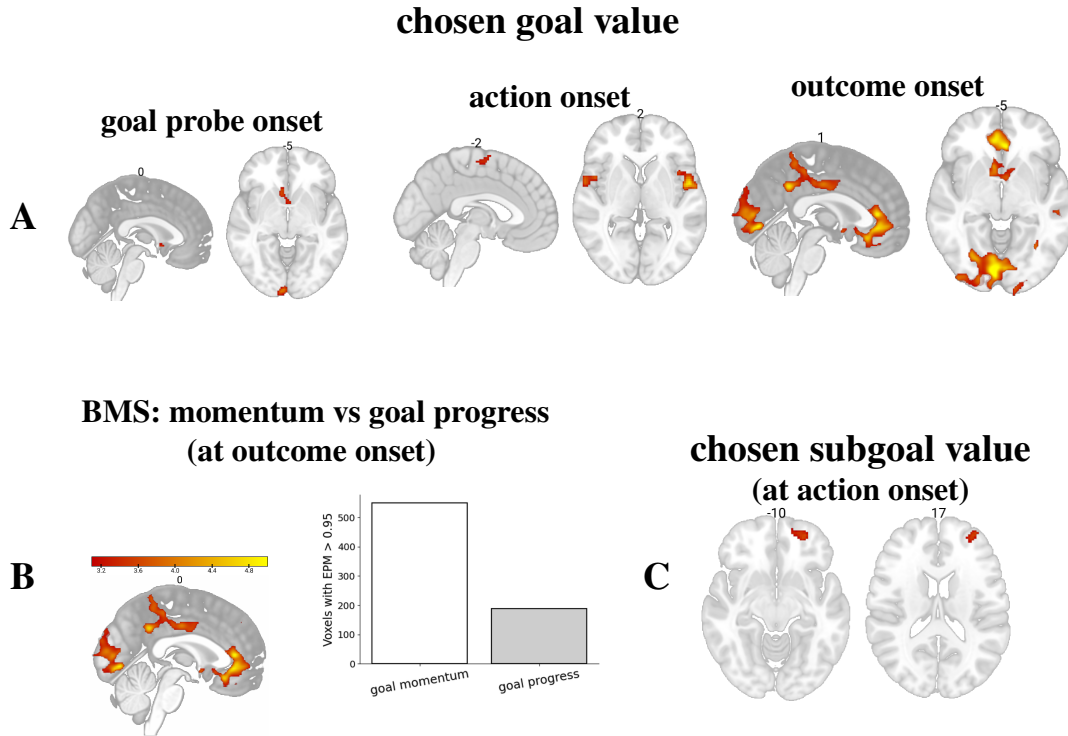


Figure 5.4: Goal valuations across task epochs. **A.** Group-level significant clusters of chosen goal value in each task epoch derived from the whole-brain GLM analysis with regressors described in Table 1. **B.** Bayesian model selection results comparing goal momentum versus progress models. Majority of the voxels in vmPFC are explained by goal momentum instead of goal progress. **C.** Chosen subgoal value represented in the frontal pole during the action selection epoch.

We next examine whether the brain also dissociates the monitoring of competing goals and subgoals during hierarchical decision-making.

Dissociation in goal and subgoal monitoring in the brain

Goal prediction errors in vmPFC and striatum:

We identified goal momentum prediction error signals in vmPFC, striatum, and frontal pole during the outcome epoch. Earlier analyses showed that vmPFC and striatum also retrospectively encoded the value of the chosen goal at outcome. This raised the possibility that these regions might additionally encode prospective goal value ($Q_{G'}(t)$), corresponding to the value of the impending goal choice, given that they already represent the previously chosen goal value ($Q_G(t-1)$) and the goal prediction error ($\delta_G(t)$).

To test this possibility, we constructed an additional GLM using the regressors described in Table 2 (Methods), but with the outcome epoch shifted such that goal

and subgoal value signals at outcome corresponded to the succeeding rather than preceding choice. Prediction error regressors were excluded from this GLM because they pertained to the preceding choice. Under this specification, prospective goal value signals emerged selectively in the striatum (Figure 5B).

Prospective goal coding in the striatum is consistent with accounts of temporal-difference learning in this region [14, 175, 75]. Within the momentum framework, values of both currently active and alternative goals may be updated through prediction-error driven learning, potentially accounting for the striatal representation of impending goal value. However, because participants often persisted with the same goal across trials, retrospective and prospective goal values were themselves highly collinear, complicating interpretation of the striatal activations. To dissociate these possibilities, we restricted analyses to a striatal mask and applied Bayesian model selection to compare retrospective and prospective value hypotheses. The majority of striatal voxels favored the prospective value model, supporting the interpretation that the striatum encodes impending goal value [153]. This interpretation is further supported by the observation that only striatum correlated with goal value during the subsequent goal-selection epoch.

Subgoal prediction errors in ACC, insula, precuneus, and frontal pole:

At the outcome epoch, subgoal momentum prediction error signals were observed in ACC [8, 151], insula [16], precuneus, and frontal pole (Figure 5C, right), with no significant activations in medial prefrontal cortex. This pattern is consistent with previous findings showing that when goal and subgoal prediction errors are elicited simultaneously, medial prefrontal regions preferentially track goal-level prediction errors [146]. Notably, subgoal prediction error activations in ACC spatially coincided with goal prediction error deactivations within the same region, suggesting a relative specificity for subgoal prediction error coding in ACC.

Lateral–medial dissociation in representations of alternative goals and subgoals in frontal pole:

Lateral frontal pole activity tracked the value of alternative goals during the outcome epoch (Figure 4A) [18, 17]. ACC and insular cortex also encoded alternative goal values during outcome processing, with ACC additionally representing alternative goals during the goal probe epoch [21]. We did not observe representations of alternative goals during action selection.

In contrast, medial frontal pole tracked alternative subgoal values during the outcome

epoch. Chosen and alternative subgoal representations were localized in close proximity within medial frontal pole (supplementary Figure A6), although chosen subgoals additionally elicited activations extending into lateral frontal pole regions.

Taken together, these findings provide evidence for dissociable monitoring of goals and subgoals, encompassing both currently active and alternative representations. The frontal pole appears to play a central role in this monitoring process by encoding prediction errors as well as values of alternative goals and subgoals. Prior work has implicated frontal pole in branching control — interrupting one task in favor of another. The present dissociation between medial and lateral frontal pole representations suggests a potential hierarchical branching mechanism, in which medial frontal pole supports within-task branching among subgoals, whereas lateral frontal pole supports branching across higher-level goals.

The preceding analyses established dissociable neural representations of goal and subgoal identity, value, and monitoring. Hierarchical behavior, however, requires not only the maintenance of persistent goal and subgoal states, but also flexible transitions between them when environmental contingencies change. Participants exhibited both sustained persistence and adaptive switching across the hierarchy. We therefore next examined the neural mechanisms associated with switching between goals and subgoals, and whether variability in momentum-related signals within these regions explains individual differences in hierarchical persistence and switching.

Goal and subgoal momentum signals dissociably account for individual differences in switching

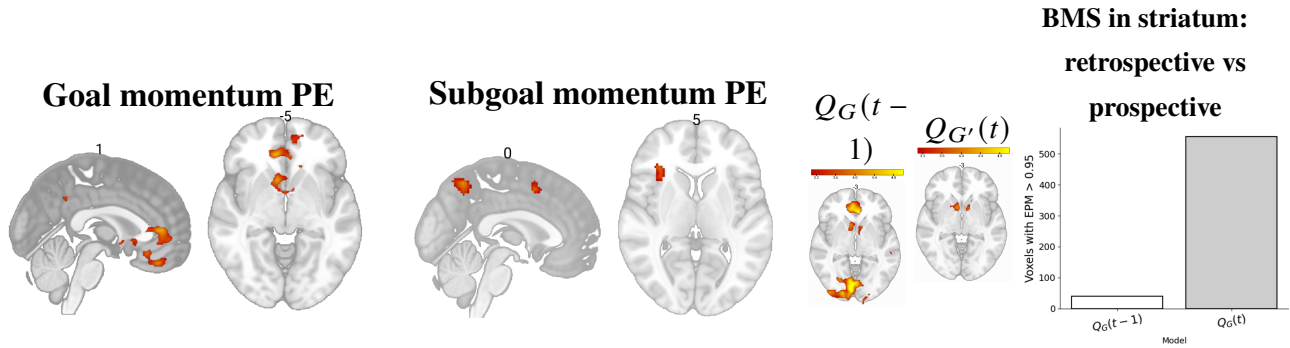
Shared neural correlates in goal and subgoal switching

To investigate the neural mechanisms associated with switching between goals and subgoals, we constructed a new GLM incorporating switch and stay events at both levels of the hierarchy. Specifically, we included the following composite event types within a single model: goal switch, goal stay–subgoal switch, and goal stay–subgoal stay. The goal-switch contrast was defined as the differential activation between goal-switch and goal-stay events, irrespective of subgoal switching. In contrast, the subgoal-switch contrast was defined conditionally on goal persistence:

$$(\text{goal stay–subgoal switch}) - (\text{goal stay–subgoal stay}).$$

These composite regressors were modeled across the three major task epochs —

A. Goal PEs in vmPFC, striatum, and subgoal PEs in ACC and insula



B. Alternate goal values and subgoal values in the frontal pole

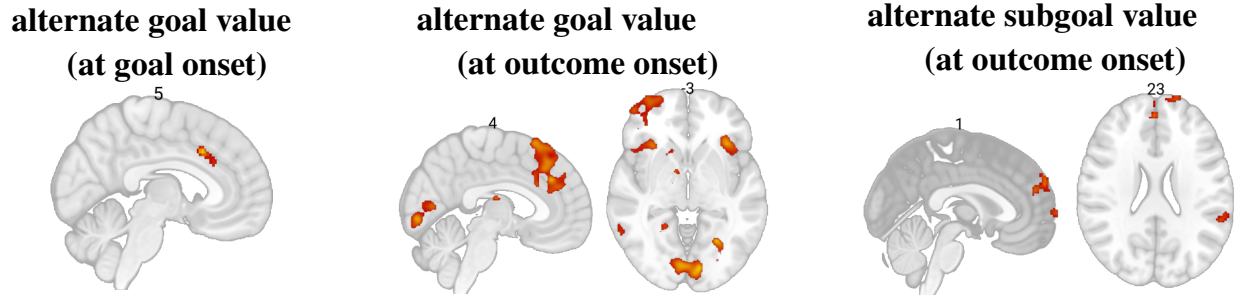
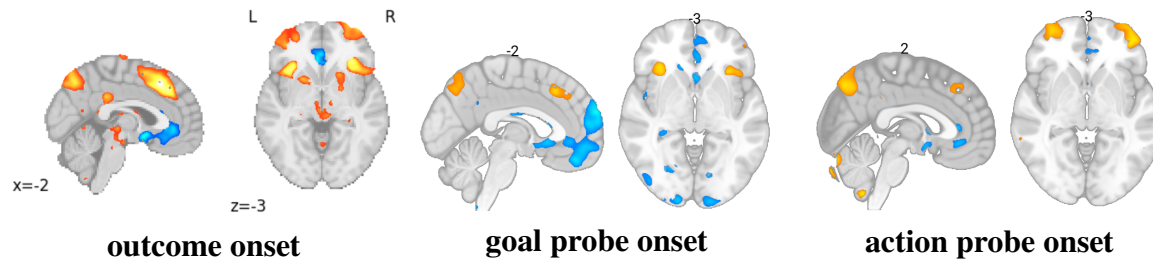


Figure 5.5: **Monitoring active and alternate goals and subgoals.** **A.** (left) Goal momentum prediction errors are found in vmPFC, striatum, and the frontal pole. (center) We found subgoal momentum predictions errors in ACC and insula. (right) Bayesian model selection on the Q_t (prospective), Q_{t-1} (retrospective) signals in the striatum, showed majority of the striatal voxels preferring the prospective value, suggesting a computational role for striatum in encoding prospective value. **B.** (left and center) We found alternate goal value signals in the frontal pole, ACC, and insula during the outcome epoch and in the ACC during the goal probe. (right) Alternate subgoal values are represented in the medial frontal pole during the outcome epoch. Statistical maps display significant clusters ($p < 0.05$, FWE-corrected).

A. Common neural signatures of goal and subgoal switching

Goal switch vs stay



Subgoal switch vs stay

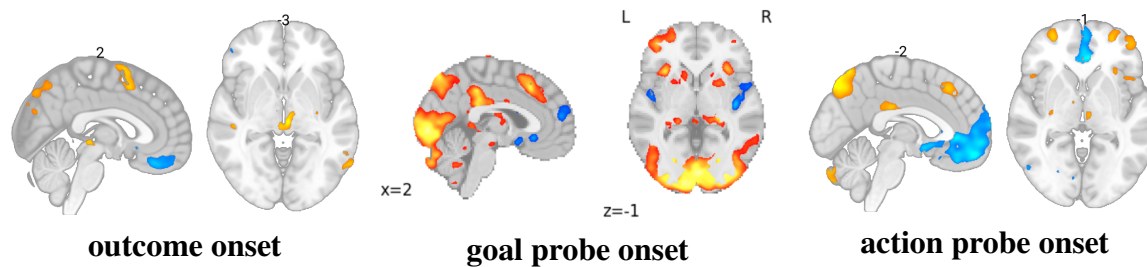


Figure 5.6: **Neural correlates of flexible control of goals and subgoals.** Common activation patterns seen in goal (top) and subgoal (bottom) switching (switch - stay contrasts). Frontal pole, ACC, insula, precuneus activate in the epoch preceding the switch, and vmPFC significantly deactivated during the time of the switch, compared to the stay condition. Statistical maps display significant clusters ($p < 0.05$, corrected).

outcome, goal probe, and action selection — yielding a total of nine regressors within a single GLM. Importantly, the outcome epoch was defined as the period preceding the subsequent switch in goals or subgoals, allowing us to identify neural activity potentially involved in initiating switching behavior.

We observed a broadly shared pattern of neural activity associated with switching at both hierarchical levels, although these signals were temporally shifted across task stages (Figure 6A). During the epoch preceding the behavioral stay/switch decision — outcome for goal switches and goal probe onset for subgoal switches — activity in ACC, frontal pole, insula, cerebellum, posterior cingulate cortex, and precuneus correlated with subsequent switching behavior. In contrast, during the epoch at which the stay/switch decision was behaviorally expressed — goal probe for goal switches and action selection for subgoal switches — vmPFC and medial frontal pole activity were more strongly associated with staying.

Elevated vmPFC activity during persistent goal and subgoal states further supports

the hypothesis that vmPFC facilitates commitment to ongoing hierarchical states. Similarly, increased activity in medial frontal pole during stay periods implicates this region in maintaining within-task persistence. In contrast, lateral frontal pole, ACC, cerebellum, insula, and posterior cingulate cortex were preferentially associated with switches toward alternative goals or subgoals.

Having identified regions associated with hierarchical switching and persistence, we next asked whether momentum-related activity within these regions accounts for individual differences in switching tendencies across participants. To address this, we extracted masks from vmPFC, lateral frontal pole, precuneus, PCC, insula, and ACC corresponding to peak switch-versus-stay contrasts. We then correlated momentum-related contrast estimates within these regions with participants' proportions of goal switching and overall hierarchical persistence (the tendency to persist with both goals and subgoals; see Table 1).

Chosen goal momentum signals in vmPFC during the goal onset epoch positively correlated with switching proportions across the hierarchy. This finding complements earlier results showing elevated vmPFC activity during stay states together with significant encoding of chosen goal momentum and perseveration signals, collectively supporting a role for vmPFC in goal commitment [82]. At the same time, individual variability in momentum-related modulation within vmPFC was associated with greater switching behavior, suggesting that stronger reliance on momentum computations may facilitate adaptive flexibility under changing environmental contingencies. This interpretation is consistent with a key property of the momentum framework: its ability to jointly support stability and flexibility in hierarchical control.

In contrast, chosen subgoal momentum signals within medial frontal pole negatively correlated with switching proportions at both hierarchical levels, indicating stronger persistence among participants exhibiting greater subgoal momentum representations in this region. Conversely, alternative goal momentum signals in lateral frontal pole positively correlated with goal switching frequency. Together, these findings further support a lateral–medial dissociation within frontal pole, whereby medial frontal pole preferentially supports maintenance of ongoing subgoal structure, whereas lateral frontal pole supports monitoring and switching toward alternative higher-level goals.

Importantly, momentum-related activity explained switching variability selectively within these regions. We did not observe comparable relationships between momentum-

related contrasts and switching behavior in other candidate regions, supporting the specificity of momentum modulation in hierarchical switching dynamics.

Table 5.1: Neural markers of individual differences in switching. Variability in beta estimates from the momentum and switch-vs-stay contrasts account for proportion of switches across individuals

Region	Contrast beta	Task epoch	Correlation type	Correlation (ρ)
vmPFC	Chosen goal value	Goal	Goal switching rate	0.41*
			Hierarchical persistence	-0.37*
Lateral FP	Alternate goal value	Goal	Goal switching rate	0.43*
Medial FP	Chosen subgoal value	Goal	Goal switching rate	-0.37*
			Hierarchical persistence	0.40*

5.3 Discussion

In the present hierarchical goal pursuit paradigm, participants pursued extended hierarchical goals by persistently maintaining both goals and subgoals over time, while simultaneously retaining the flexibility to switch to alternative options when environmental contingencies changed. Balancing stability and flexibility is a defining feature of adaptive goal pursuit, and our findings suggest that momentum computations can support such stability–flexibility tradeoffs. Goal momentum combines accumulated goal progress with the current rate of progress, thereby providing a parsimonious, model-free approximation of the value of extended goals through estimation of time-to-goal completion. We showed that momentum-based computations reasonably account for participants’ patterns of goal and subgoal selection behavior. Using participant choices together with model-derived value estimates, we then investigated how goal and subgoal momentum are represented in the brain.

Using representational similarity analysis (RSA), we first examined whether goal and subgoal identities are dissociably represented. We contrasted this dissociable coding account against alternative hypotheses involving hierarchical or conjoined representations of goals and subgoals. Our RSA results provided evidence for dissociable representations of goals, subgoals, and actions. Goal identity was represented across frontal and parietal regions, including vmPFC, PCC, precuneus, lateral frontal pole, striatum, and cerebellum. Subgoal identity, in contrast, was selectively represented in cerebellum during the action epoch after accounting for goal representations. Under less stringent correction thresholds, subgoal representations also emerged in medial frontal pole. Although these analyses do not fully rule out

alternative coding schemes, they provide evidence for at least partially dissociable neural representations of goal and subgoal identity.

The involvement of frontal and parietal regions in sustained goal representations is consistent with their established role in maintaining task-relevant information during extended cognitive control and working memory demands. Persistent goal representations in precuneus across all task epochs further suggest that this region may support sustained maintenance of higher-order task structure during hierarchical behavior. Subgoal representations in cerebellum highlight a potentially underappreciated role for cerebellar systems in organizing hierarchical control, warranting further investigation.

Goal and subgoal value representations exhibited a similar dissociation. Chosen goal momentum was represented in medial frontal and parietal regions — largely overlapping with regions encoding goal identity — during the outcome epoch. The co-occurrence of identity and value representations within these regions may facilitate maintenance of persistent goal states. Although the functional role of PCC and precuneus in hierarchical goal control remains to be fully characterized, prior work suggests that posterior cortical regions maintain task-relevant representations and coordinate with frontal systems during task execution [141, 113]. Dense reciprocal anatomical connectivity between PCC, precuneus, and vmPFC may support such interactions. Chosen subgoal momentum, in contrast, was represented in frontal pole during the action epoch, emphasizing the role of frontal pole in subgoal control. Prior theoretical work has proposed that frontal pole may support persistent control states through attractor-like dynamics [148], raising the possibility that frontal pole facilitates branching from higher-level goals to transient subgoal operations.

Goal and subgoal monitoring processes were similarly dissociable. Goal momentum prediction errors were represented in vmPFC and striatum. Striatal findings were particularly notable because striatum was the only region significantly associated with prospective goal value — that is, the value of the impending goal choice. Whereas vmPFC and parietal regions retrospectively encoded the value of the previously chosen goal during outcome processing, striatum additionally encoded future-oriented goal value representations, consistent with its established role in temporal-difference learning and reward prediction. The role of striatum in prospective planning during hierarchical goal pursuit remains an important topic for future investigation.

Subgoal prediction errors, in contrast, were represented in ACC, frontal pole, and in-

sula, dissociating subgoal monitoring from goal monitoring processes. Our findings diverge from prior work reporting subgoal prediction error representations in medial prefrontal cortex. However, they are consistent with previous observations that when goal and subgoal prediction errors are elicited simultaneously, medial prefrontal regions preferentially track goal-level prediction errors whereas ACC preferentially tracks subgoal-level prediction errors. We additionally observed a lateral–medial dissociation within frontal pole for representations of alternative goals and subgoals, with alternative goals represented laterally and alternative subgoals represented medially. Taken together, these findings strongly implicate frontal pole in hierarchical control and monitoring.

Furthermore, variability in momentum-related signals dissociably accounted for individual differences in switching tendencies. Greater chosen goal momentum variability in vmPFC positively correlated with switching proportions, suggesting that stronger reliance on momentum computations may facilitate flexible switching behavior. In contrast, chosen subgoal momentum variability in medial frontal pole positively correlated with persistence tendencies. Consistent with the observed lateral–medial dissociation, variability in alternative goal momentum signals within lateral frontal pole positively correlated with goal switching frequency.

Overall, we found evidence for dissociable neural representations supporting the encoding, valuation, and monitoring of goals and subgoals, including representations of both chosen and alternative hierarchical states. Such dissociation may be functionally advantageous for compositional behavior, particularly in contexts where subgoals are shared across multiple higher-order goals, as in the present task. Distinct subgoal schemas could then be flexibly recombined to support multiple goals. Dissociable representations supporting such reuse of subgoal structure may facilitate continual learning and extended goal pursuit across the lifespan [37].

Several limitations should nevertheless be acknowledged. Our analyses examine representational structure at the voxel level, and finer-scale investigations of neural firing patterns may reveal alternative or more integrated coding mechanisms. Nonetheless, dissociation at the voxel level remains informative regarding the distinct functional contributions of neural systems supporting hierarchical behavior.

Finally, our findings motivate a preliminary architectural account of hierarchical goal control in the brain. We propose that vmPFC and striatum encode and monitor higher-level goal momentum signals supporting persistent goal states, whereas medial frontal pole and ACC encode lower-level subgoal momentum signals supporting

subgoal persistence. ACC representations of subgoal prediction errors may interact with medial frontal pole to facilitate switching among subgoals within an ongoing goal. In contrast, lateral frontal pole may maintain representations of alternative goals that facilitate switching across higher-level tasks. PCC and precuneus may contribute to reconfiguration of task representations during switching and, through dense recurrent connectivity with vmPFC, help stabilize newly selected persistent states. At present, however, this account remains speculative and requires direct experimental validation. More broadly, our findings suggest that hierarchical behavior emerges from distributed interactions across multiple large-scale neural systems, motivating future integrative approaches to understanding hierarchical control.

5.4 Methods

Experiment and training protocol

Participants did the task in fMRI scanner over two back-to-back sessions, each lasting for 40 and 50 minutes with break of 15 minutes in between sessions. We have a total of 6 runs across two sessions with each run lasting for 11 minutes. Before the scan, participants participated in a series of training sessions, lasting for about 30-40 minutes.

Participants performed the task in the fMRI scanner over two back-to-back sessions lasting 40 and 50 minutes, respectively, with a 15-minute break between sessions. In total, participants completed 6 runs across the two sessions, with each run lasting approximately 11 minutes. Before scanning, participants completed a series of training sessions lasting approximately 30–40 minutes.

During the training session, we first introduced the different resources and aliens and instructed participants on the alien–resource mappings. Participants then played an initial training game in which they matched aliens to specific resources for 30 rounds. This game ensured that participants could correctly identify the aliens associated with each resource. We next introduced the different construction projects and the resources required to complete them. Participants then played a second training game in which, given a specific project, they selected aliens to obtain the resources needed to complete the project. In this version of the task, all aliens delivered their associated resources with 100% probability. This training phase ensured that participants understood which resources were required for each project.

In the final stage of training, participants were introduced to the structure of the main game and guided through a walkthrough demonstration of the sequence of

events within a trial. Participants then answered a series of quiz questions about the task and proceeded only after answering all questions correctly. Finally, participants completed a sample block of the main game consisting of 15 rounds before entering the scanner to begin the experiment.

In the main game, participants first viewed the resource inventory screen (1 s), followed by the goal progress screen (1 s). The goal progress screen displayed participants' progress toward each goal given the resources currently held in the inventory. Following a jitter period (randomized 0.5–3.5 s), participants reported their goal selection (*goal selection epoch*). The locations of the goals were displayed horizontally on the screen and shuffled across rounds. Participants responded using keys 1–4, with one randomly blocked key on each round because only three goal options were available. Participants had a maximum of 3 s to make a goal selection; failure to respond within this interval resulted in an instruction prompting them to respond within the allotted time, followed immediately by the onset of the next trial.

After another jitter period (0.5–3.5 s), participants selected a resource by choosing among four aliens displayed horizontally on the screen using keys 1–4 (*action selection epoch*). Alien locations were also shuffled across rounds. Following action selection, another jitter period (0.5–3.5 s) preceded presentation of the resource outcome (0.5 s). After outcome presentation, the updated resource inventory (1 s) and goal progress screens (1 s) were shown sequentially. Throughout the analyses, we treated this sequence — resource outcome, inventory, and goal progress screens — as the *outcome epoch*, lasting a total of 2.5 s. Following an additional jitter period, the next round began with the goal selection phase.

Participants completed 3 runs during the first session, with each run consisting of two game blocks. Each block contained approximately 30 rounds, although the exact number varied depending on participant response times because run duration was fixed.

In the second session, participants were introduced to a new set of aliens, each mapped to a specific resource. During the first 10 minutes of the session, while structural scans were acquired, participants completed an additional training session to learn the new alien–resource mappings. We changed the alien set between sessions to ensure that resource representations were not tied to the visual features of specific aliens, but instead reflected abstract resource representations. For example, participants seeking to collect *glass* selected different aliens across the two sessions. Participants then completed 3 additional runs using the new alien set during the

second session. Across both sessions, we obtained, on average, 360 rounds of gameplay per participant.

As a proxy measure of participant engagement, we computed goal–action congruence for each trial and calculated the overall proportion of trials in which participants selected resource actions congruent with their chosen goals. A resource selection was classified as incongruent if the selected resource was not required for the chosen goal, or if the subgoal requirement for that resource had already been satisfied.

The full game, including instructions, training tasks, and the scanner version of the experiment, is available online at:

Participants

We recruited 33 participants from the Greater Los Angeles area through the Chen Participant Center participant pool. Four participants were excluded from the study, resulting in a final sample of $n = 29$ participants for analysis. Two participants were excluded for failing to meet the engagement criterion. Specifically, we computed goal–action congruence for each participant and applied an exclusion threshold of 75%. The two excluded participants exhibited congruence rates of 52% and 57%, respectively. Two additional participants were excluded because they experienced substantial difficulty learning and performing the task and discontinued the experiment midway.

All participants provided written informed consent prior to the experiment. Participants were compensated at a rate of \$30 per hour for time spent in the scanner and \$20 per hour for training sessions, with an additional performance bonus of up to \$30.

