Intention beyond desire: Spontaneous intentional commitment regulates conflicting desires

Shaozhe Cheng[1], Minglu Zhao[3], Ning Tang[1], Yang Zhao[1], Jifan Zhou*[1], Mowei Shen*[1], Tao Gao*[2,3,4]

[1]Department of Psychology and Behavioral Sciences, Zhejiang University

[2]Department of Communication, UCLA

[3]Department of Statistics, UCLA

[4]Department of Psychology, UCLA

*Correspondence concerning this article should be addressed to be addressed to

Tao Gao, Department of Communication, UCLA, Los Angeles, CA 90095. Email: taogao@ucla.edu

Jifan Zhou or Mowei Shen, Department of Psychology and Behavioral Sciences, Zhejiang University, Xixi Campus, Hang Zhou, P.R. China 310007. Email: jifanzhou@zju.edu.cn or mwshen@zju.edu.cn

Abstract

Human mind is a mosaic composed of multiple selves with conflicting desires. How can coherent actions emerge from such conflicts? Classical desire theory argues that rational action depends on maximizing the expected utilities evaluated by all desires. In contrast, intention theory suggests that humans regulate conflicting desires with an intentional commitment that constrains action planning towards a fixed goal. Here, we designed a series of 2D navigation games in which participants were instructed to navigate to two equally desirable destinations. We focused on the critical moments in navigation to test whether humans spontaneously commit to an intention and take actions that would be qualitatively different from those of a purely desire-driven agent. Across four experiments, we found three distinctive signatures of intentional commitment that only exist in human actions: "goal perseverance" as the persistent pursuit of an original intention despite unexpected drift making the intention suboptimal; "self-binding" as the proactive binding of oneself to a committed future by avoiding a path that could lead to many futures; and "temporal leap" as the commitment to a distant future even before reaching the proximal one. These results suggest that humans spontaneously form an intention with a committed plan to quarantine conflicting desires from actions, supporting intention as a distinctive mental state beyond desire. Additionally, our findings shed light on the possible functions of intention, such as reducing computational load and making one's actions more predictable in the eyes of a third-party observer.

Keywords: conflicting desires, intention, commitment, theory of mind, planning

"... but you must bind me hard and fast, so that I cannot stir from the spot where you will stand me... and if I beg you to release me, you must tighten and add to my bonds"

—*The Odyssey*

1. Introduction

Humans are purposive agents who act to fulfill desires. However, these desires often conflict with each other, as they constantly represent competing interests. In the real world, animals and humans typically need to take a sequence of actions to fulfill a desire. But how they deal with situations where they have multiple incompatible desires is a question with deep philosophical roots. As hypothesized by the famous Buridan's ass paradox, an ass placed in the middle of two equally desirable piles of hay may end up starving to death due to indecisiveness. In everyday life, humans also constantly experience such contradictions within themselves. We suffer from conflicting desires as if we have multiple selves: parts of me may want longevity while another part may be addicted to alcohol (Schelling, 1984). This multiple-selves dilemma has long been discussed in philosophy (Schopenhauer, 1818/1969; Elster, 1987), economics (Schelling, 1984), and psychology (Freud, 1922; Lewin, 1931; Miller, 1944; Tversky & Shafir, 1992), yet the mental states involved in generating coherent actions despite the conflicts among desires are still unclear.

1.1. Desire theory of rational actions

In the classical view of rationality in philosophy, "rational decision making is a matter of selecting means that will enable us to achieve our ends. The ends are entirely a matter of what we desire." (Searle, 2003). This view is exemplified by Hume's famous claim that "reason is and ought to be the slave of the passions." Following this view, classical theory in philosophy of mind asserts that desires, despite their incompatibility, are sufficient to directly generate coherent actions when combined with beliefs (Davidson, 1963; Audi, 1974). In this account, desires as the ends are clear, and the main focus of rational action is on how to take actions that *most efficiently* satisfy those desires (Dennett, 1987). While this approach seems intuitive when the desire is simple (e.g., my desire to eat drives my action of opening the refrigerator), how could it explain complex situations where multiple desires conflict with each other?

Modern decision theory offers a solution: take the expectation of all future outcomes evaluated by all desires (Von Neumann & Morgenstern, 1944; Steele et al., 2020). Here, desires are formulated as a utility function that measures the desirability of a state. A rational agent takes actions that can maximize the expected utility (MEU) (Von Neumann & Morgenstern, 1944). More specifically for sequential actions, here we focus on one particularly important extension of MEU, the Markov Decision Process (MDP), which solves how an agent should act rationally over time (Bellman, 1957; Sutton & Barto, 1998). Following the classical view of rationality, the MDP starts with an agent driven by a well-defined reward function, analogous to a human motivated by certain desires. While the reward function defines only the immediate reward the agent can receive in one time step, the MDP is able to provide a rational policy following which the agent can take actions to maximize long-term cumulative rewards. With this definition of rational action, complex desires only need to be formulated as a complex reward function, without introducing any additional mental representation.

The MDP has become an important model in both artificial intelligence and cognitive science due to its importance in formulating sequential actions. In artificial intelligence (Russell & Norvig, 1995), advanced methods for solving complex MDPs, such as deep reinforcement learning (RL), have made remarkable breakthroughs in modeling complex intelligent behaviors. These RL models have not only achieved superhuman performance levels in many challenging games (Silver et al., 2016; Vinyals et al., 2019) but have also solved complex scientific problems, such as predicting protein structure (Jumper et al., 2021) and improving matrix multiplication algorithms (Fawzi, et al., 2022). The success of RL has led to an argument that, to model universal intelligence, “reward is enough” (Silver et al., 2021).

In cognitive science, although studies on human decision-making have repeatedly shown that humans can be irrational in various ways (Kahneman & Tversky, 1982), in theory of mind (ToM) studies that involve recognizing mental states from a sequence of actions, the principle of rationality is considered a fundamental assumption, supported by many developmental studies (Gergely et al., 1995; Gergely et al., 2002; Gergely & Csibra, 2003; Jara-Ettinger et al., 2016; Liu et al., 2017; Liu & Spelke, 2017). In this line of work, the desire theory is the default model for interpreting others’ actions in terms of their beliefs and desires (Fodor, 1992; Wellman, 1992, 2014). It is further assumed that agents choose actions that are most effective in fulfilling their desires, based on their beliefs about the world (Dennett, 1987). This has been demonstrated in an influential study showing that infants' understanding of an agent’s goal was severely undermined if the agent acted irrationally by taking a path that deviated from the optimal one (Gergely et al., 1995). More recently, it has been shown that infants can infer the value of an object based on subtle changes in the cost and reward of actions, by assuming that an agent is rational (Liu et al., 2017). These discoveries support the naïve utility calculus theory of action understanding (Jara-Ettinger et al., 2016). Due to the importance of the rationality principle, the MDP, as a model implementing rational actions driven by desires, has been widely used as the “planning engine” in recent computational modeling studies on ToM, including inferring human mental states in 2D environments (Baker et al., 2009, 2017), understanding social scenes in physics engines (Shu et al., 2021), attributing intelligence to others (Kryven et al., 2021), and modeling social interactions, such as multi-agent cooperation and communication (Kleiman-Weiner et al., 2016; Ho et al., 2016, 2021; Jiang et al., 2021). It is worth noting that most of these studies did not argue that the MDP is indeed the algorithm that humans use to make rational plans. Instead, they treated it as a solver from a practical perspective to capture the fact that human actions are largely rational, based on their beliefs and desires.

Despite this progress, these studies have not yet addressed the challenges posed by the conflicting desires. Most studies assume that the agent is driven by a single goal state as the only source of positive reward and therefore do not take into account the effects of conflicting desires. Consequently, it remains unclear whether the desire theory can effectively capture how humans act under conflicting desires. For example, in situations with two incompatible desires with equal strengths (such as the one highlighted in Buridan's ass paradox), will human decision-making differ qualitatively from the MDP policy? Exploring such deviations can test whether the desire theory is a sufficient model of human actions. It is also important to determine the extent to which MDP can be used as an approximation of human planning in a ToM framework.

1.2. Intention beyond desire: a philosophical perspective

There are many reasons to doubt the sufficiency of the desire theory in explaining human sequential actions. One classical example is demonstrated by *Odyssey.* The ancient Greek hero Ulysses desired to hear Siren’s song, but he also desired to return safely to his home without being seduced by Siren. Faced with these conflicting desires, he chose to voluntarily abandon his freedom and tied himself to a mast to resist Siren’s temptation. Why would not Ulysses just sit back and let these two conflicting desires battle it out to determine his “rational” actions? Why would Ulysses proactively and voluntarily bind himself, thereby abandoning his freedom of action?

Contemporary philosophers have argued that there is a gap between desire and rational actions (e.g., Searle & Willis, 1983; Harman, 1986; Bratman, 1987; Mele, 2003; Brand, 1984; Pacherie, 2006). For example, Searle (2003) argues that one cannot simply sit back and let desires dominate actions. He claims that only a small proportion of human actions are directly driven by desires, such as the actions of a drug addict who is overwhelmingly driven by the desire to take heroin. However, these desire-driven actions are hardly representative examples of human unique rationality. To fill the gap between desire and actions, a distinctive mental state, intention, has been proposed as an intermediate representation that regulates desires for coherent actions (Bratman, 1987). Whereas desire is a motivational mental state, intention is the deliberate state of choosing among potential desires and commitment to a course of action (Searle & Willis, 1983; Harman, 1986). Intention-based actions do not consider the expectations of all future states evaluated by all desires, but are committed to bringing about one fixed future (Bandura, 2001). Therefore, any conflicting nature of desires must be “filtered out” before forming an intention to execute actions: an agent is allowed to desire conflicting things, but not to intend conflicting things (Bratman, 1987; Searle, 2003). In this process, intention serves as a resolution to settle the debate between conflicting desires: the course of actions is committed once the intention is formed, and the execution of those actions will be shielded from the distractions of the unchosen desire.

