bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A weighted constraint satisfaction approach to human goal-directed decision making

Yuxuan Li[1*], James L. McClelland[1*]

1 Department of Psychology, Stanford University, Stanford, California, United States of America

* {liyuxuan, jlmcc}@stanford.edu

Abstract

When we choose actions aimed at achieving long-range goals, proximal information cannot be exploited in a blindly myopic way, as relevant future information must often be taken into account. However, when long-range information is irrelevant to achieving proximal subgoals, it can be desirable to focus exclusively on subgoal-relevant considerations. Here, we consider how an underlying parallel mechanism simultaneously influenced by proximal and future information may be at work when decision makers confront both types of situations. Participants were asked to find the shortest path in a simple maze where the optimal path depended on both starting-point and goal-proximal constraints. This simple task was then embedded in a more complex maze where the same two constraints, but not the final goal position, determined the optimal path to the subgoal. In both tasks, initial choice responses predominantly reflected the joint influence from relevant immediate and future constraints, yet we also found systematic deviations from optimality. We modeled initial path choice as an evidence integration process and found that participants weighted the starting-point more than the equally relevant goal in the simple task. In the complex task, there was no evidence of a separate processing stage where participants first zeroed in on the subgoal as would be expected if task decomposition occurred strictly prior to choosing a path to the subgoal. Participants again placed slightly more weight on the starting point than the subgoal as in the simple task, and also placing some weight on the irrelevant final goal. These results suggest that optimizing decision making can be viewed as adjusting the weighting of constraints toward values that favor relevant ones in a given task context, and that the dynamic re-weighting of constraints at different points in a decision process can allow an inherently parallel process to exhibit approximate emergent hierarchical structure.

Author Summary

Optimal approaches to achieving long-term goals often require considering relevant future information and, at other times, chunking a problem into subproblems that can be focused on one at a time. These two situations seemingly require separate modes of thinking. While simultaneous consideration allows proximal and future information to jointly guide our actions, tackling subgoals is often thought to require first coming up with a higher-level plan, then focusing on solving each subtask separately. In this study, we examine how both abilities might be explained by a shared mechanism. We conducted behavioral experiments and used computational modeling to understand how people weight various factors in choosing goal-reaching paths. We found that their weighting

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

of task-relevant factors allowed them to approximate optimal path choices, but they tend to place somewhat more weight on factors relevant to the immediate next action than on future considerations, and suboptimally place some weight on task-irrelevant factors. These results open up the space for considering the role of situation-dependent constraint weighting as a mechanism that allows people to integrate multiple pieces of information in decision making in a flexible, context-sensitive manner in service of optimizing performance in reaching an overall goal.

Introduction

A hallmark of human intelligence is our ability to conduct goal-directed behavior: our plans and actions are guided by long-term goals. For behavior to be effective towards achieving a goal, future information related to the goal often shapes many critical steps in our decision making. For example, when packing for an upcoming trip, the weather and our planned activities at the destination must be taken into account, even though the act of packing happens in an early stage in the global context of traveling. Studies have examined how initial-stage decision making already considers future choice points and incorporates whole-path or aggregated future information in a decision tree [1–3]. Backward reasoning stemming from known conditions about a goal is even sometimes the optimal strategy in figuring out how to achieve the goal [4].

Yet efficient planning is also marked by the ability to break a problem into smaller problems, achieving key subgoals before planning details toward the ultimate goal. In some cases, information regarding a final goal may be completely irrelevant to solving a subtask at hand. To come back to our traveling example, deciding which transportation to take to the departure airport is one subtask that can be independent to the final goal, once the flight and departure time has been settled. Indeed, studies have shown that humans engage in hierarchical planning in such situations, using learned knowledge to construct subtasks in novel problems and switching in and out of sequences of different subgoal contexts as they progress through them [5–7]. One early characterization of how people approach decomposing a problem is means-ends analysis [8], which emphasizes the application of relevant means to solve a sequence of intermediate subgoals to iteratively get closer to the final goal. More recent work provides precise computational accounts of how humans are able to optimally and efficiently discover hierarchical task structures [7,9–11]. The idea of temporarily focusing on subtasks before returning to the global task context is also at the core of hierarchical reinforcement learning, a computational framework capturing temporally abstract action plans in reward-based learning [12].

From an algorithmic perspective, humans can certainly approach solving a long-term goal by first identifying the key subgoals that must be achieved before the final goal, then devoting effort to tackle each subgoal. If the identified higher-level subgoal sequence is often close-to-optimal, this approach would naturally support efficiently solving the overall goal through the deployment of a series of lower-level, subtask decision processes. However, such a sequential mode of thinking differs from an overall parallel consideration of immediate as well as future constraints that is often beneficial in service of an overall goal and would seem to require different subprocesses to be responsible for breaking down the problem, focusing in on a subpart, then deciding the appropriate solution to each subproblem.

Here, we consider how an inherently parallel process can both account for the co-exploitation of relevant immediate and future constraints and also exhibit approximate adherence to the behavior of a more strictly sequential, hierarchically structured process. Indeed, many domains of human cognition that have often been viewed as inherently sequential, including reading, sequential action, and language production, need not rely on strictly serial processing [13–17]. Evidence of a degree of parallelism in these domains

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

comes from observations of both highly optimized, skilled behaviors as well as errors which seem to be deviations from optimality. For example, optimally typing the sequence of letters in a word can involve hand movements that prospectively prepare for future letters [17], increasing overall speed and fluency. People also produce speech errors reflecting intruding influences from words before or after the target word [13]. These behavioral patterns reveal interactions between what may otherwise be thought of as independent steps in a sequential process, rendering serially-staged characterizations of the cognitive process approximate rather than exact characterizations. This has led to the proposal that parallel processing mechanisms are the core form of computation in the brain [15]. Similar types of integrative and context-sensitive considerations, rather than modular and serial computations, may also play a role in value-based decision making [18].

We suggest that efficient goal-directed decision making can be supported by a parallel processing mechanism through the flexible weighting of multiple constraints or factors. Just as the optimal exploitation of relevant immediate and long-range information may require the simultaneous and approximately equally weighted consideration of multiple constraints, an approximation to focusing exclusively on a particular subtask can also arise out of a simultaneous and appropriately weighted consideration of subtask-relevant and subtask-irrelevant factors. Consider, for example, the goal-directed problem shown in Fig 1A. The task is simply to find the shortest path to the goal location, where the optimal path depends both on constraints proximal to the starting point (the starting position relative to the starting-point proximal wall) and on constraints proximal to the goal location (the goal position behind the goal-proximal wall). Although a sequential consideration of these two constraints is possible, the optimal approach would be to simultaneously consider and equally weight both constraints in decision making.

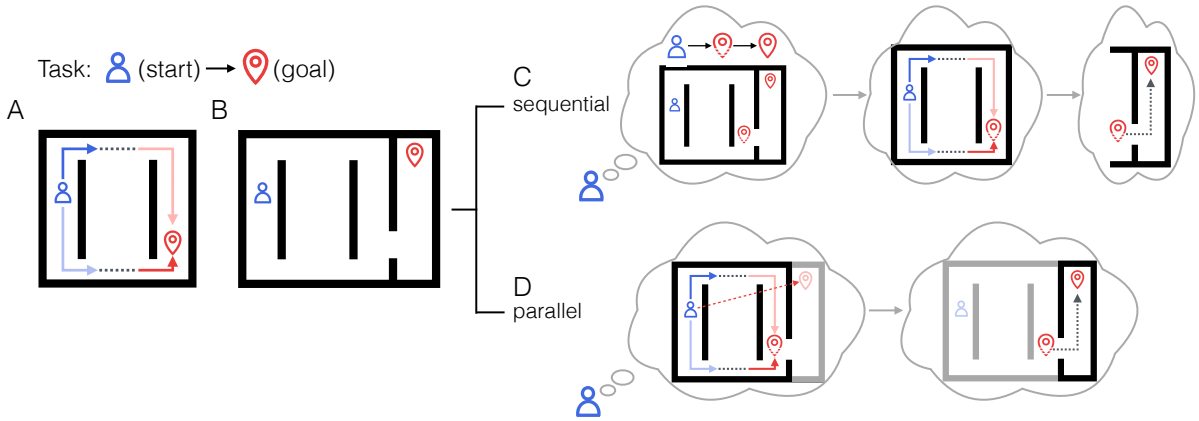

Fig 1. Sequential and parallel approaches to solving a goal hierarchy. A. Achieving the goal benefits from the joint consideration of immediate and future information. **B**. The same task in A. embedded in a larger task context leading to a critical subgoal location. **C**. A sequential approach to solving the goal hierarchy in B. by first decomposing the goal sequence, then exclusively focusing on achieving the subgoal before considering the final goal. **D**. Approximately exclusive focus on the subgoal can emerge in a parallel approach through the simultaneous but different weighting of subgoal-relevant and subgoal-irrelevant factors. However, this approach may allow suboptimal influence from out-of-subtask factors leading to deviations from strictly optimal decision making.

When the same task is embedded in a larger task context, such that what was the goal location before becomes a crucial subgoal location on the way to the final goal (Fig 1B), finding the optimal overall path still requires balanced consideration of the same two constraints (now associated with the starting point and the subgoal location). Thus, once the subgoal has been determined, an optimal decision maker should focus on the

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

subtask alone, regardless of the location of the final goal. In a sequential approach, the associated mental processes might be marked by an initial stage of explicitly processing the task structure and identifying the relevant subgoal, followed by an exclusive focus on only the two optimality-relevant constraints leading up to the subgoal (Fig 1C). But if parallel mechanisms would eventually be recruited to optimally process the two subgoal-relevant constraints, the same mechanism may also support the processing of the goal hierarchy in this case. For path selection to be close to optimal, the weighting of the various factors considered simultaneously must allow the two relevant constraints within the subtask context to exert a near equal and jointly predominant influence. But if weighting is a matter of degree, the irrelevant final goal may sometimes remain active enough to exert a subtle influence while the initial path is being chosen, perhaps biasing and/or degrading the process of choosing a path (Fig 1D).

