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 plan for long-range goals, proximal information cannot be exploited in a blindly myopic way, as relevant future information must also be considered. But when a subgoal must be resolved first, irrelevant future information should not interfere with the processing of more proximal, subgoal-relevant information. We explore the idea that flexible decision making in both situations relies on the modulation of the degree to which different pieces of information under consideration are weighted. We asked participants to find the shortest goal-reaching paths in mazes and modeled their initial path choices as a noisy, weighted information integration process. In a base task, participants weighted the starting-point and goal-proximal factors that together constrained the optimal initial path approximately equally, with participants who made overall more accurate choices exhibiting more balanced weighting. The base task was then embedded as an initial subtask in a larger maze, where the same two factors constrained the optimal path to a subgoal, and the final goal position was irrelevant to the initial path choice. In this more complex task, participants' choices reflected predominant consideration of the subgoal-relevant constraints, but also some influence from the initially-irrelevant final goal. More accurate participants again weighted the two relevant constraints more equally, and placed less weight on the optimality-irrelevant goal. Changes in the weighting of these constraining factors, rather than a longer processing time before decision making, accounted for the slowed and less accurate responses in the complex task. These findings suggest that, rather than explicitly decomposing a problem into smaller parts and solving subgoals in independent subtask contexts, humans can approximate the optimal decision outcomes produced by this ideal hierarchical approach through the task-sensitive modulation of the weighting of multiple constraints.

Author Summary

Different problems require the consideration of different information sources, including often useful long-range, future information that may impact our immediate decisions. However, when future information is irrelevant to a key subgoal, it can be desirable to focus on the subtask of achieving the subgoal first. How do decision makers flexibly select the relevant information in different situations? We suggest that, instead of engaging in hierarchical task decomposition when future information is irrelevant to an embedded subtask, humans rely on the same mechanism when confronting both types of situations, by appropriately weighting information relevant to decision making more than irrelevant ones. We conducted behavioral experiments and fitted models of different decision

processes to understand how people weight various factors in choosing the initial path in different mazes, both when a simple maze occurred alone or was embedded as an initial part in a larger maze. Our results show that people approximate the optimal outcome from a strictly hierarchical planning approach by modulating the weighting of different factors during planning, and that people who made more accurate initial path choices modulated these weightings more successfully than those who made less accurate choices.

Introduction

A hallmark of human intelligence is our ability to carry out 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 at a separate earlier time well before we reach the destination. Studies have shown that 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, focusing in succession on a sequence of subgoals on the path toward the ultimate goal [5, 6]. To come back to our traveling example, deciding whether to take the train or an Uber to the departure airport is one such smaller problem that can be independent from future aspects of the plan, once the flight and departure time has been settled. Studies have found that humans are adept at hierarchical planning, able to use learned knowledge to construct subtasks in novel problems and decompose tasks efficiently [7, 8]. Modeling work has provided precise accounts of how humans optimally group states in an environment and efficiently construct hierarchical task representations [8–11]. Furthermore, neural imaging studies have also supported the idea that the brain naturally organizes incoming stimuli based on their underlying hierarchical structure, delineating brain regions that signal differences among subgoal contexts as participants progress through them [12, 13].

On the surface, the situations discussed above may seem to call for different planning processes. When future information is relevant to decision making in an early point in a plan, selecting appropriate actions requires considering more proximal information together with the future information. But when future information is irrelevant to an immediate subgoal, optimal processing may be enhanced if we first decompose the overall task and then evoke an independent process to solve the subgoal without influences from irrelevant factors outside of the subtask context. However, many types of human sequential behavior that seem naturally characterized by a composition of independent, more atomic processes have been shown to rely on a degree of parallel consideration of multiple factors [14]. For example, optimally typing the sequence of letters in a word can involve hand movements that prospectively prepare for future letters [15], increasing overall speed and fluency. People also produce speech errors reflecting intruding influences from words after the target word [16], suggesting that current and future spoken words are being planned at the same time. More recent work has similarly argued that parallel, context-sensitive processes, rather than explicitly hierarchical or modular computations, underlie routine sequential action [17] and value-based decision making [18]. We may therefore also expect that humans solving an embedded subtask might exhibit a graded and non-exclusive focus on the subtask, with some degree of consideration of information outside of the subtask context even in situations where this information is irrelevant.

Here, we consider how a weighted constraint satisfaction process that simultaneously exploits multiple constraints can provide an alternative to the hierarchical task decomposition account of human planning and decision making. While we do not rule out that decomposition may occur in some situations, we suggest that flexible decision making in many problems may rely instead on the task-sensitive modulation of the degree to which different pieces of information are weighted in selecting the next action. For example, when solving an embedded subgoal in a larger task, a greater weighting of the subgoal-relevant information relative to the subgoal-irrelevant information can effectively approximate the optimal decision outcomes that would be generated by an explicit task decomposition process followed by exclusive subtask execution.

Our experiments explore this weighted constraint satisfaction account by considering the decision processes people engage in when selecting the first step in two related maze tasks, where the same maze navigation problem either appears as an isolated task or as an initial subtask embedded in a larger maze (Figs 1A and 1B). We next consider the two tasks in more detail, analyzing the task factors that might influence people's initial path choices and discussing the expected behavioral outcomes produced by the hierarchical task decomposition approach and the weighted constraint satisfaction approach.

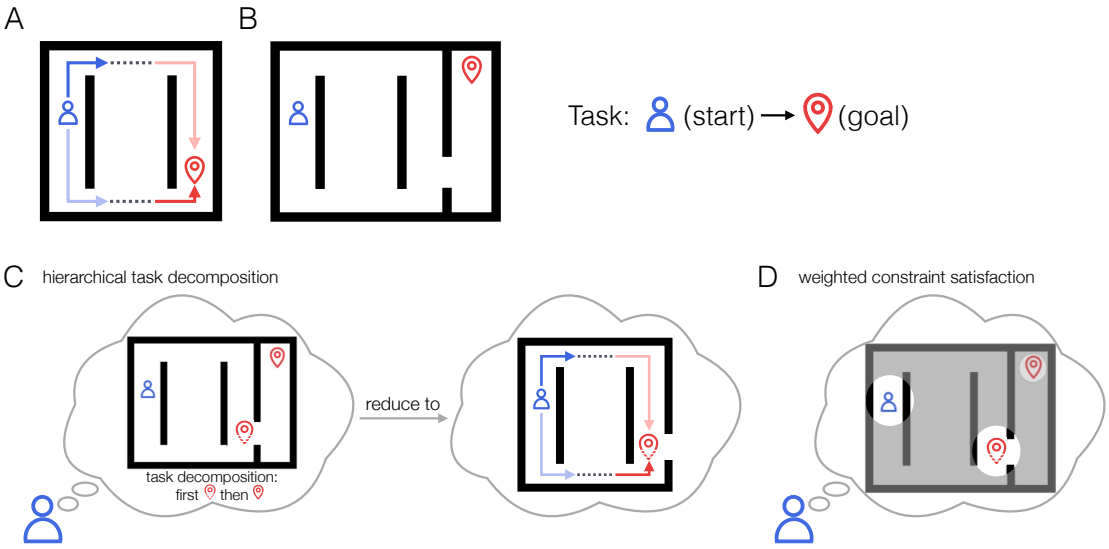

Fig 1. Hierarchical and parallel planning processes. A. A simple maze where selecting the initial direction of a best goal-reaching path benefits from the joint consideration of starting-point proximal (myopic) and goal-proximal (future) information. **B**. The task in A. embedded as an initial part of a larger maze. **C**. A hierarchical approach to selecting the initial path in B., by first reducing the task to the initial subgoal task then focusing on finding the best subgoal-reaching path. **D**. A parallel approach to selecting the initial path in B., where simultaneous but appropriately-weighted consideration of constraints both inside and outside of the subtask context generates approximately optimal choice behavior.

We consider first the base task, in which participants are rewarded for moving a token along the shortest path from a starting point to a designated goal (Fig 1A). The task can be considered to have three parts: First move out from behind the internal wall near the starting point; then traverse to the other side of the environment; then approach the goal by moving behind the internal wall near the goal. The optimal overall path, however, is determined both by a starting-point-proximal factor (the starting position relative to the internal wall near the starting point, henceforth called the *myopic* advantage) and a goal-proximal factor (the goal position relative to the internal wall near the goal, henceforth called the *future* advantage). Since the optimal paths differ in the first

action that needs to be taken, an optimal initial path choice must therefore be jointly constrained by both factors. Thus, if planning inherently involves the simultaneous consideration of multiple constraints, this task is perfectly suited to exploit it, and we would expect both the myopic and the future advantage to affect the choice of the first step and the time it takes to make it.

Fig 1B shows when the same maze in the base task is embedded as an initial part of a larger maze such that what was previously the goal location now corresponds to a subgoal location that must be reached prior to navigating to the final goal. In this more complex task, finding the optimal initial path still depends on the same two constraints, now associated with the starting point and the subgoal location. If people adopt a hierarchical task decomposition approach in this case, we may expect a longer processing time before choosing the initial action to reflect the time required to infer the subgoal and decompose the task into parts, prior to directing effort at solving the subgoal (Fig 1C, left panel). Moreover, the subsequent process of choosing the optimal initial direction within the subtask should be identical to that in the base task and unaffected by the position of the final goal (Fig 1C, right panel).