Computational model: hierarchical TD-momentum

In [3], we introduced the construct ‘goal momentum’ to account for persistence patterns in extended goals, where

$$\text{goal momentum} = \text{goal progress} * \text{rate of goal progress} \quad (5.2)$$

In , we showed that pursuing goals while optimizing for goal momentum in the task can produce similar persistence patterns seen in humans pursuing extended goals. We also introduced an algorithm, TD-momentum, that can estimate the *rate of goal progress* through experience using a temporal difference learning rule. In

hierarchical TD-momentum, momentum computations facilitate at both goal and subgoal persistence choices.

Goal and subgoal momentum

We define goal momentum for a goal g , Q_g at time t as follows:

$$Q_g(s_{g|I}^{(t)} = m_g^{(t)}) = v_g^{(t)} m_g^{(t)} + b_g^{(t)} \quad (5.3)$$

where $s_{g|I}$ indicates the state of the goal given the inventory, m_g is the fractional progress accrued in the goal, v_g is the parameter for the rate of goal progress, and b_g is the baseline progress rate for the goal. The baseline parameter, b_g , ensures the goal momentum is updated through experience even when the current progress in the goal is zero. Here, both v_g and b_g are updated from experience in progressing with the goal. v_g is computed for the overall goal, without considering individual resource availabilities specifically. Thus, momentum is a parsimonious computation that allows for approximation of potential for goal completion without the need to track individual resource probabilities.

The subgoal momentum, Q_{sg} , is defined as follows:

$$Q_{sg}(s_{sg|g,I}^{(t)} = m_{sg|g}^{(t)}) = v_{sg}^{(t)} m_{sg|g}^{(t)} + b_{sg}^{(t)} \quad (5.4)$$

where $s_{sg|g,I}$ is the state of the subgoal given the goal and the inventory, $m_{sg|g}$ indicates the fractional progress of the subgoal in the goal, v_{sg} is the parameter for subgoal progress rate and b_{sg} is the baseline subgoal progress rate.

Updating goal and subgoal momentum

For the chosen goal (G) and subgoal (SG), we updated the momentum prediction errors after observing the resource outcome as follows:

$$\delta_G = \gamma Q_G(s_{G|I}^{(t+1)}) - Q_G(s_{G|I}^{(t)}); \quad \delta_{SG} = \gamma Q_{SG}(s_{SG|G,I}^{(t+1)}) - Q_{SG}(s_{SG|G,I}^{(t)}) \quad (5.5)$$

where $s_{G|I}^t$ is the state of inventory at time t . The state advances by one resource if participants receive a resource in the round (positive prediction errors), or stays in the same state if they receive none (negative prediction errors).

Goal and subgoal momentum parameters are updated through a gradient descent of parameters in the value function as follows:

$$v_G^{(t+1)} = v_G^{(t)} + \alpha \delta_G m_G^{(t)}; \quad b_G^{(t+1)} = b_G^{(t)} + \alpha \delta_G \quad (5.6)$$

$$v_{SG} = v_{SG} + \alpha \delta_{SG} m_{SG|G}^{(t)}; \quad b_{SG}^{(t+1)} = b_{SG}^{(t)} + \alpha \delta_{SG} \quad (5.7)$$

When participants make progress in a subgoal after selecting a goal, they make progress not only in the chosen goal but also in an alternate goal that relies on the same subgoal. We posited that the updating of momentum of unchosen goals could also be influencing subsequent goal selection.

We considered the variant of TD-momentum that incorporated cross-goal transfer (TD-momentum-alt-goal) where we updated the momentum of the unchosen alternate goal that made progress on account of shared subgoals.

A hierarchical goal selection algorithm

We hierarchically modeled choices in individual trials as stay/ switch from goals and subgoals.

Goal selection: Individual goal values are computed using the momentum algorithm. The advantage of the current goal, A_G , over its alternatives is computed as follows:

$$A_G = Q_G - \max_{c \neq G} Q_c \quad (5.8)$$

The agent chooses to stay on the current goal with a probability $p_{\text{stay};G}$,

$$p_{\text{stay};G} = \frac{1}{1 + \exp(-c_g - \kappa_G - \beta_g A_G)} \quad (5.9)$$

where c_g is the goal switching cost, β_g is the softmax temperature. κ_G is the choice kernel where commitment towards the goal increases the more it is chosen in the past, as updated through the delta-rule,

$$\kappa_G = (1 - \alpha_c) \kappa_G + \alpha_c \kappa_G I_G \quad (5.10)$$

where I_G is the indicator that the goal is chosen in the round.

When the agent decides to switch away from the current goal with probability $(1 - p_{\text{stay}})$, it selects a new goal from the two alternate goals depending on the values of the two alternate goals ($\text{softmax}(Q_{alt_1}, Q_{alt_2})$)

Subgoal selection: Subgoal values are computed given the goal, assuming a hierarchical selection algorithms. The advantage of the current subgoal over its alternatives is computed from individual subgoal values as follows:

$$A_{SG|G} = Q_{SG|G} - \max_{c \neq SG} Q_{c|G} \quad (5.11)$$

The agent chooses to stay on the current subgoal with a probability $p_{\text{stay};SG|G}$,

$$p_{\text{stay};SG|G} = \frac{1}{1 + \exp(-c_{sg} - \beta_{sg} A_{SG|G})} \quad (5.12)$$

where c_{sg} is the subgoal switching cost.

Alternative hypotheses

The details on alternative hypotheses, model comparisons, model and parameter recovery are available in the supplementary section.

fMRI: data acquisition

Neuroimaging data were acquired at the Caltech Brain Imaging Center (Pasadena, CA) using a Siemens Prisma 3T scanner with a 32-channel radio frequency coil. The whole-brain functional scans were collected with a multi-band echo-planar imaging (EPI) sequence with the following parameters: 72 slices per volume, a repetition time (TR) of 1.12 s, an echo time (TE) of 30 ms, multi-band acceleration of 4, a 54° flip angle, a field of view (FoV) of 192 × 192 mm, a 96 × 96 matrix, and 2 mm isotropic resolution. After each run, positive and negative polarity, single-band EPI-based field maps were collected using similar parameters to the functional sequence, but with a TR of 5.13 s, a TE of 41.40 ms, and a 90° flip angle. T1- and T2-weighted structural images were collected for all participants, either at the end of the session or in an earlier session. Both types of structural images were collected with a 230 × 230 mm FoV, a 256 × 256 matrix, and 0.9 mm isotropic resolution. The T1-weighted images were acquired with a TR of 2.55 s, a TE of 1.63 ms, an inversion time of 1.15 s, and an 8° flip angle. The T2-weighted images were acquired with a TR of 3.2 s, a TE of 564 ms, and a 120° flip angle.

fMRI: data preprocessing

Results included in this manuscript come from preprocessing performed using fMRIprep 22.1.190 [47], which is based on Nipype 1.8.592 [65].

Preprocessing of B0 inhomogeneity mappings

A B0 fieldmap was estimated from the phase difference between two consecutive gradient-recalled echo (GRE) acquisitions. The resulting phase map was unwrapped using `prelude` (FSL 6.0.5). Additionally, a B0 fieldmap was estimated using `topup` FSL 6.0.5 based on two echo-planar imaging (EPI) reference scans with opposing phase-encoding directions.

Anatomical data preprocessing

Structural MRI data were preprocessed using `fMRIPrep`. T1-weighted (T1w) images were corrected for intensity non-uniformity using `N4BiasFieldCorrection` implemented in `ANTs` (v2.3.3) and used as the anatomical reference throughout preprocessing. Skull stripping was performed using the `antsBrainExtraction.sh` workflow with the `OASIS30ANTs` template. Brain tissue segmentation into cerebrospinal fluid (CSF), white matter (WM), and gray matter (GM) was performed using `FAST` in `FSL` (v6.0.5). Cortical surface reconstruction was carried out using `recon-all` in `FreeSurfer` (v7.2.0). Spatial normalization to `MNI152NLin2009cAsym` standard space was performed using nonlinear registration with `antsRegistration` (`ANTs` v2.3.3).

Functional data preprocessing

Functional MRI data were preprocessed using `fMRIPrep`. For each of the six functional runs, a reference BOLD volume and corresponding skull-stripped image were generated from the single-band reference images. Head-motion correction was performed using `MCFLIRT` in `FSL` (v6.0.5), yielding six motion parameters (three translations and three rotations). Fieldmap-based susceptibility distortion correction was applied by aligning the estimated fieldmap to the EPI reference image using rigid-body registration.

Functional images were co-registered to the T1w anatomical image using boundary-based registration implemented in `FreeSurfer`'s `bbregister` with six degrees of freedom. Functional data were subsequently normalized to `MNI152NLin2009cAsym` standard space using the transforms estimated during anatomical preprocessing.

Several nuisance regressors were extracted from the preprocessed BOLD data, including framewise displacement (FD), DVARS, and global signal estimates from CSF, WM, and whole-brain masks. Component-based noise correction (`CompCor`)

regressors were additionally computed following high-pass filtering (128 s cutoff). Temporal CompCor (tCompCor) components were derived from the top 2% most variable voxels within the brain mask, whereas anatomical CompCor (aCompCor) components were estimated separately within CSF and WM masks. For each decomposition, principal components explaining up to 50% of variance within the nuisance mask were retained.

Motion and physiological nuisance regressors were expanded to include temporal derivatives and quadratic terms. Volumes exceeding 0.5 mm framewise displacement or 1.5 standardized DVARS were identified as motion outliers. Additional nuisance regressors were estimated from signals along the edge of the brain using principal component analysis.

Preprocessed BOLD time series were resampled into MNI152NLin2009cAsym standard space using a single interpolation step that combined all spatial transforms. Volumetric resampling was performed using `antsApplyTransforms` with Lanczos interpolation, while surface resampling was performed using `mri_vol2surf` in FreeSurfer.

Further details regarding preprocessing workflows are available in the fMRIPrep documentation.

fMRI: trial-wise beta series modeling of task epochs

We modeled each task-epoch of individual trials separately using Least Squares-All (LSA) method. We split each trial into three task epochs — goal probe, action, and outcome — and marked them as separate events with duration as the length of the task epoch. We included all trials in a single GLM in Nilearn to generate voxel-wise beta estimates within a whole-brain mask for each trial and each task epoch. As the three epochs within the trial are sufficiently jittered (0.5 - 3.5s), we expect this model to generate reasonable trial-wise beta estimates of task-epoch activity.

We used these beta series estimates of trial-wise task epoch activity in subsequent RSA and individual differences analysis.

fMRI: RSA analysis

We used `rsatoolbox` for the RSA analysis. For each task epoch within a trial, we extracted the trial-wise beta estimates to serve as neural pattern vectors. For each trial-level beta map, we extracted the goal, subgoal, and action selected from the event log and assigned trials to conditions using composite goal–subgoal–action

labels. When analyses were restricted to a specific task epoch (goal, action, or outcome), only beta maps corresponding to that epoch were included. These composite condition labels were used to construct model RDMs.

Model RDMs

We constructed five candidate model RDMs from the unique composite condition label:

1. goal (0: same goal, 1: different goal)
2. subgoal (0: same subgoal, 1: different subgoal)
3. action (0: same action, 1: different action)
4. hierarchical (0: same goal and subgoal, 0.5: same goal/different subgoal)
5. conjoined (0: same goal and subgoal, 1: otherwise)

We reduced each model RDM to its upper triangle for analysis.

Cross-validated neural RDMs within searchlights

We performed whole-brain searchlight RSA by defining searchlight neighborhoods using rsatoolbox's volumetric searchlights with radius 5 voxels and an inclusion threshold of 0.5. For each searchlight sphere, we extracted the neural patterns and computed representational dissimilarity across trials. To reduce within-run temporal confounds and emphasize stable representational structure, we computed neural RDMs using an across-run cross-validated correlation distance metric. For each run, we computed condition-specific patterns by averaging trial patterns belonging to the same condition within that run. We defined the condition as the composite goal-subgoal-action choice, and therefore same condition would entail rounds with the same composite label. For each searchlight, and for each ordered run pair (run A < run B; 6 runs; 15 ordered pairs), we computed a condition $\times$ condition cross-run correlation matrix between run-A and run-B condition patterns (row-wise z-scoring across voxels followed by dot-product correlation). This matrix was symmetrized with the diagonal set to zero, and converted to a correlation distance RDM:

$$\text{RDM}_{AB} = 1 - \frac{1}{2} (\text{corr}(A, B) + \text{corr}(A, B)^T)$$

The final neural RDM vector for a searchlight was the average of upper-triangular RDM vectors across all run pairs. For each searchlight, the (z-scored) neural RDM vector served as the dependent variable and model RDM vectors served as predictors. We z-scored the predictors prior to regression and included an intercept in all GLMs.

Group-level inference and multiple comparisons correction

We performed the group-level inference on subject-level RSA beta maps using a second-level one-sample t-test against zero implemented in Nilearn. Prior to group analysis, subject maps were spatially smoothed with a 5 mm FWHM Gaussian kernel.

To control for multiple comparisons across the searchlight volume, nonparametric permutation testing was performed using FSL `randomise`[178]. One-sample sign-flip permutations ($n = 5000$) were used to generate the null distribution, and cluster-based inference was applied with a cluster-forming threshold of $t > 3.39$ to obtain family-wise error (FWE)-corrected statistical maps ($p < 0.05$, whole-brain).

fMRI: model-based analyses

To identify neural correlates of our computational model variables, we constructed whole-brain general linear models (GLMs) using Nilearn. For each participant, six run-wise GLMs were estimated, including regressors of interest derived from the computational model along with task-related and nuisance regressors. This yielded run-specific statistical parametric maps for each model variable. To obtain a single within-subject estimate per variable, we combined run-wise contrast effect-size and variance images using a fixed-effects model, producing subject-level contrast maps. We included the following regressors in a single GLM.

Regressors in the GLM

Regressors	Task epochs
Chosen goal value	goal, action, outcome
Best alternate goal value	goal, action, outcome
Chosen sugoal value	goal, action, outcome
Best alternate subgoal value	goal, action, outcome
Chosen goal perseveration	goal, action, outcome
Subgoal switch (given goal stay)	goal, action, outcome
Goal prediction error	outcome
Subgoal prediction error	outcome

Table 5.2: GLM for analysis described in section 2.3

At the group level, subject-specific contrast maps were entered into a second-level (one-sample) model implemented in Nilearn to generate group-level t-statistic maps. Statistical maps were thresholded at a voxel-wise threshold of $p < 0.001$ (uncorrected) and corrected for multiple comparisons at the cluster level using family-wise error (FWE) correction at $p < 0.05$.

fMRI: Bayesian model selection

To formally adjudicate between competing computational models, we performed voxelwise Bayesian model comparison using SPM25. At the first level, subject-specific general linear models were estimated using SPM's variational Bayesian framework. For each participant, six runs were modeled within a single design matrix. Model estimation used SPM's Bayesian GLM implementation, which estimates voxelwise posterior parameter distributions and log model evidence (LogEv) maps under spatial priors on regression coefficients and autoregressive noise modeling. Analyses were restricted to an anatomically defined masks to focus model comparison on regions implicated in value representation. For each subject and each candidate model, voxelwise LogEv images were computed.

At the group level, random-effects Bayesian Model Selection (RFX BMS) was conducted across subjects using SPM's Bayesian model comparison framework. For each voxel within the mask, LogEv maps from all competing models were entered into a hierarchical random-effects analysis to estimate posterior model probabilities and exceedance probabilities, accounting for inter-individual heterogeneity in model expression. Posterior probability maps thus quantify, at each voxel, the probability that a given computational model provides the best explanation of the observed neural data relative to the alternatives.

fMRI: individual differences

To examine whether regional activity was associated with stable individual differences in switching behavior, we related each participant's mean ROI contrast estimate to their overall switching tendency across the task. For each subject, fixed-effects contrast estimates were extracted from subject-specific first-level GLM contrast images and averaged within anatomically or functionally defined ROIs, yielding a single mean beta estimate per participant for each contrast of interest. Behavioral goal- and subgoal-switching rates were computed across all runs as the proportion of trials on which a switch occurred. Relationships between ROI activation and switching behavior were assessed across participants using Spearman rank

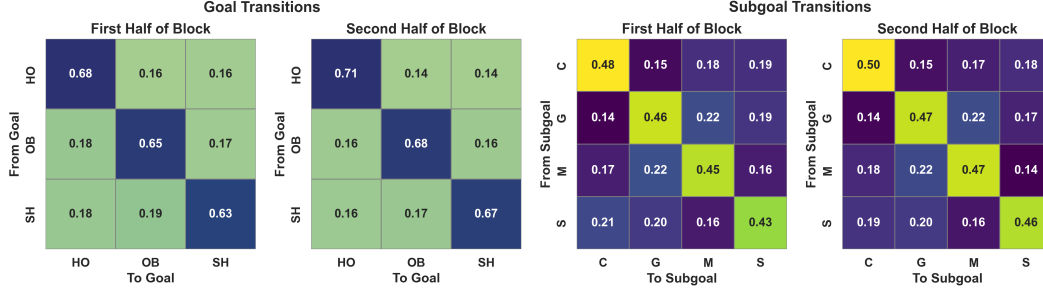


Figure .7: Participants' transitions between goals across successive rounds in first and second halves of blocks

correlation. This analysis tested whether between-subject differences in regional activation magnitude were associated with stable differences in switching propensity, providing a complementary account to the within-subject logistic analyses by capturing trait-like variability in flexible control.

.1 Computational modeling

Alternative hypotheses:

We proposed the following alternative accounts for goal and subgoal selection in the task, reproduced from [2], to compare with our hierarchical momentum model.

Prospective agent

The preference for progress can be optimal in agents that discount their future rewards. We constructed a prospective agent that optimizes for discounted expected returns in the game. The prospective agent maintains a running estimate of resource probabilities and estimates the future prospects of goal and subgoal completion using hyperbolic discounting.

The agent updates the belief about the resource probabilities using a delta learning rule:

$$M_r^{(t)} = M_r^{(t-1)} + \alpha(I_r^{(t)} - M_r^{(t-1)}) \quad (13)$$

where $M_r^{(t)}$ is the belief estimate of the probability of obtaining a specific resource r at time t , and $I_r^{(t)}$ is the indicator function which is 1 if the resource is given in a particular round. α is the learning rate of the agent.

The agent models the completion of each subgoal as a stochastic process. At each attempt, the subgoal advances with probability M_r , where r denotes the resource.

The simulation proceeds recursively until the subgoal (a specific target number for the resource r) is completed or until a maximum of 25 attempts is reached. The value of pursuing a given subgoal in a goal is given as follows:

$$\gamma = \frac{1}{1 + k * C_{sg|g}} \quad (14)$$

$$Q_{sg|g} = R_{sg|g} * \gamma \quad (15)$$

where $C_{s|g}$ is the number of rounds that the stochastic simulation process estimates would be needed to finish the subgoal given a goal, γ is the hyperbolic discount factor parameterized by k , $Q_{sg|g}$ is the value of the subgoal given the specific goal, $R_{sg|g}$ is the reward for subgoal completion. We assumed $R_{sg|g}$ to be identical across subgoals.

We compute the overall prospective value for the goal based on the estimate of the number of attempts it takes to complete the goal as follows:

$$Q_g = R \frac{1}{1 + k \sum_{sg'} C_{sg'|g}} \quad (16)$$

where R is the goal completion reward identical for all goals. $\sum_{sg'} C_{sg'|g}$ gives the number of rounds needed to finish all subgoals, and therefore the overall goal. If a specific goal has many resources already filled, the number of rounds it would take to finish pending subgoals, and then the overall goal is lower, making the goal value higher. Thus, the prospective agent gives a preference for progress when valuing goals and subgoals.

TD-persistence

TD-persistence algorithm samples resources to infer the dominant goal and persists with it until the block shifts, and repeats the process when a new block begins.

TD-persistence estimates resource probabilities (M_r) through experience, just like the prospective agent (equation 18). Goal value is estimated as the minimum resource bottleneck faced by a specific goal as follows:

$$Q_g = \min_r (M_r) \quad (17)$$

Subgoal value is estimated as its resource probability.

$$Q_{sg|g} = M_{sg} \quad (18)$$

Mixing strategies at levels of hierarchy

We considered the possibility of a combination of strategies for goal and subgoal selection. We included agents that, for instance, use TD-momentum for goal selection and use prospective computations for the subgoal agent (TD-momentum(G)-prospective(SG)).

Retrospective

We defined a retrospective algorithm that values goals and subgoals based on the number of resources needed for completion. It does not take into account the resource probabilities, so its value is related only to progress in goals and subgoals.

$$Q_{sg|g} = R_{sg|g} * \frac{1}{1 + k * \nu_{sg|g}}; \quad Q_g = R \frac{1}{1 + k \sum_{sg'} \nu_{sg'|g}} \quad (19)$$

where $\nu_{s|g}$ is the number of resources needed for the subgoal sg to be completed. Goal and subgoal choices are made as per the hierarchical goal selection algorithm.

Progress

While the retrospective algorithm values the count of resources remaining for completion, the *progress* model values fractional progress in goals and subgoals. This specifically dissociates subgoal value as defined by the retrospective agent, as subgoals in different goals have different requirements.

$$Q_{sg|g} = m_{sg|g}; \quad Q_g = m_g \quad (20)$$

where $m_{sg|g}$ is the fractional progress of subgoal sg in goal g , m_g is the fractional progress of g .

Resources

The *resources* model deviates from the hierarchical goal selection algorithm. The model finds the easiest-to-attain resources and pursues them till the resource inventory limit is reached before moving to the next one.

$$Q_{sg} = M_{sg} \quad (21)$$

Here, the subgoal value is not conditioned on the goal. Goals are selected with a value based on how closer they are to completion: $Q_g = m_g$, where m_g is fractional progress. Goal selection does not play any role in resource selection, and therefore, there is no hierarchy.

We included a choice kernel parameter to ensure perseveration of goal and subgoal selections.

Model Comparisons

We contrasted the momentum model with the alternative hypotheses described above. We optimized the model parameters to maximize the aggregate likelihood of goal and subgoal choices made by the participants. We corrected for model complexity using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) metrics. The momentum-based algorithms applied to both goal and subgoal selection provide a better fit for the aggregate choice likelihood over the alternative hypotheses (Figure). The momentum model equipped with cross-goal transfer best accounted for the choices in the experiment.

Model and parameter recovery

For parameter recovery analysis, we sampled 500 parameter vectors from a range of parameters ($\alpha \in [0, 1]$, $c_g \in [-1, 1]$, $c_{sg} \in [-1, 1]$, $\beta_g \in [0, 10]$, $\beta_{sg} \in [0, 10]$, $\gamma \in [0.6, 1.0]$, $\alpha_c \in [0, 1]$). We generate synthetic behavior for each parameter vector using the TD-momentum-learn-alt-goal model and recover the parameters by fitting the model to the simulated data. We optimized both the likelihoods of selection of the goals and subgoals while fitting the data. We plotted the fitted parameters and the generated parameters and reported correlations. Most of the parameters are recoverable with high correlation coefficients (see Fig A6). However, low discount factors closer to 0.5 do not recover well; however, the discount factors of the participant fits are mostly in the range $[0.9, 1.0]$, so we considered the range $[0.6, 1.0]$ for parameter recovery.

We note that learning rate, choice kernel, and goal switching costs show biases in parameter recovery. There is an underestimation of the learning rate and an overestimation of the choice kernel in parameter recovery. We tried eliminating the choice kernel, which did not improve the recovery of the learning rate but rather reduced the recoverability of other parameters. We employ the model with the best overall recovery over the variants of models experimented with.

We performed a model recovery analysis on the competing models. We sampled 100 random parameter vectors from parameter ranges described above. For each of the algorithms described in the manuscript, we generated choice data using the sampled parameters. We then fit the generated data to each of the models and identified the winning model through BIC score comparisons. We plotted the confusion matrices where the rows correspond to the simulated model and the columns correspond to the recovered model. The left confusion matrix indicates the probability of the fitted model given the simulated model, whereas the right matrix indicates the probability of the simulated model given the fitted model (see Fig A7).

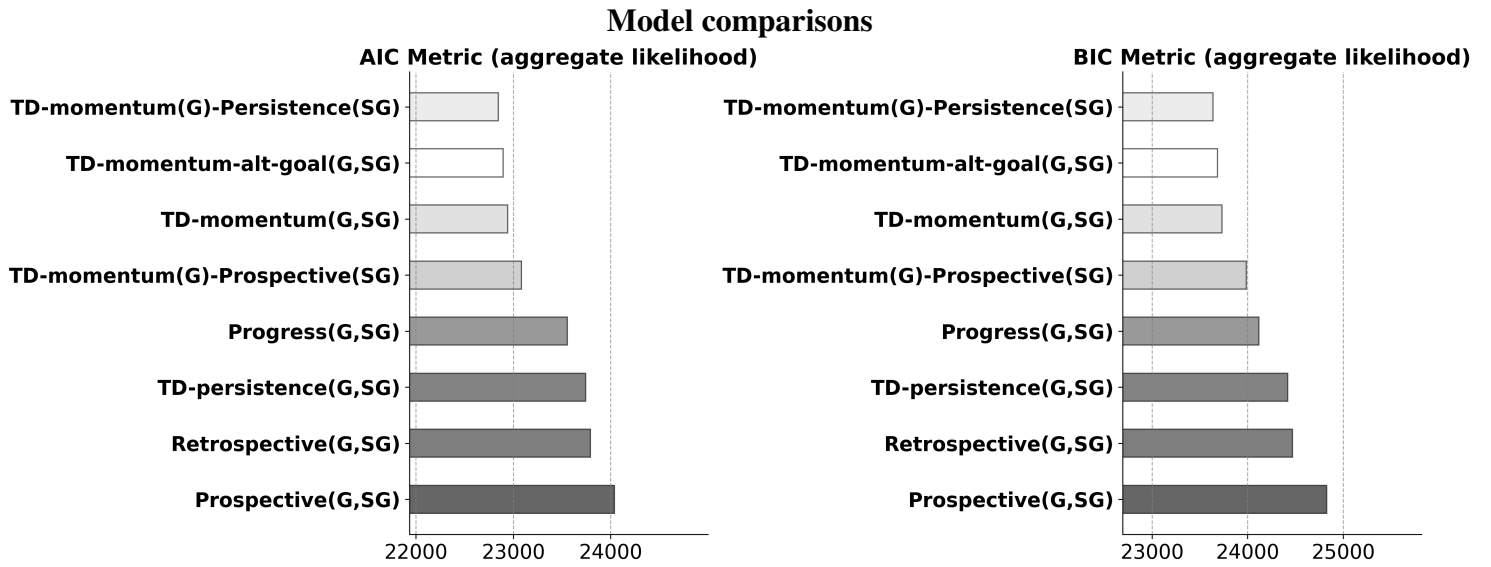


Figure .8: Model comparisons between candidate computational hypotheses. Momentum variants explain goal and subgoal selection patterns better than alternate accounts.