To explore these issues, we studied how human participants approached solving the maze problems shown in Fig 2. We compared the decision processes and outcomes when the core two-constraint task was presented as an independent problem or embedded as a subtask that had to be accomplished on the way to a final goal. We tested if participants' decision making was more in accordance with a sequential approach or a parallel weighting approach, exploiting the idea that these different approaches in this context can be modeled using an evidence integration framework [19]. Specifically, a sequentially-staged approach would lead to an increase in the initial non-decision time to process the task hierarchy, followed by the very same integration of the two optimality-relevant constraints prior to the subgoal location when the subtask was presented alone. In a parallel approach, path choices may instead reflect influences from both the subgoal-relevant constraints and the subgoal-irrelevant final goal. We first confirmed that choosing among candidate goal-reaching paths indeed involved the joint consideration of both immediate and relevant long-range constraints, though the myopic constraint tended to receive somewhat greater weight (Experiment 1). When the same task appeared in a subtask context (Experiment 2), we found no delayed onset of the integration of the subgoal-relevant constraints, and decisions were subject to degradation and biasing influence from the irrelevant final goal, consistent with predictions from the parallel approach. In the *General Discussion*, we return to the broader question of whether a more parallel, constraint-satisfaction based approach might be the biological brain's way of approximating the optimal weighting of relevant constraints in different goal-directed task contexts, through the dynamic modulation of the weighting of constraints as thought and behavior unfold.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

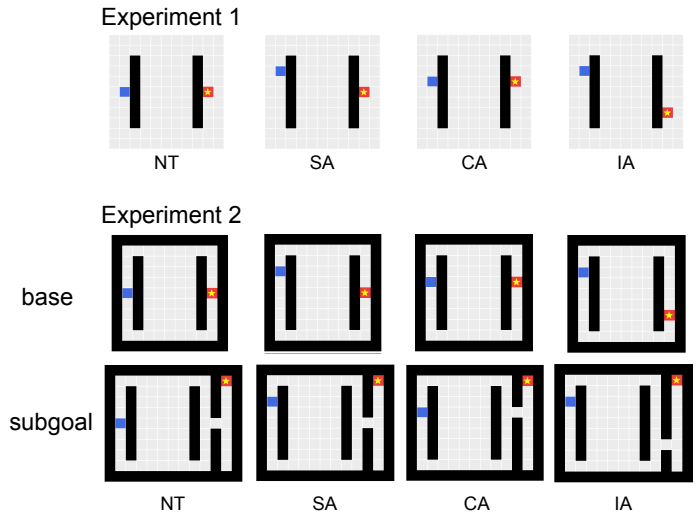

Fig 2. Experimental design. Blue block, starting point. Starred red block, goal location. Correspondence between the two optimality-relevant path advantages: NT=neutral advantage, SA=single advantage, CA=congruent advantage, IA=incongruent advantage.

Results

Experiment 1

We first examined how people take into account the immediate and future constraints in achieving a goal in a maze task where the starting-point proximal (*myopic*) advantage and the goal-proximal (*future*) advantage of candidate goal-reaching paths together determine the shortest path. We analyzed how these two path advantages shaped the decision in each trial, as revealed by the first action *choice* which determines the overall path toward the goal, and the time taken to take the first step (*response time*). We also jointly modeled the choice and response time in trials with directly competing myopic and future path advantages as a parallel, drift-diffusion process.

Methods

Ethics Statement. This study was approved by the Stanford University Institutional Review Board under protocol No.7029. Participants on Amazon Mechanical Turk gave written consent at the beginning of the experimental session.

Participants. The data from Experiment 1 comes from 100 US-based participants recruited on Amazon Mechanical Turk. To ensure data quality, each participant must have had over 92% HIT approval rate and must have completed more than 1000 approved HITs to be eligible for the study.

Task design. Participants completed a single session consisting of a brief consent, three practice trials, 184 experimental trials, and a short survey. Each trial consisted of a shortest path search task on an 11×11 grid canvas with two internal walls (Fig 2). Participants were instructed to move the blue block to the starred goal location using the minimum number of up, down, left, or right steps. A step into the walls would increase the step count but no actual movement. Participants received a base completion compensation of $1.00 and a performance-based bonus on each trial ($0.03 if a trial was solved with the minimum number of steps, $0.01 if the solution was only up to two steps more than the shortest solution, and no bonus otherwise).

The experimental trials consisted of 92 base trials and 92 filler trials. We report

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

results on the base trials in this paper. The starting location and the goal location varied across the base trials but the locations of the two length-7 walls were fixed. Each unique base trial layout was mirrored vertically, except for the neutral (NT) trial. Both the original and the mirrored trials also appeared in all four orientations, including the left-to-right orientation shown, as well as top-to-bottom, right-to-left, and bottom-to-top orientations. The filler trials randomly sampled trial orientation and block locations. The walls on the filler trials were of length- 3, 5, or 7 and were sometimes shifted from the center, but wall locations on the filler trials were never identical to the wall locations on the base trials. Each participant received the 184 experimental trials in a different randomized order.

Path advantages. The location of the starting block and the goal block were designed to vary the relative advantages of the candidate goal-reaching paths proximal to the starting location and proximal to the goal. The myopic advantage near the starting point refers to the side of the wall that the blue block can be moved toward to get out from behind the wall with fewer steps. Similarly, the future advantage near the goal refers to the side of the wall that the blue block can approach the goal from with fewer steps. Quantitatively, the myopic and the future advantages can be computed as the position offset of the starting block and the goal block relative to the center of the nearby wall. For example, in the IA trial shown in the top panel of Fig 2, the myopic advantage is 2 (towards the upper path), and the future advantage is -2 (towards the lower path). The pairing of these two advantages establish four advantage types: neutral advantage (NT), single advantage (SA), congruent advantage (CA), and incongruent advantage (IA) (Table 1). We sometimes refer to the SA and IA trials by whether the myopic or the future advantage was the larger advantage, denoted by the "-m" or "-f" tail. Note that in the NT trial layout and two of the IA trial layouts, the two global path candidates (equivalent to the upper and the lower paths in the left-to-right orientation) are equally optimal.

Adv. Type	Myopic Adv.	Future Adv.
NT	0	0
SA-m	2	0
	1	0
SA-f	0	2
	0	1
CA	1	1
IA	1	-1
	2	-2
IA-m	2	-1
	3	-1
IA-f	1	-2
	1	-3

Table 1. Myopic and future path advantage pairing. -m and -f indicates the larger advantage. All trial layouts were vertically mirrored except for the NT trial. NT=neutral advantage, SA=single advantage, CA=congruent advantage, IA=incongruent advantage.

Exclusions. We excluded data from five participants who were not able to complete the experiment due to technical reasons. Across all remaining 8740 observations (95 participants $\times$ 92 trials), we excluded trials where participants took longer than one minute to execute the first move (six trials) or took five or more steps compared to the longer canonical solution (ten trials; a canonical solution is equivalent to the upper or lower path in the left-to-right orientation, without excessive steps). We also excluded trials with ill-identified initial path direction, including trials where the first move

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

was equivalent to going left in the left-to-right orientation, or where the initial steps contradicted the overall path, e.g., an initial down action followed by a later upper path (31 trials).

Drift-diffusion modeling. We modeled the decision process in the IA trials as a noisy integration process, as these trials afford the opportunity to examine the influences of the two constraints when they each deviate from neutral and contribute to the decision outcome in opposite directions. We pooled data from all participants, first z-scoring the raw first move response time (in seconds) within each participant, then shifting the z-score distributions so that the minimum is 0.5 for each participant in order to ensure positive response time and enough non-decision time buffer for modeling purposes. All drift-diffusion model variants were fitted in a Monte-Carlo cross-validation procedure, using the density function implemented in the rtdists package [20] and the nlminb function in the stats package [21]. In each cross-validation fold, we held out data from 35 (out of 95) randomly sampled participants as the test data. Ten runs from random initial parameter values were optimized to minimize the summed negative log-likelihood (sNLL) of the training data, and the winning fit for the fold was selected based on the best training data objective. Candidate models were fitted over the same 200 cross-validation folds, to control for fold-level variability. For model selection, we fitted a linear mixed-effects model to the test objectives from all candidate model variants with fold-level random intercepts, and compared which model achieved lower or comparable test data objective with fewer free parameters.

For the winning model, we observed that some runs converged to what appeared to be a local minimum value of the fold training data sNLL with near-zero starting point variability (sz). For five of the cross-validation folds, all ten runs with different initial parameter values converged to this local minimum. In these cases, we ran 50 more runs each using another random sample of initial parameter values, in an effort to find a set that would avoid this local minimum. This was successful as the run with the smallest training sNLL across all 60 runs in all five cases had a larger sz value within the same range of values arising from the other 195 folds. The parameter estimates of the new best-fitting runs for these five folds were reported below.

Results and Discussion

In Experiment 1, our goal was to characterize the decision dynamics in a goal-directed task in which the optimal solution requires equal consideration of both immediate and long-term, goal-proximal information. We presented participants with a maze navigation task with different levels of congruent or competing advantages of the candidate goal-reaching paths, and asked how the immediate (*myopic*) and the goal-proximal (*future*) path advantages contributed to the decision leading to the selected path.

Not surprisingly, participants leveraged both the myopic and the future path advantages in deciding the shortest goal-reaching path. The selected initial direction was often optimal, but trials with congruent advantages or a single advantage led to more optimal and faster responses compared to trials with incongruent advantages: as shown in Fig 3, both path choices and response times were clearly sensitive to the level of advantage congruence, even when holding the total path difference (in steps) between the candidate paths constant. Indeed, path selection differed significantly across the nine unique advantage pairings where one of the initial directions was optimal (shown in Fig 3B), as confirmed by a mixed-effects probit regression model of trial-level path choices (with separate slopes for each advantage pairing and participant-level random intercepts), $\chi^2(8) = 301.76$, $p<0.001$. Response times (shown in Fig 3C) also differed significantly among all 12 advantage pairings based on a linear model of the medians of individual z-scored first move response times, $F(11, 1128)=22.80$, $p<0.001$. Detailed comparisons of the estimated marginal means (EMMs; with Bonferroni correction) of

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the choice model and the response time model (controlling for total path differences) confirmed that responses in the IA trials were less optimal and slower than those in the SA and CA trials (adjusted $ps<0.01$), although choice optimality rates in SA and CA trials were both near-ceiling and the difference between the associated response times was not statistically significant (adjusted $ps>0.1$ for both comparisons).