Alternatively, a weighted constraint satisfaction approach could be recruited to support decision making in the more complex task. As alluded to above, an approximation of the pattern of responding that would be produced under the hierarchical approach can be achieved by a greater weighting of the two subgoal-relevant constraints compared to the initially-irrelevant final goal in deciding the initial action (Fig 1D). For path selection to still be close to optimal, the weighting of the various factors must allow the two subgoal-relevant constraints to exert a near equal and jointly predominant influence in decision making. But a small residual weighting of the irrelevant final goal might be reflected in a subtle influence of this factor on the time course and outcome of selecting the initial path. To explore this possibility, we can therefore ask these questions: 1) To what extent do participants balance their consideration of multiple *relevant* constraints, and 2) when there is an *irrelevant* constraint, to what extent do they effectively ignore it? More generally, how does the presence of a particular constraining factor (whether relevant or irrelevant) influence the weighting of other constraining factors?

Across two experiments, we studied how human participants approached solving these maze problems (Fig 2). We assessed both the paths chosen by participants, as indicated by the direction of their first step, and the time they took to plan this step, as indexed by their reaction time. We used the drift-diffusion model (DDM) [19] to characterize how the constraining factors discussed above were used in deciding the first action on each trial. The DDM and subsequent models built on related ideas have been used to understand a wide range of human perceptual and cognitive processes [20–22], and the DDM has recently been used to account for behavior in multi-step decision making [1, 3]. Here, we leverage the DDM to characterize whether the consideration of the various factors during planning was more consistent with a hierarchical task decomposition process or a weighted constraint satisfaction process. If planning occurs according to the hierarchical task decomposition approach as described above, the process of selecting the initial action in the more complex subgoal task should resemble that in the base task, though it may begin after a longer initial delay. In contrast, in the weighted constraint satisfaction approach, we might expect simultaneous influence of the subgoal-relevant constraints and the final goal during initial path selection, albeit with relatively less weight assigned to the final goal in decision making. We subject these hypothesized decision processes to test through the joint fit of trial-level path choice data and the response time distributions.

In Experiment 1 (Fig 2, top panel), we first confirmed that initial path choices in the base task indeed reflected a joint influence from the relevant myopic and future constraints. Interestingly, we found that participants' path choices tended to reflect the

myopic advantage more strongly than the future advantage. In addition, when the myopic and future advantages were pitted against each other, each exerted a weaker influence on path choices than it did when only one of the factors was relevant to determining the optimal path. We then used the data from a subset of trials in this experiment to constrain the selection of a best-fitting DDM process for the base task. Then in Experiment 2 (Fig 2, middle and bottom panels), we compared the decision processes in the more complex task with that in the base task. We found that initial path choices were slower in the complex task compared to the base task, and that these choices were both reduced in accuracy and influenced by the final goal position. In line with these observations, DDM variants consistent with the weighted constraint satisfaction account fit the pattern of data better than DDM variants consistent with the hierarchical task decomposition account. In addition, we found that more accurate participants (those who were more likely to choose the optimal path) exhibited more balanced weighting of initial path-relevant constraints and assigned less weight to the initial-path irrelevant final goal position than less accurate participants. In the *General Discussion*, we return to the idea that a parallel, weighted constraint satisfaction approach might be the biological brain's way of planning decisions, and that our brains approximate an optimal focus on relevant information in different task contexts through the dynamic modulation of constraint weighting as thought and behavior unfold.

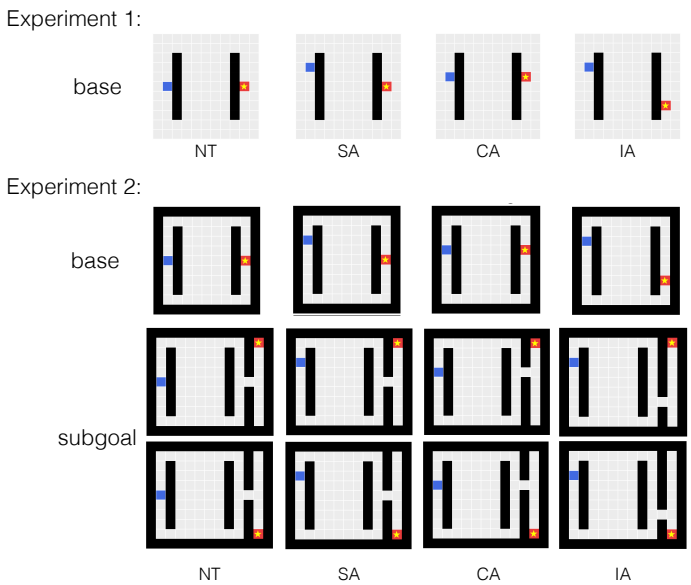

Fig 2. Experimental design. Blue block: starting point. Starred red block: goal location. Trial advantage types are based on the correspondence between the starting-point proximal constraint (myopic advantage) and the goal-/subgoal-proximal constraint (future advantage): NT=neutral advantage, SA=single advantage, CA=congruent advantage, IA=incongruent advantage. For the subgoal trials in Experiment 2, both of the possible final goal locations are shown.

Results

Experiment 1

The two main goals of Experiment 1 were 1) to examine the influence of the starting-point proximal, *myopic* advantage and the goal-proximal *future* advantage in initial

path choices in the base task and 2) to fit variants within the family of drift-diffusion models to a subset of the trials in the experiment to assess what constraint weighting scheme best captures the joint choice and response time patterns. We varied the myopic advantage and the future advantage, which together determine the shortest goal-reaching path on each trial, and analyzed participants' first action *choice* and their time taken to take the first step (*response time*).

Methods

Ethics Statement. This study was approved by the Stanford University Institutional Review Board under protocol No.7029. Participants on Amazon Mechanical Turk were given a consent page at the beginning of the experimental session in which they gave consent by continuing to the main study.

Participants. For Experiment 1, 100 US-based participants were 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 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, top panel). 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 study took around 35 minutes and participants received an average of $6.05 for completing the study.

The experimental trials included 92 base trials and 92 filler trials, presented in a different randomized order for each participant. Below we report results from the base trials, where the locations of the starting block and the goal block were designed to vary the relative advantages of the candidate goal-reaching paths near the starting location and near the goal, but the locations of the two length-7 walls were fixed (see Table 1). Each unique base trial layout was mirrored vertically, except for the neutral trial (see below). Both the original and the mirrored trials 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 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). We subset 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 main path candidates (i.e., equivalent to the upper and the lower paths in the left-to-right orientation) are equally optimal.

On each filler trial, we randomly sampled trial orientation (among all four orientations) and the length of each internal wall (3, 5, or 7). The two internal walls were also randomly shifted up or down, but were never allowed to block the upper or lower paths. Wall

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 advantages. -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.

locations on the filler trials with two length-7 walls were also never centered (i.e., never identical to the wall locations on the base trials). The location of the starting block was randomly sampled from the locations to the left of the starting-proximal wall, and the location of the goal block was randomly sampled from the locations to the right of the goal-proximal wall.

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 of the two main path candidates (ten trials; a main path candidate 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 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 path choices and planning time in the IA trials as a drift-diffusion 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. The model treats the decision making process as involving a single aggregate decision variable in which a positive value favors a path choice that satisfies the myopic advantage and a negative value favors a path choice that satisfies the future advantage. On each trial, this variable evolves randomly over time with a mean direction d given by:

$$d = md * mAdv - fd * fAdv$$

where md is the drift weight associated with the myopic advantage $mAdv$ and fd is the drift weight associated with the future advantage $fAdv$. When the value of the variable reaches an upper or lower bound at values a or $-a$, the process terminates and a response satisfying the myopic advantage is chosen if the upper bound is reached, or one satisfying the future advantage if the lower bound is reached.

In addition to the parameters md, fd, and a, the model also includes an initial non-decision time t_0 and a possible starting point bias z. We additionally modeled

inter-trial variability for the starting point (sz) and the drift rate (sd). As a baseline, we considered a model in which the weights for both advantages are equal ($md = fd$) and there was no starting bias (z=0). We tested whether the data were better accounted for with equal or different advantage weights, with or without a starting bias, and with or without the two sources of inter-trial variability.

We pooled data from all participants to fit the candidate drift-diffusion models. We first z-scored the raw first move response time (in seconds) within each participant, then shifted the z-score distributions so that the minimum is 0.5 for each participant to ensure positive response time and enough non-decision time buffer for modeling purposes. All model variants were fitted in a Monte-Carlo cross-validation procedure written in R, using the density function implemented in the rtdists package [23] and the nlminb function in the stats package [24]. In each of 200 cross-validation folds, 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 each fold was selected based on the best training sNLL. 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 sNLL on the held-out test data across all candidate models with fold-level random intercepts, and compared which model achieved lower or comparable test data sNLL with fewer free parameters.

For the winning model, we observed that some runs under each fold 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 different random samples of initial parameter values to find a set of final parameter values better than those at this local minimum. This was successful; 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 are reported below.