Correlation of momentum parameters with individual behaviors

Behavior	Parameter	Correlation	p-value
Number of goal switches	Goal switching cost (c_g)	0.84	$< 10^{-6}$
Number of subgoal switches	Subgoal switching cost (c_{sg})	0.70	$< 10^{-6}$
	Goal switching cost (c_g)	0.77	$< 10^{-6}$
Proportion of dominant choice	Gamma (γ)	0.65	< 0.0001
	Goal softmax (β_g)	0.44	< 0.05
	Action softmax (β_a)	-0.41	< 0.05
Dominant choice change proportion in high-vs-low	Gamma (γ)	0.55	< 0.01
	Learning rate (α)	0.36	< 0.05

Table .3

Parameter recovery

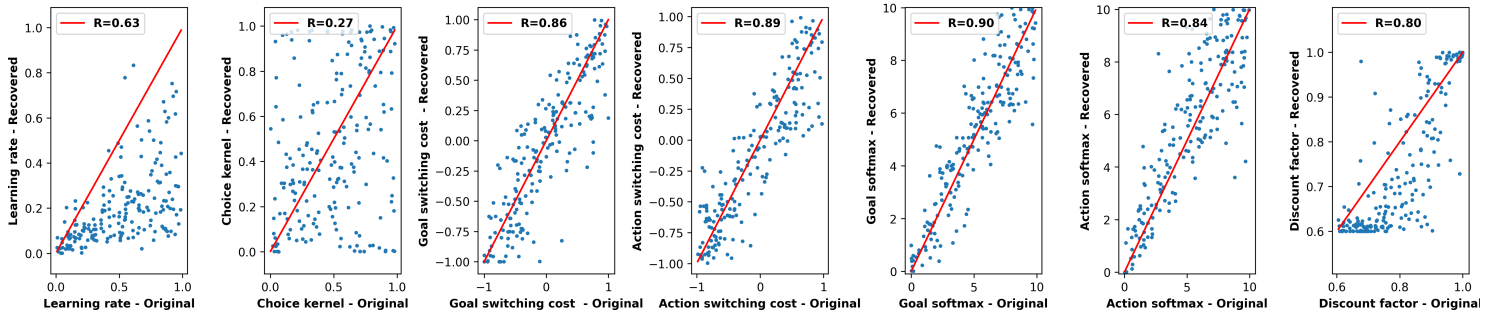


Figure .9: TD-momentum algorithm parameters recovered from independent simulations of behavior. Learning rate shows bias in recovery, and the choice kernel recovery is lower, yet significant. All other parameters recovered well.

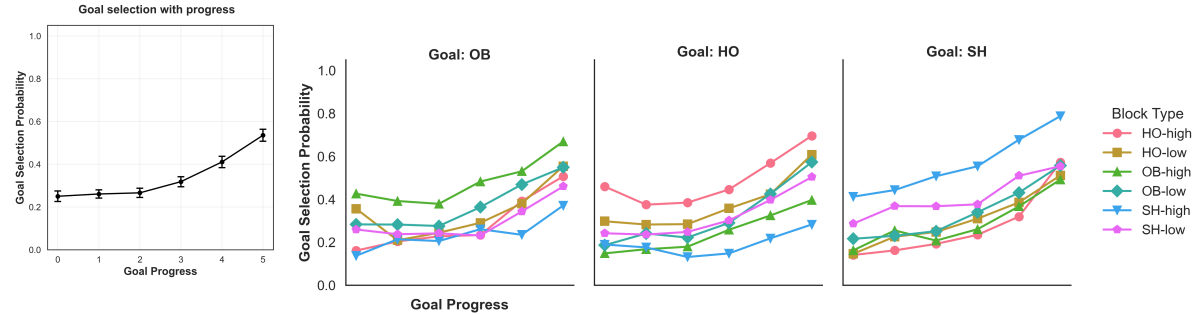
Model recovery



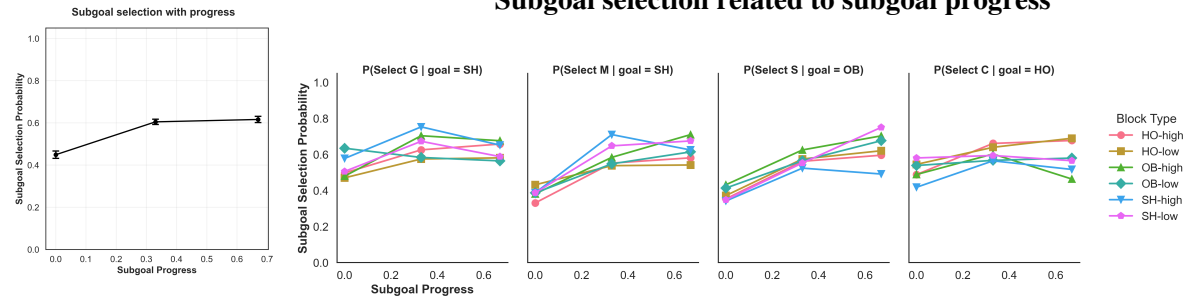
Figure .10: Model recovery of candidate computational hypotheses. TD-momentum-alt-goal and TD-momentum differ little in implementation and hence display confusion among the two. Both models, however, showed some interference with the progress model. The three momentum variants: TD-momentum-alt-goal, TD-momentum, TD-momentum-prospective showed good recovery, while other hypotheses interfered with each other. Final model used in the neural analysis is TD-momentum-alt-goal.

A. Behavior

Goal selection related to goal progress

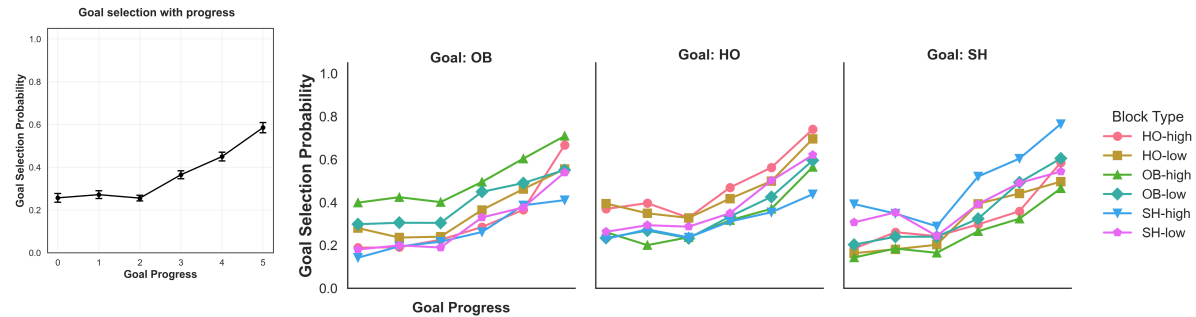


Subgoal selection related to subgoal progress



B. TD-momentum-alt-goal model simulations of participant behavior.

Goal selection related to goal progress



Subgoal selection related to subgoal progress

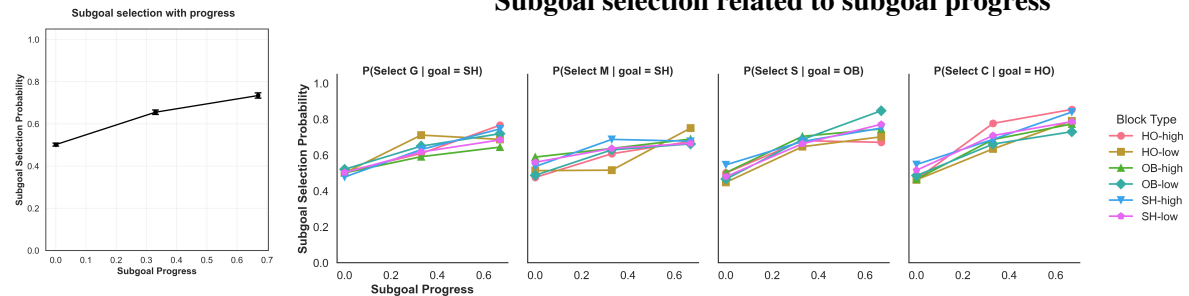


Figure .11

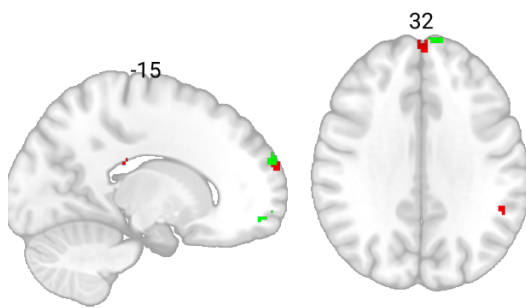


Figure .12: Chosen and alternate subgoals are represented in the medial frontal pole

Goal representations separated by task epoch

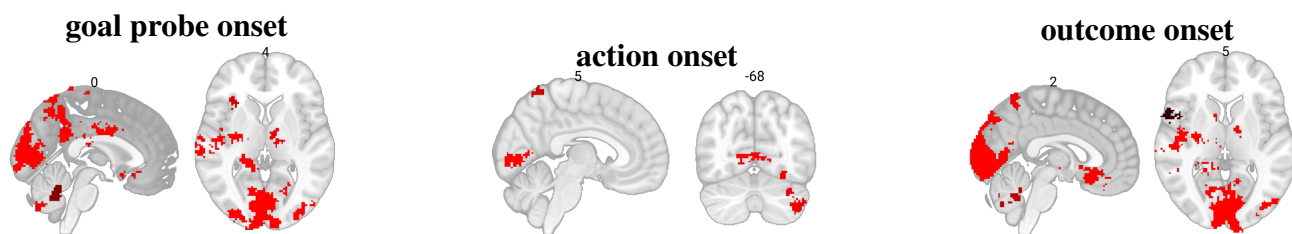


Figure .13: **Goal representations by task epoch.** Searchlight RSA analysis of goal representations (goal, subgoal, action modeled together) done separately for each task epoch. Statistical maps are corrected using permutation-based cluster correction ($n = 5000$ permutations), with a cluster-forming threshold of $t > 3.39$, $p < 0.05$ FWE-corrected, whole-brain.

Modeling RT as a separate regressor at goal onset

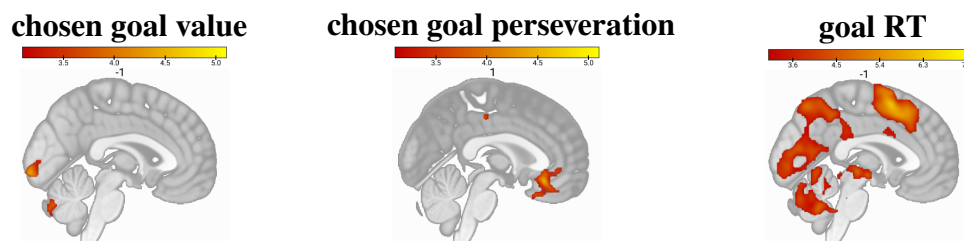


Figure .14: **Goal valuations across task epochs.** Chosen goal value, goal perseverance, and goal RT at goal probe when RT is modeled as a separate regressor rather than including it as a duration signal.

Chapter 6

GOALS AS DYNAMICAL ATTRACTORS: A MOMENTUM-BASED ACCOUNT OF STABLE AND FLEXIBLE GOAL COMMITMENT

Human goal pursuit is often marked by persistent activity toward achieving an objective, as well as flexibility in switching objectives based on environmental demands. How humans balance the stability and flexibility necessary for goal pursuit is the key question of this study. We propose that goal pursuit generates dynamic attractor modes in policy landscapes that produce stability in goal pursuit. The attractor properties are modulated through progress monitoring, allowing for the flexibility necessary to switch objectives in favor of alternative goals. Through simulations and behavioral cloning of human participants performing an extended goal selection task, we show how dynamic modes can develop in the latent spaces of recurrent neural networks trained with reinforcement learning. We develop metrics to quantitatively assess the attractor qualities of dynamic modes, validating them against synthetically generated dynamical systems, and use them to investigate the context modulation of attractor modes during goal pursuit. We then proceed to develop a circuit-level account of goal persistence incorporating self-excitation and cross-inhibition as motifs for fast, self-sustaining dynamics modulated by slow, progress-integrating momentum and context signals. Lastly, we show that the switching costs experienced while managing multiple goals are an emergent property of resistance to the intrinsic dynamics of goal pursuit, thereby contributing a fresh perspective on the dynamics of extended goal pursuit.

6.1 Introduction

We pursue multiple goals concurrently in our daily lives. Most often, however, we pursue one goal at any specific instance in time and maintain persistent activity in its favor. Say, writing a section of a manuscript requires you to sit at your desk and persistently engage with it. However, an email ping distracts you, and you switch over to the new task of checking your messages. We constantly manage multiple such goals in our daily lives, which requires us to maintain stable goal states for progressing in a task while also remaining flexible enough to switch over to alternative rewarding tasks in the environment. This fundamental aspect of

intelligence is essential to all agents aspiring toward general intelligence: the ability to manage diverse tasks under various environmental tradeoffs.

The mechanistic basis of how we maintain stability and flexibility in the pursuit of extended goals is an open question. A few theories of goal persistence address the stability aspect of the question by invoking resource-rational accounts of decision-making [60]. Goal persistence is thought to reduce policy complexity, enabling us to perform tasks by avoiding intensive computations. Some models of goal persistence incorporate policy bias or perseveration terms as inductive biases for self-sustaining activity in favor of goal pursuit. On the other hand, flexibility is thought to be a product of reinforcement learning of reward rates in the environment. However, as most extended goals deliver rewards only after goal completion, goal progress can be thought of as a pseudo-reward indicating the prospects of resolving a given goal [128]. [3, 2] proposed a momentum model of goal selection, where goal progress and its derivative (the rate of goal progress) can jointly influence decision-making, and their product, goal momentum, can serve as the drive influencing goal selection. This model was shown to capture several behavioral properties in humans pursuing extended goals: human overpersistence patterns, preference for goal progress, and increased rates of revisiting goals with greater progress after switching away from them. Such models remain good descriptive accounts of goal selection; however, they fall short of providing a mechanistic explanation of their basis in the brain, specifically how they enable the maintenance of stable goal states robust to environmental distractors and dynamic states that can be perturbed through rewards in the environment [43, 133].

The theory of dynamical systems holds great promise for generating a mechanistic explanation of extended goal pursuit, especially regarding how stability-flexibility tradeoffs are realized [134, 135, 180]. Attractor properties developed by dynamical networks are a strong analogue for architectures that can generate and maintain stable, self-sustaining intrinsic dynamics while still retaining flexibility by being pushed around by input-driven dynamics [49]. [78] made that connection by framing goals as far-from-equilibrium attractors of a dynamical system. [90] proposed an attractor-dynamics account of resolving motivational conflict by viewing behavioral switching either as a bifurcation or as a stochastic escape from an attractor basin. Attractor models of goal pursuit thus translate the metaphorical insight of persistence as a pull toward specific goals into a rigorous quantitative interpretation of decision-making as traversal through an energy landscape guided by context and rewards in

the environment.

Recurrent neural networks (RNNs) are a strong test bed for investigating dynamic modes of human decision-making, given their ability to integrate past experiences into a latent context that translates into action policies. Studying the dynamics of the latent space can therefore lend us an understanding of the behavioral dynamics seen in human decision patterns [42]. [9] put forth a modeling approach that employs tiny RNNs to discover cognitive strategies by fitting RNNs to human choices and showed that it can be a viable predictive model of human behavior. In this study, we use RNNs as both normative and descriptive accounts of human decision dynamics. We first simulate RNNs to investigate conditions under which they emulate human persistence patterns, and later we clone human behavior with RNNs to describe the dynamic modes underlying human persistence patterns.

While RNNs can be descriptive of human dynamic modes and serve as a baseline model of decision dynamics, they tell us little about the mechanistic implementation of such modes in the brain. Specific network motifs combining excitatory and inhibitory firing are known to facilitate self-sustaining behavior [174] and can facilitate selective attention in goal pursuit [82]. [107] showed in a discrete recurrent circuit that a combination of slow excitation and fast inhibition yielded persistent attractor dynamics that enabled stability and rapid switching during fly navigation. In this paper, we use a self-excitation and cross-inhibition network motif to generate fast self-sustaining dynamics that facilitate goal commitment and are dynamically modulated by slow, progress-integrating momentum signals. We investigate how well this model explains human behavior and study its dynamic modes as modulated by behavioral context. Through this work, we generate explicit mechanistic hypotheses about how goal states are stably maintained and dynamically shifted that can be tested for circuit-level neural implementation using electrophysiological recordings in animals or humans performing extended goal pursuit tasks.

6.2 Experiment and behavior

We use the extended goal pursuit paradigm introduced in [3] to model human goal selection. In the game, participants are free to pursue any of the three extended goals—collecting car, hat, or cat suits—and are free to switch between them without penalty. To finish a specific suit (goal), participants need to collect seven tokens of the same kind (Figure 1-left). The game proceeds in rounds, and participants can choose to acquire any of the three tokens at a given instance. The success of

the actions is probabilistic, and the probabilities vary throughout the task in blocks. Say, in a specific block in the task, cat tokens might be more available—for example, if you choose to acquire the cat token, you will be given the token with an 80% chance and no token with a 20% chance. On the other hand, in the same block, if you choose to acquire the hat token, you can get the token with a 20% chance and fail with an 80% chance. The task has three block types—80-20, 60-40, and 70-30—wherein the 80-20 block has one goal that gives tokens with an 80% chance, while the other two goals give tokens with a 20% chance. Participants need to infer the identity of the goal that has the highest probability within a block for optimal goal pursuit and switch adaptively when the block changes and a new goal becomes the dominant one. The objective of the game is to complete as many goals as possible. We acquired the openly available human behavior dataset from [3] to model the behavior of 44 participants who played the game.

Participants in the game maintain persistent goal states for extended durations. Participants’ aggregate transition probability matrix (Figure 1) shows a significantly higher probability of staying with the same goal across adjacent rounds. Participants also show a preference for progress, wherein their probability of selecting a goal increases with progress toward that goal. Figure 1 shows participants’ goal selection probability plotted against the progress difference between the goal with the maximum progress (max-progress goal) and the goal that gives tokens with a higher probability (max-prob goal). Participants also show sensitivity to block types. In the 80-20 block, where there is a larger incentive to pick the max-prob goal, participants show a lower preference for progress compared to the other blocks. Humans in the game were shown to overpersist beyond what is optimal for the task. [3] constructed algorithms that beat human performance (measured by the number of goals completed) by a significant margin. One of the agents prospectively estimates the time to goal completion by learning the probability structure of goals within the block while discounting future rewards. This agent shows progress preference and persistence in the task, yet substantially outperforms humans, who show patterns of overpersistence compared to the prospective agent. We now train recurrent neural networks to play the same game as participants to investigate what parameterizations can produce persistence trends similar to those observed in humans.

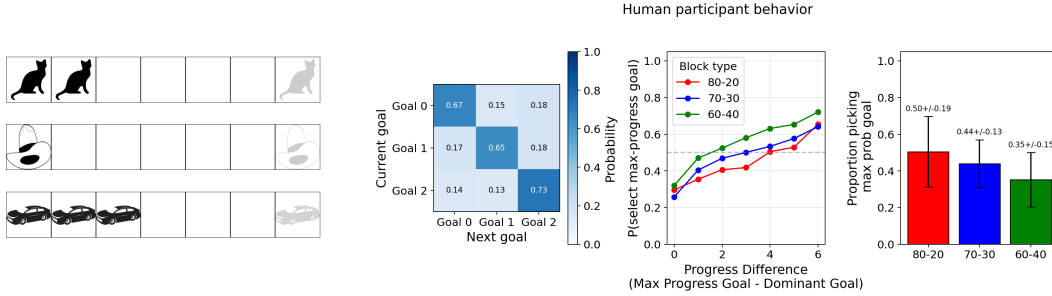


Figure 6.1: **Task and behavior.** (left) Extended goal pursuit task: Collect 7 tokens of the same kind to finish a suit and collect rewards. Participants can switch between suits (cat/ hat/ car) without penalty. (right) Human behavior in the task: (Panel 1) Probability of transitioning from one goal to the next across adjacent rounds. (Panel 2) Participants goal selection probability plotted against the progress difference between the goal with the maximum progress — max-progress goal — and the goal which gives tokens with a higher chance — max-prob / dominant goal. (Panel 3) Participants probability of selecting the max-prob goal in different block types.

6.3 Modeling the dynamics of extended goal commitment using RNNs

RNNs optimizing for discounted future returns in the task exhibit persistence and preference for progress

We trained a recurrent actor-critic network to play the extended goal pursuit task, optimizing for total rewards in the game. The network takes the following observation

$$o_t = \left[\underbrace{\tilde{p}_t}_{\text{progress}}, \underbrace{a_{t-1}^{\text{one-hot}}}_{\text{previous action}}, \underbrace{\text{token}_{t-1}}_{\text{previous outcome}}, \underbrace{g_{t-1}^{\text{one-hot}}}_{\text{goal completion}}, \underbrace{\text{block shift}}_{\text{context cue}} \right]$$

It first embeds the observation and passes it through a gated recurrent unit (GRU), which ensures the latent state integrates past observations and actions.

$$x_t = \text{ReLU}(W_{\text{in}} o_t + b_{\text{in}}); \quad h_t = \text{GRU}(x_t, h_{t-1}) \quad (6.1)$$

We then define the policy (actor) and value (critic) based of the latent state.

$$\pi_{\theta}(a_t | h_t) = \text{softmax}(W_{\pi} h_t + b_{\pi}); \quad V_{\theta}(h_t) = W_V h_t + b_V \quad (6.2)$$

We trained the network to optimize the number of goals finished in the task (details provided in Appendix Section B). The recurrent policy architecture learns to perform the task, albeit without displaying the degree of persistence seen in human participants (see Appendix Figure 4). The network also does not show the preference toward progress observed in humans. We could modulate the network’s persistence levels by tweaking the variable that discounts future returns in the game. It is widely documented that humans incorporate discounting in their valuation

[154], which partly accounts for their persistence with the max-progress goal. We therefore modified the network’s objective to optimize discounted rewards in the game ($\gamma = 0.9$), which increased persistence levels to those of human participants and also reproduced the progress preference seen in humans (Appendix Figure 4). We then let the trained network play the game and plotted the principal components (PCs) of the latent-space trajectory over time. Furthermore, the latent space shows a clear gradation of goal progress and goal progress rate, indicating that the network’s latent space adequately represents key factors necessary for playing the game. We next clone participants’ behavior using recurrent neural networks and analyze their latent space for dynamic modes of human persistence.

Behavior cloning through recurrent neural networks reveals attractor-like modes in goal-directed policy dynamics

We model each subject’s sequential decisions using a recurrent neural network (RNN) trained via behavioral cloning [168, 50]. At each trial t , the model receives an observation vector o_t encoding current goal progress, previous action, previous reward outcome, previous goal completion, and block-boundary signals.

The RNN updates a latent state h_t and produces action probabilities:

$$h_t = f_\theta(o_t, h_{t-1}), \quad (6.3)$$

$$\pi_\theta(a_t \mid o_{\leq t}) = \text{softmax}(Wh_t + b). \quad (6.4)$$

We fit the parameters independently for each subject by minimizing the negative log-likelihood of observed actions:

$$\mathcal{L}(\theta) = -\mathbb{E}_{t \in \mathcal{I}_{\text{train}}} [\log \pi_\theta(a_t \mid o_{\leq t})], \quad (6.5)$$

using an interspersed train/validation/test split (75/12.5/12.5). We evaluated performance using held-out negative log-likelihood, prediction accuracy, and agreement with empirical choice frequencies. We provide further details on the behavioral cloning procedure in Appendix Section C. The fitted hidden states $\{h_t\}$ are used for downstream dynamical analyses.

Context-Dependent Intrinsic Dynamics: To characterize the internal dynamics of the trained recurrent neural networks, we analyze their intrinsic velocity fields in a low-dimensional latent space. We first extract the hidden-state trajectory $\{h_t\}_{t=1}^T$ for each subject and project it onto the top principal components using PCA. Let $z_t \in \mathbb{R}^2$ denote the projection onto a chosen 2D plane. To define context-conditioned

dynamics, we group time points according to the previous goal and current progress level. For each context c , we construct a representative observation $\bar{o}^{(c)}$ by averaging observations within that context. Rather than evaluating dynamics only along observed trajectories, we compute the intrinsic flow field over an artificial grid of latent states spanning the PCA space. Each grid point is mapped back to a corresponding hidden state, with non-visualized principal components fixed to their mean values. Holding the context fixed via $\bar{o}^{(c)}$, we evaluate a single-step RNN transition: $h' = f_\theta(\bar{o}^{(c)}, h)$ and define the intrinsic velocity as $v(h; c) = h' - h$. This velocity is then projected into the PCA plane to obtain a vector field over the grid.

To identify candidate attractors, we locate regions of low velocity magnitude in the projected field and extract local minima of $|v(h; c)|$, which we interpret as slow points—proxies for fixed points in the underlying high-dimensional dynamics [160]. The resulting vector fields, evaluated on a grid in PCA space and visualized as quiver plots, are shown in Appendix Figure 5. This procedure (elaborated further in Appendix Section C.1) yields context-dependent dynamical landscapes that reveal how goal representations are stabilized or destabilized as a function of behavioral state. Qualitatively, the flow fields exhibit point-attractor-like structure, with trajectories converging toward compact regions in latent space. While they might not qualify as strict fixed-point attractors, these quasi-attractor-state properties are modulated by increasing goal commitment [4]. To quantify this effect, we next introduce metrics that capture the “pointiness” of these quasi-attractor modes.

Quantifying Point-Attractor Structure: To quantify the extent to which learned dynamics exhibit point-attractor structure, we analyze intrinsic velocity fields within each behavioral context (previous goal $\times$ progress level). Candidate fixed points are identified as regions of low velocity magnitude. For each candidate, we evaluate whether it satisfies the defining properties of a true point attractor using a strict composite metric. Specifically, we define a *strict point-attractor score* that combines multiple necessary dynamical criteria. This score is high only when a candidate exhibits inward flow, basin support, and full local stability. The components are:

- **Inward alignment:** Measures whether nearby vectors in the field point toward the candidate. High values indicate that the local flow is directed inward, consistent with attractor dynamics.
- **Point-specific stability:** Assesses local stability using the Jacobian of the vector field. The score penalizes both unstable directions (positive eigenval-

ues) and neutral directions (near-zero eigenvalues), ensuring that only fully stable fixed points receive high values.

- **Basin support:** Quantifies whether nearby trajectories converge toward the candidate over short rollouts. This captures whether the candidate has a meaningful basin of attraction beyond its immediate neighborhood.

These components are combined multiplicatively to yield a strict point-attractor score, such that a high score requires all properties to be simultaneously satisfied. We provide details of the metric and its validation using synthetic dynamical systems in Appendix Section D. We compute these metrics separately for flow fields separated into low- and high-progress contexts and compare them within subjects using paired statistical tests. Figure 2 (bottom) shows that high-progress states exhibit significantly higher strict point-attractor scores than low-progress states, consistent with the idea that increasing progress stabilizes goal representations by deepening attractor dynamics in latent space. While behavioral cloning of human goal selections provides insight into the context-dependent intrinsic dynamics of choice patterns, we still lack a mechanistic account of goal persistence that can point to specific circuit motifs resulting in the dynamic modes apparent in human behavior. In the next section, we formulate a circuit-level interpretable dynamical-systems account of human goal pursuit.

6.4 A momentum-based circuit model of goal pursuit

We introduce a momentum-based circuit-level implementation of goal selection, in which goal commitment emerges from recurrent competition in policy space. Persistence arises from fast competitive dynamics between goal states, modulated by progress-dependent signals, environmental context, recent choice history, and feedback from the selected action.