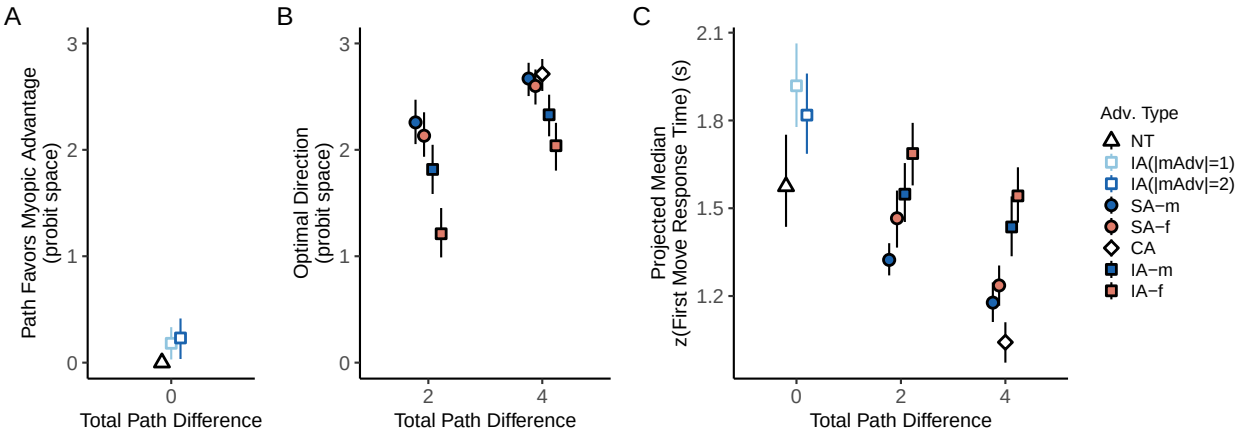

Fig 3. Path choices and response times reflected integral considerations of immediate and goal-proximal path advantages. A. Path choices in trials with equally optimal initial directions. The value for NT trials was defined to be 0 since there was neither a myopic or a future advantage on these trials. **B.** Path choices in trials where one of the initial directions was optimal. The proportion of optimal trials for each individual was converted into a probit score. If the individual probit score was larger than 3 or smaller than -3, it was capped at 3 or -3 before averaging. **C.** First move response times across all advantage pairings. The individual medians of zscored response times were projected back to the raw time space using group average mean response times and group average standard deviations. NT=neutral advantage, SA=single advantage, CA=congruent advantage, IA=incongruent advantage. mAdv=myopic advantage. -m and -f indicates the larger advantage. Error bars indicate bootstrapped 95% confidence limits.

Although facilitation from advantage congruence suggested that participants took both advantages into account in an integrative manner, responses from trials where the myopic advantage was larger compared to the counterpart trials (where the future advantage was larger) also indicated that the two constraints were not treated equally. Trials where the myopic advantage was larger resulted in significantly higher choice optimality rate (adjusted $p<0.01$, EMM comparison from the choice model) and shorter response times (adjusted $p<0.05$, from the response time model) in IA and SA trials combined. This myopic bias was seen not only at the group level (Fig 3), as the pattern was observed in participants who made overall more accurate path choices and those who made overall less accurate path choices (Fig S1). The overall myopic bias is exhibited by a majority of the individual participants, though about 30% showed the opposite tendency when the two advantages are of equal magnitude.

To understand the different role of the two path advantages in leading to the decision outcome, we next turned to the IA trials in which the two initial path choices each satisfy one of the competing advantages, using several variants of a drift-diffusion process model to account jointly for the choice probabilities and response times from these trials. This analysis suggested that the tendency to favor the path with the myopic advantage arose from a stronger weighting of the myopic advantage compared to the future advantage (Fig 4A). Specifically, we considered whether a biased starting point or different weighting of the two advantages served as the source of the myopic bias. We compared 16 candidate models with different types of inter-trial variabilities and fit pooled choice and first move response time data from all participants across 200 cross-validation folds. The

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

model with different weighting for the two advantages was better than the model with equal weighting and the model with both equal weighting and a starting bias (adjusted p's<0.001, pairwise EMM comparison based on a linear mixed-effects model of the test objectives across all models, Bonferroni corrected). Accounting for additional starting bias led to worse fit in the equal weighting case (adjusted p <0.05) and did not improve fit in the unequal weighting case (adjusted p >0.5). Models with inter-trial variabilities in both the starting point and the drift rate also consistently out-performed the others (adjusted p's<0.001).

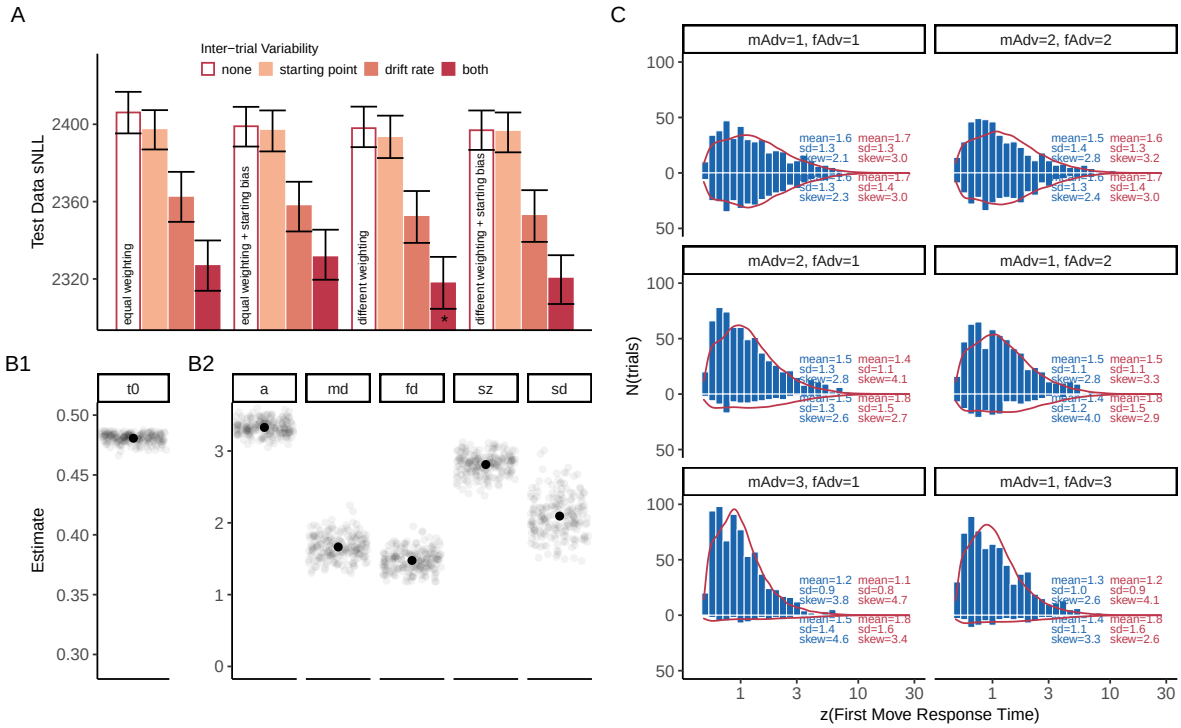

Fig 4. Simultaneous integration of myopic and future path advantages is marked by a slightly strengthened weighting of the myopic advantage. A. Test data objective (summed negative log likelihood; sNLL) of candidate drift-diffusion models across matched cross-validation folds. Asterisk marks the winning model. **B.** Parameter estimates of the winning drift-diffusion model. *t0*, non-decision time. *a*, decision bound. *md* and *fd*, myopic and future advantage weights. *sz*, inter-trial variability of the starting point. *sd*, inter-trial variability of the drift rate. **C.** Predicted response time (RT) distributions (in red) sampled from the parameter estimates of the winning fit in the fold with the best test objective, and the empirical RT distributions (in blue). The top two subpanels show the RT distributions associated with choices favoring the myopic advantage on the top and those favoring the future advantage on the bottom. The bottom four subpanels show the RT distributions for optimal responses on the top and suboptimal responses on the bottom. mAdv, myopic advantage. fAdv, future advantage. All error bars indicate 95% bootstrapped confidence limits.

The winning model estimated a ratio between the weights associated with the myopic and the future advantage at 1.13 (SD=0.02, see Fig 4B). The model fitted the combined choice and response time data fairly well, though it under-predicted fast correct responses and over-predicted the occurrence of errors, particularly when the advantage difference was plus or minus one (middle panels, Fig 4C). Fitting the model separately to data from participants with higher overall choice accuracy revealed that more equal weighting and larger weights assigned to the two advantages underlay more accurate path choices (weight ratio estimated at 1.09, SD=0.02; see Fig S2 top panel). For the group with

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

lower overall choice accuracy, the model fit suggested a larger weighting imbalance (ratio estimated at 1.19, SD=0.05) and smaller advantage weights associated with the two advantages, along with a higher drift rate variability (Fig S2 bottom panel).

Experiment 2

Experiment 1 showed that when achieving a goal, people indeed simultaneously process relevant immediate and goal-proximal, future information, albeit showing a slight myopic bias. Our main focus in Experiment 2 is to investigate how people approach the same task when embedded as a subtask in a larger problem. Experiment 2 included the same set of trials used in Experiment 1 (*base* trials) and an additional set of trials where the goal location was placed inside a chamber that participants had to reach by way of a bottleneck, creating a subgoal that must be visited prior to reaching the final goal (*subgoal* trials; see Fig 2).

We were interested in whether people approached this task setup with a strictly sequential approach, first extracting the key subgoal and then focusing exclusively on it, or with a parallel approach, as might be revealed through a graded influence of the final goal position on the decision, even when the final goal is irrelevant to optimal path selection. Building on the characterization of the decision process in the base task in Experiment 1, we examined whether the decision process in the subgoal trials was unaffected beyond a possible increase in non-decision time prior to engagement with the two relevant advantages, or whether the integration of the relevant path advantages was biased with contribution from the position of the optimality-irrelevant final goal. Experiment 2 was pre-registered through the Open Science Framework (https://osf.io/w78hu).

Methods

Ethics Statement. This study was approved by the Stanford University Institutional Review Board under protocol No.7029. Participants on Amazon Mechanical Turk gave written consent at the beginning of the experimental session.

Participants. We recruited 100 participants on Amazon Mechanical Turk with an identical set of eligibility criteria from Experiment 1 (see Experiment 1 methods), except that participants who previously participated in Experiment 1 were not eligible.

Task design. Participants completed two practice trials and 158 experimental trials in one session. Trial layouts were similar to that in Experiment 1, but the grid size was either 11×11 or 11×13 depending on the trial type (Fig 2). We added boundary walls to the grid canvas and removed the step count penalty for movements that resulted in wall collision. Participants were randomly assigned to one of the two orientation groups: group one (N=50) received trials with left-to-right and right-to-left orientations, group two (N=50) received trials with bottom-to-top and top-to-bottom orientations. Participants received a base completion compensation of $1.20 and a performance-based bonus on each trial ($0.03 for executing the shortest solution, $0.01 for a solution up to two steps longer than the shortest solution, and no bonus otherwise).