Results and Discussion

Before turning to drift-diffusion modeling, we first present a descriptive analysis of the overall choice and reaction time data to confirm that participants' choices are influenced both by myopic and future constraints and to characterize the experimental factors that influenced the degree of choice optimality and the time required to make the choice.

Overall, participants made optimal initial path choices on 91.57% of the trials (excluding the NT and IA trials with equally optimal initial directions), and they did so quickly, with a group-average median planning time of 1.34 sec. As shown in Fig 3B, participants' selection of the initial path direction favored the optimal path across all of the advantage pairings, suggesting that they must be taking both constraints into account.

Both choice optimality and response times were influenced by the relative advantage of one path over another, by whether the optimal choice was based on a single advantage or a combination of two advantages, and by whether the larger advantage was the myopic, starting-point proximal advantage or the future, goal-proximal advantage (Fig 3B and 3C). To further characterize these effects, we present a set of planned comparisons taken from a mixed-effects regression model of trial-level initial path choices (with a probit linking function), with separate parameters for each trial advantage type and participant-level random intercepts. The model confirmed a significant difference in choice optimality among different advantage pairings shown in Fig 3B, $\chi^2(8) = 301.76$, $p<0.001$. Response times (Fig 3C) also differed significantly among all 12 advantage pairings $F(11, 1128)=22.80$, $p<0.001$, based on a linear model applied to the individual

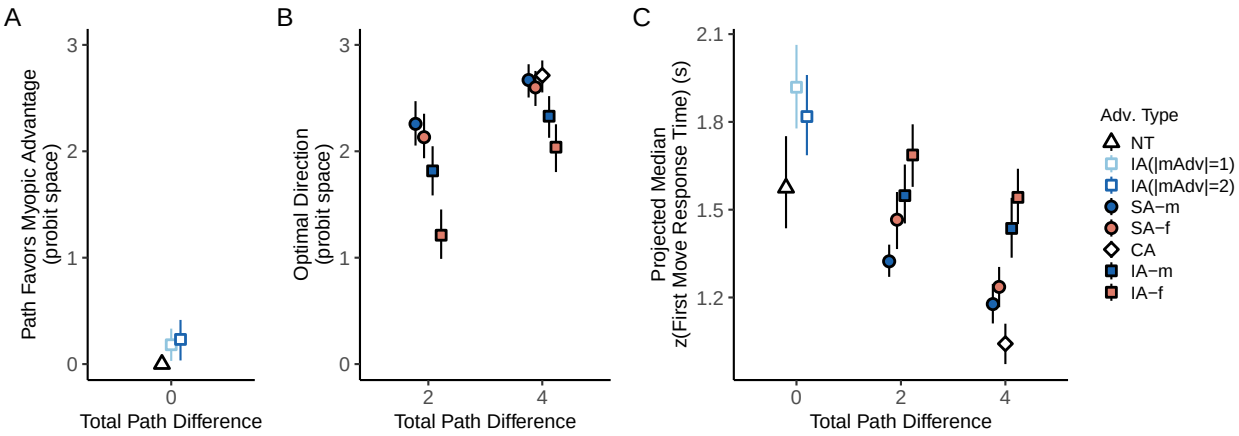

Fig 3. Initial path choices and the associated response times reflected joint consideration of myopic and future constraints. A. Path choices in trials with equally optimal initial directions. The value for NT trials was defined to be 0 since there was neither a myopic or a future advantage on these 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.** Response times associated with the first step. 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. Trial advantage types: 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.

median z-scored first move response times. Detailed comparisons of the estimated marginal means (EMMs; with Bonferroni correction) based on both the choice model and the response time model (controlling for total path differences) confirmed that responses with an overall path advantage of 4 were faster and more often in the optimal direction than responses with a path advantage of 2 (adjusted ps<0.01), and that responses in the IA trials were less optimal and slower than those in the SA and CA trials (adjusted ps<0.01). 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).

We now turn to the consideration of the relative influence of the myopic versus future advantages on speed and accuracy of initial path choices. Compared to the counterpart trials where the future advantage was larger (red points in Fig 3), trials where the myopic advantage was larger (blue points in Fig 3) had a 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. An overall myopic bias is exhibited by a majority of the individual participants in the IA trials, though about 30% showed the opposite tendency when the two advantages are of equal magnitude.

Having observed the joint consideration of myopic and future advantages in deciding the first step as well as the myopic bias, we next turn to the comparison of drift-diffusion model variants in characterizing this decision process, focusing on the IA trials in which the two advantages lead to competing decision outcomes. 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, and not from an initial bias favoring the myopic advantage (Fig 4A). The model with different weighting for the two advantages resulted in better fitting objective (summed negative log-likelihood) on the held-out test data across cross-validation folds, compared to the baseline model with equal

weighting or 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 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 the fit in the different weighting case (adjusted p >0.5). Models with inter-trial variabilities in both the starting point and the drift rate also consistently outperformed the others (adjusted p's<0.001).

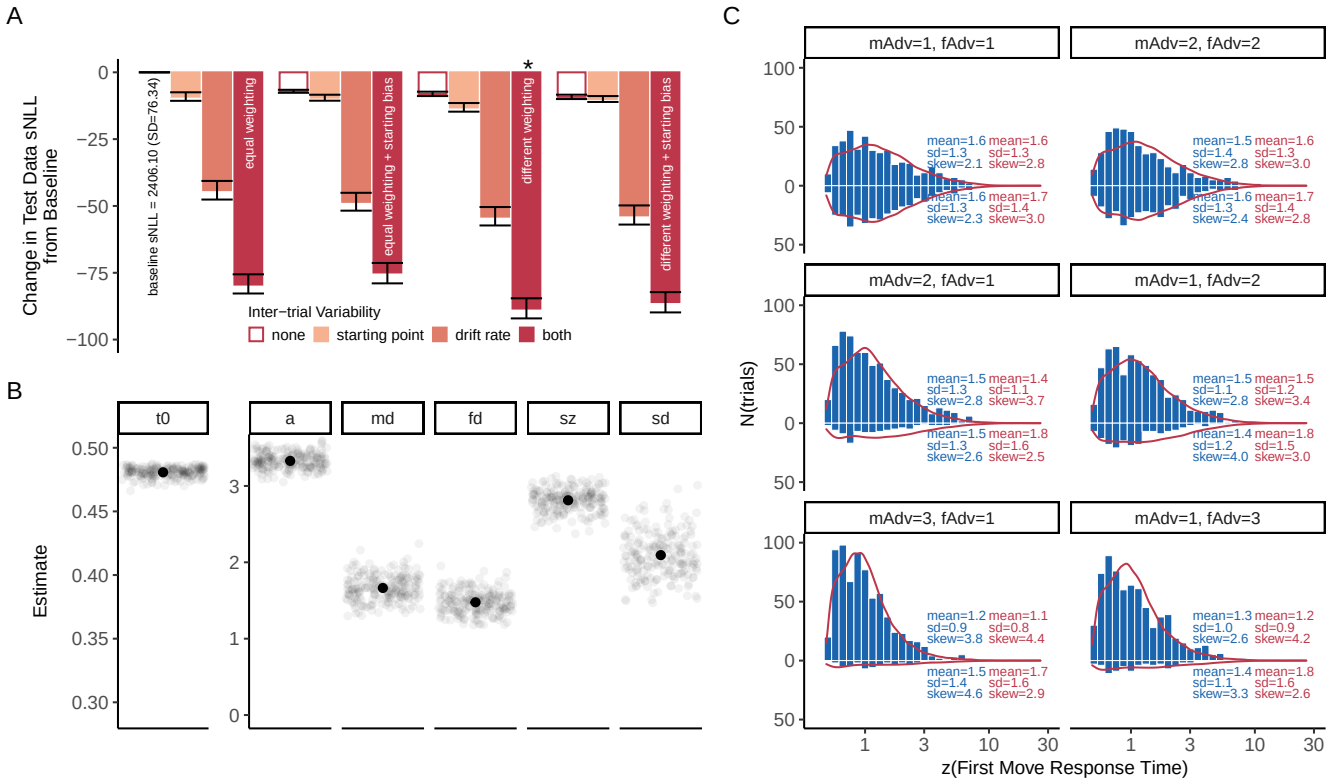

Fig 4. Weighted integration of myopic and future constraints in initial-step decision making in the base task. A. Test data objective (summed negative log-likelihood) of candidate drift-diffusion models as compared to the baseline model. Asterisk marks the winning model. **B.** Parameter estimates of the winning drift-diffusion model. t_0, 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). In C., the top two panels show the RT distributions associated with choices satisfying the myopic advantage on the top and those satisfying the future advantage on the bottom. The bottom four panels show the RT distributions for the correct responses on the top and error 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 drift 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 (Fig 4C, middle panel).