The distinctive nature of intention has been revealed by philosophical analysis of language, which shows that intentions and desires are semantically different. Unlike desires, whose strength can be quantified as weaker or stronger as defined in a utility function, an intention is a constraint that can only be dichotomously satisfied or not (Brand, 1984). Moreover, in contrast to desires that can be fulfilled in a number of ways, an intention must be “satisfied in the right way” (Searle & Willis, 1983). In a classic example (Chisholm, 1966), Tom, who wanted to poison his uncle, accidentally killed a pedestrian in a car crash on his way to buy the poison. If the pedestrian turns out to be his uncle, Tom’s desire is fulfilled; however, his intention to murder by poisoning is not fulfilled.

Intentions also function differently from desires from a pragmatic decision-making perspective. By committing to a fixed future, intentions increase the predictability of agents' actions, facilitate cooperation between multiple agents and coordinate with one's future self. On the cooperation side, consider which partner you would prefer to go to dinner with: an intention-agent who promises you to show up for dinner unconditionally, or a desire-agent who tells you that whether they show up is determined by the rational policy given how much they want dinner (reward) and how much they hate the traffic (cost). On the side of coordinating with the future self, an agent’s intention promises to bring about one future. This future stability allows an agent to concatenate multiple intentions, with one starting from the fixed future promised by the previous intention. Therefore, the agent can form a partial plan with a long horizon by leaping forward from one promised future to the next while ignoring the gaps between them (Bratman, 1987). The partial planning nature of intention frees humans from being time-slice agents who always start from scratch in their deliberations, and allows for a more distant but still foreseeable future self that can be coordinated with. One seemingly paradoxical feature of partial planning is that a distant future can be more predictable than a proximal future. For instance, while I do not have a plan for the next week, I have already decided to fly to Paris for the next Olympics, followed by a visit to the Musée du Louvre. This commitment to the future is a key aspect of intention explored in the current study.

Of course, the commitment to an intention is not perfect or irrevocable, for two reasons. First, maintaining a commitment requires mental effort, and a weakness of will can cause an intention to be abandoned (Holton, 1999) —- just think of all the New Year's resolutions that are never carried out. Second, in a changing environment, a plan can become outdated and it's necessary to be able to re-plan when new information arrives (Bratman, 1987, Chapter 5). For these reasons, theories of intention do not assume that there will be no flexibility when intention is committed. Instead, they argue that intention would make human action more inflexible than a desire theory predicts, since giving up an intentional commitment and re-planning is different from having no commitment at all.

1.3. Empirical studies of intentional commitment

Empirical studies of commitment were first explored by economists, who found that humans are not fully rational as utility-maximizers due to the fact that their preferences may change over time—referred to as the "changing tastes" problem (Strotz, 1955). To forestall the changing tastes, commitment has been proposed as a regulatory device to deal with the temporal fluctuations of preferences (Thaler, 1980; Schelling, 1980, 1984; Bryan et al., 2010). However, economists focus on explaining consumer behavior and providing high-level commitment strategies, rather than examining commitment as an intrinsic property of intention in ToM, which is important for cognitive psychology.

Psychological researchers, on the other hand, have been looking into how people manage their desires in their minds since the time of Sigmund Freud. Intentional commitment was first studied as a phenomenon of self-control. When faced with temporal fluctuations in preferences, a lack of self-control often leads to inconsistent behavior (Ainslie, 1975). One classic demonstration of self-control is children's ability to suppress current impulses in exchange for greater future interests (Mischel et al., 1972). Although children do not understand internal conflicting desires until they are at least 7 years old (Choe et al., 2005), they appear to demonstrate persistence towards a specific goal as early as infancy (Leonard et al., 2017). To ensure that long-term goals are pursued persistently, people may even impose constraints on themselves, such as setting deadlines (Muraven & Baumeister, 2000; Ariely & Wertenbroch, 2002). These findings suggest that people are able to resolve conflicting desires through self-control. However, this type of self-control has not been directly linked to the intention in a ToM for generating and inferring actions.

Studies on intention from a ToM perspective have primarily focused on investigating the semantic properties of intention. Using introspective self-reports, researchers have shown that intention is a more performable concept and more closely tied to actions than desires (Malle & Knobe, 2001; Perugini & Bagozzi, 2004). A story-comprehension task found that 5- and 7-year-olds, but not 3- and 4-year-olds, can understand that an intention is a plan of action that must be fulfilled by carrying out that plan (Schult, 2002). In addition to measuring the understanding of intentions, motivational psychology studies have investigated the role of intentions in people's goal attainment (Ajzen, 1985; Gollwitzer, 1999; Shah et al., 2002; Shah, 2005; Sheeran et al., 2005; Gollwitzer & Sheeran, 2006; Moskowitz & Grant, 2009; Ajzen et al., 2009). In this context, a goal is defined as a desire-ended point, where desire is the basic motivation to act. Intention, on the other hand, is a planning process that involves commitment to a goal and commitment to a plan of actions (Gollwitzer, 1999; Gollwitzer & Keller, 2016). Commitment to a goal is referred to as goal commitment (Shah et al., 2002), whereas commitment to the plan is often referred to as self-control and self-regulation (Sheeran et al., 2005). These studies typically instructed participants to form intentions toward a future goal and measured goal attainment using follow-up questionnaires when the future arrived. For example, one study asked participants in early December to list projects they wanted to accomplish during the Christmas holiday and measured whether participants had completed their projects four weeks after the Christmas holiday (Gollwitzer & Brandstätter, 1997). They found that participants who were instructed to make more detailed plans had a higher rate of goal attainment, indicating that intentions as plans can regulate human behavior with explicit semantic formations. The function of semantic intention is also supported by studies using priming tasks, which show that intention can also function subconsciously to shield the focal goal from alternative goals. These studies suggest that intention may serve as a self-regulatory mechanism for automatic goal pursuit in lexical decision making (Shah, 2002; Fishbach & Shah, 2006). While these behavioral studies have explored the semantic properties of intention, there is currently no direct psychophysical evidence demonstrating how intentions regulate human sequential actions in a way that differs from the desire theory in a planning task, such as the task typically used in ToM studies.

There are few studies that have discussed intention from a computational perspective. In an early work of logical artificial intelligence, intention was formalized as the selection of a goal for persistent pursuit (Cohen & Levesque, 1990). Another study showed that by modeling intention as a latent mental state, quantitative predictions can be made to capture human moral judgments on the famous trolley problem (Kleiman-Weiner et al., 2015). More recently, intention has also been modeled as optimizing the order of destinations that bring different rewards, which allows the model to focus on one destination at a time and satisfy a sequence of goals (Jara-Ettinger et al., 2020). Their model quantitatively predicted how participants inferred an agent’s subjective costs and rewards after observing its sequential navigation and also how they predicted the agent’s future actions in new situations. However, these studies have mainly focused on the capability of intention; they do not speak directly to whether a desire-belief complex is sufficient for generating human actions, especially when conflicting desires are involved.

1.4. The present study

Here, we investigated the commitment nature of intention in a series of 2D navigation games with conflicting desires, where the intention theory and desire theory should make qualitatively different predictions on human actions. We designed three games to examine three behavioral signatures of intention in human action. Overall, the experimental game setting was inspired by Buridan's ass thought experiment. The participant's task was to navigate an agent to one of two equally desirable restaurants (destinations) located apart from each other on a 2D map.

To explore whether humans spontaneously form intentional commitment, we designed our experiments according to the following principles. First, intention and commitment were not mentioned at all in the task instructions. We simply asked the participants to complete the game as quickly as possible with the least number of steps. Second, in our experimental design, commitment to an intended destination could actually hinder participants' performance. The logic of this design is similar to that of the classic Stroop effect (Stroop, 1935), which demonstrates a spontaneous process by showing how it impairs the main task. If commitment can still be observed, it would suggest that intention is a spontaneous process. Note that our goal is not to show that humans have no ability to control an established commitment, but to show that humans form a commitment even when it can hurt task performance.

Specifically, we focused on engineering critical moments in navigation, where the agent's actions will be qualitatively different depending on whether they were committed to an intended goal or simply trying to maximize the expected utility with respect to both goals. At these critical moments, the desire theory predicts that the agent will always choose the action that leads to the highest expected utility. In cases where different actions have the same expected utility, the desire theory predicts no bias in choosing between them. In contrast, the intention theory predicts that humans would bias their actions toward the previously committed destination. Such a bias can be revealed in two ways. First, among all actions having the same expected utility, humans would be biased towards choosing the one leading to the committed destinations. Second, humans may even violate the MEU principle by choosing suboptimal actions leading to the committed destinations. It is important to note that our aim here is to reveal the bias towards a committed destination, rather than a complete inflexibility of human action after commitment, as we discussed earlier that commitment can be revoked due to a weakness of will or re-planning. Nevertheless, as long as intention exists as a distinctive mental state, its commitment nature will cause a certain degree of inflexibility that can be observed in human actions as a bias toward the committed destination.

To highlight that the MDP, as an implementation of the desire theory, is incapable of capturing human action patterns in conflicting desires, we used an MDP model to run the same task in each experiment and compared its results side-by-side with those of humans. Since human actions will certainly not be perfect but will contain noise, we used a soft-max policy with a temperature parameter β that controls the amount of noise. The parameter β was fitted based on previous literature using MDP to model ToM in a 2D grid world (Baker et al., 2009, 2017), which is closest to our experimental setting (see Appendix for details). It is noteworthy that although manipulating β can change the overall performance of the MDP and the variance of the MDP's actions, it will not lead to any bias in choosing paths with equal expected utility. Therefore, our focus of using an MDP was not to fit the overall performance of human results, but to show that the MDP would not demonstrate any bias in our planning task, as predicted by the desire theory. In most cases, the results of the MDP were straightforward following the theoretical predictions; it served as a sanity check to demonstrate the validity of our experimental design, which creates two paths of equal expected utility. In other cases, where the theoretical predictions are not clear, a detailed comparison between human and MDP trajectories can enrich our understanding of how intention manifests over time.