The experimental trials consisted of base trials (× 46), subgoal trials (× 92), multi-subgoal trials (× 16), and multi-subgoal control trials (× 4). In this paper, we report data from the base and the subgoal trials. The base trials contained the full set of 12 unique trial layouts used in Experiment 1 (see Table 1). The subgoal trials used the same set of advantage pairings but replaced the goal block in the base trials with a cell leading to a bottleneck (Fig 2). In the subgoal trials, the final goal block appeared on both ends of the goal column. Similar to Experiment 1, all trials were mirrored vertically except for the NT trials.

The multi-subgoal trials have two subgoal openings on the wall prior to the final goal that were structured so that the candidate paths through the two subgoals were

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

equally optimal. On these trials, the start and the goal locations were fixed, but the subgoal locations varied and one of the subgoals was always closer to the goal. The multi-subgoal control trials were introduced where we masked the subgoal further from the goal. The layout of these control trials did not overlap with any subgoal trials.

Exclusions. All 100 participants successfully completed the experiment. At the trial level, we implemented the same set of exclusion criteria used in Experiment 1, excluding trials where participants took more than one minute to execute a first move (19 trials) or solved with five steps or more than the longer canonical solution (ten trials), as well as trials with ill-identified initial path direction (180 trials). The exclusions resulted in a total of 13595 trial observations (98.5% of original dataset) for the reported analyses.

Drift diffusion modeling. We tested whether, compared to the base trials, the path choices and response times associated with the first move in the subgoal trials can be accounted for by a longer non-decision time (as might be predicted if participants first zeroed in on the subgoal and then treated it as an isolated problem) or if the responses reflected changes to the weights of two relevant constraints plus some influence from the irrelevant final goal (as might be expected if participants engaged in a single weighted constraint process). The base model capturing the decision process in the base trials was the winning model from Experiment 1. Response variables, model fitting, and model comparison were all handled identically compared to Experiment 1 (see Experiment 1 methods), except that we used a 60/40 split for sampling the training data and the test data in the cross-validation folds.

Results and Discussion

In Experiment 2, We probed how participants approach a problem with a goal hierarchy, in a situation where a key subgoal must be achieved first and the final goal information is irrelevant to optimal initial path choice. We used drift-diffusion models to test whether the decision process reflected an additional first stage of establishing the goal hierarchy before solving the subgoal through the same initial path selection process that would have occurred without the embedded structure, or whether instead the subgoal selection process itself was affected by the final goal position.

Overall, there was both an accuracy cost and a response time cost in the subgoal trials compared to the base trials, suggesting that performance degraded when the same task was embedded in a larger task context with irrelevant final goal information outside of the core subtask context. Participants selected the optimal direction more often in the base trials (mean optimal rate 92.30%, range: 52.78%–100.0%) than in the subgoal trials (mean optimal rate 88.03%, range: 51.43%–100.0%), $t(99) = 6.61$, $p < 0.001$. The group-average median response time of the first move in the subgoal trials (mean: 1.91 sec, range: 0.55 sec–9.45 sec) was about 0.40 sec longer than that in the base trials (mean: 1.53 sec, range: 0.58 sec–6.16 sec), $t(99) = 5.86$, $p < 0.001$. The accuracy and response time costs were seen across different trial advantage types (Fig 5), and also separately in groups of different overall accuracy (Fig S3).

The increased first move response times may reflect time taken to process the final goal information as irrelevant and focus in on the subgoal, but the simultaneous change in path choice accuracy suggested that participants did not deploy the exact same computation to solve the core two-constraint problem that is identical across both trial types. However, the benefit from congruent or single optimality-relevant constraints (as opposed to incongruent ones) remained as path choices and response times showed similar patterns across the different advantage pairings, suggesting that there was still some shared treatment of the optimality-relevant path advantages across both trial types (Fig 5). This is confirmed by extending the path choice model and the median zscore response time model (used in Experiment 1) with trial type (base vs. subgoal) as a second predictor and accounting for interaction effects. We observed no significant interaction

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

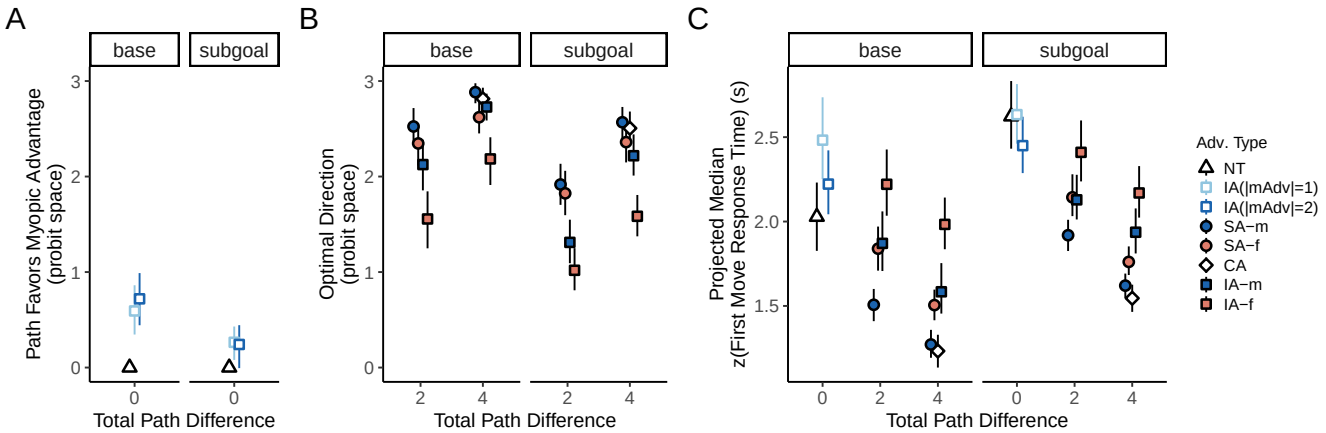

Fig 5. Accuracy and time cost when faced with a subgoal task embedded in a goal hierarchy. A. Path selection in the subgoal trials with equally optimal initial direction showed less myopic bias. As in Fig 3A, the value for NT trials was defined to be 0. **B.** Path selection across advantage pairings was less optimal in the subgoal trials, but sensitivity to the advantage pairings was preserved. As in Fig 3B, the proportion of optimal trials for each individual was converted into a probit score, with individual probit scores larger than 3 or smaller than -3 capped at 3 or -3 before averaging. **C.** Response times were overall longer in the subgoal trials. As in Fig 3C, the median zscore response times were projected back to the second's space.

in either the choice model ($\chi^2(8) = 8.19$, p=0.415, main-effects model as compared to an interaction model) or the response time model ($F(11, 2376)$=1.23, p=0.26).

Again using data from the IA trials, we next tested whether the decision process in the subgoal trials was accounted for by a lengthened non-decision time, as predicted by an initial stage of processing the goal hierarchy, or by changes to the advantage weights, as would be responsible for changes in decision outcome. As shown in Fig 6B, modeling a separate non-decision time for the subgoal trials led to very little change in data likelihood, indicating that the process of deciding between the path choices was initiated after about equal time in both trial types. Instead, we found that a proportional decrease in the weights associated with the optimality-relevant path advantages accounted for both the decreased choice optimality and slowed response time in the subgoal trials (adjusted $p < 0.001$, pairwise EMM comparisons based on a linear mixed-effects model fit to the test objectives from all candidate models, Bonferroni corrected). A proportional change accounted for the data better than an equal decrement in the two decision weights, and accounted for the data as well as two independent weight changes with one less free parameter (see *Supporting Information*). Modeling an additional separate non-decision time in the subgoal trials for the proportional weight decrease model also did not lead to significantly better test data likelihood (adjusted $p = 0.16$).

The winning model estimated about a 30% decrease to both advantage weights in the subgoal trials, while the ratio between the advantage weights (1.22, SD=0.03) was preserved (Fig 6C). This echoed the similarity in the response patterns across advantage pairings shown in Fig 5, suggesting that the decreased choice optimality and lengthened response times were a result of overall decision degradation, rather than more particular changes in the treatment of the two optimality-relevant path advantages. Interestingly, we found similar degrees of degradation in advantage weighting (∼30%) when fitting the model separately to data from the group with overall higher accuracy and the group with overall lower accuracy (Fig S5). Instead, the lower accuracy group showed reduced myopic and future advantage weights and higher degrees of drift-rate variability, consistent with that found in Experiment 1.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