Individual differences. Considered together, participants in Experiment 1 made the optimal initial choices on over 91% of trials. However, the accuracy rate varied widely across participants, from 49.23%–100.0%. Participants also differed widely in

how quickly they arrived at their first response (range of individual median response time: 0.60 sec–3.29 sec). To understand what factors differed between participants who made more and less accurate responses, we split the participant population into two groups based on their overall choice accuracy (see *Supporting Information*), with choice accuracy rates ranging from 94.44% to 100% for the higher accuracy group and from 49.23% to 94.37% for the lower accuracy group. Importantly, both groups were more likely to make the optimal choice when it satisfied the myopic advantage, though the effect was more prominent in reaction times for the more accurate group and more prominent in optimality rates in the less accurate group (Fig S1). Fitting the model separately to the higher and lower accuracy groups revealed that the more accurate group exhibited larger weights for both the myopic and future advantages and lower drift rate variability relative to the less accurate group (see *Supporting Information* and Fig S2). In addition, we found that the more accurate group tended to weight the myopic and future constraints more nearly in accordance with an optimal, equal weighting of the two constraints. For the more accurate group, the ratio of the myopic weight relative to the future weight was estimated to be 1.09 (SD=0.02); for the less accurate group, the same ratio was estimated at 1.19 (SD=0.05). The distributions of this ratio were significantly different between the model fits to the data from the two participant groups (the mean difference from 1000 bootstrap resamples was -0.11, 95% CI [-0.114, -0.098]).

Experiment 2

Both the descriptive and the model fitting analyses from Experiment 1 indicated that people rely on both the relevant myopic and future constraints in selecting initial actions in the base task. Overall, participants weighted the myopic constraint slightly more than the future constraint, and participants who were less accurate in choosing the optimal path showed a less optimal (more imbalanced) weighting of the two constraints. We next investigated how people approach the same two-constraint problem when it appears 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 base maze is the first part of a larger maze and leads to a subgoal that must be visited prior to reaching the final goal (*subgoal trials*; see Fig 2). In these subgoal trials, as illustrated in the figure, we varied the final goal location, a factor that is irrelevant to initial path choices.

We were interested in how the decision process behind the selection of the first step changes when people confront the more complex task as compared to when they solve the base task. If participants adopted a hierarchical task decomposition approach, we may expect an increase in the initial non-decision time, followed by an identical decision process to find the shortest subgoal-reaching path compared to that in the base trials, with no influence of the position of the final goal on the choice of initial path direction. If participants instead extended the weighted consideration of multiple factors to the subgoal trials, we may expect no change in non-decision time, but instead changes in the weighting of the two relevant advantages along with some weighting of the irrelevant final goal in deciding on the initial path direction. Experiment 2 was pre-registered through the Open Science Framework (https://osf.io/w78hu). The statistical tests and the drift-diffusion modeling presented in this paper were developed after the experiment was completed and the pre-registered analyses were performed (we do not report the results of the pre-registered analyses as such since the reported analyses refine and extend them).

Methods

Ethics Statement. This study was approved by the Stanford University Institutional Review Board under protocol No.7029. Participants on Amazon Mechanical Turk were

given a consent page at the beginning of the experimental session in which they gave consent by continuing to the main study.

Participants. We recruited 100 participants on Amazon Mechanical Turk with an identical set of eligibility criteria from Experiment 1, 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 task condition (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 study took around 30 minutes and participants received an average of $5.63 for completing the study.

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, as illustrated in the figure. Similar to Experiment 1, all trials were mirrored vertically except for the NT trials.

The multi-subgoal trials have two openings on the wall prior to the final goal which were structured so that the candidate paths through the two subgoals were 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. We also included the multi-subgoal control trials where we closed the subgoal further from the goal. The layout of these control trials did not overlap with any subgoal trials. See more information in the OSF repository.

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 of the two main path candidates (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, initial path choices and response times in the subgoal trials are accounted for by a different non-decision time or by changes to constraint weighting during the drift process. The baseline model capturing the decision process in the base trials was the winning model from Experiment 1, which includes six parameters (t_0, a, md, fd, sz, sd). We modeled a third addend to the trial drift rate in the subgoal trials, gd, representing the weight the final goal may carry during weighted information integration. gd drives decision toward the upper decision bound (i.e., selecting the path that satisfies the myopic advantage) when the goal location is on the same direction with an initial move that satisfies the myopic advantage, and drives decision away from the upper decision bound otherwise. We also tested whether, and if so how, the advantage weights md and fd in the subgoal trials differed from those in the base trials, including shared additive change or independent changes to md and fd. We subsequently added another variant,

modeling a shared proportional change to both md and fd. Response variables, model fitting, and model comparison were the same as in 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

We compared how participants considered the various task factors when the same maze task either appeared alone (base trials) or as an initial subtask that leads to a subgoal (subgoal trials). In the subgoal trials, optimal initial paths are constrained by the same myopic and future path advantages as in the base trials, because final goal information is initially irrelevant. Before we turn to the drift-diffusion modeling results, we first review the overall response time and choice accuracy across the two task conditions.

Overall, performance degraded in the subgoal trials with both an accuracy cost and a response time cost. Participants selected the optimal initial 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 step 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).

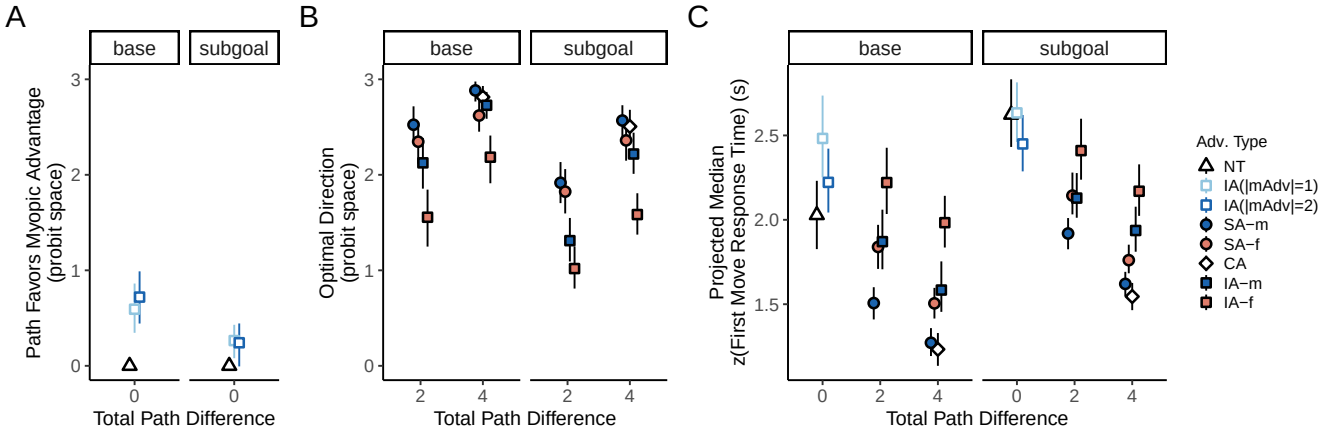

Fig 5. Accuracy and time cost in initial path choices when the two-constraint maze is embedded as a subtask. A. Path selection in the trials with equally optimal initial directions. As in Fig 3A, the value for NT trials was defined to be 0. **B.** Path selection across advantage pairings in base and subgoal trials. 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 in the base and subgoal trials. As in Fig 3C, the median zscore response times were projected back to the second's space.

The increased first move response times in the subgoal trials may reflect time taken to decompose the task and identify the final goal information as irrelevant to the subtask, but the simultaneous change in choice accuracy suggested that participants did not deploy the exact same decision process to solve the identical two-constraint task that is at core to selecting initial paths in both task conditions. We note that the benefit from congruent or single optimality-relevant constraints (as opposed to incongruent ones) was preserved, as path choices and response times showed similar patterns across the different trial advantage pairings in both task conditions (Fig 5B and 5C). This is confirmed by extending the path choice model and the median zscored response time

model from Experiment 1 with task condition (base vs. subgoal) as a second predictor and accounting for interaction effects. We observed no significant interaction in either the choice model ($\chi^2(8) = 8.19$, p=0.42, main-effects model as compared to an interaction model) or the response time model ($F(11, 2376)$=1.23, p=0.26).

Building on the winning model from Experiment 1 as a baseline, we next compared the decision processes in the subgoal IA trials to that in the base IA trials. Model comparison suggested that changes in what and how different pieces of information are weighted in the decision process, but not an increase in initial non-decision time, accounted for the slowed and less accurate initial path choices in the subgoal trials. As shown in Fig 6B, adding a goal weight (gd, see Experiment 2 *Methods*) to the drift rate resulted in a significantly better model compared to the baseline model (adjusted $p < 0.001$, pairwise EMM comparisons based on a linear mixed-effects model fit to the test objectives from all model variants, Bonferroni corrected). This echos the pattern of the initial path choices in the IA trials, which clearly showed a biasing influence from the irrelevant final goal (Fig 6A).

We also found that a proportional decrease (modeled by a weight multiplier p) in the weights of the optimality-relevant path advantages (md and fd) further accounted for the decreased choice optimality and slowed response times in the subgoal trials (Fig 6B). This echoed the similarity in the response patterns across advantage pairings shown in Fig 5, suggesting an overall degradation in the weighting of the two relevant constraints in the more complex task. The proportional weight change model was better than the model with an equal decrement in md and fd (adjusted $p < 0.001$), and on par with the independent weight change model (adjusted $p > 0.9$) with one less free parameter. Moreover, modeling a separate non-decision time in the subgoal trials for the proportional weight change model did not lead to significantly better test data likelihood (adjusted $p = 0.13$), indicating that the process of deciding between the path choices could initiate after about equal time in both task conditions, as the slowed response time was accounted for by changes to the weighting of different constraints.