2. General Method

All studies were pre-reviewed and approved by the Institutional Review Board of the Department of Psychology and Behavioral Sciences, Zhejiang University. The sample sizes were determined before data collection and no additional data were collected after the experiments began. Researchers who collected the data were blinded to the hypotheses of the study during data collection. Experiment 2b was preregistered online at https://osf.io/my92h. Samples were not intended to be representative of any population, because we assumed that the intentional nature of actions applies to all populations. All participants provided informed consent prior to participating the experiments.

2.1. Materials and Procedure

Across all the experiments, the navigation game was presented on 2D maps with 15×15 grids. Participants were instructed to control a hungry agent to reach any of the two restaurants (destinations) *as soon as possible with the least number of steps,* using the four arrow keys (up, down, left, and right) on a standard keyboard. They were instructed that they could earn additional bonuses (course credits or money) after the experiment if they followed the instructions to complete the task efficiently.

In Experiments 1, 2a, and 3, the participants performed the task individually in a single room using a laboratory computer. Each computer was equipped with a 17-inch CRT monitor (100 Hz, 1024 × 768 screen resolution). Participants sat at a viewing distance of approximately 50 cm from the monitor. After completing the experiment, all participants were given a fixed bonus. In Experiment 2b, the participants performed the task online using their own computers. They were given extra monetary bonuses, depending on their performance after the experiment (see each experiment for detailed incentive information).

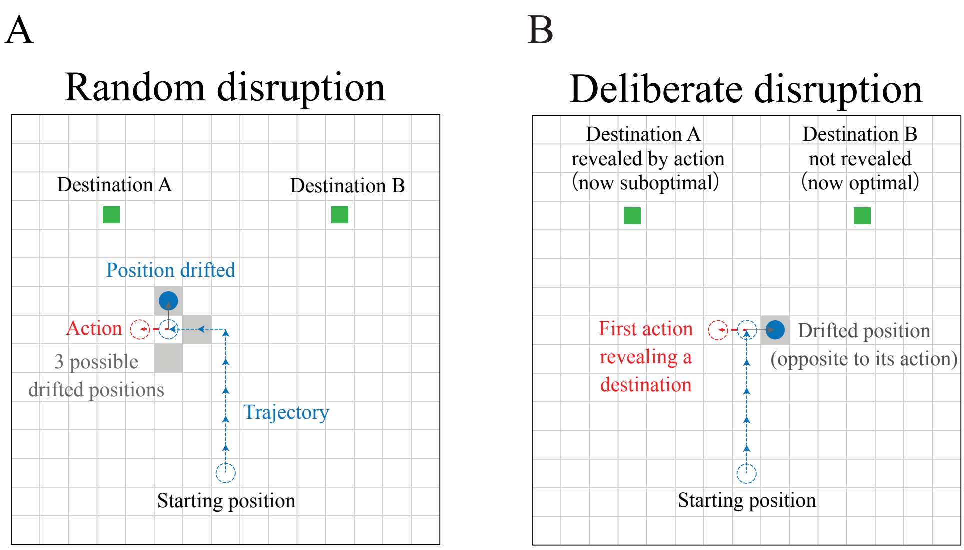
All data, codes, and materials are available at the Open Science Framework (https://osf.io/k5e69/) (Cheng et al., 2021).

3. Experiment 1: Commitment as Goal Perseverance

Experiment 1 explored the "goal perseverance" nature of intention as the persistent pursuit of one future despite unexpected disruptions and setbacks that have made that future suboptimal. A disruption was introduced as a "drift" that nullified the agent's action by placing the agent in one of the nearby cells except for its intended position (Figure 1A). In cases where a carefully engineered drift placed the agent closer to the destination that the agent was not pursuing (Figure 1B), we were particularly interested in whether the agent would continue with its previous goal or change its goal. If human actions are driven by intentions, they should tend to resist re-planning and thus be biased towards the original destination. On the other hand, if human actions are simply driven by desires, they should move towards the closer destination, as it brings higher expected utility.

Figure 1

Design of Experiment 1



Note. Panel A: Design of random disruptions. Both the time step and the direction of the disruptions were randomly sampled. Panel B: Design of deliberate disruptions. Both the time step and the direction of the disruptions were deliberately designed to push the agent away from the destination at the moment it was revealed.

3.1. Method

3.1.1. Participants

As there is no similar prior study, Experiment 1 used the sample size suggested in the social sciences (Simmons et al., 2013). Fifty undergraduate and graduate students (27 female, M_{age} = 21.3, SD = 2.1) were recruited from the participant pool at Zhejiang University. All participants were paid 10 Chinese yuan for their participation, and 5 Chinese yuan for completing the task effectively.

3.1.2. Design and Procedure

Each trial consisted of a map of an agent and two destinations. To mimic the setup of Buridan's ass, the two destinations were vertically or horizontally aligned with equal Manhattan distances to the agent's starting positions. The distances varied from 8 to 12, with a mean of 10. Each map was generated with a randomly assigned distance, and randomly rotated by an angle from [0, 90, 180, 270] degrees (see Appendix B for detailed construction of each map).

There was a total of ten trials. In the first nine trials, just as participants were instructed, the disruption randomly occurred with a 10% probability at every time step, resulting in roughly one drift per trajectory. Across trials, these "random" drifts were pseudo-randomly generated for each participant with one constraint: there were three disruptions in every three trials. There were no constraints on how these disruptions were distributed within the three trials.

In the last trial, without the participants' awareness, the disruption was not random but deliberately engineered. We defined the first-revealing-steps (FRS) as the number of steps at which the agent first revealed its destination by executing an action towards one destination while moving away from the other. The deliberate disruption was triggered at the FRS. It was against the agent's action by placing it in the cell in the opposite direction of its action (see Figure 1B). As a result, the agent ended closer to the destination, which was not revealed by its action.

Participants were explicitly informed that the environment was not fully deterministic: at every step, there was a 10% probability that the agent's action could be disrupted by a random drift that pushes the agent to a nearby cell. A trial ended once the agent reached a destination and was immediately followed by a new trial with a new map.

3.2. Results and discussions

3.2.1. Goal perseverance

The measure of "goal perseverance" was whether the agent's first revealed destination (the original destination) was consistent with their final reached destination (either the original or alternative destination). We measured the bias in choosing between the two destinations by examining whether participants' choice of the original destination was (a) significantly different from 0% when the original destination had a lower expected utility, and (b) significantly different from 50% when the two destinations had the same expected utility.

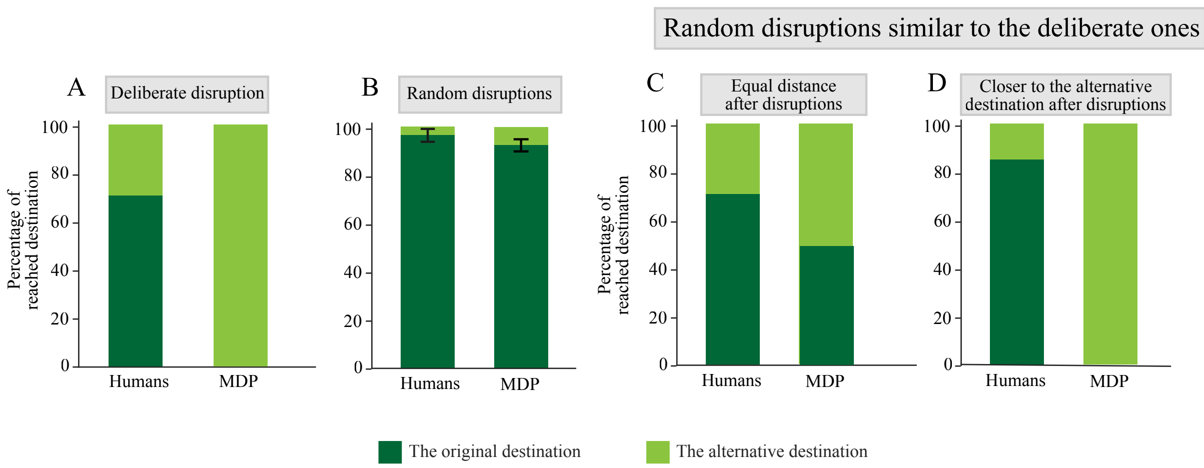
Our primary focus was on the last trial with deliberate disruption, where the original destination had lower expected utility after the disruption. As predicted by the intention theory, the majority of human participants (70%, or 35 out of 50) still tended to reach the original destination (binomial exact test, compared to 0%, $p < .001$; compared to 50%, p

= .003), while none of the MDP models (0%, or 0 out of 50 simulations) did so, which confirmed the prediction of the desire theory (Figure 2A). A direct comparison between humans and the MDP model revealed a statistically significant difference (Fisher's exact test, two-tailed $p < .001$, Cramer's $\varphi = 0.73$). These results suggest that humans, but not the MDP, were biased towards the original destination.

In the first nine trials with random disruptions, both humans (96%; 95% CI [.93, .98]) and the MDP model (92%; 95% CI [.89, .94]) reached the original destination with a high percentage (Figure 2B). This was expected, as in these trials, the original destination still typically had a higher expected utility after the disruptions. To explore whether the difference between humans and the MDP was significant, we first conducted a mixed-effects logistic regression to predict participants' choices (0 = alternative destination, 1 = original destination) from the agent-type (0 = MDP model, 1 = human), with random intercepts for individual participants. This analysis revealed a significant main effect of agent-type (β=0.68; OR = 0.51, 95% CI [0.29, 0.90], $\chi^2(1) = 5.42$, $p = .02$), suggesting that humans were more likely to choose the original destination than the MDP model in these trials. This difference was likely due to the fact that, just by chance, some of the random disruptions worked similarly to the deliberate disruptions, which caused a similar human bias.

Figure 2

Humans commit to the original destination despite disruptions



Note. Panel A: Deliberate disruption. Percentage of participants who reached the original destination in the last trial. Panel B: Random disruptions. Mean percentage reaching the original destination, averaged over the first nine trials. The error bars indicate 95% confidence intervals. Panels C and D: Percentage of original destinations reached when random disruptions were similar to the deliberate ones. Panel C: The agent was equidistant from both destinations after random disruptions. Panel D: The agent was closer to the alternative destination after random disruptions.