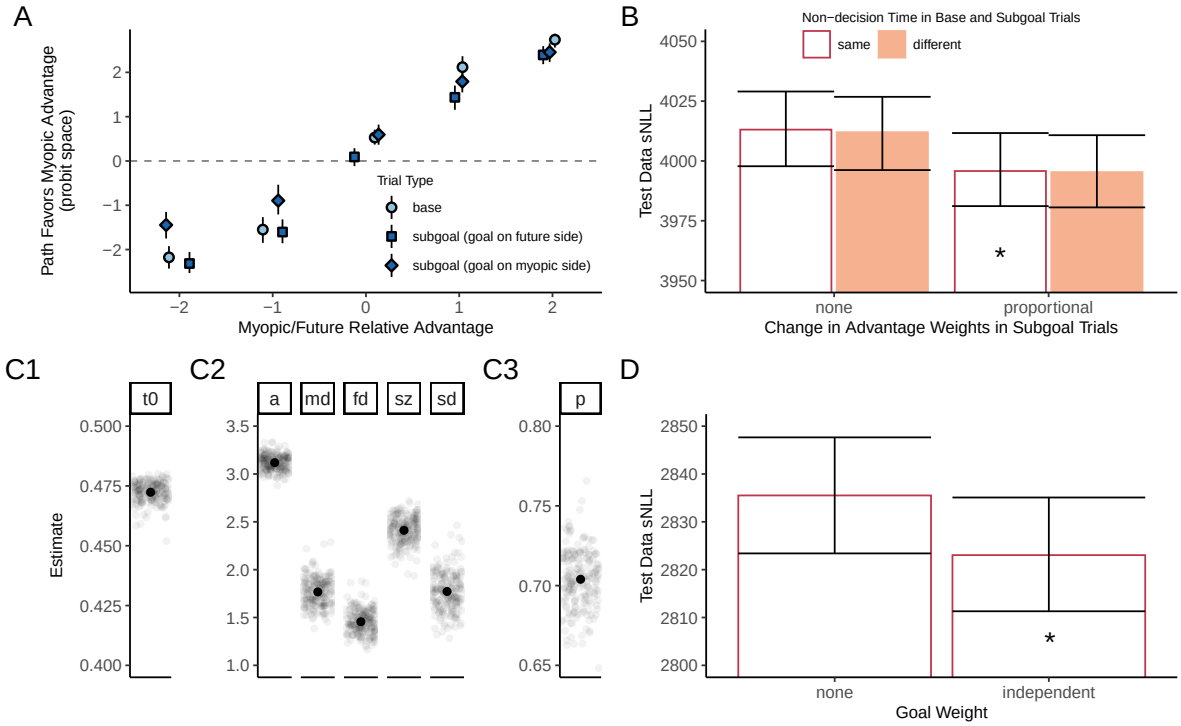

Fig 6. Decision making in the presence of a goal hierarchy showed degraded weighting of subgoal-relevant constraints and biasing influence from the irrelevant final goal, with no delayed onset of evidence integration. A. Path choices in the IA trials. As in Fig 3 and Fig 5, the proportion of optimal trials for each individual was converted into a probit score, with individual probit scores larger than 3 or smaller than -3 capped at 3 or -3 before averaging. **B.** Test data objective (summed negative log likelihood; sNLL) of the candidate drift-diffusion models. For empirical and predicted response time distributions, see Fig S4. **C.** Parameter estimates from the winning model. *t0*, non-decision time. *a*, decision bound. *md* and *fd*, myopic and future advantage weights. *sz*, inter-trial variability of the starting point. *sd*, inter-trial variability of the drift rate. *p*, shared proportional change to *md* and *fd* in the subgoal trials. **D.** Test data objective (summed negative log likelihood; sNLL) of the subgoal trials. Adding an independent weight for the final goal led to lower sNLL compared to a model with no final goal weight included. For empirical and predicted response time distributions, see Fig S6. Asterisks mark the winning models in B. and D.

Importantly, path choices were subject to biasing influence from the final goal even though this information was strictly optimality-irrelevant and outside of the core subtask context, indicating that participants did not completely cache away the final goal in choosing the path to the subgoal (Fig 6A). To understand the role of the final goal in biasing decision outcome, we modeled the final goal as a third source of influence on the decision preference in the subgoal trials. For each cross-validation fold, we froze the parameter estimates from the fit of the winning model and modeled the goal weight as an additional free parameter. Because the upper decision threshold was modeled as selecting the path that favored the myopic advantage, the goal weight encouraged reaching the upper threshold when the goal was on the myopic advantage side, and shifted the integration toward the lower threshold when the goal was on the future advantage side.

Indeed, modeling the final goal as a third source of constraint confirmed that participants engaged the final goal information in decision computation, as this further helped accounting for changes in both path choices and response times in the subgoal IA trials (Fig 6D, also see *Supporting Information*). The weight associated with the

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

irrelevant final goal was estimated at 0.30 (SD=0.05) in the winning model, a much smaller weight compared to the weights associated with the optimality-relevant path advantages (shown in Fig 6C). Notably, the influence of the final goal was estimated at markedly different levels across groups of different accuracy levels, even though the degree of overall degradation in advantage weighting was highly similar (Fig S7). For the group with higher overall accuracy, the goal carried a reliable, but very small weight (M=0.11, SD=0.04). For the group with lower overall accuracy, however, the goal weight was estimated at 0.52 (SD=0.07).

General Discussion

In this work, we have explored a parallel constraint satisfaction approach to goal-directed decision making. Within this approach, we investigated how humans weight optimality-relevant as well as optimality-irrelevant information as they decide which path to take toward a goal. When confronted with our simpler, base task, participants approximated an equal weighting of the starting-point proximal and goal-proximal constraints that are equally relevant to optimal path selection, and participants who made more accurate choices overall exhibited more nearly equal weighting. Yet participants in both accuracy groups tended to place greater weight on the starting-point proximal constraint, exhibiting a myopic bias. The base task then appeared as a subtask in a setting where reaching the equivalent of the base task goal location was embedded in the overall task of finding the path to the final goal. When confronted with this more complex task setting, participants' choices of paths to the subgoal were slowed, reduced in accuracy, and biased by the final goal location. However, these influences were relatively subtle, allowing the optimality-relevant factors to each receive more weight than the final goal that is irrelevant to optimal path selection. Participants who made more accurate choices overall placed greater and more equal weight on the subtask-relevant constraints. Yet even for these participants, we still observed a slight bias to consider the position of the irrelevant final goal in selecting the initial path. In both the simple and the more complex task settings, then, participants generally adopted appropriate weightings, but also exhibited some degree of deviation from what would be strictly optimal.

Based on these results, we suggest that parallel constraint satisfaction can serve as a framework for implementing different levels of approximation to hierarchical planning approaches in complex task settings. Rather than assuming that problem decomposition and decision making are executed by different subprocesses or occur in separate stages, our results suggest a role for parallel constraint weighting to both underlie the computation of the decision outcome and give rise to a spotlight focus on important subtasks through the strengthened weighting of relevant constraints and reduced weighting of irrelevant ones. In our tasks, participants indeed showed situation-sensitive flexibility in the weighting of the various factors when confronted with different task situations, as they exhibited nearly balanced consideration of relevant immediate and future constraints, and appropriately placed greater weight on subtask-relevant than subtask-irrelevant information. Moreover, we found no difference in the initiation time of the decision processes across the simple task and the more complex task, consistent with the idea that the extraction of the task structure and decision computation can be supported simultaneously through the weighting of the various factors and so do not necessarily follow strictly separate, consecutive stages.

Parallel consideration of multiple constraints, especially including some that are not relevant in a particular task context, can lead to suboptimal decision outcome when considered under the constraints of a specific situation. Such a tendency, however, might be viewed as a consequence of a more global, perhaps evolutionary, adaptation. Studies of optimal problem decomposition and decision suboptimalities have suggested that both

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

can result from optimizing meta-level planning, taking into consideration the cost and efficiency of representation and computation [6,7,9,11,22,23]. Strictly serially-staged and modular approaches to solving a task hierarchy may require first analyzing the problem structure, then caching away the final goal or the global task context and deploying an encapsulated subroutine to solve a subtask, which may incur an overall larger meta-computational cost compared to the parallel and appropriately-weighted processing of all exploitable factors that result in good-enough, approximate solutions. Concurrent problem decomposition and problem solving also affords the flexibility for alternative ways of decomposing the task to emerge at decision time, such that participants may not be prematurely locked into a suboptimal task decomposition, consistent with the suggestion that parallel mechanisms serve as a feature of the inherent flexibility of a cognitive system [15, 24]. Similarly, the myopic bias that is suboptimal in our task context may reflect a rational allocation of attention to the greater certainty of the near than the distant future or the tendency to plan only partially as a result of the sensitivity to meta-level planning costs [22, 23]. The suboptimal consideration of the irrelevant final goal can also be seen as a rational or at least natural exploitation for its possible computational values, considering that future-relevant considerations can often be relevant to the choices we make in the early stages of complex task situations.

Our results echo the simultaneous consideration of immediate and future constraints found in other cognitive domains (e.g., sequential action selection in typing and speech production) where it has often seemed natural to suppose that the processing of successive items is strictly sequential. As similarly shown in our results, such simultaneous consideration is revealed through both optimal and suboptimal aspects of detailed behavior in the corresponding task settings [13, 16]. The models used to account for this behavior generally weight considerations relevant to the immediate next action most heavily, with successively less weight to items more remote in the sequence. Indeed, we may understand the biases we have observed in our participants' behavior in terms of such a general tendency, in that we place the largest weight on the constraints most relevant to the immediate action at hand (accounting for the myopic bias) while allocating successively decreasing weight to future constraints (accounting for the reduced weight to the relevant future constraint and even less weight of the irrelevant final goal). Optimizing goal-directed decision making, then, can be thought of as eliciting a context-sensitive re-weighting of the constraints from such baseline weighting toward their optimal values in a given task situation.

Importantly, the finding that people's actions early in an inherently sequential process reflect future as well as immediate constraints may be relevant to planning and problem solving beyond the navigational or other everyday task domains we have considered up to this point. One such example is mathematical theorem proving, in which successful reasoning depends on constraining search for a proof based both on the givens in a problem and the statement-to-be-proven at the end of the proof sequence [25]. We may therefore expect parallel mechanisms to also underlie more advanced forms of reasoning, with immediate and long-range information (e.g., known conditions and the goal statement) simultaneously used to construct the right path to achieve the overall goal (e.g., prove the theorem) and facilitate the finding of important intermediate steps. Within the global parallel constraint satisfaction process to solve a complex problem, relatively strengthened weighting of the right subset of constraints can then allow a temporary spotlight focus on important subgoals (e.g., proving a lemma).

It is important to note that some aspects of our findings may depend at least in part on particular features of the experimental design. The interleaved nature of the base and subgoal trials in our second experiment may have contributed to some of the details of the performance we observed in the subgoal trials. On one hand, the repeated display of similar problems across the two task conditions may have encouraged the realization that

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

essentially the same two constraints were the only ones relevant to selecting the shortest path in both task settings. This may have contributed to the equal non-decision time cost found across both trial types. On the other hand, the final goal on the base trials had a direct computational value because the future path advantage was defined around it. Participants may therefore have learned to place considerable weight on the final goal on the base trials, possibly increasing attention towards the final goal on the subgoal trials. Another aspect of the current study is the visual accessibility of the problem space and the consistency of the placement of the start and goal along opposite sides of the environment. Finally, the visual salience of the goal block may have strengthened its biasing influence when irrelevant.

An important next step will be to examine whether similar parallel weighting mechanisms would also support the approximation to optimal representation and computation when humans solve a wider range of structured problems, such as tasks that require a larger number of intermediate steps, tasks in which optimally decomposing a task hierarchy requires learning or relies on information stored in memory, or tasks requiring more abstract forms of reasoning. As the problem becomes more abstract and its depth and complexity grows, we may expect a greater degree of approximately sequential processing, using shifts in the weighting of various factors to alternate between more focused and more global considerations as we tackle more complex task settings.

Another important future direction is to consider the relation between the dynamic weighting of explicit constraints in the problem space and alternative approaches to capturing goal-directedness in humans, such as the reinforcement learning framework for human action planning and reward maximization [26], in which hierarchical representation and computation also play a key role [12, 27]. To this end, the unification of deep neural networks, a prominent parallel processing model, and reinforcement learning algorithms, both having theoretical motivations rooted in psychology and neuroscience, have enabled the successful learning of sophisticated planning and human-like behavior in complex tasks [28, 29]. There is thus an exciting potential to use reinforcement learning as a means for learning how to optimize parallel weighting mechanisms as an implementation of human sequential planning and decision making [30].

