How to divide and conquer the world, one step at a time

Reka Daniel, Nicolas W. Schuck, and Yael Niv[1]
Psychology Department and Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08542

Many human everyday decisions are parts of plans to reach a higher-level goal; unfortunately, the computational cost of planning increases steeply with the length of the sequence to be planned. So how do we manage to plan, often near optimally, given the limited capacity of our brains? In PNAS, Huys et al. (1) suggest that we might achieve our goals by cleverly fragmenting the decision tree into subpaths and retrieving frequently used subpaths from memory. However, they also warn that prematurely discarding paths that traverse unpleasant states might lead to strategies that are suboptimal overall.

Imagine you live at Charing Cross and want to take the London Tube (subway) to watch a film in Notting Hill (Fig. 1*A*). What will you do? One option is to hop on the yellow Circle Line that takes you directly there. Alternatively, you could change trains at Oxford Circus and optimize physical distance while trading the notoriously unreliable Circle Line for the unbearably crowded Oxford Circus. Solving such multistep planning problems is computationally expensive, yet humans solve them routinely and with seemingly minimal effort. Often, we make up plans as we go, and interlace them with our other plans (as unpleasant as it is to change at Oxford Circus, it will give you a chance to drop off a pair of trousers that need mending at a shop nearby).

Huys et al. shed light on this fundamental and intriguing puzzle of human cognition by showing that we spontaneously and flexibly combine different heuristics to minimize computational cost. In a task with a structure that is rather similar to the tube example (Fig. 1*B*), they placed participants in a location randomly and rewarded them (or levied costs) for each transition to another location. Importantly, although most transitions resulted in only small gains or losses, three of the transitions were punished heavily (−70) and one provided a large reward (+140). The participants' task was to plan a sequence of three to five moves that would maximize their overall returns.

One unique feature of the study of Huys et al. is that they did not instruct participants as to how to achieve this goal but rather just recorded their behavior. This way the authors could observe and model how humans spontaneously generate and combine different heuristics to solve a complex (but naturalistic) planning task. They found evidence for three different strategies: pruning, fragmentation, and memoization. The first one, pruning, has been observed before (2) and refers to the heuristic of heavily discounting rewards that come after a large punishment, such that paths that include negative events are effectively ignored. In our example, you might have had a few really bad experiences wading through the masses at Oxford Circus and therefore decided to stay away from paths that include this hub regardless of how good they are. The second heuristic is fragmentation; when using this strategy, the task at hand is split up hierarchically into several subgoals, thus allowing shorter-horizon planning to each subgoal. Fragmentation is intimately related to the third strategy, memoization, which refers to storing and remembering past paths rather than computing the same old paths anew every time. For example, if you work near Tottenham Court Road, you might use your memoized fragment of how to get there when developing your intention to go to the cinema. On the other hand, if you are more frequently at Victoria station, you might be tempted to use the Circle Line. Huys et al. (1) show that, even in their simplified task, different individuals reuse different fragments, depending on their early experiences with the task.

In addition to delineating these different heuristics in a convincing way, Huys et al. use computational modeling to show that the fragmentation used by participants was near optimal: the decision tree was fragmented such that computational costs were minimized while still allowing participants to choose the best option. However, beyond