Further analysis of the random disruptions revealed that indeed, there were around 10% of these disruptions that made the expected utility of the original destination equal or lower than the alternative destination. First, in cases where the original destination and the alternative destination have equal expected utility after the disruption, humans chose the original destination more often than the 50% chance level, (73.2% or 30/41 trials; binomial test, $p < .01$, Figure 2C), whereas the MDP model showed no preference between the two destinations (50% or 26/52 trials). Second, in cases where the original destination had a lower expected utility after the disruption, humans still showed an above-chance preference for the original destination (85.7% or 12/14 trials; binomial test, $p < .01$, Figure 2D), while the MDP model did not (0%, or 0/10 trials).

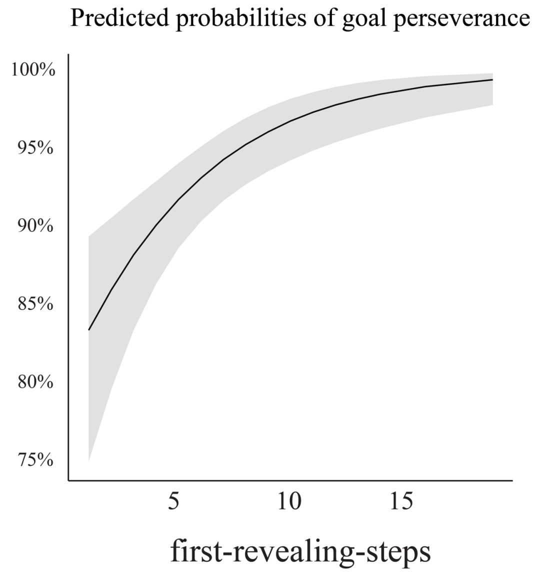
These results from the deliberate and random disruptions together demonstrate the “goal perseverance” nature of human intention as committing to a destination in an uncertain environment.

3.2.2. Does the timing of intention manifestation affect goal perseverance?

We further explored how human goal perseverance interacts with FRS, which measures how long it takes humans to manifest an intention. Intention is a deliberative process that takes time and effort; therefore, the strength of goal perseverance may depend on FRS, with a larger FRS reflecting longer deliberation. A fitted mixed-effects logistic model (predicting goal perseverance from FRS, with random intercepts for subject and map) revealed a significant main effect of FRS ($\beta = 0.20$; OR = 1.22, 95% CI [1.11, 1.34]; $p < .001$), with a larger FRS showing a stronger goal perseverance (see Figure 3). This result is consistent with the theory that intention requires deliberation (Harman, 1986). Our results further show that the duration of deliberation has an impact on the strength of commitment; the longer the deliberation, the stronger the commitment.

Figure 3

Human's goal perseverance increases when their intentions were manifested later



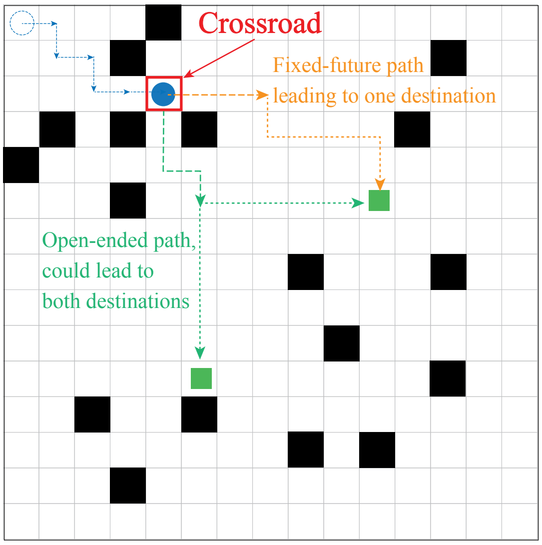
Note. Humans' goal perseverance as a function of the first-revealing-steps. The curve was plotted by the fitted mixed-effect logistic regression model. The error shadows reflect 95% confidence intervals.

4. Experiment 2a: Commitment as self-binding

Experiment 2a investigated the self-binding nature of intention, inspired by the story of Ulysses in Greek mythology who bound himself to a mast to resist the temptation of Sirens' song. While this self-binding phenomenon has inspired philosophical discussions of intentional commitment (e.g., Elster, 2000), it is not typically considered a common aspect of intention and is thought to occur only in special cases (Bratman, 1987, p. 21). In this experiment, we explored whether humans would proactively constrain their freedom by avoiding a path that could lead to many potential futures. The opportunity for “self-binding” was presented at a crossroad with two paths: an open-ended path that could lead to two destinations, or a fixed-future path that leads to only one destination (see Figure 4 for an example map). A desire-driven agent should show no preference, as the expected utilities of taking these two paths are identical. An intention-driven agent may prefer the fixed-future path, which allows the agent to demonstrate its commitment to an intention, thereby shielding the agent from potential distractions. As intentional commitment can take time and effort (Harman, 1986; Bratman, 1987), we also manipulated when the agent would face the crossroad by varying the number of steps it took the agent to reach the crossroad (Figure 5D). We predicted that the self-binding effect would only be observed if there were enough steps for humans to form an intention before reaching a crossroad. As the new map demands more complicated navigations, we also aim to replicate the “goal perseverance” results with these richer map sets by adding a deliberate disruption at the end of all trials, just like Experiment 1 (see Figure 5C for an example map).

Figure 4

Design of Experiment 2



Note. An agent at a crossroad. Agents can either choose a fixed-future path that leads to one destination (yellow arrow) or an open-ended path that could lead to both destinations (green arrow).

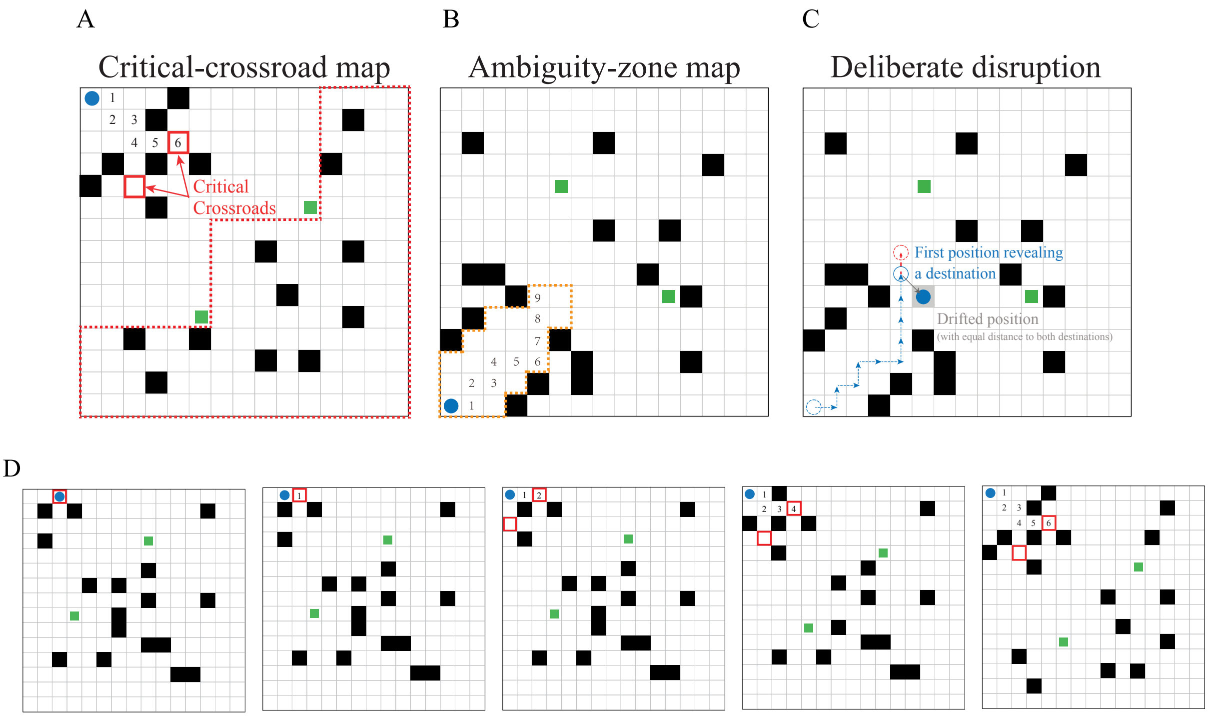
4.1. Method

4.1.1. Participants

A power analysis (*power* = .8, *alpha* = .05) using the effect size (Cramer's φ = 0.73) in Experiment 1 indicated that 20 human participants were sufficient for Experiment 2. Twenty participants (15 females; M_{age} = 20.9 years, *SD* = 2.0 years) were recruited in the same fashion as in Experiment 1. All participants were paid 20 Chinese yuan for their participation, and 5 Chinese yuan for completing the task effectively.

Figure 5

Design of Experiment 2a



Note. Panel A: Certain barriers form two critical crossroads. This is a 6-steps-to-crossroad condition, defined as the length of the shortest path (a sampled shortest path is marked by the numbers) between the agent's starting position and either of the critical crossroads is 6. Extra barriers are scattered within the red region marked by the dashed line in the map. Barriers in this region do not block any of the shortest paths to any destination. Panel B: Barriers that form an ambiguity-zone, marked by the orange dashed line. The numbers represent a sampled trajectory in this zone. Within this zone, the trajectory cannot reveal the agent's destination. Leaving this zone within nine steps lead to a suboptimal path, whereas leaving at step 10 reveals the agent's destination. Panel C: Design of the deliberate disruption. Once an agent revealed its destination, it was immediately pushed back into the ambiguity-zone at a position equally distant between the two destinations. All dashed lines in the maps are only used to illustrate the design and were not visible in the experimental display. Panel D: Sampled maps for all five steps-to-crossroad conditions. From left to right: steps 0, 1, 2, 4, 6.