We represent goal commitment as a distribution over three goals, $\mathbf{x}_t \in \Delta^2$, where the 2-simplex is defined as:

$$\Delta^2 = \left\{ \mathbf{x} \in \mathbb{R}^3 : x_i \geq 0, \sum_{i=1}^3 x_i = 1 \right\}. \quad (6.6)$$

Drive and policy: At each time step, the circuit computes a drive vector:

$$\mathbf{d}_t = W\mathbf{x}_t + \beta_{\text{ctx}}\mathbf{c}_t + \beta_{\text{vel}}\mathbf{v}_t + \beta_{\text{mom}}(\mathbf{p}_t \odot \mathbf{v}_t) + \beta_{\text{prog}}\mathbf{p}_t + \beta_{\text{prev}}\mathbf{a}_{t-1} + \mathbf{b}, \quad (6.7)$$

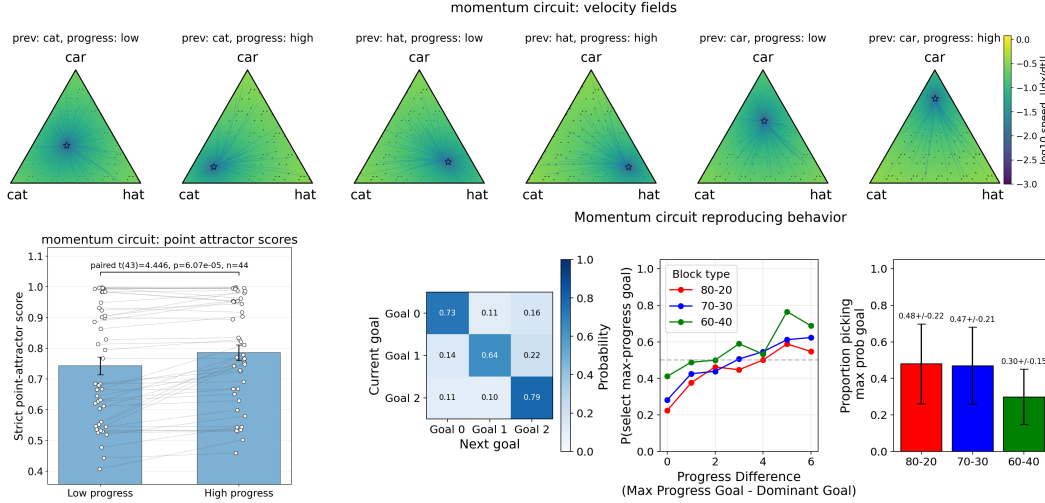


Figure 6.2: (top) Context-dependent intrinsic velocity fields of the momentum-circuit model of a single participant in the simplex space showing point-attractor-like structure. Each panel shows the velocity field for a specific previous goal and progress level. (bottom) (left) Significantly higher point-attractor scores in high-progress contexts compared to low-progress contexts showing attractor-structure evolution with environmental context. (right) Momentum-circuit model reproducing human persistence patterns and progress preferences.

where $W \in \mathbb{R}^{3 \times 3}$ encodes recurrent competition, $\mathbf{c}_t$ is a context estimate, $\mathbf{p}_t \in [0, 1]^3$ is goal progress before choice, $\mathbf{v}_t$ is an asymmetric temporally integrated progress velocity, $\mathbf{a}_{t-1}$ is a one-hot vector encoding the previous action, $\mathbf{b}$ is a goal-specific bias, and $\odot$ denotes elementwise multiplication. Progress therefore enters both directly through $\beta_{\text{prog}} \mathbf{p}_t$ and indirectly through velocity and progress-gated momentum.

Recurrent competition: The recurrent interaction matrix implements self-excitation and cross-inhibition between goal states [122].

$$W_{ij} = \begin{cases} w_{\text{self}}, & i = j, \\ -w_{\text{cross}}, & i \neq j. \end{cases} \quad (6.8)$$

Equivalently, the recurrent contribution can be written componentwise as

$$(W\mathbf{x}_t)_i = w_{\text{self}}x_{t,i} - w_{\text{cross}} \sum_{j \neq i} x_{t,j}. \quad (6.9)$$

Asymmetric momentum computation: Humans pursuing goals care not only about how much progress they have made, but also about how fast they continue to make progress. To capture this, we define an asymmetric leaky progress-velocity trace. First, instantaneous progress change is computed as

$$\Delta \mathbf{p}_t = \mathbf{p}_t - \mathbf{p}_{t-1}, \quad (6.10)$$

with $\Delta \mathbf{p}_0 = \mathbf{0}$. Positive and negative changes are separated as

$$\Delta \mathbf{p}_t^+ = \max(\Delta \mathbf{p}_t, \mathbf{0}), \quad \Delta \mathbf{p}_t^- = \max(-\Delta \mathbf{p}_t, \mathbf{0}), \quad (6.11)$$

where the maximum is applied elementwise. The velocity trace is then updated according to

$$\mathbf{v}_t = \lambda_v \mathbf{v}_{t-1} + \eta_v^+ \Delta \mathbf{p}_t^+ - \eta_v^- \Delta \mathbf{p}_t^-, \quad (6.12)$$

where $0 < \lambda_v < 1$ is the velocity decay and $\eta_v^+, \eta_v^- > 0$ are learned gains for positive and negative progress changes. This allows positive progress and negative/no-progress evidence to exert different effects on the velocity trace. This leaky updating of progress gradients induces a slow-integrating momentum signal [45], akin to momentum in gradient descent updates [162].

Another momentum-like computation that further builds inertia at high-progress levels is the progress-gated velocity signal: $\mathbf{m}_t = \mathbf{p}_t \odot \mathbf{v}_t$. Thus, $\mathbf{v}_t$ captures a temporally smoothed and valence-asymmetric estimate of recent progress change, while $\mathbf{p}_t \odot \mathbf{v}_t$ amplifies commitment when progress is high and recently increasing.

Context computation: While slow-integrating momentum signals provide an approximate measure of ongoing goal approach, they do not by themselves encode broader environmental statistics that can inform goal selection, such as block structure in the task. The context vector $\mathbf{c}_t \in \mathbb{R}^3$ encodes a running estimate of goal-specific success rates within each block. Let $S_{t,i}$ and $F_{t,i}$ denote the number of successes and failures for goal i up to time t . We compute:

$$\hat{p}_{t,i} = \frac{S_{t,i} + \alpha_0}{S_{t,i} + F_{t,i} + \alpha_0 + \beta_0}, \quad c_{t,i} = \hat{p}_{t,i} - \frac{1}{3} \sum_j \hat{p}_{t,j}, \quad (6.13)$$

so that $\mathbf{c}_t$ reflects the relative advantage of each goal within a given block.

Action policy: Actions are sampled from a softmax policy, where $\tau > 0$ is the softmax temperature.

$$\boldsymbol{\pi}_t = \text{softmax}\left(\frac{\mathbf{d}_t}{\tau}\right), \quad (6.14)$$

State dynamics. The latent goal state evolves via leaky integration toward the current policy, with an additional choice-feedback term that pulls the latent state toward the action selected on the current trial. Let $\mathbf{a}_t$ denote a one-hot vector encoding the chosen action at time t . The unnormalized next state is

$$\tilde{\mathbf{x}}_{t+1} = (1 - \alpha_x) \mathbf{x}_t + \alpha_x \boldsymbol{\pi}_t + \gamma_{\text{choice}} \mathbf{a}_t, \quad (6.15)$$

and the simplex-constrained latent state is obtained by rectification and normalization:

$$\mathbf{x}_{t+1} = \frac{\max(\tilde{\mathbf{x}}_{t+1}, \epsilon)}{\sum_i \max(\tilde{x}_{t+1,i}, \epsilon)}, \quad (6.16)$$

where $0 < \alpha_x < 1$, $\gamma_{\text{choice}} \geq 0$, and ϵ is a small numerical floor.

The model defines a dynamical system in which recurrent interactions $W\mathbf{x}_t$ generate attractor structure over goals, while context, direct progress, asymmetric progress velocity, progress-gated momentum, recent action history, and choice feedback modulate this landscape based on ongoing task performance. This interaction gives rise to persistent goal states and switching behavior as emergent properties of the underlying dynamics. We fit the model independently for each subject using maximum likelihood and analyze the resulting dynamics in the simplex policy landscape.

Simplex Landscape of Goal Dynamics. To visualize the dynamical modes of the fitted circuit model, we construct vector fields over the goal simplex. Given a goal state $\mathbf{x}$, the model updates it based on recurrent dynamics and fixed task-dependent inputs, inducing movement within simplex policy space. For a fixed input condition $(\mathbf{c}, \mathbf{p}, \mathbf{v}, \mathbf{a}_{\text{prev}})$ and drive, $\mathbf{d}$, defined as per equation (7), define

$$\pi(\mathbf{x}; \mathbf{c}, \mathbf{p}, \mathbf{v}, \mathbf{a}_{\text{prev}}) = \text{softmax}\left(\frac{\mathbf{d}}{\tau}\right) \quad (6.17)$$

The corresponding policy-driven update direction is

$$\dot{\mathbf{x}} = \pi(\mathbf{x}; \mathbf{c}, \mathbf{p}, \mathbf{v}, \mathbf{a}_{\text{prev}}) - \mathbf{x}. \quad (6.18)$$

When visualizing the full latent-state update including choice feedback, the vector field can instead be defined using the normalized update map

$$\dot{\mathbf{x}} = \mathcal{N}((1 - \alpha_x)\mathbf{x} + \alpha_x \pi(\mathbf{x}; \mathbf{c}, \mathbf{p}, \mathbf{v}, \mathbf{a}_{\text{prev}}) + \gamma_{\text{choice}}\mathbf{a}) - \mathbf{x}, \quad (6.19)$$

where $\mathcal{N}(\mathbf{z}) = \max(\mathbf{z}, \epsilon) / \sum_i \max(z_i, \epsilon)$. In the intrinsic policy-landscape analyses, we fix the behavioral context by holding all non-state inputs constant. Specifically, for each condition, defined by previous goal identity and progress level, we compute representative values of progress, velocity, momentum, previous action, and context signals by averaging over matching trials. These anchor variables define a fixed input condition, allowing us to analyze the induced dynamical system over the goal state alone. We evaluate $\dot{\mathbf{x}}$ on a dense grid over the simplex and visualize vector directions, speed $\|\dot{\mathbf{x}}\|$, and simulated trajectories from random initial states.

Candidate attractors are identified by simulating trajectories under the vector field and clustering their endpoints. Stable goal states appear as convergence points in the simplex.

Figure 3 (top) shows such a simplex visualization for one participant with the asymmetric progress-velocity momentum circuit model fitted to behavioral choices. The fitted flow fields show policy dynamics biased toward the previously chosen option, thereby inducing persistence. The flow toward the previously chosen goal becomes stronger in high-progress contexts. We repeated the same analyses used to quantify the point-attractor score, this time estimating the score in simplex policy space. Figure 3 (bottom) shows a significant increase in point-attractor scores for flow fields derived from high-progress contexts compared with low-progress contexts, indicating a progress-modulated shift in the dynamic landscape guiding policy transitions. We thus show a circuit-level model of goal selection that can induce persistence through dynamic modes in policy space, with dynamics shaped by environmental rewards, goal progress, progress velocity, choice history, and choice feedback.

Ablation experiments: We performed ablation experiments on the momentum-circuit model to understand the contribution of each component to the model’s performance. We ablated the following terms one at a time: progress (β_{prog}), progress velocity (β_{vel}), interaction (β_{mom}), context vector (β_{ctx}), and velocity asymmetry (collapsing η_v^+ , η_v^-). Table 1 shows the results of the ablations evaluated on the validation dataset. We report the model’s negative log-likelihood of choices (val NLL) as well as its performance on key behavioral measures. We evaluated mean-squared errors (MSE) on three behavioral metrics corresponding to the plots shown in Figure 3 (bottom): conditional switching probabilities $P(\text{switch} \mid a_{t-1})$ for each previous action, progress-conditioned switching $P(\text{switch} \mid p_t^{(a_{t-1})})$ across low, medium, and high progress bins, and dominant-goal choice probability $P(a_t = g_t^* \mid \text{block})$ across block conditions (80–20, 70–30, 60–40). The full momentum-circuit model successfully recovers these behavioral metrics, as validated through simulations (details in Appendix Section E2, Appendix Figure 8). The full model performs better in val NLL scores as well as aggregate MSE scores on behavioral measures. Moreover, the momentum-circuit model outperforms the behavior-cloned RNN and the momentum model described in [3] across all metrics.

Table 6.1: Ablation study of the momentum-circuit architecture on held-out validation data. Performance is evaluated using validation negative log-likelihood (val NLL) and several switch-prediction metrics $\sqrt{\text{MSE}}$

Model	Val NLL	Conditional switch $\sqrt{\text{MSE}}$	Progress conditioned switch $\sqrt{\text{MSE}}$	Max prob choice given block $\sqrt{\text{MSE}}$	Aggregate $\sqrt{\text{MSE}}$
Momentum-circuit: Full model	1967.1	0.0327	0.0370	0.0262	0.0323
Momentum-circuit: Interaction ablated	1971.7	0.0328	0.0374	0.0266	0.0326
Momentum-circuit: Progress ablated	1981.4	0.0356	0.0348	0.0268	0.0326
Momentum-circuit: Velocity asymmetry ablated	1989.1	0.0328	0.0371	0.0282	0.0329
Momentum-circuit: Context ablated	2004.7	0.0325	0.0372	0.0375	0.0358
Momentum-circuit: Velocity ablated	2013.1	0.0340	0.0584	0.0290	0.0425
Behavior cloning (RNN)	2047.5	0.0501	0.0526	0.0514	0.0514
Momentum model from [3]	2135.7	0.0569	0.0601	0.0700	0.0626

6.5 Switching Costs as Resistance to Intrinsic Dynamics

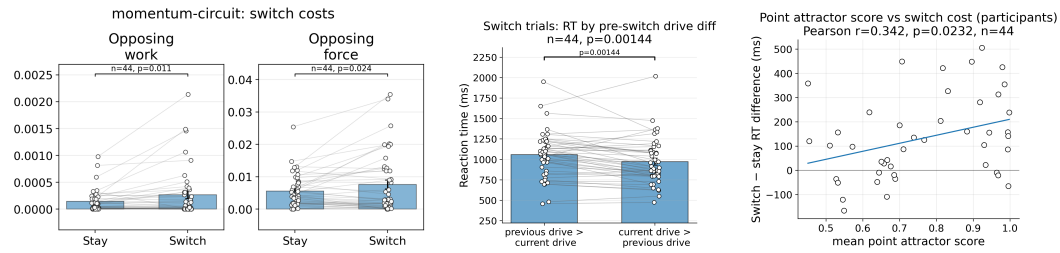


Figure 6.3: **Switching costs as resistance to intrinsic dynamics:** (left) Work done opposing the intrinsic dynamics and force opposing the state update are significantly greater in the switch condition than in the stay condition in the latent dynamics of momentum-circuit model, accounting for the emergence of switching costs. (center) Reaction time for switch trials is significantly greater when the previous goal drive is greater than the current goal drive compared to when the current goal drive is greater than the previous goal drive. (right) Participants’ switching cost is positively correlated with their point attractor score, indicating that individual differences in behavioral switching are explained by differences in the dynamical stability of goal policies.

We formalize switching costs [158] as resistance to the intrinsic flow field of the learned dynamical system. Let $x_t \in \Delta^2 \subset \mathbb{R}^3$ denote the goal state. The model induces a velocity field: $v(x_t) = \pi(x_t) - x_t$, where $\pi(x_t)$ is the policy implied by the circuit. Given an empirical trajectory $x_t \rightarrow x_{t+1}$, we define the observed displacement: $\Delta x_t = x_{t+1} - x_t$.

We measure how much this movement opposes the intrinsic dynamics via projection onto the velocity field. The instantaneous opposing force is defined as

$$F_{\text{opp}}(t) = \max(0, -\Delta x_t^\top v(x_t)). \quad (6.20)$$

This quantity is zero when the trajectory follows the flow and positive when it moves against it. We define the total opposing work along a trajectory as $W_{\text{opp}} =$

$\sum_t F_{\text{opp}}(t)$, which captures the cumulative resistance required to deviate from the system’s natural dynamics.

To quantify switching cost, we partition timesteps into stay ($a_{t+1} = a_t$) and switch ($a_{t+1} \neq a_t$) events, and compute the mean opposing work in each condition. Switching cost emerges when: $\bar{W}_{\text{switch}} > \bar{W}_{\text{stay}}$, indicating that switching requires greater work against the intrinsic flow field (Figure 3). When $\bar{W}_{\text{switch}} < \bar{W}_{\text{stay}}$, where the drive to switch is stronger than staying, switching costs do not emerge. This explains why participants often revisit high-progress goals after switching away from them [3], and why some task switches are easier than others [131, 91]. Figure 3 (center) shows that participants’ reaction times in switch trials are significantly higher when the previous goal’s drive exceeds the current goal’s drive, indicating higher switching costs when working against the intrinsic dynamics of the system. Participants’ switching costs are also positively correlated with their mean point-attractor score (Pearson correlation $R = 0.34, p = 0.02$), indicating that individual differences in switching are explained by differences in the dynamical stability of goal policies.

6.6 Discussion

Human goal pursuit is a mix of stability and flexibility, stability to ensure goals make consistent progress toward their targets, and flexibility to facilitate switching to better alternatives when available. Previous work shows that human goal commitment is significantly modulated by goal progress, where decision-making vigor increases closer to goal resolution, and by goal progress rates, where repeated failures increase the drive to switch away from the goal [3, 2, 86, 99]. In a normative account, we show that simply incorporating the objective of optimizing discounted returns is sufficient to elicit human-like persistence in RNN actor-critic models whose latent spaces show clear gradations of goal progress and progress rates. We then investigated the latent-space dynamics of human goal commitment by cloning behavioral policies with RNNs. By analyzing velocity flow fields in latent space, we show the emergence of point-attractor-like dynamic modes whose “pointiness” is modulated by goal progress. This suggests a shifting of the energy landscape of neural activity patterns toward sustaining stable goal states in high-progress contexts. We quantified metrics for estimating point-attractor scores that could serve as estimates of the degree of goal commitment.

We introduced a circuit-level mechanistic account of goal commitment alongside a descriptive benchmark of RNN behavior cloning. We incorporated a recurrent

competition network motif with self-excitation and cross-inhibition that is modulated by a momentum-generating drive circuit. Momentum signals are computed from the interaction between current progress and a temporally smoothed estimate of recent progress change, and we show that it reasonably approximates time-to-goal-completion, an estimate of goal value for agents optimizing discounted future returns. At the implementational level, we envision a circuit where the goal-progress signal is compared to a time-delayed version to compute progress differences, which are passed through an asymmetric leaky integrator to estimate progress velocity. The resulting velocity signal is then multiplied with the current progress signal to drive progress-dependent escalation of commitment. We show that this recurrent circuit produces similar point-attractor-like dynamic modes in its policy landscape, with higher point-attractor scores in high-progress goal contexts. We show that this model reproduces participants' behavioral patterns better than the behavior-cloning RNN and simpler behavioral models from previous studies [3]. We further performed ablation experiments on this model to ensure all its components contribute to explaining human behavior. Nevertheless, the mechanistic account put forth in this study is fairly limited, lacking precise neural correlates of momentum computations. A likely candidate is the prefrontal cortex, which is known to use grid-like computations that track goal progress and progress rates in abstract goal navigation, just as they compute spatial progress and velocity during spatial navigation [58].

One of the key interpretations of this dynamical-systems perspective on goal commitment is the origin of switching costs [182, 147]. We claim that switching costs appear not as a consequence of switching goals, but because of the resistance that the intrinsic dynamics of the system offer to state updates. This resistance is significantly higher while switching goals, though not always. At times, switching goals align with the intrinsic dynamics, where staying with the goal is more costly. This accounts for why it is difficult to maintain goal commitment in the presence of distracting tasks, where switching is more amenable to the intrinsic dynamics of the system.

One of the major limitations of this study is that we cannot establish whether the attractor modes are manifestations of behavioral commitment or causally sustain commitment through the mechanistic process we describe. This remains promising future work [173]. Moreover, we studied extended goals in a simplistic setting, whereas naturalistic goal pursuit is far more complex and influenced by several factors other than goal progress and progress rates. Most goals are also hierarchical and

compositional, where the complexity involved in memory retrieval and management influences goal selection and maintenance. These limitations form the basis for fresh investigations in this area, which is ripe for new insights into goal-directed agency in the real world, a question pertinent to both natural and artificial intelligence.

Task Environment

We consider a three-goal sequential decision-making task. At each time step t , the agent selects an action $a_t \in \{0, 1, 2\}$ corresponding to pursuing one of three goals: *cat*, *hat*, *car*.

Each goal i maintains a progress state $p_t^{(i)} \in \{0, \dots, 7\}$, where 7 is the number of tokens required to complete the goal. Upon completion ($p_t^{(i)} = 7$), the agent receives a reward R_{goal} and the progress resets: $p_{t+1}^{(i)} = 0$.

At each step, a token is obtained with probability determined by the current block:

$$\Pr(\text{token}_t = 1 \mid a_t = i) = \pi_b^{\text{env}}(i),$$

where b indexes the current block. Blocks last for a fixed number of steps and define reward probability structures (e.g., 80-20, 60-40).

The reward function is:

$$r_t = \begin{cases} R_{\text{goal}} & \text{if goal completed,} \\ \beta_{\text{prog}} & \text{if token obtained (optional shaping),} \\ 0 & \text{otherwise,} \end{cases}$$

The observation o_t consists of:

$$o_t = \left[\underbrace{\tilde{p}_t}_{\text{progress}}, \underbrace{a_{t-1}^{\text{one-hot}}}_{\text{previous action}}, \underbrace{\text{token}_{t-1}}_{\text{previous outcome}}, \underbrace{g_{t-1}^{\text{one-hot}}}_{\text{goal completion}}, \underbrace{\text{block shift}}_{\text{context cue}} \right],$$

Recurrent Policy Architecture

The agent is parameterized by a recurrent actor-critic network. Observations are first embedded:

$$x_t = \text{ReLU}(W_{\text{in}} o_t + b_{\text{in}}),$$

and passed through a gated recurrent unit (GRU):

$$h_t = \text{GRU}(x_t, h_{t-1}).$$

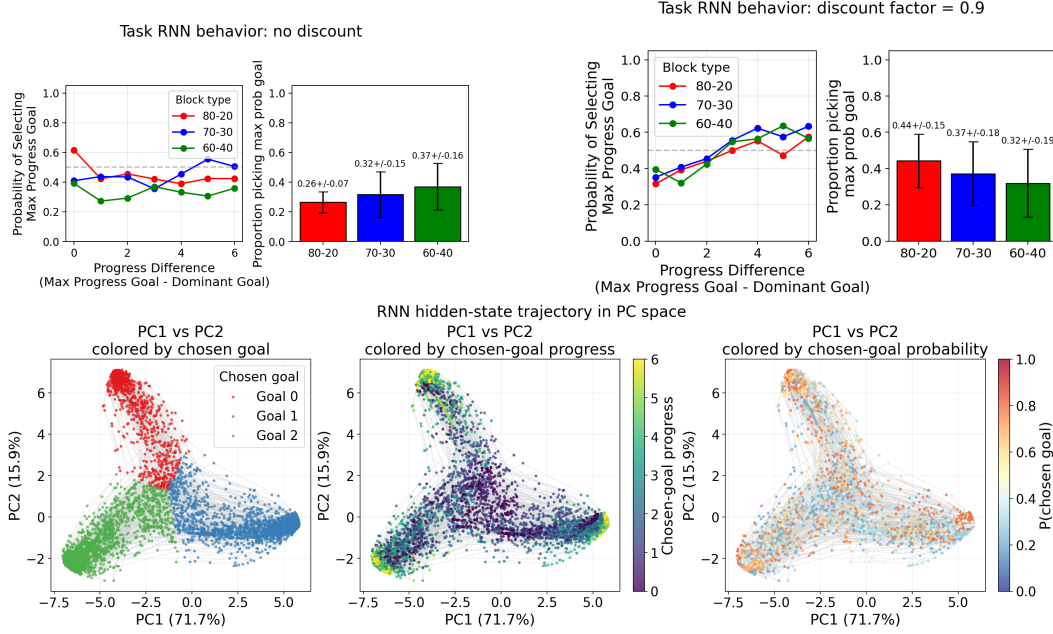


Figure 6.4: (top)(left) RNN actor-critic models trained on the task show no progress preference when rewards are not discounted. (right) RNN actor-critic models reproduce human block sensitivity and progress preferences when rewards are discounted. (bottom) Latent state of the RNN model (with discounted rewards) visualized in PC space. (panel 1) colored by chosen goal. (panel 2) colored by goal progress. (panel 3) colored by goal progress rate.

The policy (actor) and value (critic) are defined as:

$$\pi_{\theta}(a_t | h_t) = \text{softmax}(W_{\pi}h_t + b_{\pi}),$$

$$V_{\theta}(h_t) = W_Vh_t + b_V.$$

The recurrent hidden state h_t enables integration over past observations and actions, allowing the policy to represent latent goal states and environmental context.

Trajectory Collection: Training proceeds by sampling trajectories $\tau = (o_1, a_1, r_1, \dots, o_T)$ from the environment. Actions are sampled from the policy:

$$a_t \sim \pi_{\theta}(\cdot | h_t),$$

with additional ϵ -greedy exploration:

$$a_t = \begin{cases} \text{random action} & \text{with prob. } \epsilon, \\ \text{policy sample} & \text{otherwise.} \end{cases}$$

Return and Advantage Estimation: Discounted returns are computed as: $G_t = \sum_{k=0}^{T-t} \gamma^k r_{t+k}$

The critic is trained to predict normalized returns: $\hat{G}_t = \frac{G_t - \mu}{\sigma}$, where μ and σ are running estimates of return mean and variance.

Advantages are computed using the denormalized value estimate: $A_t = G_t - V_\theta(h_t)$

Advantages are normalized across the trajectory.

Optimization Objective: The actor loss is:

$$\mathcal{L}_{\text{actor}} = -\mathbb{E}_t [\log \pi_\theta(a_t | h_t) A_t] - \beta_{\text{ent}} \mathbb{E}_t [\mathcal{H}(\pi_\theta(\cdot | h_t))] - \beta_{\text{div}} \mathbb{E}_t [b_t \log \pi_\theta(a_t | h_t)],$$

where:

- $\mathcal{H}$ is policy entropy (encouraging exploration),
- $b_t = \frac{1}{1 + \text{count}(a_t)}$ is a diversity bonus promoting selection of under-sampled goals.

The critic loss is: $\mathcal{L}_{\text{critic}} = \mathbb{E}_t \left[\left(V_\theta(h_t) - \hat{G}_t \right)^2 \right]$

The total loss is: $\mathcal{L} = \mathcal{L}_{\text{actor}} + \lambda_V \mathcal{L}_{\text{critic}}$

Parameters are optimized using Adam with gradient clipping.

Training Procedure: Agents are trained for multiple episodes, each consisting of a sequence of blocks with changing reward structures. At each episode:

1. A trajectory is sampled from the environment.
2. Returns and advantages are computed.
3. Actor and critic losses are evaluated.
4. Parameters are updated via stochastic gradient descent.

Over training, the recurrent dynamics adapt to encode both goal persistence and sensitivity to changing environmental contingencies, giving rise to structured dynamical regimes in hidden-state space.

All analyses were performed on a high-performance computing (HPC) cluster utilizing parallelized workloads distributed across multiple CPU cores.