The winning model estimated that the weight placed on the irrelevant final goal (gd) was about 0.31 (SD=0.06; Fig 6C). It is a much smaller weight compared to the weights associated with the optimality-relevant myopic and future advantages in the subgoal trials, which were estimated at md=1.27 (SD=0.12), fd=1.04 (SD=0.10) after about a 30% (SD=2%) decrease compared to their values in the base trials, with their ratio preserved at 1.22 (SD=0.03).

Individual differences. As in Experiment 1, there were large individual differences in participants' overall accuracy (Fig S3). As before, the qualitative pattern of effects exhibited in the combined results across all participants was also exhibited by participants in both the higher accuracy group (accuracy rate: 91.67%–100.0%) and the lower accuracy group (accuracy rate: 53.85%–91.59%). When fitting the model separately to data from the group with overall higher accuracy and the group with overall lower accuracy (see *Supporting Information*), the lower accuracy group showed a higher degree of drift-rate variability (Fig S5), consistent with that found in Experiment 1.

The groups also differed in the relative weighting of the different task factors, with the more accurate group showing a more nearly optimal (i.e., equal) weighting of the initially-relevant myopic and future constraints, and a tendency to place smaller weight on the initially-irrelevant final goal location (Fig 6D). For the more accurate group, the ratio of myopic to future weights was 1.17 (SD = 0.02), which was significantly lower than that estimated for the less accurate group (M=1.30, SD = 0.06; the mean difference from 1000 bootstrap resamples was -0.14, 95% CI [-0.146, -0.129]). More strikingly, the influence of the final goal was estimated at markedly different levels. For the group with higher overall accuracy, the goal carried a reliable, but very small weight (M=0.08, SD=0.03). For the group with lower overall accuracy, however, the goal weight

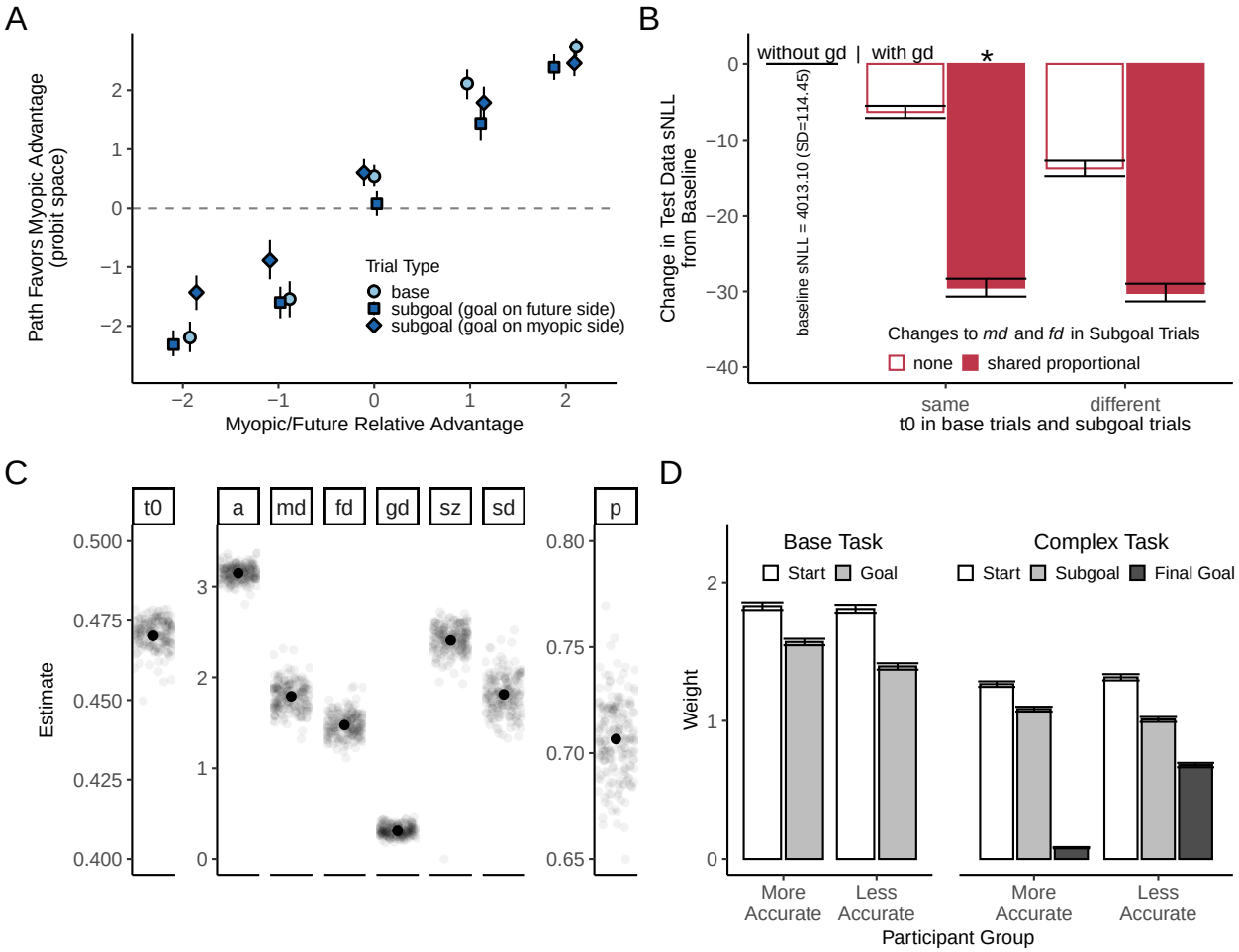

Fig 6. Appropriately-weighted consideration of subgoal-relevant and -irrelevant constraints approximated optimal choice behavior when the base task was embedded as an initial subgoal task. A. Path choices in the IA trials. As in Fig 3 and Fig 5, the proportion of optimal trials for each individual was 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) of the candidate drift-diffusion models compared to the baseline model. Asterisk marks the winning model. For empirical and predicted response time distributions, see Fig S4. **C.** Parameter estimates from the winning model. t_0, non-decision time. a, decision bound. md and fd, myopic and future advantage weights. gd, goal weight. sz, inter-trial variability of the starting point. sd, inter-trial variability of the drift rate. p, proportional change to md and fd in the subgoal trials. **D.** Drift weight estimates for overall more accurate participants and less accurate participants.

was estimated at 0.68 (SD=0.12), which was significantly higher than that in the more accurate group (the mean difference from 1000 bootstrap resamples was 0.60, 95% CI [0.582, 0.615]). Both groups showed around 30% of degradation in the weighting of the optimality-relevant myopic and future advantages in the complex task (more accurate group: M=30.89%, SD=1.94%; less accurate group: M=27.49%, SD=2.65%). Although the weighting degradation was also significantly different across the two groups (the mean difference from 1000 bootstrap resamples was 3.40%, 95% CI [2.95%, 3.91%]), the difference was small.

General Discussion

In this work, we have proposed a weighted constraint satisfaction approach to goal-directed decision making, contrasting this with a hierarchical task decomposition approach. If participants where optimal, task decomposing planners, we would expect that their decision processes in the complex maze task would correspond to finding the best subgoal-reaching initial path and thus resemble the choices made in the base maze task. Instead, we found that participants' initial path choices were less likely to be optimal in the complex task, and that their choices were influenced by both the subgoal-relevant constraints and a subgoal-irrelevant factor (the location of the final goal).

We used a drift-diffusion modeling framework to jointly account for path choices and response times in our experiments. The results of the modeling effort revealed several things. First, when confronted with the base task, participants' decisions were influenced by both the relevant starting-point proximal (myopic) constraint and the goal-proximal (future) constraint, but to differing degrees. This was accounted for in the drift-diffusion model by assigning a greater weight for the myopic over the future constraint in determining the overall drift direction of the diffusion process. In the more complex task, the weighting of the same two optimality-relevant constraints was reduced, and participants' weighting of the final goal was strong enough to lead to suboptimal initial step choices on some trials. Responses were overall slower in the more complex task, a finding that, taken by itself, might seem to reflect the time required to focus in on the embedded subgoal in a hierarchical task decomposition process. However, changes in the degree to which the optimality-relevant or -irrelevant task factors were weighted, rather than a simple additive offset on the total time to make the decision, accounted simultaneously for the reduction in choice accuracy and for the slowed responses in the more complex task. Overall, the results support the weighted constraint satisfaction approach to understanding how participants made their initial decisions in our experiments.