4.1.2. Design and Procedure

The task and environment of Experiment 2a were identical to those of Experiment 1, except as noted here. Depending on the design of the barriers, there were three types of maps: the critical-crossroad maps, the ambiguity-zone maps, and the random-barriers maps. The total number of barriers in each map was fixed at 18 across all maps.

Critical-crossroad maps. These maps were designed to test "self-binding." At certain cells, the barriers form critical crossroads, where the agent can choose between a fixed-future path and an open-ended path. To explore the development of "self-binding" as a function of time, we manipulated the steps-to-crossroad condition as the number of steps required to reach the critical crossroad from the starting positions, varying from [0, 1, 2, 4, 6] steps (see Figure 5A for a 6-steps-to-crossroad map; see Appendix B for details on how to arrange the blocks to achieve this). To make the critical crossroad less salient and to cover the purpose of the experiment, we also randomly scattered other barriers on the map with the constraint that they did not block the shortest path to any destination. The starting position of the agent was fixed at the corner of the map so that the steps-to-crossroad could be controlled (see Appendix B: Figure S2 for example maps).

Ambiguity-zone maps. These maps were designed to test "goal perseverance." Barriers were carefully placed to create an ambiguity-zone. Within this zone, the agent always lies at an equal distance from the two destinations and, therefore, cannot reveal its destination. The agent can leave this zone at step 10 (see Figure 5B). Similar to Experiment

1, a deliberate disruption was implemented in the last trial, which dragged the agent back into the ambiguity zone once it left the zone. In this way, the agent was again placed in a position with equal distance to the two destinations (Figure 5C). The agent's starting position was fixed at the corner of the map, so that the steps the agent needed to take to leave the ambiguity-zone were fixed.

The map of the last trial with deliberate disruption was fixed as an ambiguity-zone map. All previous trials had a 1/12 chance of using the ambiguity-zone map with random disruptions. The intermixing of different types of maps also makes the critical crossroads less salient.

Random-barriers maps. To further increase the variety of maps so that both the critical crossroad and ambiguity-zone became less salient, we added random-barriers maps, in which all barriers were placed randomly with only one constraint: the agent can reach at least one destination without being trapped by barriers. The starting position of the agent was randomly sampled from all grids.

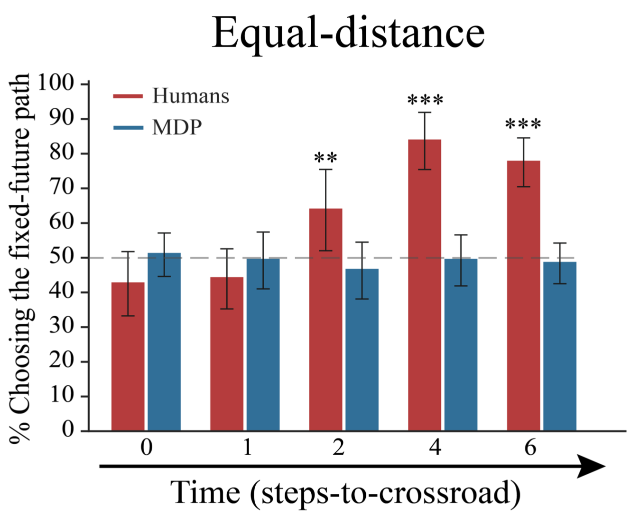
To avoid participants randomly picking one destination without looking at the barriers on the map, the distances between the agent's starting position and the two destinations were not always identical and could vary slightly with the distance-difference sampled from [0, 1, 2]. In all types of maps, constrained by the distance-difference, the positions of the two destinations were sampled with their Manhattan distances to the agent varying from 10 to 20, with 15 as the mean.

Experiment 2 consisted of 289 trials. The first 288 trials were designed to test the effect of "self-binding." These trials consisted of 240 critical-crossroad maps with 48 trials in each of the five steps-to-crossroad conditions (48 × 5 = 240), plus another 24 trials of ambiguity-zone maps and 24 trials of random-barriers maps (240 + 24 + 24 = 288). Similar to Experiment 1, in the first 288 trials, the random disruptions occurred with a probability of

1/15 (roughly 1 drift per trajectory), to push the agent to one of the 7 nearby cells. The last trial was designed to test the "goal perseverance" with deliberate disruption, using an ambiguity-zone map.

Figure 6

Humans constrained their freedom with a bias towards the fixed-future path



Note. Percentage of fixed-future paths chosen as a function of steps-to-crossroad on equal-distance maps. The error bars indicate 95% confidence intervals. $^{**}p < .01$, $^{***}p < .001$. The error bars indicate 95% confidence intervals.

4.2. Results and discussion

4.2.1. Self-binding

"Self-binding" measures whether participants preferred the fixed-future path to the open-ended path. We aimed to test whether the participants' choice of the fixed-future path was significantly different from 50%. We focused on analyzing critical-crossroad maps of equal-distance destinations. Figure 6 shows a pattern in which humans displayed a preference for the fixed-future path, with a mean percentage of choices above the chance level of 50%. Furthermore, human preferences developed over time, reaching a plateau at 4-steps-to-crossroad.

To test this pattern, we conducted a mixed-effects logistic regression model to predict participants' choices (0=open-ended path, 1=fixed-future path) from steps-to-crossroad, with random intercepts included for individual participants and individual maps. The fitted model indicated that overall, human participants chose the fixed-future path more often than the 50% chance ($\beta_{intercept}$ = 0.29, OR = 1.34, 95% CI:1.08-1.66, p = .007). In addition, the main effect of steps-to-crossroad was significant ($\chi^2(4)$ = 128.5, p < .001), suggesting that this preference varied with different steps-to-crossroad.

To further inspect this main effect, we first tested participants' choices against chance level for each step-to-crossroad condition, using estimated marginal means with Bonferroni corrections. This analysis revealed that human participants chose the fixed-future path more often than chance in the 2,4,6 steps-to-crossroad conditions (predicted probability: 2-steps: 65%, 95%CI = [.57, .72]; z = 3.63, p = .003; 4-steps: 86%, 95%CI = [.79, .90]; z = 8.51, p < .001; 6-steps: 80%, 95%CI = [.73, .85]; z = 7.15, p < .001). Post-hoc analysis using follow-up paired comparison with Tukey's correction further indicated that this preference was much stronger in steps-4 and steps-6, compared than in steps-2 (both ps < .001). The same mixed-effects logistic regression was also conducted for the MDP model. As predicted, it showed no preferences in any of the steps-to-crossroad conditions (all ps > .5).

These results show that humans prefer a path that is locked to a fixed destination. Moreover, this preference was not instantly revealed but developed gradually. This is consistent with the theory that, unlike desires, which often arise effortlessly, intention is a deliberate process that requires time and effort (Harman, 1986, Bratman, 1987).

4.2.2. The consequence of self-binding: faster BToM inference

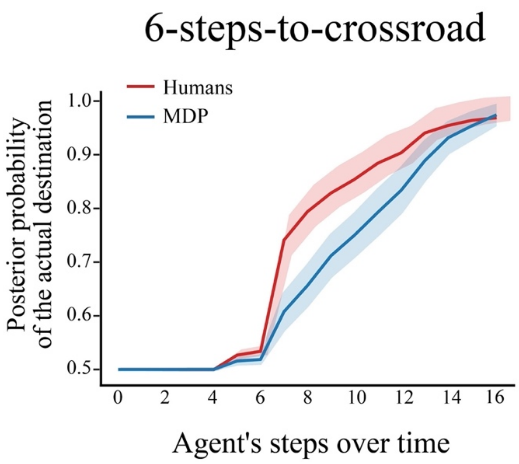
By preferring the fixed-future path, humans essentially exclude paths that might render their potential destinations unclear. One consequence of this self-binding could be that human intentions are easier to read from the perspective of third-party observers. To confirm this, we implemented a BToM model as the observer (Baker et al., 2009), which infers destinations from trajectories in real time, following the inverse planning principle:

$$P(Goal\ |Action_{1:t}\ , Enviroment)\ \propto p(Action_{1:t}\ |\ Goal, Enviroment)\ P(Goal\ |\ Enviroment)$$

The BToM model infers the posterior probability of each destination as the goal that the agent is pursuing, given a sequence of actions successively executed from the beginning to the current time *t*. We focused on the 6-steps-to-crossroad condition, as it has sufficient time steps for humans to form an intention and establish commitment. We predicted that the posterior of goal inference would converge faster when observing human trajectories than when observing the MDP trajectories. The BToM inferred posterior of the actual destination (finally reached) in the equal-distance maps is shown in Figure 7. As predicted, humans revealed their actual destination much faster than the MDP model (cluster-based permutation tests (Maris & Oostenveld, 2007) identified a significant gap between humans and the MDP model from steps 6 to 11, $p < .001$).

Figure 7

Faster BToM inference in human trajectories



Note. The posterior of the BToM inference for the agent's actual destination as a function of the number of steps in the 6-steps-to-crossroad condition. The error shadows reflect 95% confidence intervals.

As this was an individual navigation task without an observer, a faster BToM inference in human trajectories can be viewed as a self-demonstration of intention. This implies that when forcing people to make a stand at a crossroad, intentional commitment will lead to a path that can facilitate the recognition of intention from a third-party's perspective. The implications of this self-demonstration interpretation are elaborated in the General Discussion.

4.2.3. Goal perseverance

The "goal perseverance" result was consistent with Experiment 1: in the last trial with a deliberate disruption, human participants demonstrated a bias towards the original destination (95% or 19/20) by reaching it more often than the 50% chance (binomial test, $p < .001$). As expected, the MDP model was consistent with the prediction of the desire theory by showing no such bias (55% or 11/20) (binomial test, $p = .82$).

In summary, this experiment showed two behavioral signatures of intentional commitment in human sequential actions: "self-binding" and "goal perseverance."