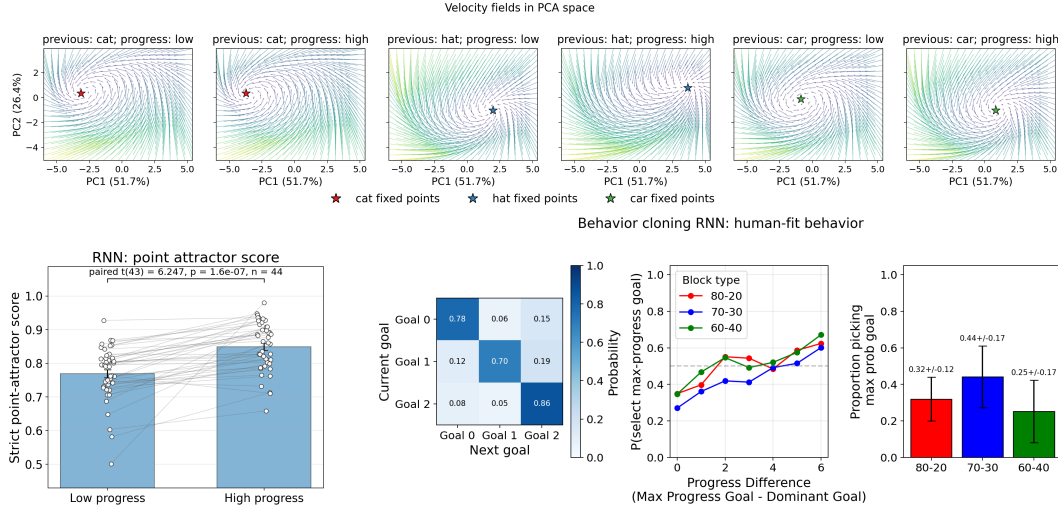


Figure 6.5: (top) Context-dependent intrinsic velocity fields of behavior-cloned RNN model of a single participant in PC space showing point-attractor-like structure. Each panel shows the velocity field for a specific previous goal and progress level. (bottom) (left) Significantly higher point-attractor scores in high-progress contexts compared to low-progress contexts showing attractor-structure evolution with environmental context. (right) Behavioral cloning through RNN models reproducing human persistence patterns and progress preferences.

Behavioral Cloning with Recurrent Neural Networks

Input Representation: For each subject, we construct a single sequence of trials $t = 1, \dots, T$. Each observation $o_t \in \mathbb{R}^{11}$ consists of:

- normalized goal progress $p_t \in \mathbb{R}^3$,
- previous action encoded as a one-hot vector $\text{onehot}(a_{t-1}) \in \{0, 1\}^3$,
- previous reward outcome $r_{t-1} \in \{0, 1\}$,
- previous goal completion indicator $g_{t-1} \in \{0, 1\}^3$,
- block boundary indicator $b_t \in \{0, 1\}$.

Thus,

$$o_t = [p_t, \text{onehot}(a_{t-1}), r_{t-1}, g_{t-1}, b_t]. \quad (6.21)$$

Actions are discrete choices $a_t \in \{0, 1, 2\}$.

Model Architecture: We use a recurrent neural network with hidden state $h_t \in \mathbb{R}^H$:

$$x_t = \text{ReLU}(W_{\text{in}}o_t + b_{\text{in}}), \quad (6.22)$$

$$h_t = \text{RNN}(x_t, h_{t-1}), \quad (6.23)$$

$$\ell_t = W_{\text{out}}h_t + b_{\text{out}}, \quad (6.24)$$

$$\pi_{\theta}(a_t \mid o_{\leq t}) = \text{softmax}(\ell_t). \quad (6.25)$$

Training Objective: We fit parameters θ via maximum likelihood estimation:

$$\mathcal{L}(\theta) = -\frac{1}{|\mathcal{I}_{\text{train}}|} \sum_{t \in \mathcal{I}_{\text{train}}} \log \pi_{\theta}(a_t \mid o_{\leq t}). \quad (6.26)$$

Optimization is performed using Adam with gradient clipping. Early stopping is applied based on validation negative log-likelihood.

Data Splitting: Trials are randomly partitioned into disjoint sets:

$$\mathcal{I}_{\text{train}}, \quad \mathcal{I}_{\text{val}}, \quad \mathcal{I}_{\text{test}}, \quad (6.27)$$

with proportions 75%, 12.5%, and 12.5%, respectively. Splitting is interspersed across time to preserve coverage of the full behavioral sequence.

Evaluation Metrics: We evaluate performance on held-out trials $\mathcal{I}$ using:

$$\text{NLL} = -\frac{1}{|\mathcal{I}|} \sum_{t \in \mathcal{I}} \log \pi_{\theta}(a_t \mid o_{\leq t}), \quad (6.28)$$

$$\text{Accuracy} = \frac{1}{|\mathcal{I}|} \sum_{t \in \mathcal{I}} \mathbb{I}[\arg \max_a \pi_{\theta}(a \mid o_{\leq t}) = a_t]. \quad (6.29)$$

To assess agreement in marginal choice structure, we compute:

$$\text{MSE}_{\text{freq}} = \frac{1}{3} \sum_{a=1}^3 \left(f_a^{\text{human}} - f_a^{\text{model}} \right)^2, \quad (6.30)$$

where f_a denotes empirical choice frequencies.

Hidden State Extraction: After training, we compute the hidden trajectory $\{h_t\}_{t=1}^T$ under the fitted model. These latent states are used for subsequent dynamical systems analyses (e.g., vector fields and attractor structure).

Per-Subject Fitting: All models are fit independently per subject, yielding parameters $\theta^{(s)}$. Group-level statistics are obtained by aggregating performance metrics across subjects.

Intrinsic Velocity Field Analysis

Hidden State Extraction and Dimensionality Reduction: For each subject, we compute the hidden state trajectory $\{h_t\}_{t=1}^T$ from the fitted RNN. We perform principal component analysis (PCA) on $\{h_t\}$ and obtain a low-dimensional embedding:

$$z_t = Ph_t, \quad (6.31)$$

where $P \in \mathbb{R}^{d \times H}$ contains the top 2 principal components.

Context Definition: We define discrete behavioral contexts based on:

- previous goal identity $g \in \{0, 1, 2\}$,
- goal progress level (low vs. high).

Formally, a context $c = (g, \text{bin})$ selects timepoints satisfying:

$$a_{t-1} = g, \quad (6.32)$$

$$p_{t,g} \leq \frac{2}{7} \quad (\text{low}), \quad (6.33)$$

$$p_{t,g} \geq \frac{5}{7} \quad (\text{high}). \quad (6.34)$$

For each context, we compute a prototype observation:

$$\bar{o}^{(c)} = \frac{1}{|\mathcal{T}_c|} \sum_{t \in \mathcal{T}_c} o_t. \quad (6.35)$$

Intrinsic Velocity Field: We define the intrinsic dynamics by evaluating the RNN transition from arbitrary latent states under a fixed context:

$$h' = f_\theta(\bar{o}^{(c)}, h), \quad (6.36)$$

$$v(h; c) = h' - h. \quad (6.37)$$

We construct a grid $\{z_i\}$ in PCA space and map each point back to latent space:

$$h_i = P^\top z_i + \mu, \quad (6.38)$$

where μ is the PCA mean.

The velocity field is then computed as:

$$v_i^{(c)} = f_\theta(\bar{o}^{(c)}, h_i) - h_i, \quad (6.39)$$

and projected back to PCA coordinates:

$$\tilde{v}_i^{(c)} = P v_i^{(c)}. \quad (6.40)$$

The projected field $\tilde{v}_i^{(c)}$ is visualized as a quiver plot over the PCA grid: $(z_i, \tilde{v}_i^{(c)})$. Velocity magnitude $\|\tilde{v}_i^{(c)}\|$ is used for color coding.

Fixed Point Estimation: We estimate candidate fixed points as local minima of the velocity magnitude:

$$\|v(h; c)\| \approx 0. \quad (6.41)$$

Specifically:

1. Compute speeds $s_i = \|\tilde{v}_i^{(c)}\|$.
2. Select grid points below a quantile threshold.
3. Identify local minima in the grid neighborhood.
4. Cluster nearby minima using a distance threshold.

Cluster centers define candidate fixed points:

$$h^* \approx \arg \min_h \|v(h; c)\|. \quad (6.42)$$

Strict Point-Attractor Metric and Synthetic Validation

Definition of the Strict Point-Attractor Metric

To quantify the presence of a true point attractor in a vector field, we define a strict metric that combines multiple necessary dynamical properties. The goal is to assign a high score only when a candidate point exhibits inward flow, basin support, and full local stability.

Candidate selection. Candidate points are identified as locations where the vector field magnitude is low. Specifically, we select grid points whose speed lies below a chosen quantile threshold, with an additional constraint that selected candidates are spatially separated.

Inward convergence. For each candidate point x_0 , we measure whether nearby vectors point toward it. For neighboring points x , let $v(x)$ denote the vector field. The inward convergence score is defined as:

$$S_{\text{inward}} = \mathbb{E} \left[\frac{(x_0 - x)^\top v(x)}{|x_0 - x|, |v(x)|} \right], \quad (6.43)$$

where the expectation is taken over a local neighborhood of x_0 .

This quantity reflects whether the local flow is directed toward the candidate point.

Basin support. To assess whether the candidate has a meaningful basin of attraction, we evaluate short trajectories initialized in a surrounding region. For each initial point x , we compute the change in distance to the candidate after a fixed number of steps:

$$\Delta(x) = |x - x_0| - |x_T - x_0|, \quad (6.44)$$

where x_T is obtained by iterating the dynamics.

We define the basin score as a combination of:

- the fraction of points with $\Delta(x) > 0$, and
- the fraction of points that enter a local neighborhood of x_0 .

This captures whether trajectories converge toward the candidate beyond its immediate vicinity.

Point-specific stability. We assess local stability using the Jacobian of the vector field:

$$J(x_0) = \frac{\partial F}{\partial x} \Big|_{x_0} \quad (6.45)$$

Let λ_i denote the real parts of the eigenvalues of $J(x_0)$. We define a point-specific stability score as:

$$S * \text{stability} = \prod_i \frac{1}{1 + \exp(\kappa(\lambda_i + \delta))}, \quad (6.46)$$

where $\kappa > 0$ controls sharpness and $\delta > 0$ enforces a margin.

This formulation penalizes both unstable directions ($\lambda_i > 0$) and neutral directions ($\lambda_i \approx 0$), ensuring that line attractors do not receive high scores.

Final metric. The strict point-attractor score is defined as:

$$S_{\text{point}} = (S_{\text{inward}})^\alpha (S_{\text{basin}})^\beta (S_{\text{stability}})^\gamma, \quad (6.47)$$

where $\alpha = 0.35, \beta = 0.16, \gamma = 0.48$ are weighting exponents chosen to maximize the score's ability to distinguish point attractors from other dynamical regimes.

This multiplicative form ensures that all components must be simultaneously present for a high score.

Synthetic Validation

To validate the metric, we constructed a set of synthetic two-dimensional dynamical systems with known ground-truth structure. Each system defines a vector field:

$$\dot{x} = F(x), \quad x \in \mathbb{R}^2. \quad (6.48)$$

We generated multiple classes of systems:

- **Point attractors:** Stable fixed points were generated using perturbed linear contraction dynamics:

$$\dot{x} = A(x - c) + \left(1 - \exp\left(-\frac{|x - c|^2}{2\sigma^2}\right)\right) \epsilon(x). \quad (6.49)$$

Here, c is the attractor location, A has negative eigenvalues, and $\epsilon(x)$ is a smooth perturbation that vanishes near $x = c$.

- **Limit cycles:** Limit cycles were generated using radial stabilization and tangential rotation:

$$\dot{x} = M \left[\omega \begin{pmatrix} -y_2 \\ y_1 \end{pmatrix} + k(r_0 - |y|)y \right] + \epsilon(x), \quad (6.50)$$

$$y = M^{-1}(x - c). \quad (6.51)$$

Here, r_0 is the target radius, k controls radial attraction, and ω sets angular velocity.

- **Line attractors:** Line attractors were generated by separating dynamics along tangent and normal directions:

$$\dot{x} = (-k_n v)n + (-k_t u + \delta)t + \eta(v)\epsilon(x). \quad (6.52)$$

The scalar coordinates are:

$$u = (x - c)'t, \quad (6.53)$$

$$v = (x - c)'n. \quad (6.54)$$

Here, $k_n > 0$ enforces contraction toward the line, while k_t is close to zero, giving approximately neutral dynamics along the line.

- **Repellers:** Repellers were generated as unstable linear systems:

$$\dot{x} = A(x - c) + \left(1 - \exp\left(-\frac{|x - c|^2}{2\sigma^2}\right)\right) \epsilon(x). \quad (6.55)$$

Here, A has positive eigenvalues, producing outward flow in all directions.

- **Saddles:** Saddles were generated using one unstable and one stable direction:

$$\dot{x} = R \begin{pmatrix} \lambda_1 & 0 \\ 0 & -\lambda_2 \end{pmatrix} R'(x - c) + \left(1 - \exp\left(-\frac{|x - c|^2}{2\sigma^2}\right)\right) \epsilon(x). \quad (6.56)$$

Here, $\lambda_1 > 0$ and $\lambda_2 > 0$.

- **Random flows:** Random flows were generated as smooth Fourier vector fields:

$$\dot{x} = b + Ax + \sum_{k=1}^K \left[a_k \sin(w'_k x + \phi_k) + d_k \cos(w'_k x + \phi_k) \right]. \quad (6.57)$$

The coefficients, frequencies, and phases were sampled randomly, producing smooth dynamics without a planted attractor.

Across all systems, parameters such as location, strength, anisotropy, and rotation were randomized. In addition, smooth spatial perturbations were added to introduce variability while preserving the qualitative structure of each dynamical class.

Evaluation procedure. For each synthetic field:

1. Candidate points were identified using low-speed regions.
2. Each candidate was scored using the strict metric.
3. The highest-scoring candidate was selected.

This procedure was repeated across multiple randomized replicates for each class.

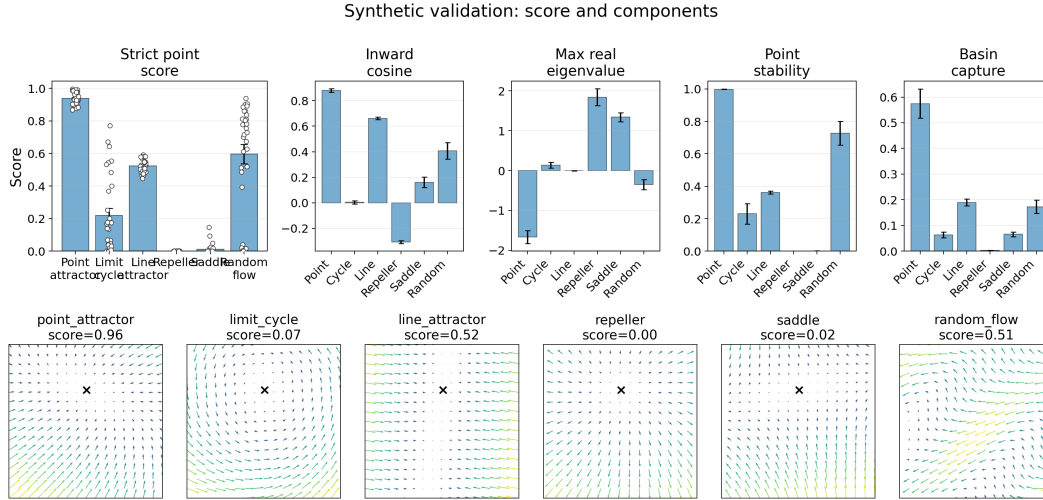


Figure 6.6: **Synthetic validation of the strict point-attractor score:** (top) Summary of point attractor scores by dynamical system class. (bottom) Example velocity fields and their corresponding point attractor scores.

Validation criterion. A valid metric should assign:

- high scores to point attractors,
- low scores to limit cycles and random flows,
- reduced scores to line attractors due to neutral directions,
- low scores to repellers and saddles due to instability.

The strict metric reliably distinguished point attractors from all other dynamical regimes. In particular, the point-specific stability component was critical in separating point attractors from line attractors, which otherwise exhibit local convergence but lack a unique stable equilibrium.

These results demonstrate that the metric captures the defining properties of point attractors and is not driven by artifacts such as slow dynamics or tangential flow (see appendix Figure 6).

Momentum-Circuit Model

State and dynamics: The model maintains a latent goal state:

$$\mathbf{x}_t \in \Delta^2 = \left\{ \mathbf{x} \in \mathbb{R}^3 : x_g \geq 0, \sum_{g=1}^3 x_g = 1 \right\}. \quad (6.58)$$

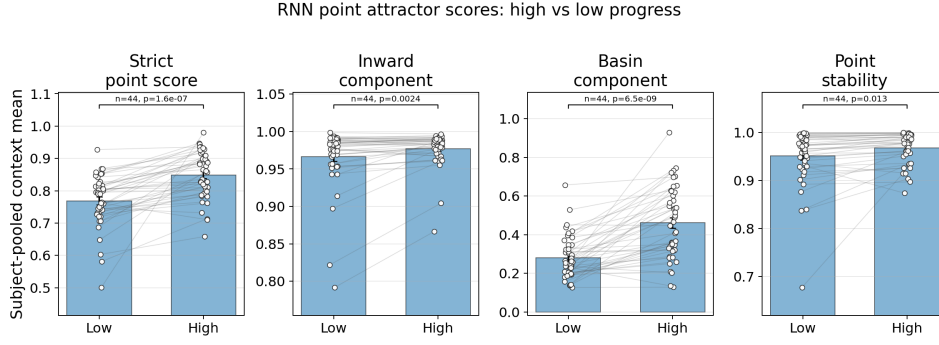


Figure 6.7: Significantly higher point-attractor scores in high-progress contexts compared to low-progress contexts showing attractor-structure evolution with environmental context in Behavior-cloned RNN model.

Recurrent circuit: The recurrent interaction matrix $W \in \mathbb{R}^{3 \times 3}$ is defined as:

$$W_{gh} = \begin{cases} w_{\text{self}}, & g = h, \\ -w_{\text{cross}}, & g \neq h. \end{cases} \quad (6.59)$$

$$(W\mathbf{x}_t)_g = w_{\text{self}}x_{g,t} - w_{\text{cross}} \sum_{h \neq g} x_{h,t}. \quad (6.60)$$

Asymmetric momentum computations: Let $\mathbf{p}_t \in [0, 1]^3$ denote normalized progress.

First define instantaneous progress change:

$$\Delta \mathbf{p}_t = \mathbf{p}_t - \mathbf{p}_{t-1}. \quad (6.61)$$

We separate positive and negative progress changes:

$$\Delta \mathbf{p}_t^+ = \max(\Delta \mathbf{p}_t, 0), \quad \Delta \mathbf{p}_t^- = \max(-\Delta \mathbf{p}_t, 0), \quad (6.62)$$

where the max is taken elementwise.

Velocity is computed via a leaky integration with asymmetric gains:

$$\mathbf{v}_t = \gamma_v \mathbf{v}_{t-1} + \eta_+ \Delta \mathbf{p}_t^+ - \eta_- \Delta \mathbf{p}_t^-. \quad (6.63)$$

This allows positive progress and negative (or stalled) progress to have distinct impacts on the velocity signal.

Drive and policy: The total drive now includes additional inputs:

$$\mathbf{D}_t = W\mathbf{x}_t + \beta_{\text{ctx}}\mathbf{c}_t + \beta_{\text{vel}}\mathbf{v}_t + \beta_{\text{mom}}p_t \odot \mathbf{v}_t + \beta_{\text{prog}}\mathbf{p}_t + \beta_{\text{prev}}\mathbf{a}_{t-1} + \mathbf{b}, \quad (6.64)$$

where:

- $\mathbf{p}_t$: direct progress input
- $\mathbf{a}_{t-1}$: one-hot vector of the previous action

Choice probabilities follow:

$$\boldsymbol{\pi}_t = \text{softmax} \left(\frac{\mathbf{D}_t}{\tau} \right). \quad (6.65)$$

State update with choice feedback: The latent state evolves as:

$$\tilde{\mathbf{x}}_{t+1} = (1 - \alpha_x) \mathbf{x}_t + \alpha_x \boldsymbol{\pi}_t + \kappa \mathbf{a}_t, \quad (6.66)$$

where $\mathbf{a}_t$ is the chosen action (one-hot) and κ is the choice feedback gain.

The state is then renormalized onto the simplex:

$$\mathbf{x}_{t+1} = \frac{\tilde{\mathbf{x}}_{t+1}}{\sum_g \tilde{x}_{g,t+1}}. \quad (6.67)$$

The model decomposes goal commitment into:

- **Recurrent dynamics:** $W\mathbf{x}_t$ defines attractor structure.
- **Asymmetric velocity:** $\mathbf{v}_t$ distinguishes gains from losses in progress.
- **Momentum:** $\mathbf{p}_t \odot \mathbf{v}_t$ amplifies sustained success.
- **Direct progress drive:** $\mathbf{p}_t$ biases toward advanced goals.
- **Action persistence:** $\mathbf{a}_{t-1}$ and $\mathbf{a}_t$ reinforce recent choices.

This yields a dynamical system where persistence arises from both attractor structure and history-dependent progress signals, while switching reflects competition between accumulated evidence streams.

Momentum as an Approximation of Time-to-Goal Completion

Let $x(t) \in [0, 1]$ denote fractional progress toward a goal, with $x = 1$ corresponding to goal completion, and let $\dot{x}(t)$ denote the rate of progress. We define a momentum-like quantity:

$$m(t) = x(t) \dot{x}(t). \quad (6.68)$$

The exact remaining time to goal completion is given by:

$$T_{\text{goal}} = \int_{x(t)}^1 \frac{dx}{\dot{x}(x)}. \quad (6.69)$$

Under a local constant-velocity approximation, where $\dot{x}$ is assumed to vary slowly over the remaining trajectory, this simplifies to:

$$T_{\text{goal}} \approx \frac{1-x}{\dot{x}}. \quad (6.70)$$

Substituting $\dot{x} = \frac{m}{x}$ yields:

$$T_{\text{goal}} \approx \frac{x(1-x)}{m}. \quad (6.71)$$

Thus, momentum is inversely related to time-to-goal completion, up to a state-dependent scaling factor:

$$m \approx \frac{x(1-x)}{T_{\text{goal}}}. \quad (6.72)$$

This relationship shows that m is not a direct estimator of time-to-goal, but rather a *progress-weighted inverse time-to-completion*. The multiplicative factor $x(1-x)$ introduces a characteristic dependence on the current state: it is maximal at intermediate progress ($x \approx 0.5$) and vanishes near the start and end of the trajectory. This could potentially explain human motivation patterns pertaining to goal pursuit, where they lose motivation at the last stretch of goal-directed activity, wherein the drive to finish the last 1% is substantially lower than the drive mid-progress.

Consequently, m provides a locally accurate proxy for inverse time-to-goal in mid-progress regimes, while systematically underestimating it near boundary states. Despite this limitation, the quantity $m = x\dot{x}$ captures a behaviorally meaningful signal, reflecting the interaction between current commitment (progress) and ongoing motion (rate of progress), and thus serves as a compact summary of goal-directed dynamics.

Behavior Recovery Analysis

To complement parameter recovery, we evaluated whether the model can reproduce *behavioral statistics* of synthetic agents generated from known ground-truth parameters. This provides a functionally relevant validation, as different parameter configurations can produce similar observable behavior.

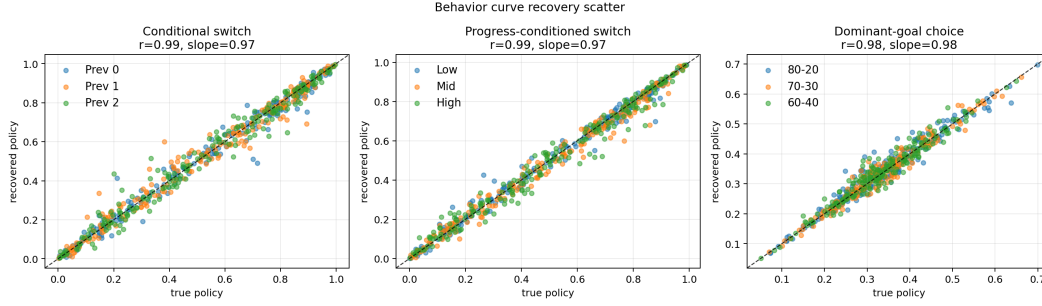


Figure 6.8: Behavioral recovery of structured decision statistics across simulations. Each panel compares ground-truth (x-axis) and recovered (y-axis) values computed from the fitted model’s stochastic policy. Left: conditional switching probabilities $P(\text{switch} \mid a_{t-1})$ for each previous action. Middle: progress-conditioned switching $P(\text{switch} \mid p_t^{(a_{t-1})})$ across low, medium, and high progress bins. Right: dominant-goal choice probability $P(a_t = g_t^* \mid \text{block})$ across block conditions (80-20, 70-30, 60-40). Each point corresponds to one simulation; colors denote conditioning levels. Dashed lines indicate identity.

Synthetic Data Generation: For each recovery simulation, we sampled ground-truth parameters from predefined ranges and instantiated the momentum-circuit model. Using the empirical task structure (progress and context sequences from a human subject), we generated synthetic choice sequences by rolling out the model policy:

$$\pi_t = \text{softmax} \left(\frac{\mathbf{D}_t}{\tau} \right), \quad (6.73)$$

where the drive is given by:

$$\mathbf{D}_t = W\mathbf{x}_t + \beta_{\text{ctx}}\mathbf{c}_t + \beta_{\text{vel}}\mathbf{v}_t + \beta_{\text{mom}}\mathbf{m}_t + \beta_{\text{prog}}\mathbf{p}_t + \beta_{\text{prev}}\mathbf{a}_{t-1} + \mathbf{b}. \quad (6.74)$$

Actions were sampled stochastically from this policy to preserve realistic variability. To ensure identifiability, we rejected degenerate simulations with low entropy or highly imbalanced action frequencies.

Behavioral Metrics: We summarized behavior using task-relevant conditional statistics:

Dominant-goal choice probability (by block) For each block condition $b \in \{80-20, 70-30, 60-40\}$, we compute the probability of selecting the dominant goal:

$$P(a_t = g_t^* \mid b_t = b) = \frac{1}{|T_b|} \sum_{t \in T_b} \mathbf{1}(a_t = g_t^*), \quad (6.75)$$

where g_t^* denotes the dominant (highest-reward-probability) goal on trial t , and T_b is the set of trials belonging to block b .

Conditional switching by previous action We quantify action-dependent persistence by computing:

$$P(\text{switch} \mid a_{t-1} = i) = \mathbb{E}[\mathbf{1}(a_t \neq a_{t-1}) \mid a_{t-1} = i], \quad (6.76)$$

for each action $i \in \{1, 2, 3\}$.

Progress-conditioned switching We measure how switching depends on goal completion progress:

$$P(\text{switch} \mid p_t^{(a_{t-1})} \in \mathcal{B}_k), \quad (6.77)$$

where $p_t^{(a_{t-1})}$ denotes the progress of the previously selected goal, and $\mathcal{B}_k$ are discrete progress bins (low, medium, high).

Recovery Procedure: For each synthetic dataset, we refit the model using the same optimization procedure as in the main analysis. After fitting, we computed the recovered behavioral statistics using:

- **Stochastic policy estimate:** behavioral metrics computed directly from the model’s predicted action probabilities π_t .
- **Sampled rollout:** behavioral metrics computed from new sequences generated by sampling from the recovered policy.

Behavior Recovery Evaluation: For each synthetic dataset, we fit the model to the generated choice sequence using the same optimization procedure as in the main analysis. We then evaluated whether the recovered model reproduced the *behavioral structure* of the synthetic agent by comparing behavioral statistics computed from the recovered stochastic policy to those computed from the ground-truth generating policy. Specifically, recovered behavioral metrics were computed directly from the model’s predicted action probabilities π_t , conditioned on the same task covariates and synthetic action history used during generation, rather than from deterministic rollouts.