While participants' responses usually favored the optimal path, they did not always do so, and there were individual differences in how closely participants approximated optimal choice behavior in our tasks. The drift-diffusion model implementation of our weighted constraint satisfaction account can exhibit varying degrees of optimality in the predicted initial path selection, allowing us to capture these variations across participants. When the weights assigned to optimality-relevant constraints are strong compared the magnitude of sources of variability, and when the decision criterion is chosen sufficiently conservatively, the optimal path will be selected by a drift-diffusion process with a probability approaching 1. We explored what factors differed between participants who were more or less accurate in selecting the optimal path and found that more accurate participants showed less trial-to-trial variability in the overall drift direction and placed greater weight on the relevant constraints than less accurate participants. These differences were also accompanied by differences in how well the participants conformed to the ideal pattern of constraint weighting in our experiments (Fig 6D). More accurate participants showed more balanced weighting of the relevant myopic and future constraints, and placed less weight on the irrelevant final goal compared to the less accurate participants. Thus, greater approximation to optimality was not only associated with a stronger overall signal-to-noise profile in the decision process, but also with a greater degree of optimality in weighting of different constraints.

The overall pattern of responding as well as the pattern of individual differences support the view that goal-directed planning can usefully be viewed as a weighted constraint satisfaction process, and that optimizing planning can be viewed as occurring through adjusting the weighting of constraints toward values that favor ones relevant to selecting the optimal action. In this context, it is important that even participants who made overall highly accurate choices showed some imbalance in their weighting of the

myopic and future constraints and placed some weight on the irrelevant final goal. This consistency across groups of higher overall accuracy and lower overall accuracy again suggests that both groups relied on the same constraint weighting mechanism, but that participants in the more accurate group modulated their constraint weightings more successfully.

We have focused on understanding the weighting of different constraints leading to the very first action, but our account naturally extends to multi-step decision making over time. In the more complex task, the initially-irrelevant goal information would eventually constrain the path after the bottleneck location. In general, generating consecutive steps in this task may be supported by a weighting shift as decision making progresses, so that, for example, as a person proceeds to and through the bottleneck leading to the passageway containing the goal, they will come to weight the final goal more strongly to support subsequent action selection.

Our results echo the simultaneous consideration of immediate and future constraints found in other cognitive domains (e.g., typing and speech production), as similarly revealed through both optimal and suboptimal aspects of participants' behavior when planning and executing a sequence of actions [16, 25]. 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 at least the myopic bias 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 proximal to the immediate action at hand (accounting for the myopic bias) while allocating successively decreasing weight to future constraints (accounting for the reduced weight of the relevant future constraint and even less weight of the irrelevant final goal). Optimizing goal-directed planning can thus be further thought of as eliciting an adjustment from a default or baseline future-discounting weighting toward the optimal weighting values in a given task setting.

It should be noted that the time course and outcomes that are captured by the drift-diffusion model could arise from a process that involves several alternations of attention to the different constraining factors in a situation, as constraint influences accumulate before the decision bound is reached in the model. Some models of preferential choice include such alternations [26, 27], and eye movement data provides evidence that such alternations do occur [20]. Viewed in this light, the overall weightings we obtained from the DDM could reflect in part the fraction of the decision time devoted to attending to one or the other sources of constraint, with the variability parameters of the DDM approximately capturing the variability in the selection and timing of attentional shifts. In this context, it is interesting to note that attention did not shift in an all-or-none fashion with shifts in eye position in the preferential choice between alternative food items [20]. Instead, the weighting of the fixated choice alternative was greater than the weighting of the other alternative, so that the extent of the weightings increased and decreased as the eyes moved. In future work, it would be possible to investigate whether a similar effect would be observed in planning situations such as the one we have explored in our experiments. In any case, it is worth emphasising that, even if there is attention shifting underlying the combined influences of different sources of information, the overall outcome of the decision process still reflects the weighted influences of multiple constraining factors.

A weighted consideration of multiple aspects of a situation, especially including some that are irrelevant, can sometimes lead to suboptimal decision outcomes in the context of a particular task. Such a tendency, however, might be viewed as a consequence of a more global, perhaps evolutionary, adaptation. Studies have separately suggested that optimal task decomposition and decision suboptimalities can both result from optimizing meta-level costs associated with representation and computation [7–9, 11, 28, 29]. One of

the motivations for the idea that humans might consistently adopt a weighted constraint satisfaction approach stems from a similar consideration. The computational cost associated with switching between parallel or hierarchical planning processes when facing different situations may in fact be higher compared to the consistent use of parallel processing of multiple exploitable factors, which would result in approximately optimal behavior as long as information is appropriately weighted. The simultaneous consideration of multiple pieces of potentially relevant information can also be useful in problems where an exact hierarchical decomposition may not be possible, and also support the flexible incorporation of new information as it arises in a dynamically-changing environment. The myopic bias that is suboptimal in our task context may also 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 sensitivity to meta-level planning costs [29, 30]. The suboptimal consideration of the irrelevant final goal can also be seen as a rational or at least natural exploitation for its possible values, considering that future information can often be relevant to the choices we make in the early stages of complex task situations, such as the packing stage of planning for a long trip, as mentioned in the introduction.

An important future step will be to examine whether a weighted constraint satisfaction mechanism also supports more complex human problem solving beyond the navigational or other everyday tasks we have considered up to this point, for example, in tasks that require a larger number of intermediate steps or more abstract forms of reasoning. 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 [31]. We may therefore expect a weighted constraint satisfaction mechanism to also underlie more advanced forms of reasoning, with immediate and long-range information (e.g., known conditions and the goal statement) being simultaneously processed to facilitate the finding of important intermediate steps, while a strengthened weighting of a subset of the constraints marking a temporarily graded focus on an intermediate subtask (e.g., proving a lemma).

It is worth noting 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 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 the same two constraints were the only ones relevant to selecting the initial step. This may have contributed to the equal non-decision time cost found across both task conditions. On the other hand, the final goal on the base trials directly defines the future constraint. 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. The visual salience of the final goal may have also strengthened its biasing influence. Future work is needed to fully disentangle the semantic salience of a future goal from its perceptual properties.

The drift-diffusion framework has illuminated human cognition across multiple domains, including perception, memory, and decision making [1, 3, 20–22]. In our tasks, it has provided a useful characterization of the constraint weighting dynamics of the underlying decision process. Existing work on planning and hierarchical computation in a variety of tasks has leveraged probabilistic models or the reinforcement learning framework to model the hierarchical task decomposition approach [6, 8, 10, 32]. One particularly exciting direction for our own future investigations will be the exploration of contemporary attention-based neural network architectures as a mechanistic instantiation of the process of assigning and using weightings to various factors in a decision making context [33, 34]. These architectures have been used for machine translation, in which

weighted attention to relevant words in a source sentence in one language are used to guide choices of words to produce in the translation [34]. Such models adapt their attention weights for each successive item in an output sequence and acquire sensitivity to latent hierarchical structures [35]. We are excited by the prospect of adapting such models to the goal of understanding how humans generate sequential actions and acquire sensitivity to latent hierarchical structures in a wide range of task contexts, and of the potential convergence between such an approach and other frameworks [36].

Conclusion

In summary, we propose that goal-directed decision making can be characterized as a parallel, weighted constraint satisfaction process. In our experiments, simultaneous and appropriately-weighted consideration of multiple constraints helped account for the detailed choice and response time patterns from human participants, both when a simple maze navigation task was presented alone and when it was embedded in a larger maze. In both task settings, participants adapted their weighting of the relevant myopic and future constraints in the direction of optimizing their initial path choices, but deviations from optimality revealed gradations in the weighting of optimality-relevant constraints and revealed suboptimal consideration of optimality-irrelevant factors which biased decision outcomes. These results suggest that human choice behavior can be usefully viewed as a context-sensitive modulation of the weighting of multiple constraints, which can both exploit multiple pieces of relevant information in decision making and produce an approximation to the optimal choice outcomes predicted by a hierarchical task decomposition process.

Supporting Information

Analysis of individual differences. 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 cross validation procedure for fitting the model to data from the two groups 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 analyses were based on the full range of parameter estimates

across the 200 cross-validation folds.