4.2.4. Early commitment avoidance?

In addition, we also noticed an unexpected but intriguing pattern showing that the numerical values of the choice of the fixed-future path were actually lower than 50% in both steps-0 (42.8%) and steps-1 (43.95%), suggesting a possible "commitment avoidance" effect when participants did not have enough time to deliberate. Nevertheless, this numerical difference was not statistically significant, as shown by a fitted mixed-effects logistic regression model with only an intercept (steps-0: $\beta_{intercept} = -0.30$, OR = 0.74, 95% CI:0.51-1.09, $p = .126$; steps-1: $\beta_{intercept} = -0.26$, OR = 0.77, 95% CI:0.54-1.10, $p = .15$). Whether "commitment avoidance" is a real phenomenon was further explored in the following experiment.

5. Experiment 2b: Replicating the self-binding

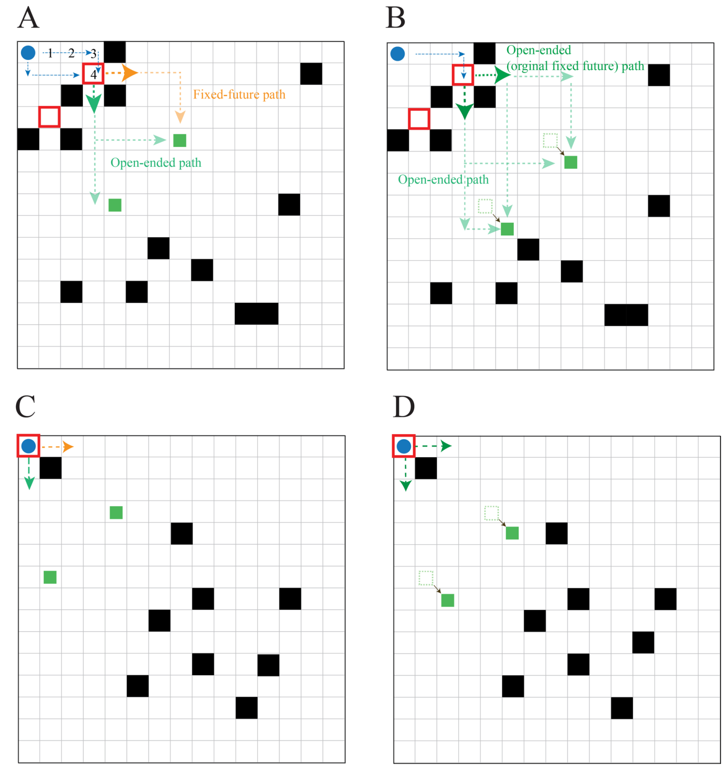
Considering that "self-binding" has only been considered as a function of intention in special cases (Bratman, 1987), it's striking that Experiment 2a showed a spontaneous "self-binding" effect in a simple navigation task without any instructions regarding intention. Given the importance of this finding, in this experiment with a different group of participants (from Amazon Mechanical Turk), we aimed to both replicate the "self-binding" effect with a 4-steps-to-crossrad condition and to explore whether the "commitment avoidance" was a real phenomenon with a 0-steps-to-crossrad condition.

Furthermore, we also aimed to show that the bias towards one of the two paths (open-ended vs. fixed-future) was indeed caused by whether a path allows an agent to commit, rather than other low-level heuristics based on the features of the map. We did this by using a between-participants design. In the Open-Fixed Crossroad condition (Figure 8A), the design of the crossroad was the same as that of Experiment 2a, with one open-ended and one fixed-future path. In the Open-Open Crossroad condition (Figure 8B), the configuration of the map was the same, except that the positions of the destinations were shifted slightly so that both paths at the crossroad were open ended. An intentional commitment interpretation would predict that the "self-binding" effect would only be observed in the Open-Fixed condition, whereas the map heuristic predicted the bias should be observed in both conditions. In addition, we updated the layout of the crossroads to increase the variety of paths from the starting position to the crossroad to test whether the "self-binding" effect could be generalized.

Due to the time constraints of running online experiments, we focused on two time-step conditions that showed categorical differences in Experiment 2a, the 0-steps-to-crossroad condition and the 4-steps-to-crossroad condition. This experiment was preregistered on the Open Science Framework https://osf.io/my92h

Figure 8

Design of Experiment 2b



Note. Panel A: A 4-steps Open-Fixed Crossroad map. It takes the agent (blue circle) at least four steps to reach a crossroad (red rectangle). At the crossroad, the agent has two alternative paths: a fixed-future path (orange dashed arrow) that leads to only one destination, or an open-ended path (green dashed arrow) that could lead to both destinations. Panel B: A 4-steps Open-Open Crossroad map. All the barriers, including those forming the crossroads are identical to those in Panel A, except that the positions of the two destinations all moved one grid horizontally and one grid vertically away from the crossroad, so that the two alternative paths at the crossroad are now both open-ended. As a result, the original fixed-future path (moving right at the crossroad) in Panel A has now become an open-ended path. Panel C: A 0-step Open-Fixed crossroad map. Panel D: A 0-step Open-Open crossroad map.

5.1. Method

5.1.1. Participants

The sample size was determined based on an a priori power analysis using G*Power 3.1, with a presumed medium effect (Cohen's $g = 0.25$) for a two-tailed binomial test (*power*=0.8, *alpha*=.05). We planned to recruit 30 participants for each between-participants condition. Using Amazon Mechanical Turk, we ended up recruiting 31 participants for the Open-Fixed condition and 30 participants for the Open-Open condition. The experiment lasted for approximately 15 minutes. Participants were paid $1 for their participation, with the possibility of receiving an additional bonus of $0.5 to $1.

5.1.2. Design

We used a 2×2 mixed design in this experiment. The crossroad-type is a between-participants factor with two conditions: Open-Fixed and Open-Open. The Open-Fixed condition was identical to Experiment 2a, with the following exceptions. First, there were only two steps-to-crossroad conditions: 0- and 4-steps-to-crossroad. Second, by adjusting the configuration of the crossroad, there were many optimal paths that could lead to a crossroad, with the last movement being either horizontal or vertical (Figure 8A). There were a total of 24 trials from which data were collected, 8 trials for each of the two steps-to-crossroad conditions, plus 8 trials with random maps. The order of these 24 trials was randomized for each participant.

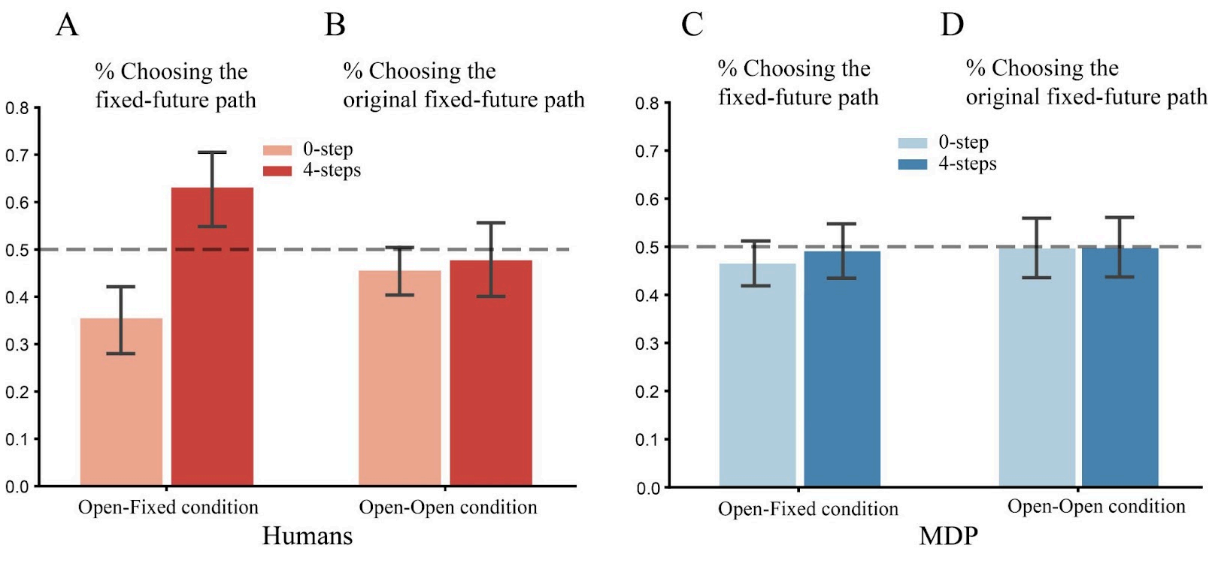
The Open-Open condition was exactly the same, except that the positions of the two destinations were slightly shifted away from the crossroad — one grid horizontally and one grid vertically, making both paths at the crossroads open-ended (see Figure 8B).

5.1.3. Procedure

The procedure of this experiment was identical to that of Experiment 2a, with the following adaptations for online participants: First, participants were asked to complete three practice trials before proceeding to the 24 formal trials. The practice trials were followed by a four-question test to check participants' understanding of the task. Participants who failed any of the questions were excluded from the analysis and replaced. Second, to maintain participants' focus during the experiment, they were given extra monetary bonuses (up to $1) depending on the steps of their trajectories, relative to those of the optimal trajectories. In addition, if participants took a trajectory that was more than 10 steps longer than the length of the optimal trajectory in a trial, a red warning text ("Took too many steps! Please use fewer steps") was displayed on the screen, and the participants had to press a button next to the text to continue the task. We also recorded the time in addition to the total number of steps taken by the participants to complete each trial. We excluded trials from further analysis if these two measures deviated by more than three standard deviations from the mean. We removed 2% (8/496) trials in the Open-Fixed condition and 3% (14/480) trials in the Open-Open condition.

Figure 9

Results of Experiment 2b



Note. Panel A: Mean percentage of human participants choosing the fixed-future paths in the Open-Fixed condition. Panel B: Human results in the Open-Open condition. Panels C and D: Results of the MDP model in the Open-Fixed and Open-Open conditions. The error bars indicate 95% confidence intervals.

5.2. Results and Discussion

5.2.1. Self-binding

The measure of the "self-binding" effect in the Open-Fixed condition was the same as in Experiment 2a. The analysis of the Open-Open condition was identical to that of the Open-Fixed condition, except that the "fixed-future path" was labeled as the "original fixed-future path," as this path was now open-ended in the new map.