Behavior recovery focused on three task-relevant statistics: (1) conditional switching probabilities given the previous action, (2) progress-conditioned switching probabilities across low, medium, and high progress states, and (3) dominant-goal choice probabilities across block conditions. Recovery quality was quantified across simulations using Pearson correlation, linear regression slope, and root mean squared

error (RMSE) between ground-truth and recovered behavioral statistics. Successful recovery therefore indicates that the fitted model reproduces the functional mapping from task state variables (progress, context, and action history) to action-selection behavior, rather than merely matching individual parameter values.

Simplex Landscape Analysis

Simplex Representation: The goal state lies on the simplex:

$$\Delta^2 = \{x \in \mathbb{R}^3 : x_i \geq 0, \sum_i x_i = 1\}. \quad (6.78)$$

We embed x into 2D coordinates for visualization using barycentric projection.

Context Anchoring: For each condition $c = (\text{previous goal, progress bin})$, we construct anchor variables by averaging over trials matching the condition:

$$v^{(c)} = \mathbb{E}[v_t | c], \quad m^{(c)} = \mathbb{E}[m_t | c], \quad \tilde{c}^{(c)} = \mathbb{E}[c_t | c], \quad p^{(c)} = \mathbb{E}[p_t | c], \quad a^{(c)} = \mathbb{E}[a_{t-1} | c]. \quad (6.79)$$

Here, v_t and m_t denote velocity and momentum, c_t is the inferred context, p_t is the progress vector across goals, and a_{t-1} is the previous action (one-hot encoded).

These anchors define a fixed input condition under which the dynamical system is evaluated.

Vector Field Definition: The drive is given by:

$$d(x; c) = w_{\text{self}} x - w_{\text{cross}} (\mathbf{1}^\top x - x) + \beta_{\text{ctx}} \tilde{c}^{(c)} + \beta_{\text{vel}} v^{(c)} + \beta_{\text{mom}} m^{(c)} + \beta_{\text{prog}} p^{(c)} + \beta_{\text{prev}} a^{(c)} + b. \quad (6.80)$$

The policy is:

$$\pi(x; c) = \text{softmax}\left(\frac{d(x; c)}{\tau}\right). \quad (6.81)$$

The induced vector field is:

$$\dot{x} = \alpha_x (\pi(x; c) - x). \quad (6.82)$$

To ensure the dynamics remain within the simplex, we project onto the tangent space:

$$\dot{x} \leftarrow \dot{x} - \frac{1}{3} (\mathbf{1}^\top \dot{x}) \mathbf{1}. \quad (6.83)$$

Numerical Integration: We simulate trajectories using Euler updates:

$$x_{t+1} = x_t + \Delta t \dot{x}_t, \quad (6.84)$$

$$x_{t+1} = \frac{\max(x_{t+1}, \epsilon)}{\mathbf{1}^\top \max(x_{t+1}, \epsilon)}. \quad (6.85)$$

Grid Construction: We discretize the simplex using a triangular grid:

$$x_i = \left(\frac{i}{N}, \frac{j}{N}, \frac{k}{N} \right), \quad i + j + k = N. \quad (6.86)$$

The vector field is evaluated at each grid point. We visualize the projected vector field $\dot{x}$ (quiver plot), the scalar field $\log \|\dot{x}\|$, and trajectories from random initial conditions.

Attractor Detection: We simulate trajectories $\{x_t^{(i)}\}$ from random initial conditions and extract endpoints:

$$x_{\text{end}}^{(i)} = x_T^{(i)}. \quad (6.87)$$

Endpoints are clustered using a distance threshold ϵ :

$$\{x_k^*\} = \text{Cluster}(\{x_{\text{end}}^{(i)}\}). \quad (6.88)$$

The resulting landscape characterizes the number and location of attractors, basin geometry, and convergence strength via $\|\dot{x}\|$. Differences across conditions reveal how velocity, momentum, progress, contextual signals, and action history jointly shape the stability and geometry of goal representations.

Strict Point-Attractor Analysis for the Momentum-Circuit Model

We apply the same strict point-attractor metric described in the appendix section D to the momentum-circuit model. Here we describe how the underlying dynamical system differs and how the metric is instantiated in this setting.

Dynamical System on the Simplex: Unlike the RNN, whose dynamics evolve in a high-dimensional latent space, the momentum-circuit model defines an explicit dynamical system over the goal simplex:

$$x \in \Delta^2 = \{x \in \mathbb{R}^3 : x_i \geq 0, \sum_i x_i = 1\}. \quad (6.89)$$

For a fixed behavioral context c , the vector field is given analytically by:

$$\dot{x} = \pi(x; c) - x, \quad (6.90)$$

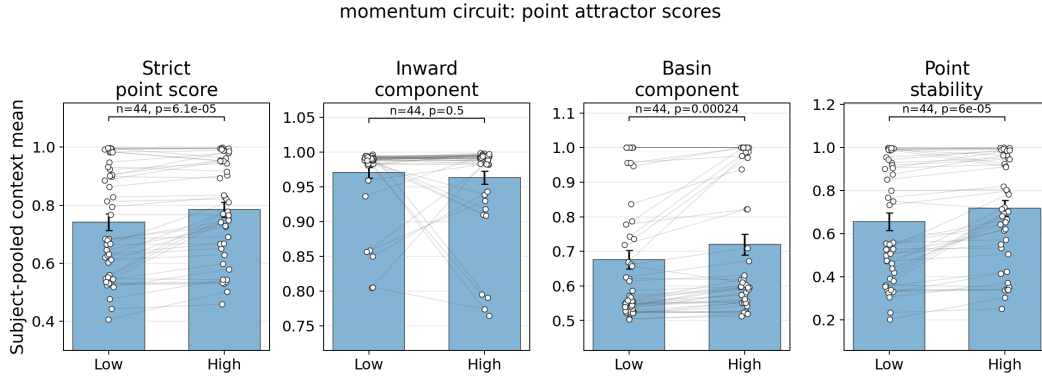


Figure 6.9: Significantly higher point-attractor scores in high-progress contexts compared to low-progress contexts showing attractor-structure evolution with environmental context for the momentum-circuit model.

where $\pi(x; c)$ is the softmax policy induced by the model drive.

Thus, the momentum-circuit model provides a closed-form vector field, rather than requiring rollout-based estimation as in the RNN.

Context Conditioning: Contexts are defined identically to the RNN analysis, using:

- previous goal identity,
- goal progress (low vs. high).

For each context c , we compute anchor variables by averaging over trials matching the condition:

$$\bar{u}^{(c)} = \left(v^{(c)}, m^{(c)}, \tilde{c}^{(c)}, p^{(c)}, a_{\text{prev}}^{(c)} \right), \quad (6.91)$$

where

$$v^{(c)} = \mathbb{E}[v_t | c] \quad (\text{progress velocity}), \quad (6.92)$$

$$m^{(c)} = \mathbb{E}[m_t | c] \quad (\text{momentum}), \quad (6.93)$$

$$\tilde{c}^{(c)} = \mathbb{E}[\text{context}_t | c] \quad (\text{context estimate}), \quad (6.94)$$

$$p^{(c)} = \mathbb{E}[p_t | c] \quad (\text{goal progress}), \quad (6.95)$$

$$a_{\text{prev}}^{(c)} = \mathbb{E}[a_{t-1} | c] \quad (\text{previous-action state}). \quad (6.96)$$

These anchors define a fixed input condition under which the vector field is evaluated:

$$\dot{x} = f(x; \bar{u}^{(c)}). \quad (6.97)$$

Candidate Fixed Points: As in the RNN analysis, candidate fixed points are identified as low-speed regions of the vector field evaluated on a simplex grid:

$$s(x) = \|\dot{x}\|. \quad (6.98)$$

Points below a quantile threshold are selected and filtered to enforce spatial separation.

Metric Evaluation: All strict point-attractor components (slowness, inward convergence, radial contraction, local stability, and basin support) are computed exactly as in the RNN case, but now directly on the simplex vector field.

A key difference is that:

- distances and directions are computed in simplex coordinates,
- the Jacobian is evaluated in the simplex tangent plane,
- rollouts are performed using explicit Euler updates with projection onto the simplex.

Jacobian in the Tangent Plane: To respect the simplex constraint $\sum_i x_i = 1$, we compute the Jacobian in the tangent subspace:

$$T = \{v \in \mathbb{R}^3 : \mathbf{1}^\top v = 0\}. \quad (6.99)$$

The Jacobian J is estimated numerically via finite differences along a basis of T , and stability is determined from the maximum real eigenvalue.

Basin Analysis: Basin support is evaluated by short rollouts:

$$x_{t+1} = \Pi_\Delta(x_t + \eta \dot{x}_t), \quad (6.100)$$

where Π_Δ projects onto the simplex.

This yields estimates of:

- radial contraction over multiple steps,
- basin capture probability.

Aggregation and Comparison: Metrics are computed across all contexts and aggregated within subjects:

$$\bar{s}_{\text{low}}, \quad \bar{s}_{\text{high}}. \quad (6.101)$$

We compare low and high progress using paired tests across subjects, identical to the RNN analysis.

The momentum-circuit model thus provides a mechanistic instantiation of attractor dynamics, allowing direct interpretation of:

- how recurrent competition shapes attractor geometry,
- how momentum (β_{mom}) stabilizes goal states,
- how progress (β_{prog}) modulates attractor strength.

Because the vector field is analytically defined, this analysis provides a direct link between model parameters and the emergence of point-attractor structure.

Emergence of switching costs

We quantify switching cost as movement against the intrinsic velocity field of the learned dynamical system.

Velocity field. Let $x_t \in \Delta^2 \subset \mathbb{R}^3$ denote the goal state on the simplex. The model defines a velocity field:

$$v(x_t) = \frac{dx}{dt} = \pi(x_t) - x_t, \quad (6.102)$$

where $\pi(x_t) \in \Delta^2$ is the softmax policy induced by the circuit.

Observed state displacement. Given the empirical trajectory $x_t \rightarrow x_{t+1}$, we define:

$$\Delta x_t = x_{t+1} - x_t. \quad (6.103)$$

Alignment with the flow field. We measure alignment between observed movement and intrinsic dynamics:

$$\langle \Delta x_t, v(x_t) \rangle. \quad (6.104)$$

Opposing force. We define the instantaneous opposing force as:

$$F_{\text{opp}}(t) = \max(0, -\Delta x_t^\top v(x_t)). \quad (6.105)$$

Accumulated opposing work. Total opposing work along a trajectory is:

$$W_{\text{opp}} = \sum_{t=1}^{T-1} F_{\text{opp}}(t). \quad (6.106)$$

This corresponds to the continuous form:

$$W_{\text{opp}} = \int \max(0, -\dot{x}(t)^\top v(x(t))) dt. \quad (6.107)$$

Stay vs. switch decomposition. We partition timesteps into:

$$\mathcal{T}_{\text{stay}} = \{t : a_{t+1} = a_t\}, \quad (6.108)$$

$$\mathcal{T}_{\text{switch}} = \{t : a_{t+1} \neq a_t\}. \quad (6.109)$$

Mean opposing work is then:

$$\bar{W}_{\text{stay}} = \frac{1}{|\mathcal{T}_{\text{stay}}|} \sum_{t \in \mathcal{T}_{\text{stay}}} F_{\text{opp}}(t), \quad (6.110)$$

$$\bar{W}_{\text{switch}} = \frac{1}{|\mathcal{T}_{\text{switch}}|} \sum_{t \in \mathcal{T}_{\text{switch}}} F_{\text{opp}}(t). \quad (6.111)$$

Switching cost emerges if:

$$\bar{W}_{\text{switch}} > \bar{W}_{\text{stay}}, \quad (6.112)$$

indicating that switching requires greater work against the intrinsic flow field of the dynamical system.

Chapter 7

CONCLUSION

There are a lot of debates ongoing in the community as I conclude my thesis. How can human intelligence reconcile with itself when faced with other forms of intelligence vastly outpacing it? Confronted with the rise of artificial intelligence, I have a different reaction. I'm convinced more than ever before of the human singularity. How long this conviction will hold, I cannot say. Perhaps I'll give it up when I have my own personal C3PO and R2D2. Until then, I retain my right to human hubris.

Humans might be subpar at any one task, but their superpower lies in handling diverse tasks and successfully managing them over their lifetime by setting goals with agency. They are autotelic, having a purpose unto themselves, and are ever evolving cognitively and culturally. Current AI can perhaps manage the vast corpus of human knowledge that exists in the world, but humans still retain the credit for creating that knowledge out of nothing. The invention of the printed word and the rise of civilizations all arose from the singular feature of goal setting in order to manipulate and control the world to efficiently procure rewards. Goal-directed control is central to this human evolution and deserves extensive attention.

In my thesis, I focused on a narrow framing of goal-directed control. In keeping with the neuroscience tradition of reductionist approaches, I proceeded to study a complex integrative construct in its reductionist form. The games that I built for my thesis require extended and persistent activity for resolution. The one complexity that I afforded the paradigm is the hierarchy in goal pursuit that we observe in the real world. This vastly reduced form seems childishly simple compared to the goals we set in daily life and the factors that influence their selection and maintenance. Baby steps, but that is how we start.

In my vastly simplified toy setting, I found that humans have a tendency to over-persist compared to artificially designed agents. However, in the light of diversity considerations, this suboptimality might have its benefits. The momentum computational framework that I introduced in this thesis can account for this suboptimality, as it tries to estimate goal prospects by trading off efficiency with parsimony. I put forth other considerations as to why we tolerate this so-called suboptimality. Goal

persistence can help reduce delays and enable us to build skills that can increase efficiency in the long run by compromising short-term efficiency. Again, the brain is considering the big picture here.

Looking inside the brain, I find correlates of momentum computations distinctly organized across the hierarchy. I found that goal and subgoal pursuit engage distinct subsystems that maintain, monitor, and adapt goals in a dissociable manner. While we only demonstrated this dissociation at a coarser resolution through fMRI, the nature of this dissociation needs to be investigated further at a finer implementational level. I made a small attempt toward that understanding by building a circuit-level model of goal commitment.

I venture that the prefrontal cortex maintains persistence states through attractor dynamics. Its attractor states are triggered by inputs from the parietal cortex, which can interface with memory systems for maintaining and executing complex long-term goals. This is in line with the prefrontal-cortex-is-the-neural-workspace theory, where representations needed for handling the workspace are buffered through interaction with the parietal areas. These theories are speculative at this point and can form the basis for my future investigations.

As I focused heavily on higher cognition, it is difficult to gain mechanistic understanding due to the lack of animal homologues. Some of the regions that I identified as critical to long-term goal pursuit in humans—say, the frontal pole and the precuneus—are vastly more evolved in humans compared to primates, making it unsatisfactory to use animal models. However, I would wager that some of the principles, such as momentum computations, can translate across species and are worth investigating in animals. As animals pursue varied goals during their lifetimes and are known to show similar persistence patterns, studying the phenomenon of goal commitment in animals with appropriately designed tasks can be fruitful, given the amenability of their neurons to perturbations. The rare opportunity afforded in humans through single-unit electrophysiology is not to be missed either.

As for my future steps, I look toward unpacking the complexity involved in goal-directed agency further. There are several questions that remain unanswered regarding extended goal pursuit. The immediate question that I started to answer through the course of my dissertation is how the brain abstracts context from the environment to support goal pursuit. Further questions that emerge from this are: how is context represented and maintained over several timescales of cognition? Multi-scale memory management for goal pursuit is the next big question that I plan to

tackle. Another question that materialized as I worked through my thesis is how the brain abstracts away goal-irrelevant information and efficiently codes goal-relevant features to facilitate skilled execution of goals. Another question currently on my docket is how humans discover subgoals for efficient goal pursuit when learning new skills—specifically, how they chunk motor sequences as they become proficient at typing when the goal is to type quickly and accurately. I am excited to study more about the layered control architectures in the brain that enable us to pursue various goals under different tradeoffs.

As I work through this incredibly challenging question, I plan to stay plugged into developments in the AI world, filled with both amazement and self-satisfied smugness. Like any other tool that we have developed through the course of human civilization, I am sure it has its benefits and downsides depending on how it is used. However, I'm certain that it will help advance our understanding of our own agency by being an efficient sidekick. There will be some upheaval in the structures that we built for ourselves, structures that did not account for another species. But we will adapt, as we always do. Our agency is both robust and flexible.

For all those humans worried about AI taking over, I say: fear not. We may not compute as fast or remember as much, but we know how to stay in charge, and how to be in control.

Maybe I'm hopelessly delusional. Time will tell.

BIBLIOGRAPHY

- [1] RESCORLA R. A. “A theory of Pavlovian conditioning : Variations in the effectiveness of reinforcement and non-reinforcement”. In: *Classical conditioning, Current research and theory 2* (1972), pp. 64–69. URL: <https://cir.nii.ac.jp/crid/1572543025504096640>.
- [2] Sneha Aenugu and John O’Doherty. “Why we dive deep: A momentum-based account of depth-first persistence in hierarchically organized goals”. In: (Feb. 2026). DOI: [10.31234/OSF.IO/Y8ZTU_v2](https://doi.org/10.31234/OSF.IO/Y8ZTU_v2). URL: https://osf.io/y8ztu_v2.
- [3] Sneha Aenugu and John P. O’Doherty. “Building momentum: A computational account of persistence toward long-term goals”. In: *PLOS Computational Biology* 21.5 (May 2025), e1013054. ISSN: 1553-7358. DOI: [10.1371/JOURNAL.PCBI.1013054](https://doi.org/10.1371/JOURNAL.PCBI.1013054). URL: <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1013054>.
- [4] Matthew Ainsworth et al. “Transient frontopolar cortex stimulation induces prolonged disruption to counterfactual processing”. In: *PLOS Biology* 23.11 (Nov. 2025), e3003495. ISSN: 1545-7885. DOI: [10.1371/JOURNAL.PBIO.3003495](https://doi.org/10.1371/JOURNAL.PBIO.3003495). URL: <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3003495>.
- [5] Takuya Akiba et al. “Optuna: A Next-generation Hyperparameter Optimization Framework”. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (July 2019), pp. 2623–2631. DOI: [10.1145/3292500.3330701](https://doi.org/10.1145/3292500.3330701). URL: <https://dl.acm.org/doi/10.1145/3292500.3330701>.
- [6] Takuya Akiba et al. “Optuna: A Next-generation Hyperparameter Optimization Framework”. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (July 2019), pp. 2623–2631. DOI: [10.1145/3292500.3330701](https://doi.org/10.1145/3292500.3330701). URL: <https://arxiv.org/abs/1907.10902v1>.
- [7] Ricardo J. Alejandro and Clay B. Holroyd. “Hierarchical control over foraging behavior by anterior cingulate cortex”. In: *Neuroscience and biobehavioral reviews* 160 (May 2024). ISSN: 1873-7528. DOI: [10.1016/J.NEUBIOREV.2024.105623](https://doi.org/10.1016/J.NEUBIOREV.2024.105623). URL: <https://pubmed.ncbi.nlm.nih.gov/38490499/>.
- [8] William H. Alexander and Joshua W. Brown. “The Role of the Anterior Cingulate Cortex in Prediction Error and Signaling Surprise”. In: *Topics in Cognitive Science* 11.1 (Jan. 2019), pp. 119–135. ISSN: 1756-8765. DOI: [10.1111/TOPS.12307](https://doi.org/10.1111/TOPS.12307). URL: <https://onlinelibrary.wiley.com/doi/full/10.1111/tops.12307%20https://onlinelibrary.wiley>.

com/doi/abs/10.1111/tops.12307%20https://onlinelibrary.wiley.com/doi/10.1111/tops.12307.

- [9] Li Ji-An, Marcus K. Benna, and Marcelo G. Mattar. “Discovering cognitive strategies with tiny recurrent neural networks”. In: *Nature* 2025 644:8078 644.8078 (July 2025), pp. 993–1001. ISSN: 1476-4687. DOI: 10.1038/s41586-025-09142-4. URL: <https://www.nature.com/articles/s41586-025-09142-4>.
- [10] Hal R Arkes and Catherine Blumer. “The Psychology of Sunk Cost”. In: *BEHAVIOR AND HUMAN DECISION PROCESSES* 35 (1985), pp. 124–140.
- [11] Hal R. Arkes and Peter Ayton. “The sunk cost and concorde effects: Are humans less rational than lower animals?” In: *Psychological Bulletin* 125 (5 1999), pp. 591–600. ISSN: 00332909. DOI: 10.1037/0033-2909.125.5.591.
- [12] Nathaniel J.S. Ashby and Kinneret Teodorescu. “The effect of switching costs on choice-inertia and its consequences”. In: *PLOS ONE* 14.3 (Mar. 2019), e0214098. ISSN: 1932-6203. DOI: 10.1371/JOURNAL.PONE.0214098. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0214098>.
- [13] César Ávila et al. “Frontostriatal response to set switching is moderated by reward sensitivity”. In: *Social cognitive and affective neuroscience* 7.4 (Apr. 2012), pp. 423–430. ISSN: 1749-5024. DOI: 10.1093/SCAN/NSR028. URL: <https://pubmed.ncbi.nlm.nih.gov/21737433/>.
- [14] Ian Ballard et al. “Beyond Reward Prediction Errors: Human Striatum Updates Rule Values During Learning”. In: *Cerebral Cortex (New York, NY)* 28.11 (Nov. 2017), p. 3965. ISSN: 14602199. DOI: 10.1093/cercor/bhx259. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC6685076/>.
- [15] Russell A. Barkley. “Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD”. In: *Psychological bulletin* 121 (1 1997), pp. 65–94. ISSN: 0033-2909. DOI: 10.1037/0033-2909.121.1.65. URL: <https://pubmed.ncbi.nlm.nih.gov/9000892/>.
- [16] Pablo Billeke et al. “Human Anterior Insula Encodes Performance Feedback and Relays Prediction Error to the Medial Prefrontal Cortex”. In: *Cerebral Cortex* 30.7 (June 2020), pp. 4011–4025. ISSN: 1047-3211. DOI: 10.1093/CERCOR/BHAA017. URL: <https://dx.doi.org/10.1093/cercor/bhaa017>.
- [17] Erie D. Boorman, Timothy E. Behrens, and Matthew F. Rushworth. “Counterfactual Choice and Learning in a Neural Network Centered on Human Lateral Frontopolar Cortex”. In: *PLOS Biology* 9.6 (June 2011), e1001093. ISSN: 15449173. DOI: 10.1371/journal.pbio.1001093. URL: <https://>

journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001093.

- [18] Erie D. Boorman et al. “How Green Is the Grass on the Other Side? Frontopolar Cortex and the Evidence in Favor of Alternative Courses of Action”. In: *Neuron* 62.5 (June 2009), pp. 733–743. ISSN: 0896-6273. DOI: 10.1016/J.NEURON.2009.05.014.
- [19] Erica A. Boschini, Carinne Piekema, and Mark J. Buckley. “Essential functions of primate frontopolar cortex in cognition”. In: *Proceedings of the National Academy of Sciences of the United States of America* 112.9 (Mar. 2015), E1020–E1027. ISSN: 10916490. DOI: 10.1073/PNAS.1419649112/SUPPL{ _ }FILE/PNAS.201419649SI.PDF. URL: <https://www.pnas.org/doi/abs/10.1073/pnas.1419649112>.
- [20] Matthew M. Botvinick. “Hierarchical models of behavior and prefrontal function”. In: *Trends in Cognitive Sciences* 12.5 (May 2008), pp. 201–208. ISSN: 1364-6613. DOI: 10.1016/J.TICS.2008.02.009.
- [21] Matthew M. Botvinick, Jonathan D. Cohen, and Cameron S. Carter. “Conflict monitoring and anterior cingulate cortex: An update”. In: *Trends in Cognitive Sciences* 8.12 (Dec. 2004), pp. 539–546. ISSN: 13646613. DOI: 10.1016/J.TICS.2004.10.003/ASSET/BABBB42D-CAE8-419E-A1BB-0A65FCA84E9A/MAIN.ASSETS/GR3.JPG. URL: [https://www.cell.com/action/showFullText?pii=S1364661304002657%20https://www.cell.com/action/showAbstract?pii=S1364661304002657%20https://www.cell.com/trends/cognitive-sciences/abstract/S1364-6613\(04\)00265-7](https://www.cell.com/action/showFullText?pii=S1364661304002657%20https://www.cell.com/action/showAbstract?pii=S1364661304002657%20https://www.cell.com/trends/cognitive-sciences/abstract/S1364-6613(04)00265-7).
- [22] Matthew M. Botvinick, Yael Niv, and Andrew C. Barto. “Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective”. In: *Cognition* 113.3 (Dec. 2009), pp. 262–280. ISSN: 0010-0277. DOI: 10.1016/J.COGNITION.2008.08.011.
- [23] Veronika Brandstätter and Katharina Bernecker. “Persistence and Disengagement in Personal Goal Pursuit”. In: <https://doi.org/10.1146/annurev-psych-020821-110710> 73 (Jan. 2022), pp. 271–299. ISSN: 15452085. DOI: 10.1146/ANNUREV-PSYCH-020821-110710. URL: <https://www.annualreviews.org/doi/abs/10.1146/annurev-psych-020821-110710>.
- [24] Todd S. Braver. “The variable nature of cognitive control: a dual mechanisms framework”. In: *Trends in cognitive sciences* 16.2 (Feb. 2012), pp. 106–113. ISSN: 1879-307X. DOI: 10.1016/J.TICS.2011.12.010. URL: <https://pubmed.ncbi.nlm.nih.gov/22245618/>.
- [25] Todd S. Braver et al. “Flexible neural mechanisms of cognitive control within human prefrontal cortex”. In: *Proceedings of the National Academy of Sciences of the United States of America* 106.18 (May 2009), pp. 7351–

7356. ISSN: 00278424. DOI: 10.1073/PNAS.0808187106/SUPPL{_}FILE/0808187106SI.PDF. URL: www.pnas.org/cgi/content/full/10.1073/PNAS.0808187106/SUPPL{_}FILE/0808187106SI.PDF.
- [26] Walid Briki and Keith D. Markman. “Psychological momentum: The phenomenology of goal pursuit”. In: *Social and Personality Psychology Compass* 12 (9 Sept. 2018), e12412. ISSN: 1751-9004. DOI: 10.1111/SPC3.12412. URL: <https://onlinelibrary.wiley.com/doi/full/10.1111/spc3.12412> <https://onlinelibrary.wiley.com/doi/abs/10.1111/spc3.12412> <https://compass.onlinelibrary.wiley.com/doi/10.1111/spc3.12412>.
 - [27] Timothy J. Buschman and Earl K. Miller. “Goal-direction and top-down control”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 369.1655 (Nov. 2014), p. 20130471. ISSN: 14712970. DOI: 10.1098/RSTB.2013.0471. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC4186225/>.
 - [28] Ilaria Carta et al. “Cerebellar modulation of the reward circuitry and social behavior”. In: *Science* 363.6424 (Jan. 2019). ISSN: 10959203. DOI: 10.1126/SCIENCE.AAV0581/ASSET/446297CC-7304-4D40-8053-2B0EB3B72BCE/ASSETS/GRAPHIC/363{_}AAV0581{_}F7.JPEG. URL: http://dx.doi.org/10.1126/SCIENCE.AAV0581/ASSET/446297CC-7304-4D40-8053-2B0EB3B72BCE/ASSETS/GRAPHIC/363{_}AAV0581{_}F7.JPEG.
 - [29] Matthew V. Chafee and Patricia S. Goldman-Rakic. “Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task”. In: *Journal of neurophysiology* 79.6 (1998), pp. 2919–2940. ISSN: 0022-3077. DOI: 10.1152/JN.1998.79.6.2919. URL: <https://pubmed.ncbi.nlm.nih.gov/9636098/>.
 - [30] Eric L. Charnov. “Optimal foraging, the marginal value theorem”. In: *Theoretical Population Biology* 9 (2 Apr. 1976), pp. 129–136. ISSN: 0040-5809. DOI: 10.1016/0040-5809(76)90040-X.
 - [31] Sylvain Charron and Etienne Koechlin. “Divided representation of concurrent goals in the human frontal lobes”. In: *Science* 328.5976 (Apr. 2010), pp. 360–363. ISSN: 00368075. DOI: 10.1126/SCIENCE.1183614/SUPPL{_}FILE/CHARRON.SOM.PDF. URL: <https://www.science.org/doi/10.1126/science.1183614>.
 - [32] Shaozhe Cheng et al. “Intention beyond desire: Spontaneous intentional commitment regulates conflicting desires”. In: *Cognition* 238 (Sept. 2023), p. 105513. ISSN: 0010-0277. DOI: 10.1016/J.COGNITION.2023.105513.
 - [33] John A. Clithero and Antonio Rangel. “Informatic parcellation of the network involved in the computation of subjective value”. In: *Social Cognitive and Affective Neuroscience* 9.9 (Sept. 2014), pp. 1289–1302. ISSN: 1749-5016. DOI: 10.1093/SCAN/NST106. URL: <https://dx.doi.org/10.1093/scan/nst106>.