Conclusion

In summary, we propose that goal-directed decision making can be understood in a weighted constraint satisfaction framework. In our experiments, simultaneous, weighted consideration of graded constraints helped account for the detailed choices and response times from human participants, both when a simple maze navigation task was presented alone and when it was nested in a larger maze. Participants adapted their weighting of relevant proximal and future constraints when faced with different task settings in the direction of optimizing their choice of paths toward a goal or subgoal, but deviations from optimality revealed gradations in the weighting of both optimality-relevant and optimality-irrelevant factors that biased decision outcomes. These results support the view that optimization in human choice behavior results from a context-sensitive weighting of multiple constraints to facilitate co-exploitation of immediate and long-range information as well as to allow an approximation to optimal focus on important subtasks.

Supporting Information

Experiment 2 model variants. We considered two additional model variants as potential forms of degraded integration, modeling independent additive changes to each advantage weight or a shared additive change to the two advantage weights. We added

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the proportional weight change model as a post-hoc variant after observing that the independent change model was better than the shared change model (adjusted $p < 0.001$). The proportional change model was significantly better than the shared change model (adjusted $p < 0.001$) and on par with the independent change model (adjusted $p > 0.9$) with one less free parameter.

We also considered variants of the final goal weight during evidence integration in the subgoal trials. Modeling an independent weight for the final goal regardless of its position resulted in better test data likelihood compared to two other models we tested: the model with different additive weights depending on which side (the myopic or future advantage side) the goal appeared was not better than the independent weight model (adjusted $p = 0.20$, pairwise EMM comparisons with Bonferroni correction), and the model in which the goal weight was dependent on the advantage weights was worse (adjusted $p < .001$).

Analysis of individual difference. We conducted individual difference analyses to investigate how the decision processes differed among participants with different levels of overall accuracy. To do this, we first split all participants into two groups based on the group median path choice accuracy, where individual accuracy scores were computed on all trials with a unique optimal initial direction (see Fig 3B). For Experiment 1, this resulted in N=51 in the higher accuracy group and N=44 in the lower accuracy group. For Experiment 2, this resulted in N=50 in each of the higher and lower accuracy group. The individual difference analyses were handled identically compared to the group analyses (see *Methods*), but the test data were re-sampled within the subset of participants in each group. For Experiment 1's higher accuracy group, we used a 36/15 train/test split. For Experiment 1's lower accuracy group, we used a 31/13 train/test split. For both groups in Experiment 2, we used a 35/15 train/test split.

For data from Experiment 1's higher accuracy group, we used 50 independent runs (instead of 10) for each fold in order to robustly find large *sz*'s, as in practice a large *sz* always resulted in better data likelihood compared to a near-zero *sz*. Similar to the group analysis (see Experiment 1 *Methods*), we conducted additional rounds of model fitting (50 runs at a time) for the winning model in a few folds that converged to local minima with near-zero sz's (eight folds for Experiment 1's higher accuracy group, three folds for Experiment 1's lower accuracy group), and plotted the parameter estimates from the best-fitting run among these additional runs. As in Fig 4C, we used parameter estimates from the fold-level model with the best test objective to show the model predicted response time distribution against the empirical distribution, but all model comparison and parameter estimate analyses were based on the full range of parameter estimates across the 200 cross-validation folds.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

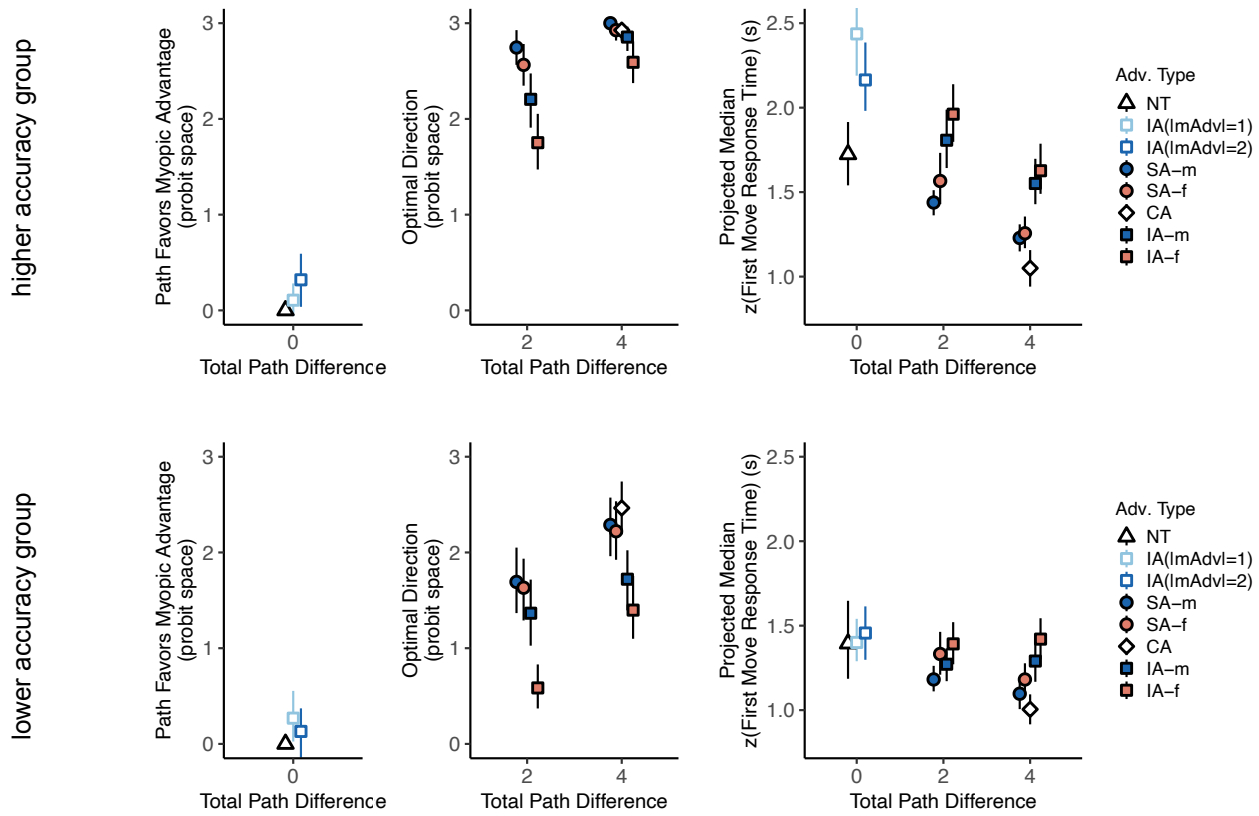

Fig S1. Experiment 1 path choices and response times for groups of higher and lower accuracy. Visualization and notation as in Fig 3. The individual median zscore response times were projected to the second's space using the subgroup mean mean response time and mean standard deviation.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

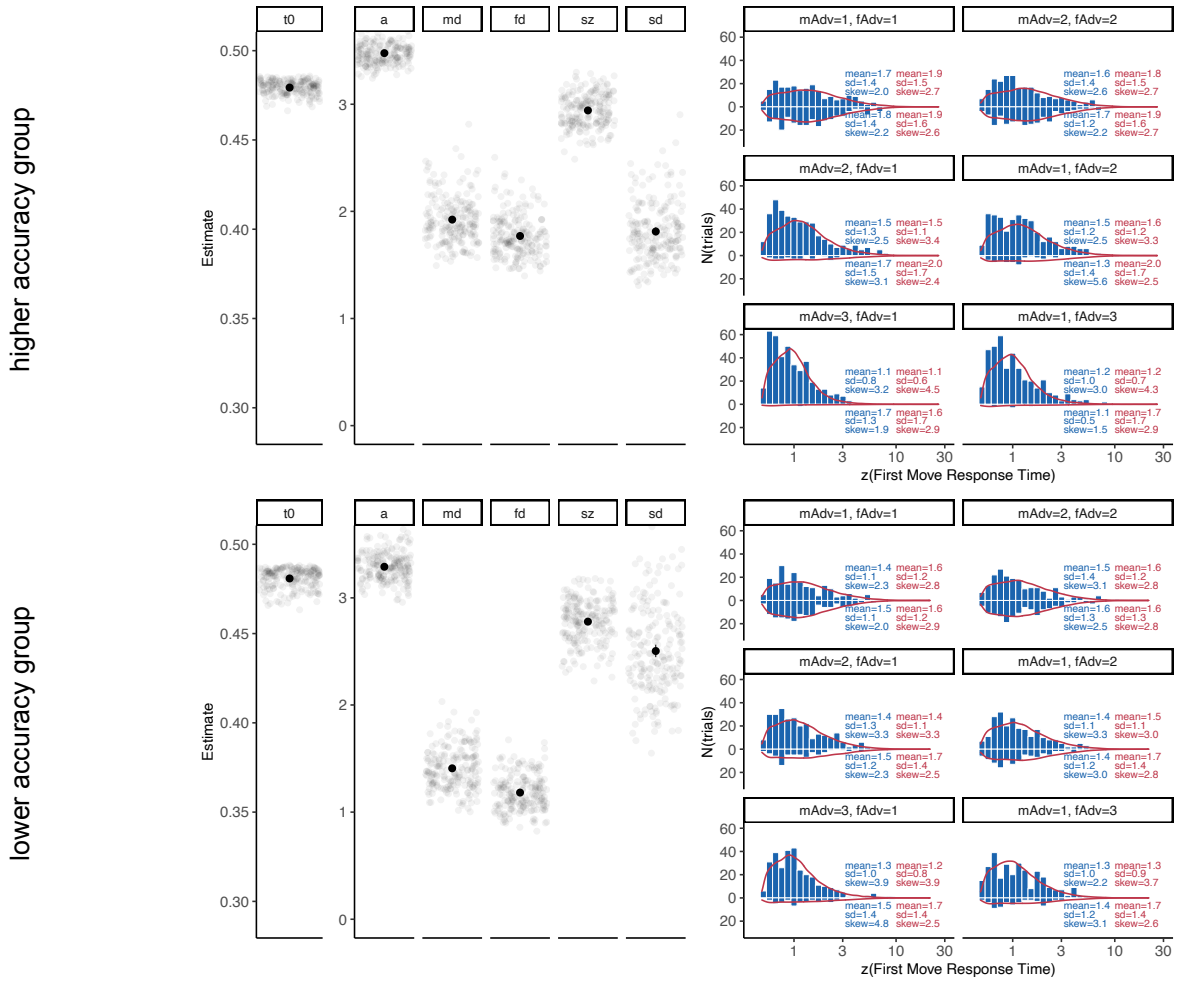

Fig S2. Experiment 1 IA trial responses and drift-diffusion model fits among groups of higher and lower accuracy. Visualization and notation as in Fig 4B, C.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