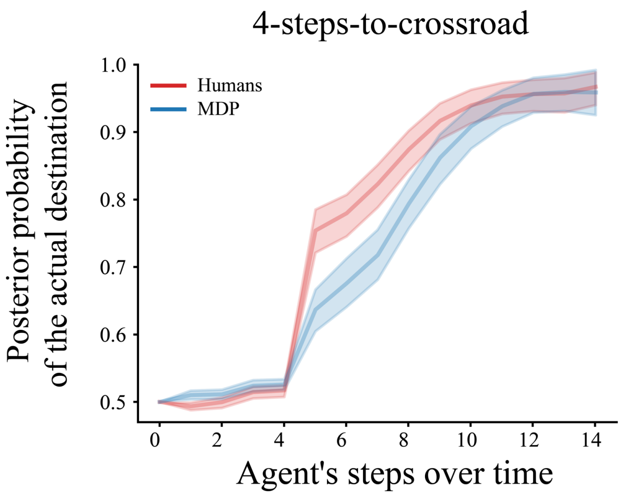
Figure 9 shows that in the Open-Fixed condition, the mean percentage of human participants choosing the fixed-future path was above the 50% chance for the 4-steps-to-crossroad condition, but below 50% for the 0-steps-to-crossroad condition (Figure 9A). In contrast, in the Open-Open condition, the percentage of participants choosing the originally fixed future path was at chance for both 4- and 0-steps-to-crossroad (Figure 9B). To test the statistical significance of this observation, we conducted a mixed-effects logistic regression model to predict participants' choices (0=open-ended path, 1=fixed-future path), with fixed effects for crossroad-type, steps-to-crossroad, and their interactions and random intercepts for individual participants and individual maps. We found a significant interaction between crossroad-type and steps-to-crossroad ($\chi^2(1) = 16.56$, $p < .001$). We further inspected this interaction by conducting a post-hoc analysis using the estimated marginal means with Bonferroni correction. The analysis yielded several interesting results. First, in the Open-Fixed condition with the 4-steps-to-crossroad, the percentage of participants choosing the fixed future path was significantly higher than chance ($M = 0.64$, 95% CI = [0.56, 0.71], $z = 3.45$, $p = .002$). This replicated the "self-binding" effect in Experiment 2a. Second, in the same crossroad condition with 0-steps-to-crossroad, the percentage of participants choosing the fixed future path was significantly lower than chance ($M = 0.34$, 95% CI = [0.27, 0.42], $z = 3.91$, $p < .001$). This result confirmed the numerical observation in Experiment 2a, reflecting a phenomenon that we interpreted as a type of "commitment avoidance." That is, when people did not have sufficient time to deliberate, instead of choosing paths randomly, they preferred a path that allowed them to keep their choices open. Third, in the Open-Open condition, in both 4-steps- and 0-steps-to-crossroad, the percentage of choosing the original fixed-future path did not differ from chance (0-steps-to, $M = 0.45$, 95% CI = [0.38, 0.53], $z = 3.91$, $p = .96$; 4-steps, $M = 0.48$, 95% CI = [0.4, 0.56], $z = 3.91$, $p = 1$). These results suggest that the different bias observed in the Open-Fixed condition was not due to low-level features of the crossroad configurations, but due to whether a path could lead to a fixed destination. The same statistical analysis was also conducted for the MDP model. As predicted by the desire theory, the MDP showed no preference for any path in all steps-to-crossroad conditions (all ps > .5).

5.2.2. The consequence of self-binding: faster BToM inference

Following the same strategy as in Experiment 2a, we also conducted a BToM analysis on the Open-Fixed condition to reveal the consequence of the "self-binding" from a third-party perspective. The BToM inferred posterior of the actual destination (finally reached) was plotted in Figure 10. As predicted, humans revealed their actual destination much faster than the MDP model in the 4-steps-to-crossroad condition (cluster-based permutation tests identified a significant gap between humans and the MDP model from steps 4 to 9, $p < .001$).

Figure 10

Faster BToM inference in human trajectories in Open-Fixed condition



Note. The posterior of the BToM inference for the agent's actual destination as a function of the number of steps in the 4-steps-to-crossroad condition. The error shadows reflect 95% confidence intervals.

Taken together, the "self-binding" and "commitment avoidance" effects demonstrate the richness of how intentional commitment shapes human sequential actions over time. The findings suggest that, at a crossroad, when humans have sufficient time to consider their options, they are more likely to bind themselves to a path that can better demonstrate their commitment. On the other hand, when confronted with a decision that must be made quickly, such as at the beginning of navigation, people tend to choose a path that allows them to retain their options.

6. Experiment 3: Commitment as temporal leap

In this experiment, we explored the partial planning nature as one important function of intention (Bratman, 1987). Agents with partial planning are capable of forming future goals even when the current goal has not yet been achieved. Thus, we hypothesized a "temporal leap" prediction that the agent will be biased towards the very next promised future in the chain of commitments, even though new emerging opportunities at that time have made that future suboptimal.

We tested the “temporal leap” hypothesis with a Pac-Man-like task that required participants to pursue a stream of destinations, with the constraint that only two destinations were available at any given time. This was achieved by removing a destination from the map once it had been reached by the agent, and simultaneously adding a new destination at a random location. In this way, there were always two destinations on the map: one left over from the previous trial (the old destination) and one newly added destination (the new destination). If humans spontaneously make partial plans that include not only the current destination but also the subsequent destination, then a commitment to the old destination has already been formed before the presentation of the new destination. Similar to Experiment 1, such a commitment can be revealed by a bias towards the old destination, when the old and new destinations are equidistant from the agent. The intention theory predicts that humans would tend to resist re-planning and demonstrate a bias toward the old destination. In contrast, the desire theory predicts no bias since the expected utilities of reaching both destinations are the same. In addition, the bias can also be revealed by the human reaction time for initiating the next navigation after reaching the current destination. Only the intention theory, and not the desire theory, predicts that the reaction time of moving towards the new destination should be slower due to the resistance to re-planning that comes with commitment.

6.1. Method

6.1.1. Participants

The sample size was determined in the same way as in Experiment 2a. Twenty participants (12 females; M_{age} = 21.3 years, SD = 2.1 years) were recruited for this experiment.

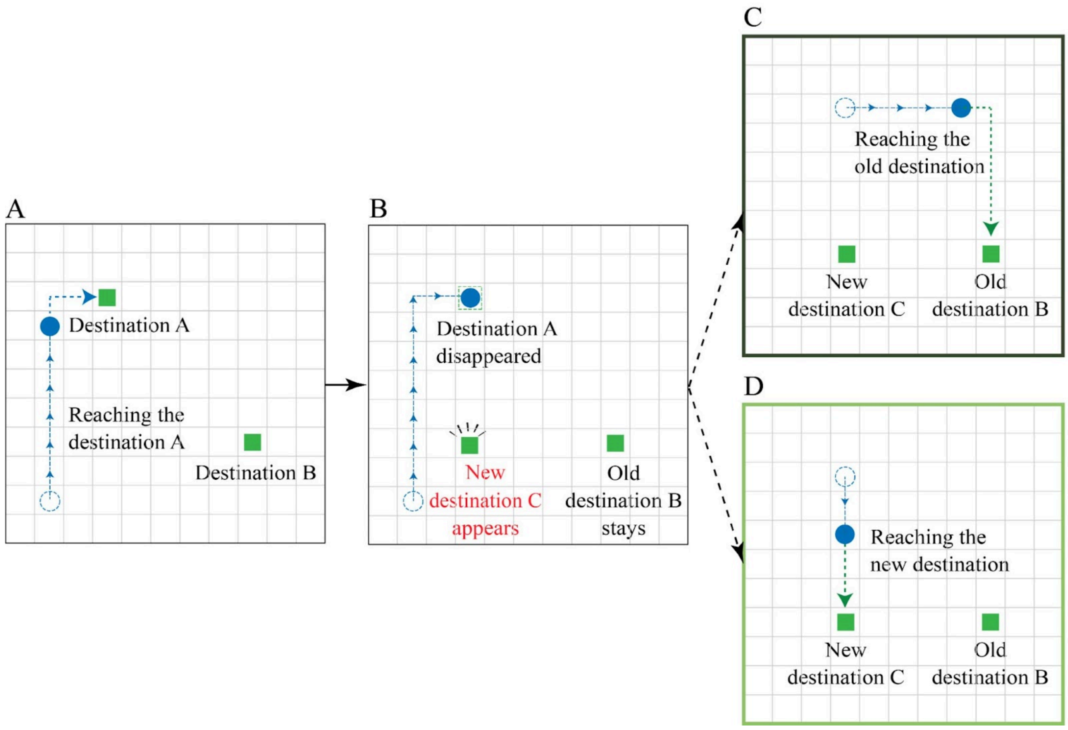
All participants were paid 20 Chinese yuan for their participation, and 5 Chinese yuan for completing the task effectively.

6.1.2. Design and Procedure

The maps in Experiment 3 were similar to those in Experiment 1, with only one agent and two destinations on the map and no barriers. The main difference was that this experiment required continuous navigation through a stream of destinations, with the reached destination disappearing and a new destination appearing, similar to the Pac-Man game. At the start of the experiment, an agent was positioned on a grid with two destinations at an equal Manhattan distance away from it. A trial ended when the agent reached one destination, followed immediately by another trial. In the next trial, the unchosen destination from the previous trial remained in the same location, and a new destination was added to the map (see Figure 11).

Figure 11

Navigating to a chain of two destinations in Experiment 3



Note. Panel A: At any moment, there were always two destinations (green squares) for an agent (blue circles) to choose from. Panel B: Once the agent reached a destination (A), this destination disappeared. The other destination (B) stayed as the old destination, and a new destination (C) appeared at a new location. The distance-difference in this map is -5 (5-10). Panel C: The agent chose to pursue the old destination B. Panel D: Alternatively, the agent chose to pursue the new destination C.