- [34] Sara M. Constantino and Nathaniel D. Daw. “Learning the opportunity cost of time in a patch-foraging task”. In: *Cognitive, affective & behavioral neuroscience* 15 (4 Dec. 2015), pp. 837–853. ISSN: 1531-135X. DOI: 10.3758/S13415-015-0350-Y. URL: <https://pubmed.ncbi.nlm.nih.gov/25917000/>.
- [35] Julia Cox and Ilana B. Witten. “Striatal circuits for reward learning and decision-making”. In: *Nature Reviews Neuroscience* 2019 20:8 20.8 (June 2019), pp. 482–494. ISSN: 14710048. DOI: 10.1038/s41583-019-0189-2. URL: <https://www.nature.com/articles/s41583-019-0189-2>.
- [36] Clayton E. Curtis and Mark D’Esposito. “Persistent activity in the prefrontal cortex during working memory”. In: *Trends in Cognitive Sciences* 7.9 (Sept. 2003), pp. 415–423. ISSN: 13646613. DOI: 10.1016/S1364-6613(03)00197-9 / ASSET / 9D126EBB - D627 - 4878 - 91FA - E0106D1614A9 / MAIN.ASSETS / GR4.GIF. URL: [https://www.cell.com/action/showFullText?pii=S1364661303001979%20https://www.cell.com/action/showAbstract?pii=S1364661303001979%20https://www.cell.com/trends/cognitive-sciences/abstract/S1364-6613\(03\)00197-9](https://www.cell.com/action/showFullText?pii=S1364661303001979%20https://www.cell.com/action/showAbstract?pii=S1364661303001979%20https://www.cell.com/trends/cognitive-sciences/abstract/S1364-6613(03)00197-9).
- [37] Guy Davidson and Todd M Gureckis. *Toward Complex and Structured Goals in Reinforcement Learning*.
- [38] Joanne M. Dickson et al. “An integrative study of motivation and goal regulation processes in subclinical anxiety, depression and hypomania”. In: *Psychiatry Research* 256 (Oct. 2017), pp. 6–12. ISSN: 0165-1781. DOI: 10.1016/J.PSYCHRES.2017.06.002.
- [39] Joanne M. Dickson et al. “Goal Fluency, Pessimism and Disengagement in Depression”. In: *PLOS ONE* 11 (11 Nov. 2016), e0166259. ISSN: 1932-6203. DOI: 10.1371/JOURNAL.PONE.0166259. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0166259>.
- [40] Linda Douw et al. “State-dependent variability of dynamic functional connectivity between frontoparietal and default networks relates to cognitive flexibility”. In: *Neuroscience* 339 (Dec. 2016), pp. 12–21. ISSN: 0306-4522. DOI: 10.1016/J.NEUROSCIENCE.2016.09.034.
- [41] Anja Dove et al. “Prefrontal cortex activation in task switching: an event-related fMRI study”. In: *Cognitive Brain Research* 9.1 (Jan. 2000), pp. 103–109. ISSN: 0926-6410. DOI: 10.1016/S0926-6410(99)00029-4.
- [42] Laura N. Driscoll, Krishna Shenoy, and David Sussillo. “Flexible multitask computation in recurrent networks utilizes shared dynamical motifs”. In: *Nature Neuroscience* 2024 27:7 27.7 (July 2024), pp. 1349–1363. ISSN: 1546-1726. DOI: 10.1038/s41593-024-01668-6. URL: <https://www.nature.com/articles/s41593-024-01668-6>.

- [43] Daniel Durstewitz, Jeremy K. Seamans, and Terrence J. Sejnowski. “Neuro-computational models of working memory”. In: *Nature Neuroscience* 3:11 (2000), pp. 1184–1191. ISSN: 1546-1726. DOI: 10.1038/81460. URL: https://www.nature.com/articles/nn1100_1184.
- [44] Eran Eldar et al. “Mood as Representation of Momentum”. In: *Trends in Cognitive Sciences* 20 (1 Jan. 2016), pp. 15–24. ISSN: 1364-6613. DOI: 10.1016/J.TICS.2015.07.010.
- [45] Eran Eldar et al. “Mood as Representation of Momentum”. In: *Trends in Cognitive Sciences* 20.1 (Jan. 2016), p. 15. ISSN: 1879307X. DOI: 10.1016/J.TICS.2015.07.010. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC4703769/>.
- [46] K. Anders Ericsson. “Deliberate practice and acquisition of expert performance: a general overview”. In: *Academic emergency medicine : official journal of the Society for Academic Emergency Medicine* 15.11 (Nov. 2008), pp. 988–994. ISSN: 1553-2712. DOI: 10.1111/J.1553-2712.2008.00227.X. URL: <https://pubmed.ncbi.nlm.nih.gov/18778378/>.
- [47] Oscar Esteban et al. “fMRIPrep: a robust preprocessing pipeline for functional MRI”. In: *Nature Methods* 2018 16:1 16.1 (Dec. 2018), pp. 111–116. ISSN: 1548-7105. DOI: 10.1038/s41592-018-0235-4. URL: <https://www.nature.com/articles/s41592-018-0235-4>.
- [48] Edward F. Ester, Thomas C. Sprague, and John T. Serences. “Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory”. In: *Neuron* 87.4 (Aug. 2015), pp. 893–905. ISSN: 1097-4199. DOI: 10.1016/J.NEURON.2015.07.013. URL: <https://pubmed.ncbi.nlm.nih.gov/26257053/>.
- [49] Tala Fakhoury et al. “Models of attractor dynamics in the brain”. In: (May 2025). URL: <https://arxiv.org/abs/2505.01098v2>.
- [50] Dylan J. Foster, Adam Block, and Dipendra Misra. “Is Behavior Cloning All You Need? Understanding Horizon in Imitation Learning”. In: *Advances in Neural Information Processing Systems* 37 (July 2024). ISSN: 10495258. DOI: 10.52202/079017-3834. URL: <https://arxiv.org/abs/2407.15007v2>.
- [51] Michael J. Frank, Bryan Loughry, and Randall C. O’Reilly. “Interactions between frontal cortex and basal ganglia in working memory: a computational model”. In: *Cognitive, affective & behavioral neuroscience* 1.2 (2001), pp. 137–160. ISSN: 1530-7026. DOI: 10.3758/CABN.1.2.137. URL: <https://pubmed.ncbi.nlm.nih.gov/12467110/>.
- [52] Shane Frederick, George Loewenstein, and Ted O’Donoghue. “Time Discounting and Time Preference: A Critical Review”. In: *Journal of Economic Literature* 40 (2 Jan. 2002), pp. 351–401. ISSN: 0022-0515. DOI: 10.1257/002205102320161311.

- [53] Jennifer J. Freyd and Ronald A. Finke. “Representational momentum”. In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10 (1 Jan. 1984), pp. 126–132. ISSN: 02787393. DOI: 10.1037/0278-7393.10.1.126.
- [54] Paola Fuentes-Claramonte et al. “Reward Sensitivity Modulates Brain Activity in the Prefrontal Cortex, ACC and Striatum during Task Switching”. In: *PLOS ONE* 10.4 (Apr. 2015), e0123073. ISSN: 1932-6203. DOI: 10.1371/JOURNAL.PONE.0123073. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0123073>.
- [55] S. Funahashi, C. J. Bruce, and P. S. Goldman-Rakic. “Mnemonic coding of visual space in the monkey’s dorsolateral prefrontal cortex”. In: *Journal of neurophysiology* 61.2 (1989), pp. 331–349. ISSN: 0022-3077. DOI: 10.1152/JN.1989.61.2.331. URL: <https://pubmed.ncbi.nlm.nih.gov/2918358/>.
- [56] Joaquín M. Fuster. “The prefrontal cortex—an update: time is of the essence”. In: *Neuron* 30.2 (2001), pp. 319–333. ISSN: 0896-6273. DOI: 10.1016/S0896-6273(01)00285-9. URL: <https://pubmed.ncbi.nlm.nih.gov/11394996/>.
- [57] Joaquin M. Fuster and Garrett E. Alexander. “Neuron Activity Related to Short-Term Memory”. In: *Science* 173.3997 (Aug. 1971), pp. 652–654. ISSN: 00368075. DOI: 10.1126/SCIENCE.173.3997.652. URL: <https://www.science.org/doi/10.1126/science.173.3997.652>.
- [58] Mohamady El-Gaby et al. “A cellular basis for mapping behavioural structure”. In: *Nature* 2024 636:8043 636.8043 (Nov. 2024), pp. 671–680. ISSN: 1476-4687. DOI: 10.1038/s41586-024-08145-x. URL: <https://www.nature.com/articles/s41586-024-08145-x>.
- [59] Neil Garrett and Nathaniel D. Daw. “Biased belief updating and suboptimal choice in foraging decisions”. In: *Nature Communications* 2020 11:1 11 (1 July 2020), pp. 1–12. ISSN: 2041-1723. DOI: 10.1038/s41467-020-16964-5. URL: <https://www.nature.com/articles/s41467-020-16964-5>.
- [60] Samuel J. Gershman. “Origin of perseveration in the trade-off between reward and complexity”. In: *Cognition* 204 (Nov. 2020). ISSN: 1873-7838. DOI: 10.1016/J.COGNITION.2020.104394. URL: <https://pubmed.ncbi.nlm.nih.gov/32679270/>.
- [61] Ali Ghazizadeh, Simon Hong, and Okihide Hikosaka. “Prefrontal Cortex Represents Long-Term Memory of Object Values for Months”. In: *Current biology : CB* 28.14 (July 2018), pp. 2206–2217. ISSN: 1879-0445. DOI: 10.1016/J.CUB.2018.05.017. URL: <https://pubmed.ncbi.nlm.nih.gov/30056855/>.

- [62] Sasha E.B. Gibbs and Mark D’Esposito. “Individual capacity differences predict working memory performance and prefrontal activity following dopamine receptor stimulation”. In: *Cognitive, Affective and Behavioral Neuroscience* 5.2 (June 2005), pp. 212–221. ISSN: 15307026. DOI: 10.3758/CABN.5.2.212/METRICS. URL: <https://link.springer.com/article/10.3758/CABN.5.2.212>.
- [63] Erving Goffman. “The Presentation of Self in Everyday Life”. In: *Social Theory Re-Wired* (June 2023), pp. 450–459. DOI: 10.4324/9781003320609-59. URL: <https://www.taylorfrancis.com/chapters/edit/10.4324/9781003320609-59/presentation-self-everyday-life-erving-goffman>.
- [64] P. S. Goldman-Rakic. “Cellular basis of working memory”. In: *Neuron* 14.3 (1995), pp. 477–485. ISSN: 0896-6273. DOI: 10.1016/0896-6273(95)90304-6. URL: <https://pubmed.ncbi.nlm.nih.gov/7695894/>.
- [65] Krzysztof Gorgolewski et al. “Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python”. In: *Frontiers in Neuroinformatics* 5 (Aug. 2011), p. 12318. ISSN: 16625196. DOI: 10.3389/FNINF.2011.00013/TEXT. URL: www.frontiersin.org.
- [66] Thomas Goschke and Julius Kuhl. “Representation of Intentions: Persisting Activation in Memory”. In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19.5 (1993), pp. 1197–1210. ISSN: 02787393. DOI: 10.1037/0278-7393.19.5.1211.
- [67] Ivan Grahek et al. “Control adjustment costs limit goal flexibility: Empirical evidence and a computational account”. In: *bioRxiv* (Feb. 2025), p. 2023.08.22.554296. DOI: 10.1101/2023.08.22.554296. URL: <https://www.biorxiv.org/content/10.1101/2023.08.22.554296v4>
<https://www.biorxiv.org/content/10.1101/2023.08.22.554296v4.abstract>.
- [68] Ann M. Graybiel. “Habits, rituals, and the evaluative brain”. In: *Annual review of neuroscience* 31 (2008), pp. 359–387. ISSN: 0147-006X. DOI: 10.1146/ANNUREV.NEURO.29.051605.112851. URL: <https://pubmed.ncbi.nlm.nih.gov/18558860/>.
- [69] Cooper D. Grossman, Vincent Man, and John P. O’Doherty. “The representation and valuation of subgoals in the human brain during model-based hierarchical behavior”. In: *Neuron* 0.0 (2026). ISSN: 0896-6273. DOI: 10.1016/J.NEURON.2025.12.023. URL: [https://www.cell.com/action/showFullText?pii=S0896627325009791%20https://www.cell.com/action/showAbstract?pii=S0896627325009791%20https://www.cell.com/neuron/abstract/S0896-6273\(25\)00979-1](https://www.cell.com/action/showFullText?pii=S0896627325009791%20https://www.cell.com/action/showAbstract?pii=S0896627325009791%20https://www.cell.com/neuron/abstract/S0896-6273(25)00979-1).

- [70] Daniel E. Gustavson et al. “Genetic relations among procrastination, impulsivity, and goal-management ability: Implications for the evolutionary origin of procrastination”. In: *Psychological Science* 25 (6 Apr. 2014), pp. 1178–1188. ISSN: 14679280. DOI: 10.1177/0956797614526260/ASSET/IMAGES/LARGE/10.1177_0956797614526260-FIG3.JPEG. URL: https://journals.sagepub.com/doi/full/10.1177/0956797614526260?casa_token=CD5h5i1xVIMAAAAA%3A4HMioUzvoFVGSqVriTJdfeli3juAhpRxR0pns7Wi0JmdquUGUFhiRFs10KRPx0JzNfK3dw.
- [71] Adrian M. Haith and John W. Krakauer. “The multiple effects of practice: skill, habit and reduced cognitive load”. In: *Current opinion in behavioral sciences* 20 (Apr. 2018), p. 196. ISSN: 23521546. DOI: 10.1016/J.COBEHA.2018.01.015. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC6443249/>.
- [72] Grace E. Hallenbeck et al. “Working memory representations in visual cortex mediate distraction effects”. In: *Nature Communications* 2021 12:1 12.1 (Aug. 2021), pp. 4714–. ISSN: 2041-1723. DOI: 10.1038/s41467-021-24973-1. URL: <https://www.nature.com/articles/s41467-021-24973-1>.
- [73] Fred H. Hamker, Javier Baladron, and Lieneke K. Janssen. “Interacting cortico-basal ganglia-thalamocortical loops shape behavioral control through cognitive maps and shortcuts”. In: *Trends in Neurosciences* 48.11 (Nov. 2025), pp. 841–852. ISSN: 0166-2236. DOI: 10.1016/J.TINS.2025.09.006.
- [74] Jean Harb et al. “When Waiting Is Not an Option: Learning Options with a Deliberation Cost”. In: (). DOI: 10.5555/3504035.3504422. URL: <https://dl.acm.org/doi/10.5555/3504035.3504422>.
- [75] Todd A. Hare et al. “Dissociating the Role of the Orbitofrontal Cortex and the Striatum in the Computation of Goal Values and Prediction Errors”. In: *Journal of Neuroscience* 28.22 (May 2008), pp. 5623–5630. ISSN: 02706474. DOI: 10.1523/JNEUROSCI.1309-08.2008. URL: <https://www.jneurosci.org/content/28/22/5623%20https://www.jneurosci.org/content/28/22/5623.abstract>.
- [76] Seyed Reza Hashemirad, Mojtaba Abbaszadeh, and Ali Ghazizadeh. “Pre-frontal cortex temporally multiplexes slow and fast dynamics in value learning and memory”. In: *Nature Communications* 2025 16:1 16.1 (Dec. 2025), pp. 11081–. ISSN: 2041-1723. DOI: 10.1038/s41467-025-66081-4. URL: <https://www.nature.com/articles/s41467-025-66081-4>.
- [77] Seth A. Herd et al. “A neural network model of individual differences in task switching abilities”. In: *Neuropsychologia* 62 (Sept. 2014), pp. 375–389. ISSN: 0028-3932. DOI: 10.1016/J.NEUROPSYCHOLOGIA.2014.04.014.

- [78] Francis Heylighen. “The meaning and origin of goal-directedness: a dynamical systems perspective”. In: *Biological Journal of the Linnean Society* 139.4 (Aug. 2023), pp. 370–387. ISSN: 0024-4066. DOI: 10.1093/BIOLINNEAN/BLAC060. URL: <https://dx.doi.org/10.1093/biolinnean/blac060>.
- [79] Lee Hogarth, Henry W. Chase, and Kathleen Baess. “Impaired goal-directed behavioural control in human impulsivity”. In: *Quarterly Journal of Experimental Psychology* 65 (2 Feb. 2012), pp. 305–316. ISSN: 17470226. DOI: 10.1080/17470218.2010.518242/ASSET/IMAGES/LARGE/10.1080_17470218.2010.518242-FIG1.JPEG. URL: <https://journals.sagepub.com/doi/full/10.1080/17470218.2010.518242>.
- [80] Clay B. Holroyd and Nick Yeung. “An Integrative Theory of Anterior Cingulate Cortex Function: Option Selection in Hierarchical Reinforcement Learning”. In: *Neural Basis of Motivational and Cognitive Control* (Oct. 2011), pp. 333–350. DOI: 10.7551/MITPRESS/8791.003.0024. URL: <https://direct.mit.edu/books/edited-volume/3322/chapter/107640/An-Integrative-Theory-of-Anterior-Cingulate-Cortex>.
- [81] Eleanor Holton et al. “Goal commitment is supported by vmPFC through selective attention”. In: *Nature Human Behaviour* 2024 8:7 8 (7 Apr. 2024), pp. 1351–1365. ISSN: 2397-3374. DOI: 10.1038/s41562-024-01844-5. URL: <https://www.nature.com/articles/s41562-024-01844-5>.
- [82] Eleanor Holton et al. “Goal commitment is supported by vmPFC through selective attention”. In: *Nature Human Behaviour* 2024 8:7 8.7 (Apr. 2024), pp. 1351–1365. ISSN: 2397-3374. DOI: 10.1038/s41562-024-01844-5. URL: <https://www.nature.com/articles/s41562-024-01844-5>.
- [83] Christopher J. Honey, Abhijit Mahabal, and Buddhika Bellana. “Psychological Momentum”. In: <https://doi.org/10.1177/09637214221143053> 32 (4 May 2023), pp. 284–292. ISSN: 14678721. DOI: 10.1177/09637214221143053. URL: <https://journals.sagepub.com/doi/abs/10.1177/09637214221143053>.
- [84] James D. Howard et al. “Identity-specific coding of future rewards in the human orbitofrontal cortex”. In: *Proceedings of the National Academy of Sciences of the United States of America* 112.16 (Apr. 2015), pp. 5195–5200. ISSN: 10916490. DOI: 10.1073/PNAS.1503550112/SUPPL{_}FILE/PNAS.201503550SI.PDF. URL: <https://www.pnas.org/doi/abs/10.1073/pnas.1503550112>.
- [85] Timothy L. Hubbard. “Representational momentum and related displacements in spatial memory: A review of the findings”. In: *Psychonomic Bulletin and Review* 12 (5 2005), pp. 822–851. ISSN: 10699384. DOI: 10.3758/BF03196775/METRICS. URL: <https://link.springer.com/article/10.3758/BF03196775>.

- [86] C. L. Hull. “The goal-gradient hypothesis and maze learning”. In: *Psychological Review* 39.1 (Jan. 1932), pp. 25–43. ISSN: 0033295X. DOI: 10.1037/H0072640.
- [87] Alexandre Hyafil, Christopher Summerfield, and Etienne Koechlin. “Two Mechanisms for Task Switching in the Prefrontal Cortex”. In: *The Journal of Neuroscience* 29.16 (Apr. 2009), p. 5135. ISSN: 15292401. DOI: 10.1523/JNEUROSCI.2828-08.2009. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC6665449/>.
- [88] Akiko Ikka and Clayton E. Curtis. “Common neural mechanisms supporting spatial working memory, attention and motor intention”. In: *Neuropsychologia* 49.6 (May 2011), pp. 1428–1434. ISSN: 1873-3514. DOI: 10.1016/J.NEUROPSYCHOLOGIA.2010.12.020. URL: <https://pubmed.ncbi.nlm.nih.gov/21182852/>.
- [89] Shamsi T. Iqbal and Eric Horvitz. “Disruption and recovery of computing tasks: Field study, analysis, and directions”. In: *Conference on Human Factors in Computing Systems - Proceedings* (2007), pp. 677–686. DOI: 10.1145/1240624.1240730/SUPPL{_}FILE/1240730.MP4. URL: <https://dl.acm.org/doi/10.1145/1240624.1240730>.
- [90] Alejandro Jimenez-Rodriguez et al. “A Framework for Resolving Motivational Conflict via Attractor Dynamics”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 12413 LNAI (2020), pp. 192–203. ISSN: 1611-3349. DOI: 10.1007/978-3-030-64313-3{_}19. URL: https://link.springer.com/chapter/10.1007/978-3-030-64313-3_19.
- [91] Kerstin Jost, Vera Hennecke, and Iring Koch. “Task dominance determines backward inhibition in task switching”. In: *Frontiers in Psychology* 8.MAY (May 2017), p. 257339. ISSN: 16641078. DOI: 10.3389/FPSYG.2017.00755/TEXT. URL: www.frontiersin.org.
- [92] Daniel Kahneman, Jack L. Knetsch, and Richard H. Thaler. “Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias”. In: *Journal of Economic Perspectives* 5 (1 1991), pp. 193–206. ISSN: 0895-3309. DOI: 10.1257/JEP.5.1.193.
- [93] Moon Sun Kang and Yu Chin Chiu. “Proactive and reactive metacontrol in task switching”. In: *Memory and Cognition* 49.8 (Nov. 2021), pp. 1617–1632. ISSN: 15325946. DOI: 10.3758/S13421-021-01189-8/TABLES/2. URL: <https://link.springer.com/article/10.3758/s13421-021-01189-8>.
- [94] Riley K. Kendall and Andrew M. Wikenheiser. “Quitting While You’re Ahead: Patch Foraging and Temporal Cognition”. In: *Behavioral Neuroscience* 136 (5 July 2022), pp. 467–478. ISSN: 19390084. DOI: 10.1037/BNE0000526.

- [95] Dirk Kerzel. “Representational momentum beyond internalized physics: Embodied mechanisms of anticipation cause errors in visual short-term memory”. In: *Current Directions in Psychological Science* 14 (4 Aug. 2005), pp. 180–184. ISSN: 09637214. DOI: 10.1111/J.0963-7214.2005.00360.X/ASSET/IMAGES/LARGE/10.1111_J.0963-7214.2005.00360.X-FIG4.JPEG. URL: <https://journals.sagepub.com/doi/full/10.1111/j.0963-7214.2005.00360.x>.
- [96] Andrea Kiesel et al. “Control and interference in task switching—a review”. In: *Psychological bulletin* 136.5 (Sept. 2010), pp. 849–874. ISSN: 1939-1455. DOI: 10.1037/A0019842. URL: <https://pubmed.ncbi.nlm.nih.gov/20804238/>.
- [97] Chobok Kim et al. “Domain general and domain preferential brain regions associated with different types of task switching: A Meta-Analysis”. In: *Human Brain Mapping* 33.1 (Jan. 2012), pp. 130–142. ISSN: 1097-0193. DOI: 10.1002/HBM.21199. URL: <https://onlinelibrary.wiley.com/doi/full/10.1002/hbm.21199%20https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.21199%20https://onlinelibrary.wiley.com/doi/10.1002/hbm.21199>.
- [98] Hyoung F. Kim and Okihide Hikosaka. “Distinct basal ganglia circuits controlling behaviors guided by flexible and stable values”. In: *Neuron* 79.5 (Sept. 2013), pp. 1001–1010. ISSN: 1097-4199. DOI: 10.1016/J.NEURON.2013.06.044. URL: <https://pubmed.ncbi.nlm.nih.gov/23954031/>.
- [99] Ran Kivetz, Oleg Urminsky, and Yuhuang Zheng. “The goal-gradient hypothesis resurrected: Purchase acceleration, illusionary goal progress, and customer retention”. In: *Journal of Marketing Research* 43.1 (2006), pp. 39–58. ISSN: 00222437. DOI: 10.1509/JMKR.43.1.39.
- [100] Etienne Koechlin and Alexandre Hyafil. “Anterior prefrontal function and the limits of human decision-making”. In: *Science* 318.5850 (Oct. 2007), pp. 594–598. ISSN: 00368075. DOI: 10.1126/SCIENCE.1142995/SUPPL{_}FILE/KOECHLIN{_}SOM.PDF. URL: <https://www.science.org/doi/10.1126/science.1142995>.
- [101] Etienne Koechlin et al. “The role of the anterior prefrontal cortex in human cognition”. In: *Nature* 399.6732 (May 1999), pp. 148–151. ISSN: 0028-0836. DOI: 10.1038/20178. URL: <https://pubmed.ncbi.nlm.nih.gov/10335843/>.
- [102] Nils Kolling et al. “Neural mechanisms of foraging”. In: *Science* 335.6077 (Apr. 2012), pp. 95–98. ISSN: 10959203. DOI: 10.1126/SCIENCE.1216930/SUPPL{_}FILE/1216930.KOLLING.SM.PDF. URL: <https://www.science.org/doi/10.1126/science.1216930>.