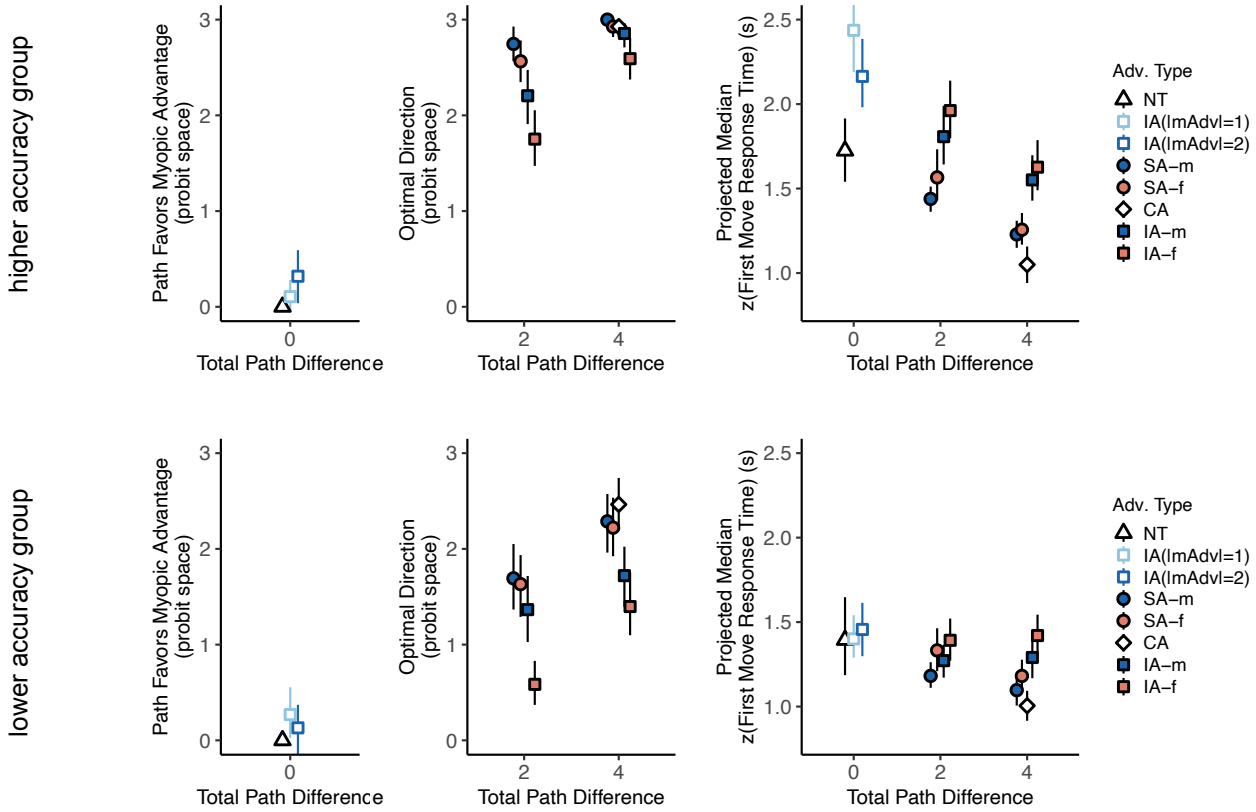

Fig S1. Experiment 1 path choices and response times for groups of higher and lower accuracy. Visualization and notation as in Fig 3. The individual median zscore response times were projected to the second's space using the subgroup-average of mean response time and standard deviation.

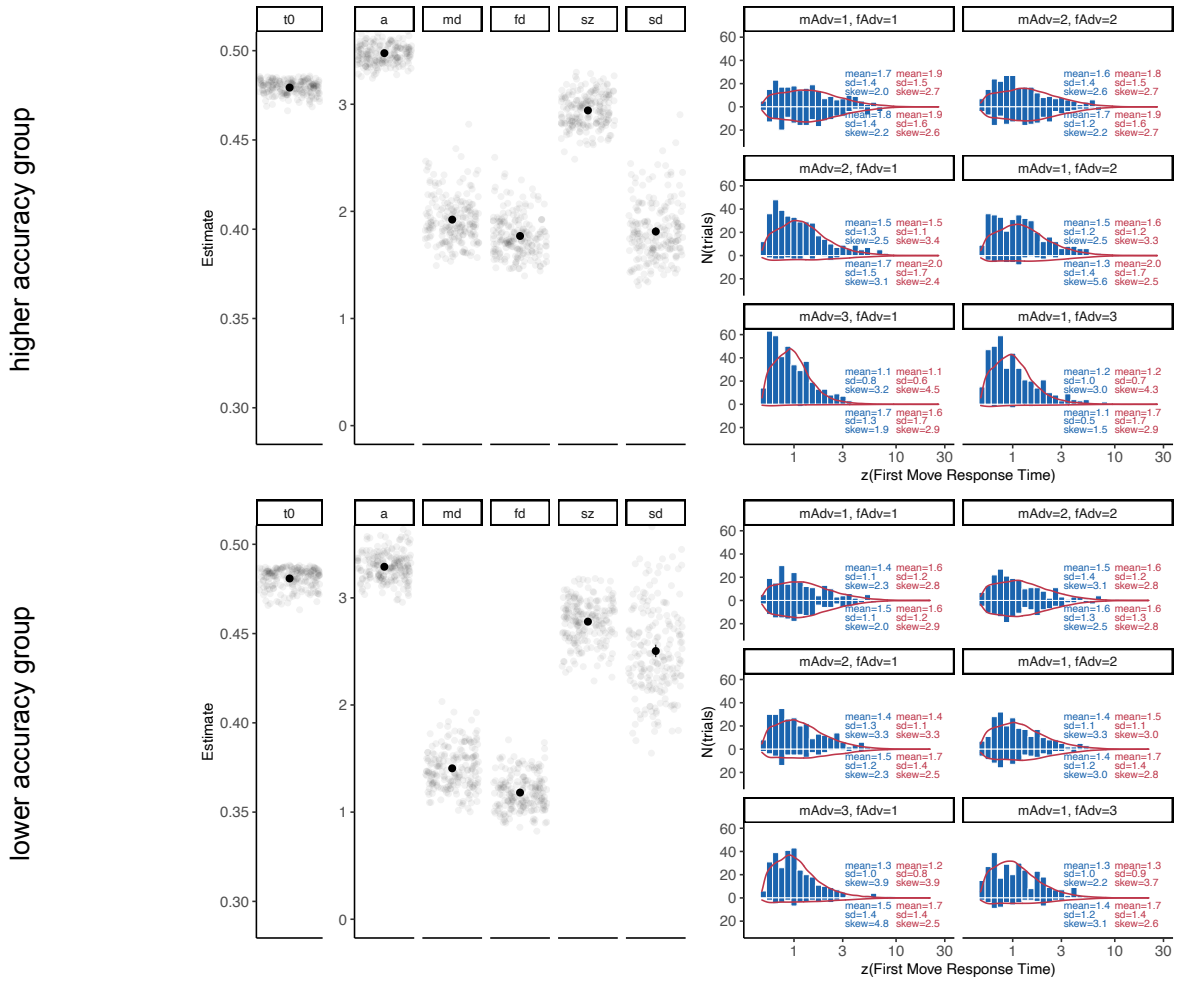

Fig S2. Experiment 1 IA trial responses and drift-diffusion model fits among groups of higher and lower accuracy. Visualization and notation as in Fig 4B and 4C.

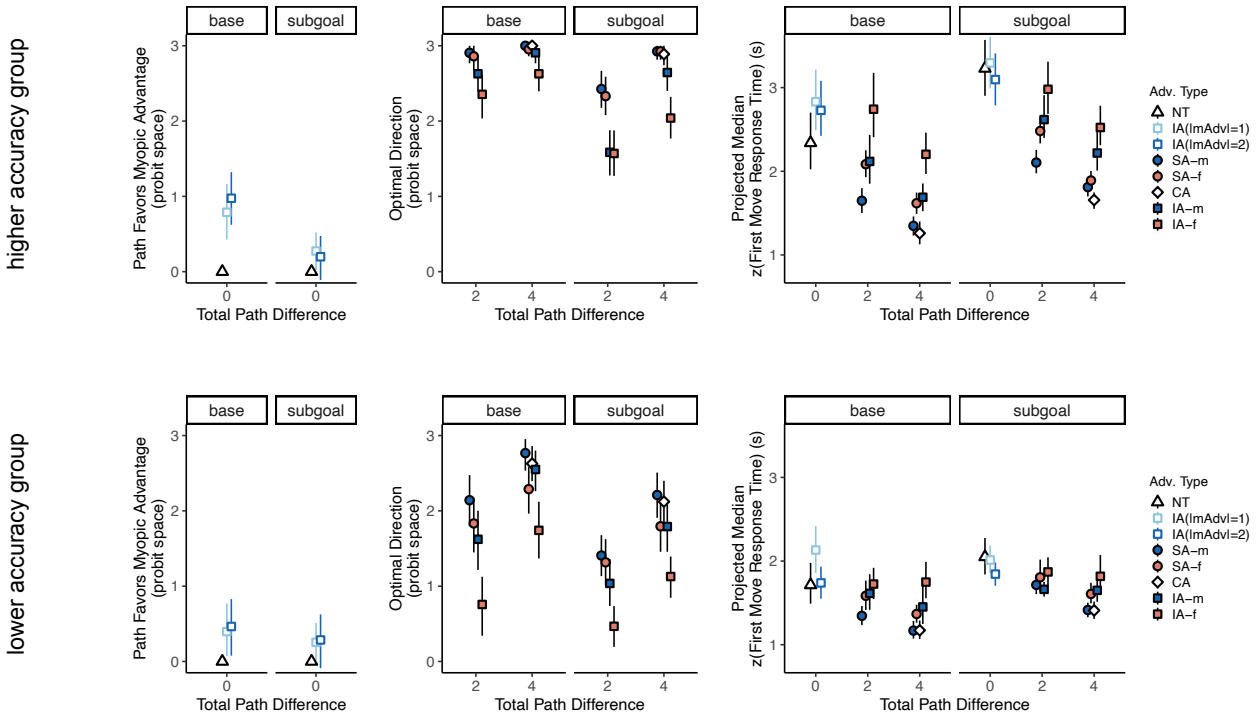

Fig S3. Experiment 2 path choices and response times for groups of higher and lower accuracy. Visualization as in Fig 5. The individual median zscore response times were projected to the second's space using the subgroup-average of mean response time and standard deviation.