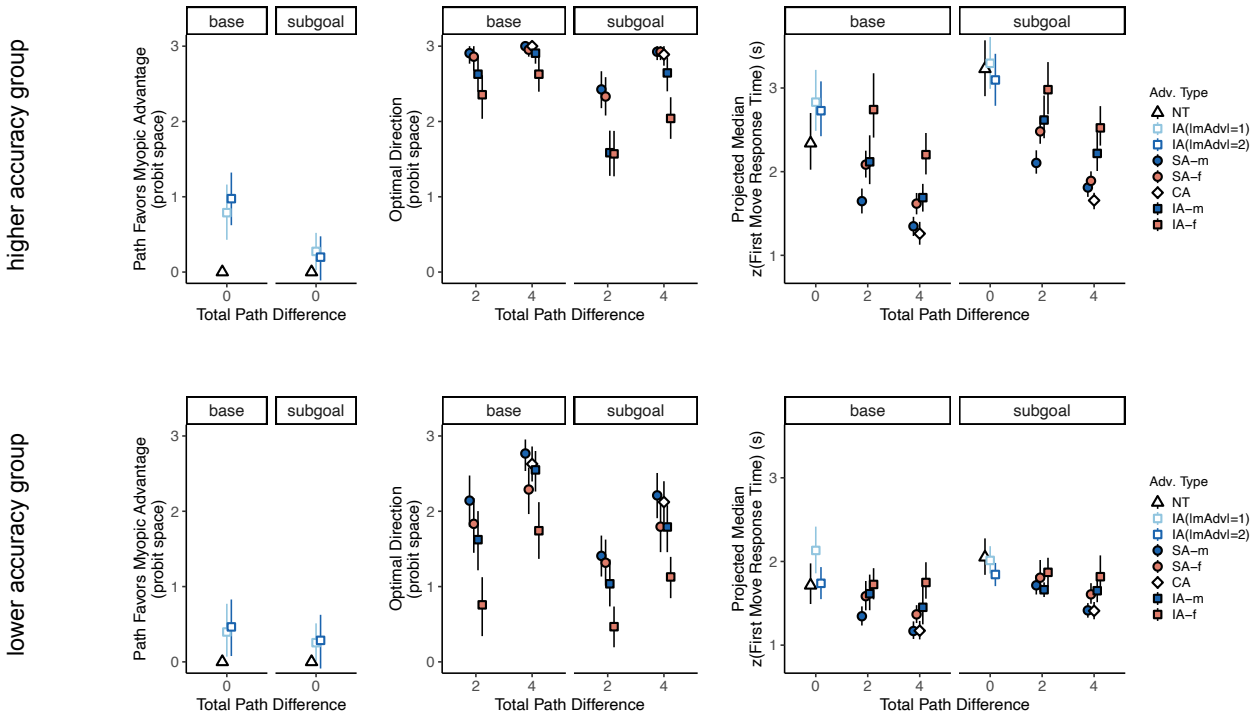

Fig S3. Experiment 2 path choices and response times for groups of higher and lower accuracy. Visualization as in Fig 5. The individual median zscore response times were projected to the second's space using the subgroup mean mean response time and mean standard deviation.

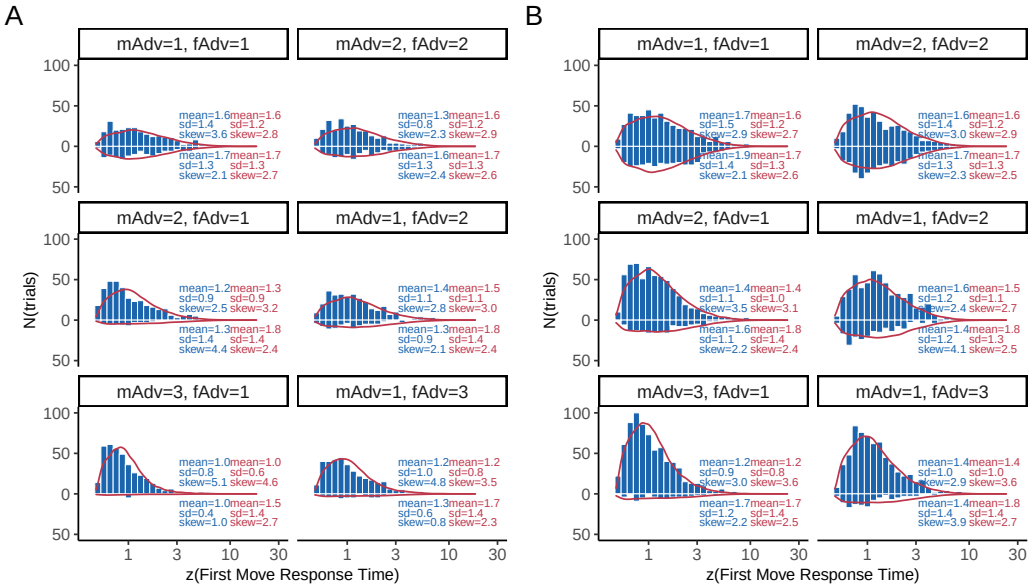

Fig S4. Experiment 2 model-predicted response times (in red) and empirical response times (in blue). A. Base trials. **B.** Subgoal trials. Visualization and notations as in Fig 4C.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

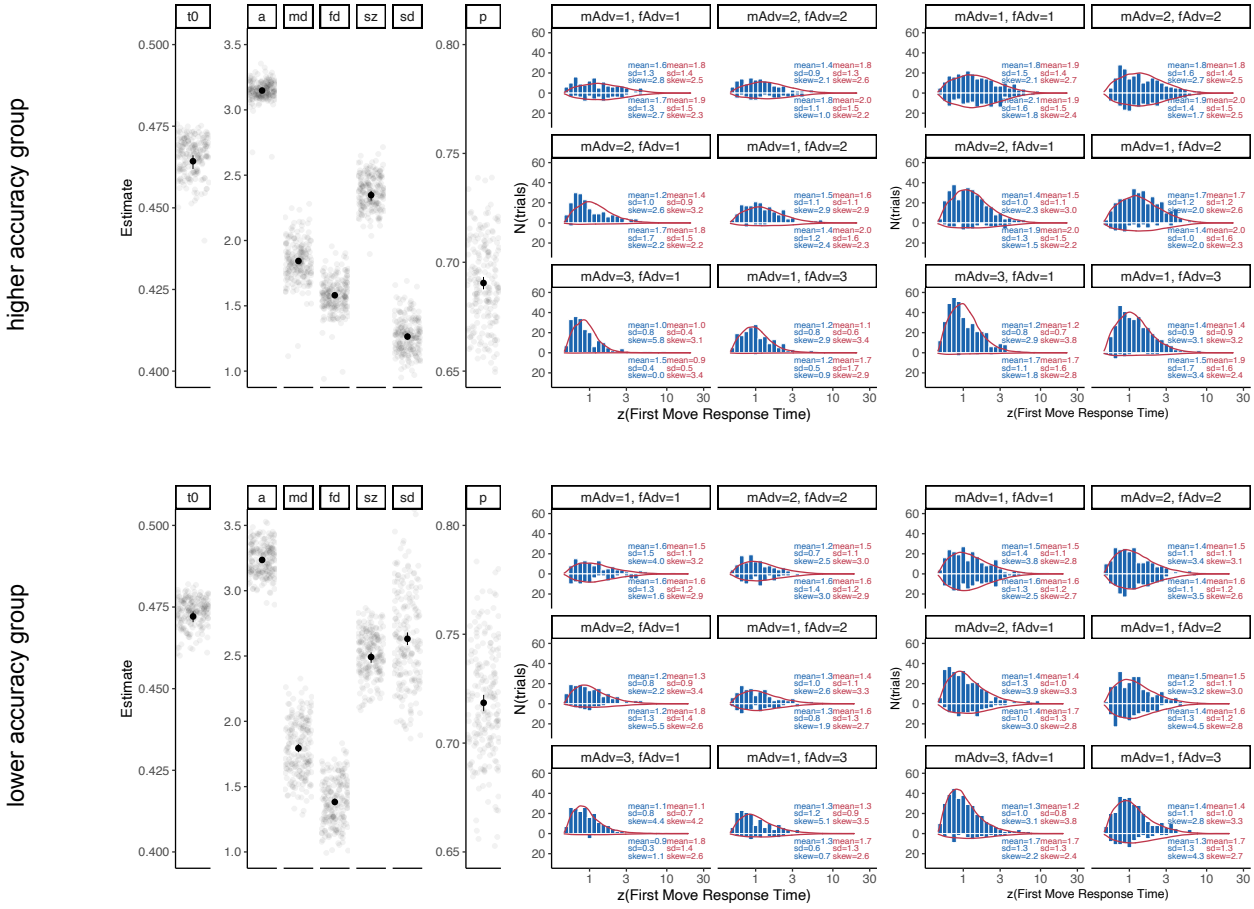

Fig S5. Experiment 2 IA trial responses and drift-diffusion model fits for groups of higher and lower accuracy. Left panel, model parameter estimates. Middle panel, base trials. Right panel, subgoal trials. Visualization and notations as in Fig 6C and S4.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

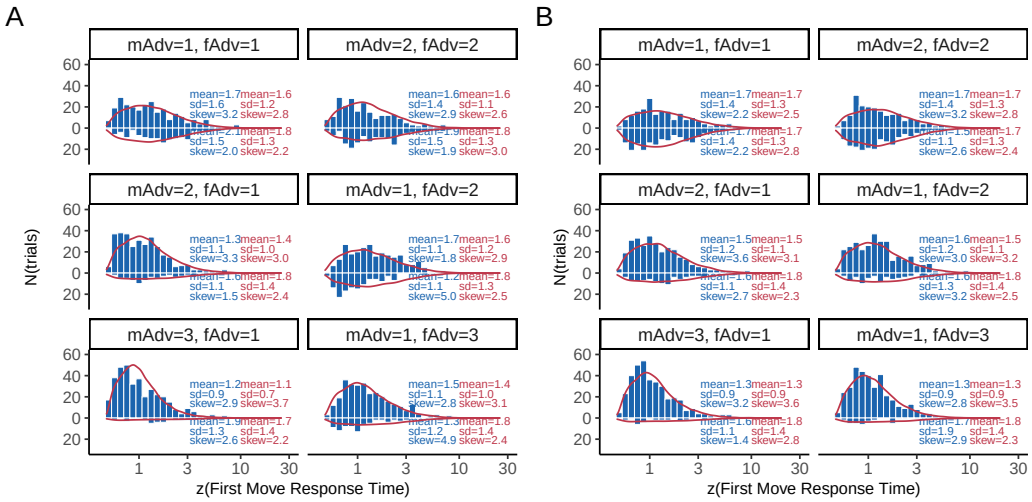

Fig S6. Experiment 2 IA subgoal trial responses and model-predicted responses split by the final goal location. A. Trials where the final goal was on the side of the myopic advantage. **B**. Trials where the final goal was on the side of the future advantage. Visualization and notation as in Fig 4C.