- [103] Nils Kolling et al. “Value, search, persistence and model updating in anterior cingulate cortex”. In: *Nature neuroscience* 19.10 (Oct. 2016), pp. 1280–1285. ISSN: 1546-1726. DOI: 10.1038/NN.4382. URL: <https://pubmed.ncbi.nlm.nih.gov/27669988/>.
- [104] Katharina Koppe and Klaus Rothermund. “Let it go: Depression facilitates disengagement from unattainable goals”. In: *Journal of Behavior Therapy and Experimental Psychiatry* 54 (Mar. 2017), pp. 278–284. ISSN: 0005-7916. DOI: 10.1016/J.JBTEP.2016.10.003.
- [105] Christoph W. Korn and Dominik R. Bach. “Heuristic and optimal policy computations in the human brain during sequential decision-making”. In: *Nature Communications* 2018 9:1 9.1 (Jan. 2018), pp. 325–. ISSN: 2041-1723. DOI: 10.1038/s41467-017-02750-3. URL: <https://www.nature.com/articles/s41467-017-02750-3>.
- [106] Robert Kurzban et al. “An opportunity cost model of subjective effort and task performance”. In: *The Behavioral and brain sciences* 36.6 (Dec. 2013), 10.1017/S0140525X12003196. ISSN: 0140525X. DOI: 10.1017/S0140525X12003196. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC3856320/>.
- [107] Aaron J. Lanz et al. “Disinhibition of a recurrent attractor gates a persistent goal signal for navigation”. In: *bioRxiv* (Apr. 2026), p. 2025.10.07.681003. ISSN: 2692-8205. DOI: 10.1101/2025.10.07.681003. URL: <https://www.biorxiv.org/content/10.1101/2025.10.07.681003v2%20https://www.biorxiv.org/content/10.1101/2025.10.07.681003v2.abstract>.
- [108] Antonio H. Lara and Jonathan D. Wallis. “The Role of Prefrontal Cortex in Working Memory: A Mini Review”. In: *Frontiers in Systems Neuroscience* 9.DEC (Dec. 2015), p. 173. ISSN: 16625137. DOI: 10.3389/FNSYS.2015.00173. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC4683174/>.
- [109] K S Lashley. “The Problem of Serial Order in Behavior”. In: ().
- [110] Daniella Laureiro-Martínez et al. “Frontopolar cortex and decision-making efficiency: comparing brain activity of experts with different professional background during an exploration-exploitation task”. In: *Frontiers in Human Neuroscience* 7.JAN (Jan. 2014), p. 927. ISSN: 16625161. DOI: 10.3389/FNHUM.2013.00927. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC3897871/>.
- [111] Gordon D. Logan. “Toward an Instance Theory of Automatization”. In: *Psychological Review* 95.4 (1988), pp. 492–527. ISSN: 0033295X. DOI: 10.1037/0033-295X.95.4.492.

- [112] Alizée Lopez-Persem, Philippe Domenech, and Mathias Pessiglione. “How prior preferences determine decision-making frames and biases in the human brain”. In: *eLife* 5.NOVEMBER2016 (Nov. 2016). ISSN: 2050084X. DOI: 10.7554/ELIFE.20317.
- [113] Brian Nils Lundstrom, Martin Ingvar, and Karl Magnus Petersson. “The role of precuneus and left inferior frontal cortex during source memory episodic retrieval”. In: *NeuroImage* 27.4 (Oct. 2005), pp. 824–834. ISSN: 1053-8119. DOI: 10.1016/J.NEUROIMAGE.2005.05.008. URL: <https://pubmed.ncbi.nlm.nih.gov/15982902/>.
- [114] F. Charles Mace et al. “BEHAVIORAL MOMENTUM IN THE TREATMENT OF NONCOMPLIANCE”. In: *Journal of Applied Behavior Analysis* 21 (2 June 1988), pp. 123–141. ISSN: 1938-3703. DOI: 10.1901/JABA.1988.21-123. URL: <https://onlinelibrary.wiley.com/doi/full/10.1901/jaba.1988.21-123%20https://onlinelibrary.wiley.com/doi/abs/10.1901/jaba.1988.21-123%20https://onlinelibrary.wiley.com/doi/10.1901/jaba.1988.21-123>.
- [115] Robert G. Mair et al. “Where Actions Meet Outcomes: Medial Prefrontal Cortex, Central Thalamus, and the Basal Ganglia”. In: *Frontiers in Behavioral Neuroscience* 16 (July 2022), p. 928610. ISSN: 16625153. DOI: 10.3389/FNBEH.2022.928610. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9294389/>.
- [116] Sanjay Manohar et al. “Reduced decision bias and more rational decision making following ventromedial prefrontal cortex damage”. In: *Cortex; a journal devoted to the study of the nervous system and behavior* 138 (May 2021), pp. 24–37. ISSN: 1973-8102. DOI: 10.1016/J.CORTEX.2021.01.015. URL: <https://pubmed.ncbi.nlm.nih.gov/33677325/>.
- [117] Farshad A. Mansouri, Keiji Tanaka, and Mark J. Buckley. “Conflict-induced behavioural adjustment: a clue to the executive functions of the prefrontal cortex”. In: *Nature reviews. Neuroscience* 10.2 (Feb. 2009), pp. 141–152. ISSN: 1471-0048. DOI: 10.1038/NRN2538. URL: <https://pubmed.ncbi.nlm.nih.gov/19153577/>.
- [118] Farshad Alizadeh Mansouri, David J. Freedman, and Mark J. Buckley. “Emergence of abstract rules in the primate brain”. In: *Nature Reviews Neuroscience* 2020 21:11 21.11 (Sept. 2020), pp. 595–610. ISSN: 1471-0048. DOI: 10.1038/s41583-020-0364-5. URL: <https://www.nature.com/articles/s41583-020-0364-5>.
- [119] Farshad Alizadeh Mansouri et al. “Managing competing goals — a key role for the frontopolar cortex”. In: *Nature Reviews Neuroscience* 2017 18:11 18.11 (Sept. 2017), pp. 645–657. ISSN: 1471-0048. DOI: 10.1038/nrn.2017.111. URL: <https://www.nature.com/articles/nrn.2017.111>.

- [120] Keith D Markman and Corey L Guenther. “Psychological Momentum: Intuitive Physics and Naive Beliefs”. In: *PSPB* 33 (6 2007), pp. 800–812. DOI: 10.1177/0146167207301026.
- [121] Marcelo G. Mattar and Máté Lengyel. “Planning in the brain”. In: *Neuron* 110.6 (Mar. 2022), pp. 914–934. ISSN: 0896-6273. DOI: 10.1016/J.NEURON.2021.12.018.
- [122] Ulrich Mayr and Steven W. Keele. “Changing internal constraints on action: the role of backward inhibition”. In: *Journal of experimental psychology. General* 129.1 (2000), pp. 4–26. ISSN: 0096-3445. DOI: 10.1037/0096-3445.129.1.4. URL: <https://pubmed.ncbi.nlm.nih.gov/10756484/>.
- [123] Juri Minxha et al. “Flexible recruitment of memory-based choice representations by the human medial frontal cortex”. In: *Science* 368.6498 (June 2020). ISSN: 10959203. DOI: 10.1126/SCIENCE.ABA3313/SUPPL{_}FILE/ABA3313{_}S1.MOV. URL: <https://www.science.org/doi/10.1126/science.aba3313>.
- [124] Karin Mogg and Brendan P. Bradley. “Attentional bias in generalized anxiety disorder versus depressive disorder”. In: *Cognitive Therapy and Research* 29 (1 Feb. 2005), pp. 29–45. ISSN: 01475916. DOI: 10.1007/S10608-005-1646-Y/METRICS. URL: <https://link.springer.com/article/10.1007/s10608-005-1646-y>.
- [125] Gaia Molinaro and Anne G.E. Collins. “A goal-centric outlook on learning”. In: *Trends in Cognitive Sciences* (Sept. 2023). ISSN: 1364-6613. DOI: 10.1016/J.TICS.2023.08.011.
- [126] Gaia Molinaro, Aly Lidayan, and Anne G E Collins. *Investigating the Role of Representation Switching Costs in Goal Persistence Bias*. URL: <https://www.prolific.com/>.
- [127] Gaia Molinaro et al. “Latent Learning Progress Drives Autonomous Goal Selection in Human Reinforcement Learning”. In: (June 2024). DOI: 10.31234/OSF.IO/5DTX3. URL: <https://osf.io/5dtx3>.
- [128] Gaia Molinaro et al. “Latent Learning Progress Drives Autonomous Goal Selection in Human Reinforcement Learning”. In: *Advances in Neural Information Processing Systems* 37 (Dec. 2024), pp. 32251–32280. ISSN: 10495258. DOI: 10.52202/079017-1013.
- [129] Stephen Monsell. “Task switching”. In: *Trends in Cognitive Sciences* 7.3 (Mar. 2003), pp. 134–140. ISSN: 13646613. DOI: 10.1016/S1364-6613(03)00028-7. URL: <https://pubmed.ncbi.nlm.nih.gov/12639695/>.
- [130] Stephen Monsell and Guy A. Mizon. “Can the task-cuing paradigm measure an endogenous task-set reconfiguration process?” In: *Journal of experimental psychology. Human perception and performance* 32.3 (2006), pp. 493–

516. ISSN: 0096-1523. DOI: 10.1037/0096-1523.32.3.493. URL: <https://pubmed.ncbi.nlm.nih.gov/16822121/>.
- [131] Stephen Monsell, Nick Yeung, and Rayna Azuma. “Reconfiguration of task-set: is it easier to switch to the weaker task?” In: *Psychological research* 63.3-4 (2000), pp. 250–264. ISSN: 0340-0727. DOI: 10.1007/S004269900005. URL: <https://pubmed.ncbi.nlm.nih.gov/11004879/>.
- [132] Paul S. Muhle-Karbe, Wouter De Baene, and Marcel Brass. “Do tasks matter in task switching? Dissociating domain-general from context-specific brain activity”. In: *NeuroImage* 99 (Oct. 2014), pp. 332–341. ISSN: 1053-8119. DOI: 10.1016/J.NEUROIMAGE.2014.05.058.
- [133] John D. Murray et al. “Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex”. In: *Proceedings of the National Academy of Sciences of the United States of America* 114.2 (Jan. 2017), pp. 394–399. ISSN: 10916490. DOI: 10.1073/PNAS.1619449114/SUPPL{_}FILE/PNAS.201619449SI.PDF. URL: <https://www.pnas.org/doi/abs/10.1073/pnas.1619449114>.
- [134] Sebastian Musslick and Anastasia Bizyaeva. “Examining cognitive flexibility and stability through the lens of dynamical systems”. In: *Current Opinion in Behavioral Sciences* 57 (June 2024), p. 101375. ISSN: 2352-1546. DOI: 10.1016/J.COBCHA.2024.101375.
- [135] Aditya Nair et al. “An approximate line attractor in the hypothalamus encodes an aggressive state”. In: *Cell* 186.1 (Jan. 2023), pp. 178–193. ISSN: 10974172. DOI: 10.1016/J.CELL.2022.11.027/ASSET/FD78DC3B-A851-4049-9E3D-095019B8FCC6/MAIN.ASSETS/GR6.JPG. URL: [https://www.cell.com/action/showFullText?pii=S0092867422014714%20https://www.cell.com/action/showAbstract?pii=S0092867422014714%20https://www.cell.com/cell/abstract/S0092-8674\(22\)01471-4](https://www.cell.com/action/showFullText?pii=S0092867422014714%20https://www.cell.com/action/showAbstract?pii=S0092867422014714%20https://www.cell.com/cell/abstract/S0092-8674(22)01471-4).
- [136] John A. Nevin and Randolph C. Grace. “Behavioral momentum and the Law of Effect”. In: *Behavioral and Brain Sciences* 23 (1 2000), pp. 73–90. ISSN: 1469-1825. DOI: 10.1017/S0140525X00002405. URL: <https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/behavioral-momentum-and-the-law-of-effect/6BBA6C4A6C7B6E678609F9467AA86E98>.
- [137] John A. Nevin, Charlotte Mandell, and Jean R. Atak. “THE ANALYSIS OF BEHAVIORAL MOMENTUM”. In: *Journal of the Experimental Analysis of Behavior* 39 (1 Jan. 1983), pp. 49–59. ISSN: 1938-3711. DOI: 10.1901/JEAB.1983.39-49. URL: <https://onlinelibrary.wiley.com/doi/full/10.1901/jeab.1983.39-49%20https://onlinelibrary.wiley.com/doi/abs/10.1901/jeab.1983.39-49%20https://onlinelibrary.wiley.com/doi/10.1901/jeab.1983.39-49>.

- [138] John A. Nevin and Timothy A. Shahan. “BEHAVIORAL MOMENTUM THEORY: EQUATIONS AND APPLICATIONS”. In: *Journal of Applied Behavior Analysis* 44 (4 Dec. 2011), pp. 877–895. ISSN: 1938-3703. DOI: 10.1901/JABA.2011.44-877. URL: <https://onlinelibrary.wiley.com/doi/full/10.1901/jaba.2011.44-877> <https://onlinelibrary.wiley.com/doi/abs/10.1901/jaba.2011.44-877> <https://onlinelibrary.wiley.com/doi/10.1901/jaba.2011.44-877>.
- [139] John O’Doherty et al. “Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning”. In: *Science* 304.5669 (Apr. 2004), pp. 452–454. ISSN: 00368075. DOI: 10.1126/SCIENCE.1094285/SUPPL{_}FILE/O{_}DOHERTY.PDF. URL: <https://www.science.org/doi/10.1126/science.1094285>.
- [140] Randall C. O’Reilly and Michael J. Frank. “Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia”. In: *Neural computation* 18.2 (Feb. 2006), pp. 283–328. ISSN: 0899-7667. DOI: 10.1162/089976606775093909. URL: <https://pubmed.ncbi.nlm.nih.gov/16378516/>.
- [141] John M. Pearson et al. “Posterior Cingulate Cortex: Adapting Behavior to a Changing World”. In: *Trends in cognitive sciences* 15.4 (Apr. 2011), p. 143. ISSN: 13646613. DOI: 10.1016/J.TICS.2011.02.002. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC3070780/>.
- [142] Alex Pine et al. “Encoding of Marginal Utility across Time in the Human Brain”. In: *Journal of Neuroscience* 29 (30 July 2009), pp. 9575–9581. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.1126-09.2009. URL: <https://www.jneurosci.org/content/29/30/9575> <https://www.jneurosci.org/content/29/30/9575.abstract>.
- [143] Joseph J. Plaud, George A. Gaither, and Jason B. Lawrence. “Operant schedule transformations and human behavioral momentum”. In: *Journal of Behavior Therapy and Experimental Psychiatry* 28 (3 Sept. 1997), pp. 169–179. ISSN: 0005-7916. DOI: 10.1016/S0005-7916(97)00007-4.
- [144] Francesco Poli et al. “Contributions of expected learning progress and perceptual novelty to curiosity-driven exploration”. In: *Cognition* 225 (Aug. 2022), p. 105119. ISSN: 0010-0277. DOI: 10.1016/J.COGNITION.2022.105119.
- [145] José J.F. Ribas-Fernandes et al. “A Neural Signature of Hierarchical Reinforcement Learning”. In: *Neuron* 71.2 (July 2011), p. 370. ISSN: 08966273. DOI: 10.1016/J.NEURON.2011.05.042. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC3145918/>.

- [146] José J.F. Ribas-Fernandes et al. “Subgoal- and Goal-related Reward Prediction Errors in Medial Prefrontal Cortex”. In: *Journal of cognitive neuroscience* 31.1 (2019), pp. 8–23. ISSN: 1530-8898. DOI: 10.1162/JOCN{_}_A{_}_01341. URL: <https://pubmed.ncbi.nlm.nih.gov/30240308/>.
- [147] Robert D. Rogers and Stephen Monsell. “Costs of a Predictable Switch Between Simple Cognitive Tasks”. In: *Journal of Experimental Psychology: General* 124.2 (1995), pp. 207–231. ISSN: 00963445. DOI: 10.1037/0096-3445.124.2.207.
- [148] Edmund T. Rolls et al. “The connectivity of the human frontal pole cortex, and a theory of its involvement in exploit versus explore”. In: *Cerebral cortex (New York, N.Y. : 1991)* 34.1 (Jan. 2024). ISSN: 14602199. DOI: 10.1093/cercor/bhad416. URL: <https://pubmed.ncbi.nlm.nih.gov/37991264/>.
- [149] H. Enger Rosvold et al. “A continuous performance test of brain damage”. In: *Journal of Consulting Psychology* 20.5 (Oct. 1956), pp. 343–350. ISSN: 00958891. DOI: 10.1037/H0043220.
- [150] M. F.S. Rushworth et al. “The Effect of Cingulate Cortex Lesions on Task Switching and Working Memory”. In: *Journal of Cognitive Neuroscience* 15.3 (Apr. 2003), pp. 338–353. ISSN: 0898-929X. DOI: 10.1162/089892903321593072. URL: <https://dx.doi.org/10.1162/089892903321593072>.
- [151] Matthew F S Rushworth and Timothy E J Behrens. “Choice, uncertainty and value in prefrontal and cingulate cortex”. In: *NATURE NEUROSCIENCE VOLUME* 11.4 (Apr. 2008), pp. 389–397. DOI: 10.1038/nn2066. URL: <http://www.nature.com/natureneuroscience>.
- [152] Robb B. Rutledge et al. “A computational and neural model of momentary subjective well-being”. In: *Proceedings of the National Academy of Sciences of the United States of America* 111 (33 Aug. 2014), pp. 12252–12257. ISSN: 10916490. DOI: 10.1073/PNAS.1407535111/SUPPL_FILE/PNAS.201407535SI.PDF. URL: <https://www.pnas.org/doi/abs/10.1073/pnas.1407535111>.
- [153] Kazuyuki Samejima et al. “Neuroscience: Representation of action-specific reward values in the striatum”. In: *Science* 310.5752 (Nov. 2005), pp. 1337–1340. ISSN: 00368075. DOI: 10.1126/science.1115270. URL: [/doi/pdf/10.1126/science.1115270?download=true](https://doi/pdf/10.1126/science.1115270?download=true).
- [154] Paul A. Samuelson. “A note on measurement of utility”. In: *Review of Economic Studies* 4.2 (1937), pp. 155–161. ISSN: 1467937X. DOI: 10.2307/2967612.
- [155] Darryl W. Schneider and Gordon D. Logan. “Chunking away task-switch costs: a test of the chunk-point hypothesis”. In: *Psychonomic Bulletin and*

- Review* 22.3 (June 2015), pp. 884–889. ISSN: 15315320. DOI: 10.3758/S13423-014-0721-3.
- [156] Cognitive Sciences et al. “The adaptive value of stubborn goals”. In: *Trends in Cognitive Sciences* 0.0 (Oct. 2025). ISSN: 1364-6613. DOI: 10.1016/J.TICS.2025.10.003. URL: [https://www.cell.com/action/showFullText?pii=S1364661325002785%20https://www.cell.com/action/showAbstract?pii=S1364661325002785%20https://www.cell.com/trends/cognitive-sciences/abstract/S1364-6613\(25\)00278-5](https://www.cell.com/action/showFullText?pii=S1364661325002785%20https://www.cell.com/action/showAbstract?pii=S1364661325002785%20https://www.cell.com/trends/cognitive-sciences/abstract/S1364-6613(25)00278-5).
 - [157] Amitai Shenhav, Matthew M. Botvinick, and Jonathan D. Cohen. “The Expected Value of Control: An Integrative Theory of Anterior Cingulate Cortex Function”. In: *Neuron* 79.2 (July 2013), pp. 217–240. ISSN: 0896-6273. DOI: 10.1016/J.NEURON.2013.07.007.
 - [158] *Shifting intentional set: Exploring the dynamic control of tasks*. URL: <https://psycnet.apa.org/record/1994-98352-017>.
 - [159] Nikolay Sukhov et al. “When to Keep Trying and When to Let Go: Benchmarking Optimal Quitting”. In: (). DOI: 10.31234/OSF.IO/GJUCY. URL: <https://psyarxiv.com/gjucy/>.
 - [160] David Sussillo and Omri Barak. “Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks”. In: *Neural computation* 25.3 (2013), pp. 626–649. ISSN: 1530-888X. DOI: 10.1162/NECO{_}A{_}00409. URL: <https://pubmed.ncbi.nlm.nih.gov/23272922/>.
 - [161] Ilya Sutskever et al. *On the importance of initialization and momentum in deep learning*. May 2013. URL: <https://proceedings.mlr.press/v28/sutskever13.html>.
 - [162] Ilya Sutskever et al. *On the importance of initialization and momentum in deep learning*. May 2013. URL: <https://proceedings.mlr.press/v28/sutskever13.html>.
 - [163] Richard S. Sutton, Doina Precup, and Satinder Singh. “Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning”. In: *Artificial Intelligence* 112 (1-2 Aug. 1999), pp. 181–211. ISSN: 0004-3702. DOI: 10.1016/S0004-3702(99)00052-1.
 - [164] Richard S. Sutton, Doina Precup, and Satinder Singh. “Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning”. In: *Artificial Intelligence* 112.1-2 (Aug. 1999), pp. 181–211. ISSN: 0004-3702. DOI: 10.1016/S0004-3702(99)00052-1.
 - [165] Shosuke Suzuki et al. “Distinct regions of the striatum underlying effort, movement initiation and effort discounting”. In: *Nature Human Behaviour* 2020 5:3 5.3 (Nov. 2020), pp. 378–388. ISSN: 2397-3374. DOI: 10.1038/

- s41562-020-00972-y. URL: <https://www.nature.com/articles/s41562-020-00972-y>.
- [166] Brian M. Sweis et al. “Sensitivity to “sunk costs” in mice, rats, and humans”. In: *Science* 361 (6398 July 2018), pp. 178–181. ISSN: 10959203. DOI: 10.1126/SCIENCE.AAR8644/SUPPL_FILE/AAR8644_SWEIS_SM.PDF. URL: <https://www.science.org/doi/10.1126/science.aar8644>.
 - [167] Ching Yune C. Sylvester et al. “Switching attention and resolving interference: fMRI measures of executive functions”. In: *Neuropsychologia* 41.3 (Jan. 2003), pp. 357–370. ISSN: 0028-3932. DOI: 10.1016/S0028-3932(02)00167-7.
 - [168] Faraz Torabi, Garrett Warnell, and Peter Stone. “Behavioral Cloning from Observation”. In: *IJCAI International Joint Conference on Artificial Intelligence* 2018-July (May 2018), pp. 4950–4957. ISSN: 10450823. DOI: 10.24963/ijcai.2018/687. URL: <https://arxiv.org/abs/1805.01954v2>.
 - [169] Robert Louis Treuting et al. “Adaptive reinforcement learning is causally supported by anterior cingulate cortex and striatum”. In: *Neuron* 113.16 (Aug. 2025), pp. 2708–2723. ISSN: 10974199. DOI: 10.1016/J.NEURON.2025.05.018/ASSET/81913BB1-8F19-4A75-916D-581D86BB39B3/MAIN.ASSETS/GR7{_}LRG.JPG. URL: [https://www.cell.com/action/showFullText?pii=S0896627325003654%20https://www.cell.com/action/showAbstract?pii=S0896627325003654%20https://www.cell.com/neuron/abstract/S0896-6273\(25\)00365-4](https://www.cell.com/action/showFullText?pii=S0896627325003654%20https://www.cell.com/action/showAbstract?pii=S0896627325003654%20https://www.cell.com/neuron/abstract/S0896-6273(25)00365-4).
 - [170] John N Tsitsiklis and Benjamin Van Roy. “Analysis of Temporal-Difference Learning with Function Approximation”. In: *Advances in Neural Information Processing Systems* 9 (1996).
 - [171] Lucina Q. Uddin. “Cognitive and behavioural flexibility: neural mechanisms and clinical considerations”. In: *Nature Reviews Neuroscience* 2021 22:3 22.3 (Feb. 2021), pp. 167–179. ISSN: 1471-0048. DOI: 10.1038/s41583-021-00428-w. URL: <https://www.nature.com/articles/s41583-021-00428-w>.
 - [172] Wouter Van Den Bos et al. “Adolescent impatience decreases with increased frontostriatal connectivity”. In: *Proceedings of the National Academy of Sciences of the United States of America* 112.29 (July 2015), E3765–E3774. ISSN: 1091-6490. DOI: 10.1073/PNAS.1423095112. URL: <https://pubmed.ncbi.nlm.nih.gov/26100897/>.
 - [173] Amit Vinograd et al. “Causal evidence of a line attractor encoding an affective state”. In: *Nature* 2024 634:8035 634.8035 (Aug. 2024), pp. 910–918. ISSN: 1476-4687. DOI: 10.1038/s41586-024-07915-x. URL: <https://www.nature.com/articles/s41586-024-07915-x>.

- [174] Xiao Jing Wang. “Probabilistic Decision Making by Slow Reverberation in Cortical Circuits”. In: *Neuron* 36.5 (Dec. 2002), pp. 955–968. ISSN: 0896-6273. DOI: 10.1016/S0896-6273(02)01092-9.
- [175] Chentao Wen, Yukiko Ogura, and Toshiya Matsushima. “Striatal and Tegmental Neurons Code Critical Signals for Temporal-Difference Learning of State Value in Domestic Chicks”. In: *Frontiers in neuroscience* 10.NOV (2016). ISSN: 1662453X. DOI: 10.3389/fnins.2016.00476. URL: <https://pubmed.ncbi.nlm.nih.gov/27877100/>.
- [176] Andrew Westbrook, Daria Kester, and Todd S. Braver. “What is the subjective cost of cognitive effort? Load, trait, and aging effects revealed by economic preference”. In: *PloS one* 8.7 (July 2013). ISSN: 1932-6203. DOI: 10.1371/JOURNAL.PONE.0068210. URL: <https://pubmed.ncbi.nlm.nih.gov/23894295/>.
- [177] Robert C Wilson and Anne GE Collins. “Ten simple rules for the computational modeling of behavioral data”. In: *eLife* 8 (Nov. 2019). Ed. by Timothy E Behrens, e49547. ISSN: 2050-084X. DOI: 10.7554/eLife.49547. URL: <https://doi.org/10.7554/eLife.49547>.
- [178] Anderson M. Winkler et al. “Permutation inference for the general linear model”. In: *NeuroImage* 92.100 (May 2014), pp. 381–397. ISSN: 1095-9572. DOI: 10.1016/J.NEUROIMAGE.2014.01.060. URL: <https://pubmed.ncbi.nlm.nih.gov/24530839/>.
- [179] Ryan A. Wirt et al. “Temporal information in the anterior cingulate cortex relates to accumulated experiences”. In: *Current Biology* 34.13 (July 2024), pp. 2921–2931. ISSN: 0960-9822. DOI: 10.1016/J.CUB.2024.05.045.
- [180] Kong Fatt Wong and Xiao Jing Wang. “A recurrent network mechanism of time integration in perceptual decisions”. In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 26.4 (Jan. 2006), pp. 1314–1328. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.3733-05.2006. URL: <https://pubmed.ncbi.nlm.nih.gov/16436619/>.
- [181] Carsten Wrosch et al. “Adaptive Self-Regulation of Unattainable Goals: Goal Disengagement, Goal Reengagement, and Subjective Well-Being”. In: <http://dx.doi.org/10.1177/0146167203256921> 29 (12 Dec. 2003), pp. 1494–1508. ISSN: 01461672. DOI: 10.1177/0146167203256921. URL: https://journals.sagepub.com/doi/abs/10.1177/0146167203256921?casa_token=ygTSUI-vpDEAAAAA%3AnyASpnz877TqNcNbNGTQ6y5qNfPusg6wVC6mnnKbJs97IegofAtqy-lsISg.
- [182] Glenn Wylie and Alan Allport. “Task switching and the measurement of “switch costs””. In: *Psychological Research* 63.3-4 (2000), pp. 212–233. ISSN: 14302772. DOI: 10.1007/S004269900003/METRICS. URL: <https://link.springer.com/article/10.1007/s004269900003>.

- [183] Henry H. Yin and Barbara J. Knowlton. “The role of the basal ganglia in habit formation”. In: *Nature Reviews Neuroscience* 2006 7:6 7.6 (June 2006), pp. 464–476. ISSN: 1471-0048. DOI: 10.1038/nrn1919. URL: <https://www.nature.com/articles/nrn1919>.
- [184] Mengya Zhang and Qing Yu. “The representation of abstract goals in working memory is supported by task-congruent neural geometry”. In: *PLOS Biology* 22.12 (Dec. 2024), e3002461. ISSN: 1545-7885. DOI: 10.1371/JOURNAL.PBIO.3002461. URL: <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3002461>.