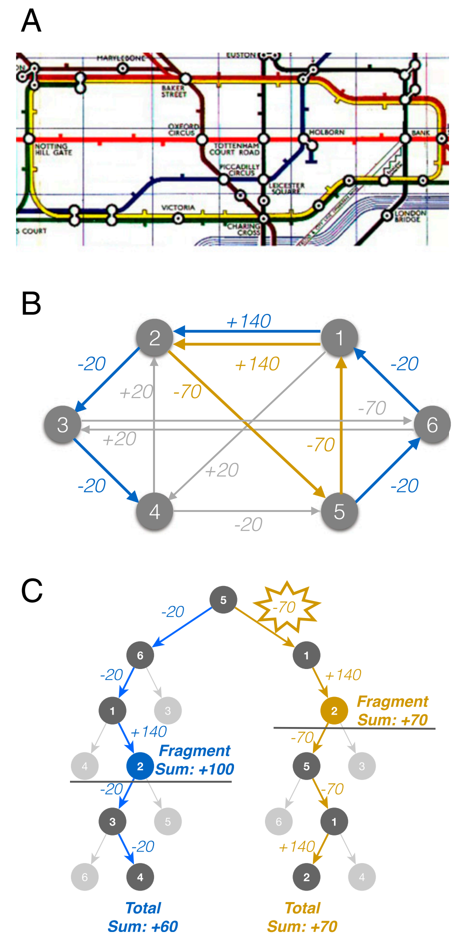

Fig. 1. (*A*) A schematic overview of the London tube (subway), showing the stations mentioned in the text. Navigating a city using public transit is a complex planning problem that humans solve with apparent ease. © Transport for London. Reproduced by kind permission of Transport for London. (*B*) The task structure that Huys et al. use to explore the heuristics that humans spontaneously use to solve complex planning problems. At the beginning of each trial, participants were placed in one of six states (gray circles) and asked to make three to five transitions that would maximize their overall gains (rewards and costs are indicated by the numbers on the arrows). Yellow and blue lines indicate the decision tree described in *C*. (*C*) An example of a five-step decision tree for a trial starting at state 5. The blue path indicates the solution that was preferred by participants. Although this path leads to slightly lower rewards than the optimal yellow path (60 instead of 70 points), the used heuristics (pruning, fragmentation, and memoization; see text) allowed participants to greatly reduce computational cost of navigating the virtual city.

Author contributions: R.D., N.W.S., and Y.N. wrote the paper.

The authors declare no conflict of interest.

See companion article on page 3098.

[1]To whom correspondence should be addressed. Email: yael@princeton.edu.

their computational advantages, these strategies were slightly suboptimal in terms of reaping rewards. For example, imagine starting at state 5 and being asked to plan a five-step journey (Fig. 1*B*). The optimal trajectory (shown by the yellow arrows) is 5-1-2-5-1-2, which travels from state 1 to state 2 twice, thereby winning +280 that offsets the costs associated with this route. However, participants in the experiment most often chose the sequence shown in blue. Huys et al. provide several explanations for why the blue sequence might seem more attractive than the yellow sequence. First, pruning the parts of the tree that involve large negative events (the star in Fig. 1*C*) would mean that planning the yellow sequence would often not go beyond the first step. Second, participants predominantly fragmented the task into subpaths that end with the highly rewarding 1-2 transition (Fig. 1*C*). From this perspective, the blue sequence also seems superior because the total reward until the end of the first fragment is larger than in the yellow sequence. Interestingly, the blue sequence defies "Pavlovian approach behavior" (3, 4), a form of automatic behavior that would lead participants to take the shortest path to the highest reward (5), suggesting that this particular suboptimal choice was due to deliberation and not impulsivity (although in other cases, Pavlovian behavior was sometimes observed) (1, 2). Indeed, to give full credit to the study participants and their computationally frugal mental heuristics, this suboptimal strategy is not far from optimal—it is the second best sequence of states, which is a very respectable result given the complexity of the task.

With their elegant task design, Huys et al. open the door to a host of follow-up questions. For example, the heuristics they discuss all rely on some salient subgoals; that is, they can be characterized by sentences like "avoid going through the large loss" or "aim for the large win." One central question is: How exactly do humans decide how to fragment their environment? Hierarchical fragmentation is an excellent computational shortcut (6), but figuring out the optimal hierarchical decomposition of the task is a computationally formidable task in itself (7). In the study of Huys et al., participants used the salient +140 reward to guide this decision. However, it is unclear if fragmentation would still be near optimal in the absence of such salient cues, or rather, would alternative heuristics emerge instead. Relatedly, it is interesting to ask how these salient cues relate to subgoals and bottleneck states (states that are gateways to other parts of the task space)—concepts frequently discussed in the literature on solving decision problems with a hierarchical structure (8). Finally, behavior and reaction times in the task indicated that if a useful fragment was not available at the current location, participants' strategy was to move one station further along the circle and try to plan from there. One might worry that this strategy is specific to the task structure used here (or to the London Tube!); however, such a "divide and conquer" strategy might actually be more generally useful. Generalizing the interesting results of Huys et al. to other planning domains is an obvious next step.

Huys et al. suggest that we might achieve our goals by cleverly fragmenting the decision tree into subpaths and retrieving frequently used subpaths from memory.

ACKNOWLEDGMENTS. Funding was provided by the Templeton Foundation and Army Research Office Award W911NF-14-1-0101.

1 Huys QJM, et al. (2015) Interplay of approximate planning strategies. *Proc Natl Acad Sci USA* 112:3098–3103.

2 Huys QJM, et al. (2012) Bonsai trees in your head: How the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLOS Comput Biol* 8(3):e1002410.

3 Rescorla RA (1988) Pavlovian conditioning. It's not what you think it is. *Am Psychol* 43(3):151–160.

4 Bray S, Rangel A, Shimojo S, Balleine B, O'Doherty JP (2008) The neural mechanisms underlying the influence of pavlovian cues on human decision making. *J Neurosci* 28(22):5861–5866.

5 Dayan P, Niv Y, Seymour B, Daw ND (2006) The misbehavior of value and the discipline of the will. *Neural Netw* 19(8):1153–1160.

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

7 Solway A, et al. (2014) Optimal behavioral hierarchy. *PLOS Comput Biol* 10(8):e1003779.

8 Botvinick MM (2012) Hierarchical reinforcement learning and decision making. *Curr Opin Neurobiol* 22(6):956–962.