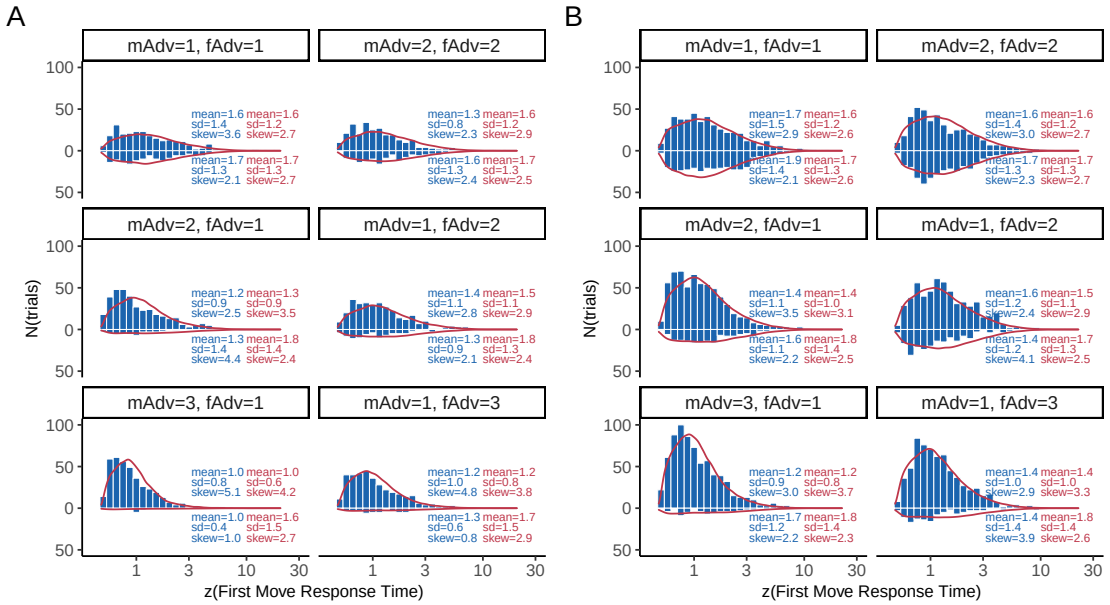

Fig S4. Experiment 2 model-predicted response times (in red) and empirical response times (in blue). A. Base trials. **B.** Subgoal trials. Visualization and notations as in Fig 4C.

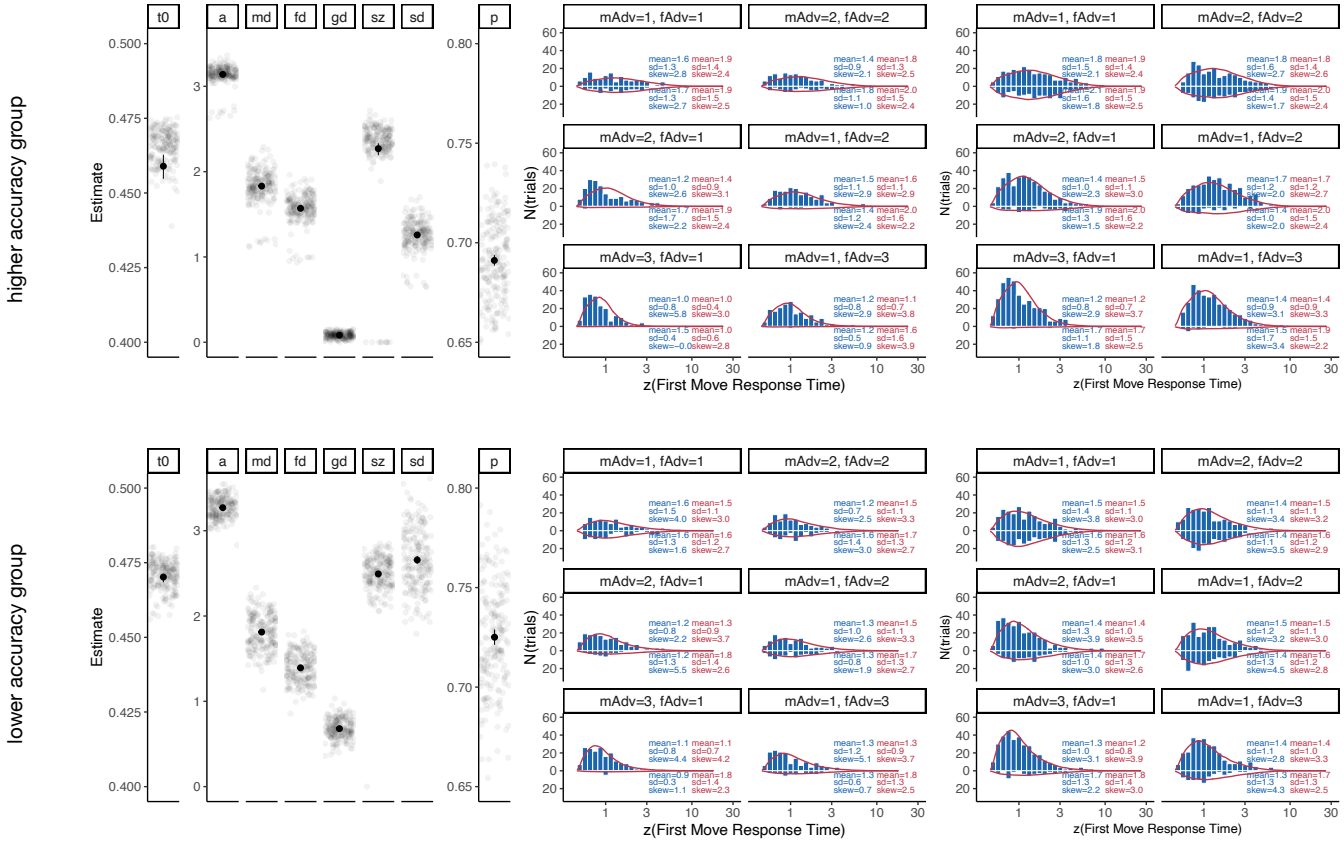

Fig S5. Experiment 2 IA trial responses and drift-diffusion model fits for groups of higher and lower accuracy. Left panel, model parameter estimates. Middle panel, base trials. Right panel, subgoal trials. Visualization and notations as in Fig 6C and S4.

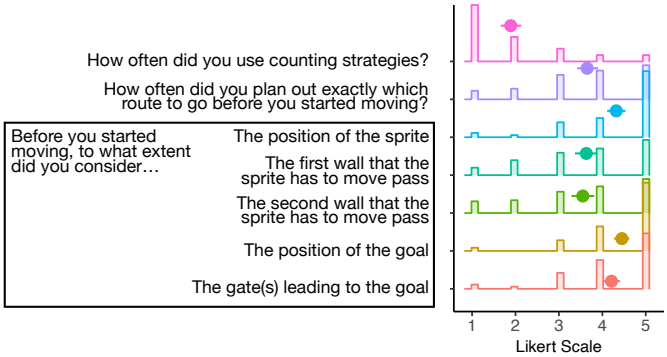

Fig S6. Experiment 2 survey responses. Participants' self-reported considerations of different task elements was in line with the behavioral analyses presented above. In free-form responses, when asked how they came up with the route to reach the goal, their responses consistently suggested that the 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.

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

6. 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.

7. 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.

8. 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.

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. 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.

13. Schapiro AC, Rogers TT, Cordova NI, Turk-Browne NB, Botvinick MM. Neural representations of events arise from temporal community structure. Nature Neuroscience. 2013;16:486–492. doi:10.1038/nn.3331.

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

15. 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.

16. 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.

17. Botvinick M, Plaut DC. Doing without schema hierarchies: a recurrent connectionist approach to normal and impaired routine sequential action. Psychological review. 2004;111(2):395.

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. Krajbich I, Armel C, Rangel A. Visual fixations and the computation and comparison of value in simple choice. Nature neuroscience. 2010;13(10):1292–1298.

21. Polyn SM, Norman KA, Kahana MJ. A Context Maintenance and Retrieval Model of Organizational Processes in Free Recall. Psychological Review. 2009;116:129–156. doi:10.1037/a0014420.

22. Usher M, McClelland JL. The time course of perceptual choice: the leaky, competing accumulator model. Psychological review. 2001;108(3):550.

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

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

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

26. Roe RM, Busemeyer JR, Townsend JT. Multialternative decision field theory: A dynamic connectionst model of decision making. Psychological review. 2001;108(2):370.

27. Usher M, McClelland JL. Loss aversion and inhibition in dynamical models of multialternative choice. Psychological review. 2004;111(3):757.

28. Correa CG, Ho MK, Callaway F, Griffiths TL. Resource-rational task decomposition to minimize planning costs. 2020;.

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

30. 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`.

31. 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.

32. 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.

33. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:14090473. 2014;.

34. Vaswani A, Brain G, Shazeer N, Parmar N, Uszkoreit J, Jones L, et al. Attention Is All You Need. Advances in neural information processing systems. 2017; p. 5998–6008.

35. Manning CD, Clark K, Hewitt J, Khandelwal U, Levy O. Emergent linguistic structure in artificial neural networks trained by self-supervision. Proceedings of the National Academy of Sciences. 2020;117(48):30046–30054.

36. 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.