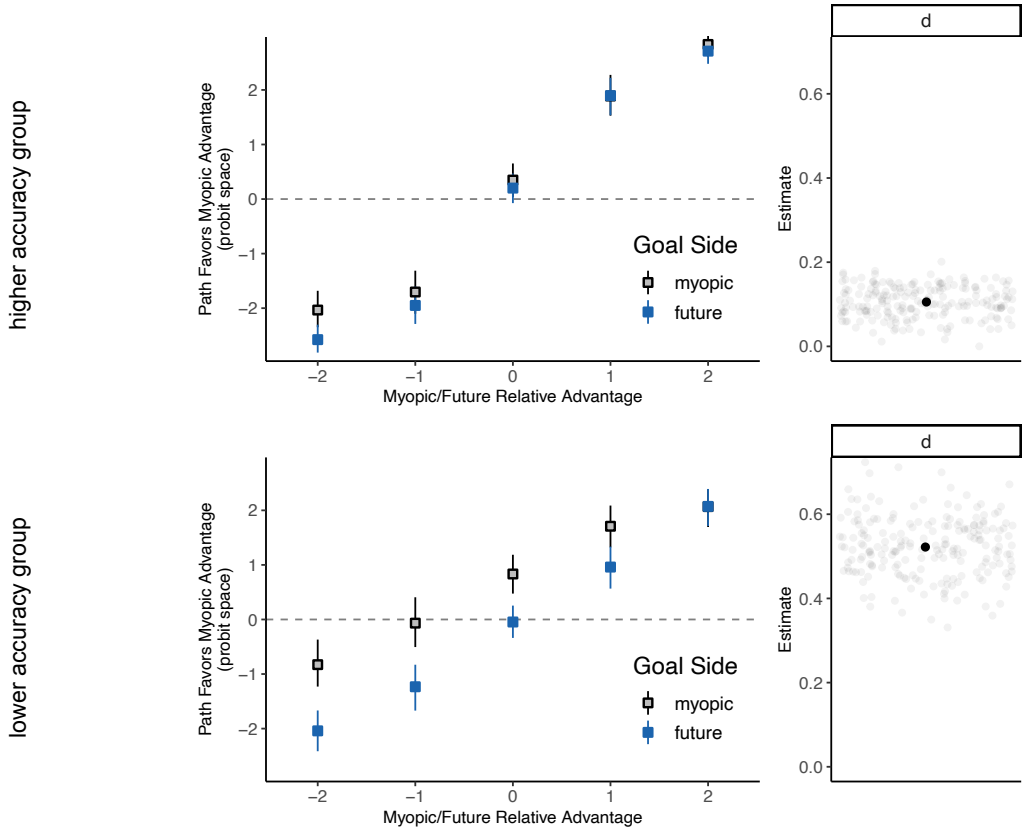

Fig S7. Experiment 2 contribution of the final goal in decision making for groups of higher and lower accuracy. Right panel, goal weight estimates for the higher accuracy group (top) were much smaller than goal weight estimates for the lower accuracy group (bottom).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

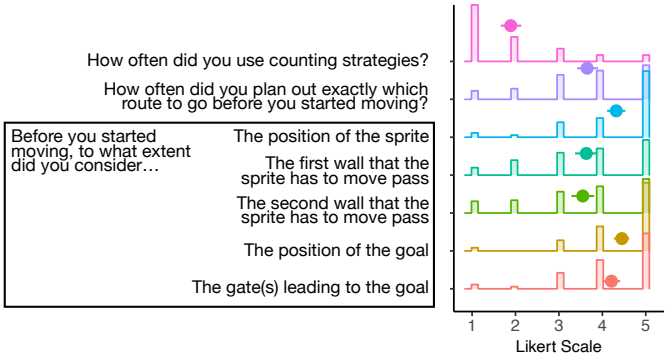

Fig S8. Experiment 2 survey responses. Participants' self-reported considerations of different task elements was in line with the behavioral analyses presented above. In free-form responses, when asked how they came up with the route to reach the goal, their responses consistently suggested that the reasoning process was rapid and intuitive. Likert scale: 1=never, 2=rarely, 3=sometimes, 4=often, 5=always. Sprite denotes the movable blue block. Error bars indicate bootstrapped 95% confidence limits.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Acknowledgments

We thank Stanford University for supporting the effort of both authors on this project, and we thank the members of the second author's lab group for useful discussions.

References

1. Calderon CB, Dewulf M, Gevers W, Verguts T. Continuous track paths reveal additive evidence integration in multistep decision making. Proceedings of the National Academy of Sciences of the United States of America. 2017;114:10618–10623. doi:10.1073/pnas.1710913114.

2. Kaplan R, King J, Koster R, Penny WD, Burgess N, Friston KJ. The Neural Representation of Prospective Choice during Spatial Planning and Decisions. PLoS Biology. 2017;15:1–26. doi:10.1371/journal.pbio.1002588.

3. Solway A, Botvinick MM. Evidence integration in model-based tree search. Proceedings of the National Academy of Sciences of the United States of America. 2015;112:11708–11713. doi:10.1073/pnas.1505483112.

4. Wickelgren WA. How to solve problems: Elements of a theory of problems and problem solving. WH Freeman; 1974.

5. Balaguer J, Spiers H, Hassabis D, Summerfield C. Neural mechanisms of hierarchical planning in a virtual subway network. Neuron. 2016;90:893–903. doi:10.1016/j.neuron.2016.03.037.

6. Huys QJM, Lally N, Faulkner P, Eshel N, Seifritz E, Gershman SJ, et al. Interplay of approximate planning strategies. Proceedings of the National Academy of Sciences of the United States of America. 2015;112:3098–3103. doi:10.1073/pnas.1414219112.

7. Tomov MS, Yagati S, Kumar A, Yang W, Gershman SJ. Discovery of hierarchical representations for efficient planning. PLoS Computational Biology. 2020;16:1–42. doi:10.1371/journal.pcbi.1007594.

8. Newell A, Simon HA. Human problem solving. Prentice-hall; 1972.

9. Maisto D, Donnarumma F, Pezzulo G. Divide et impera: Subgoaling reduces the complexity of probabilistic inference and problem solving. Journal of the Royal Society Interface. 2015;12. doi:10.1098/rsif.2014.1335.

10. Solway A, Diuk C, Córdova N, Yee D, Barto AG, Niv Y, et al. Optimal behavioral hierarchy. PLoS Computational Biology. 2014;10. doi:10.1371/journal.pcbi.1003779.

11. Dijk SGV, Polani D. Informational constraints-driven organization in goal-directed behavior. Advances in Complex Systems. 2013;16. doi:10.1142/S0219525913500161.

12. Botvinick MM, Niv Y, Barto AC. Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. Cognition. 2009;113:262–280. doi:10.1016/j.cognition.2008.08.011.

13. Dell GS, Burger LK, Svec WR. Language production and serial order: A functional analysis and a model. Psychological Review. 1997;104:123–147. doi:10.1037/0033-295X.104.1.123.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.13.464177; this version posted October 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

14. McClelland JL, Rumelhart DE. An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. Psychological Review. 1981;88:193–199. doi:10.1016/S0022-0728(02)01421-3.

15. McClelland JL, Rumelhart DE, Hinton GE. The appeal of parallel distributed processing; 1986.

16. Rumelhart DE. Toward an interactive model of reading; 1985.

17. Rumelhart DE, Norman DA. Simulating a skilled typist: a study of skilled cognitive-motor performance. Cognitive Science. 1982;6:1–36. doi:10.1016/S0364-0213(82)80004-9.

18. Suri G, Gross JJ, McClelland JL. Value-based decision making: An interactive activation perspective. Psychological Review. 2020;127:153–185. doi:10.1037/rev0000164.

19. Ratcliff R, McKoon G. The diffusion decision model: Theory and data for two-choice decision tasks. Neural Computation. 2008;20:873–922. doi:10.1162/neco.2008.12-06-420.

20. Singmann H, Brown S, Gretton M, Heathcote A. rtdists: Response Time Distributions; 2020. Available from: `https://CRAN.R-project.org/package=rtdists`.

21. R Core Team. R: A Language and Environment for Statistical Computing; 2019. Available from: `https://www.R-project.org/`.

22. Gershman SJ, Bhui R. Rationally inattentive intertemporal choice. Nature Communications. 2020;11. doi:10.1038/s41467-020-16852-y.

23. Ho MK, Abel D, Cohen JD, Littman ML, Griffiths TL. The efficiency of human cognition reflects planned information processing; 2020.Available from: `http://arxiv.org/abs/2002.05769`.

24. Dell GS. A spreading-activation theory of retrieval in sentence production. Psychological Review. 1986;93:283–321. doi:10.1037/0033-295X.93.3.283.

25. Koedinger KR, Anderson JR. Abstract planning and perceptual chunks: Elements of expertise in geometry. Cognitive Science. 1990;14:511–550. doi:10.1016/0364-0213(90)90008-K.

26. Dolan RJ, Dayan P. Goals and habits in the brain. Neuron. 2013;80:312–325. doi:10.1016/j.neuron.2013.09.007.

27. Botvinick M, Weinstein A. Model-based hierarchical reinforcement learning and human action control. Philosophical Transactions of the Royal Society B: Biological Sciences. 2014;369. doi:10.1098/rstb.2013.0480.

28. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Driessche GVD, et al. Mastering the game of Go with deep neural networks and tree search. Nature. 2016;529:484–489. doi:10.1038/nature16961.

29. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. Nature. 2015;518:529–533. doi:10.1038/nature14236.

30. Botvinick M, Wang JX, Dabney W, Miller KJ, Kurth-Nelson Z. Deep Reinforcement Learning and Its Neuroscientific Implications. Neuron. 2020;107:603–616. doi:10.1016/j.neuron.2020.06.014.