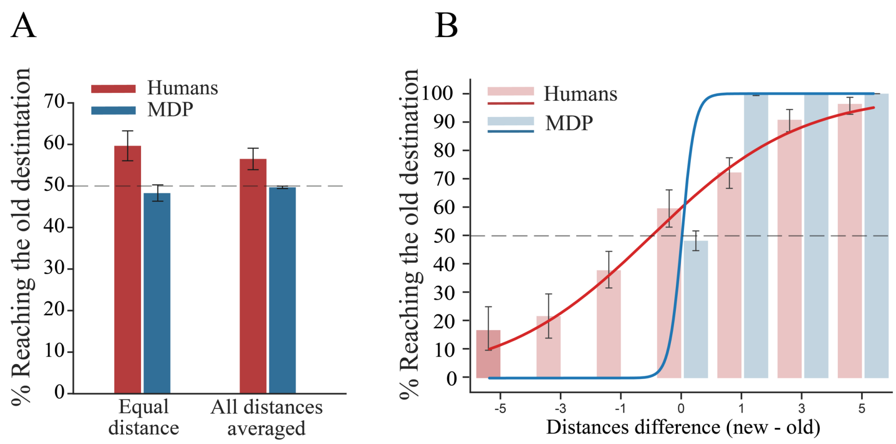
The position of each newly presented destination was systematically manipulated according to the difference between the distance from the agent to the new destination and the distance from the agent to the original destination. Each trial was pseudo-randomly assigned to one of seven distance-difference conditions: [-5, -3, -1, 0, 1, 3, 5], with positive values indicating that the old destination was closer. In addition to the distance-difference constraint, the position of the new destination was chosen to encourage the paths to the two destinations to diverge as much as possible by maximizing the angle 'old destination - agent - new destination.' The environment of this game was deterministic, without any random or deliberate disruptions.

There were 45 trials for each distance-difference condition, for a total of 315 trials, which were randomly grouped into 7 blocks, with each block continuing to present 45 random trials. The first trial in each block was excluded from the analysis as it started the block with two simultaneously presented destinations. There was a self-paced break between the blocks for the participants to rest.

The instructions given to the participants emphasized the importance of task performance in terms of reaching the destinations as fast as possible, without mentioning any difference between the old and new destinations. Specifically, participants were told that they would play a Pac-Man-like game with the following instructions: First, they would control an agent who needs to eat *as many dots as possible* in a given period of time. Second, there are only two dots on the screen at any given time. When a dot is eaten, a new dot appears in a new location simultaneously. Third, they would earn an additional monetary bonus if they followed the instructions by achieving high performance in all seven blocks.

Figure 12

Humans commit to a distant future with a bias towards reaching old destinations



Note. Panel A: Percentage of trials in which agents reached the old destination from the equal-distance condition and averaged across all distance conditions. Panel B: Result of all distance conditions. Percentage of choosing the old destination and curves fitted by logistic regression. The error bars indicate 95% confidence intervals.

6.2. Results and Discussion

6.2.1. Temporal Leap

As shown in Figure 12, human participants chose the old destination more often than the new destination both when focusing on the equal-distance condition (distance-difference=0) and when including all conditions (with the mean distance-difference being 0). To test the statistical significance of these results, we first analyzed the data from the equal-distances condition, using a mixed-effects logistic regression model to predict human participants' choices (0= new destination, 1=old destination) with a random intercept for individual participants. This analysis revealed a significant positive intercept coefficient

($\beta_{\text{intercept}}$ = 0.293, OR = 1.34, 95% CI:1.08-1.65, p = .007), suggesting that participants chose the old destination more often than 50%.

We then analyzed data from all distance-difference conditions with a similar mixed-effects logistic regression model, which also included distance-difference as a fixed effect. This analysis revealed that distance-difference had a significant main effect on participants' destination choices (β = 0.56, OR = 1.76, 95% CI:1.70-1.81, p < .001), showing that as the distance difference between two destinations increased, people tended to choose the closer destination more often. This suggests that humans' choices between the old and new destinations were largely rational in terms of utility maximization. However, the old and new destinations were indeed treated differently, as evidenced by a significant positive intercept coefficient ($\beta_{\text{intercept}}$ = 0.47, OR = 1.59, 95% CI:1.14-2.23, p = .006), suggesting that participants chose the old destination more often than 50% chance. Moreover, as shown in the fitted curve (Figure 12B), with the same absolute value of distance-difference, humans were more likely to choose the further destination when it was the old destination (negative distance-difference) than when it was the new destination (positive distance-difference). We highlighted this asymmetry by focusing on the most extreme distance-difference conditions (-5 and 5). When the far destination was the old one (distance-difference = -5), it was chosen in 17% of the trials; whereas when the far destination was the new one (distance-difference = 5), it was chosen in only 4% of the trials. The difference was significant (paired t-test, $t(19)$=2.87, p = .001, d = 1.07).

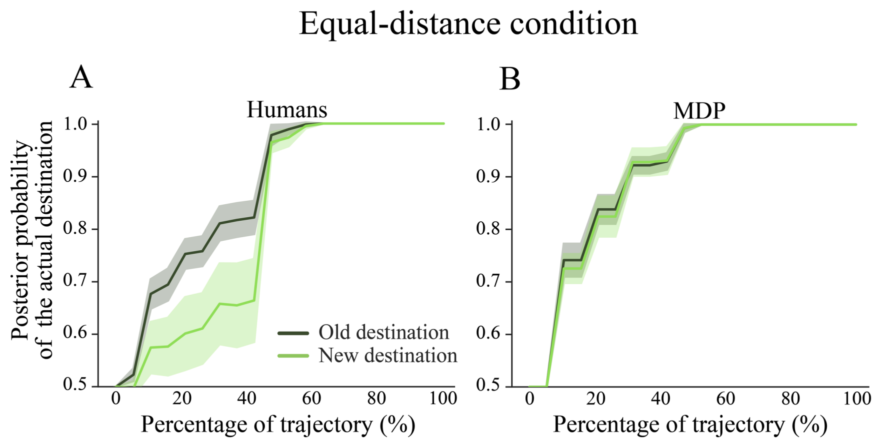
The same mixed-effects logistic regressions were also conducted for the MDP model. As predicted by the desire theory, there was no bias towards the old or new destination when analyzing the equal-distance condition ($\beta_{\text{intercept}}$ = -0.07, OR = 0.93, 95% CI:0.8-1.09, p = .37) or when analyzing all conditions ($\beta_{\text{intercept}}$ = -0.07, OR = 0.92, 95% CI:0.79-1.08, p = .314).

The human bias was also observed in their reaction time of initiating a new navigation. In the equal-distance condition, participants took less time to initiate their first action towards the old destinations (353ms, 95% CI [291ms, 416ms]) than towards the new destinations (405ms, 95% CI [336ms, 475ms]) (paired t-test, $t(19) = -2.33$, $p = .031$, $d = 0.37$). Similar results were also found when analyzing all conditions (to the old destination: 347ms; 95% CI [291ms, 405ms]; to the new destination: 406ms; 95% CI [346 ms, 465ms]; paired t-test, $t(19) = -5.65$, $p < .001$, Cohen's $d = 0.47$). These results suggest that humans were more prepared to choose the old destinations.

We further explored whether this bias could affect a third-party observer's inference of humans' destination by observing their trajectories. If humans were indeed more prepared to pursue the old destination through partial planning, they might also reveal their destinations more quickly when they were reaching the old destination. We tested this hypothesis by applying the BToM model to infer the destinations of human trajectories. We focused on the equal-distance condition, as the length of the trajectory to the two destinations was about the same, we can easily demonstrate their posterior differences over the course of the navigation. As shown in Figure 13A, humans revealed their destinations faster when they were the old rather than the new destinations (cluster-based permutation tests identified a significant gap from 6.7% to 46.7% of the trajectory, $p = .018$). The same BToM analysis was also conducted for the MDP model, as predicted by the desire theory, it showed no such bias (Figure 13B). The BToM results for all distance-difference conditions showed similar patterns, albeit with larger variance (see Appendix, Figure S3). Taken together, the BToM results showed that when humans navigated to a destination consistent with their partial plans, the speed of revealing their intentions was accelerated.

Figure 13

Faster BToM inference of the old destination in human trajectories



Note. The posterior of the BToM inference of the final destination reached by an agent over the time course of the trajectories from the equal-distance condition. Panel A: Human trajectories. Panel B: MDP trajectories. The error shading reflects 95% confidence intervals.

In summary, this experiment demonstrated the "temporal leap" nature of human intention as forming partial plans with a spontaneous commitment to a distant future, even before reaching a proximal one. This commitment can be revoked, so that humans can re-plan, with the percentage of choosing the closer destination increases when the distance difference between the two destinations increases. However, revoking a partial plan must meet the resistance of the intentional commitment, which can still cause inertia that biases humans towards the old destination. This bias makes them choose the old destination more often, initiate the navigation more quickly, and reveal this pre-committed intention through their actions more promptly.

7. General Discussion

Our results showed that humans spontaneously commit to intention when facing conflicting desires with three distinctive behavioral signatures. The first one is "goal perseverance." Experiments 1 and 2 showed that humans pursued their goals persistently by fighting the disruption and continually pursuing the original one. The second behavior signature is “self-binding,” where humans preferred a fixed-future path without the opportunity to be distracted. It is also noteworthy that this preference gradually emerged over time, which is consistent with the prediction that intentional commitment takes time and effort (Harman, 1986). The last characteristic is the “temporal leap” in partial planning, with which humans plan their actions with a long-term horizon—a future goal is formed while the current goal is still in pursuit.

7.1. Comparing humans with MDP

Throughout the study, we analyzed humans and the MDP results side-by-side in the same way. In most cases, the MDP results confirmed the straightforward predictions of the desire theory that a desire-driven agent should not show “goal perseverance,” “self-binding,” and “temporal leap.” From a technical perspective, this confirmation is important due to the sequential nature of the navigation task. In sequential decision-making, a rational action in a state is chosen based on this action’s expected *cumulative* reward, which is the sum of all future rewards by taking future actions following the agents’ policy (Sutton & Barto, 1998). Therefore, the setup in one region of the map may subtly influence the actions far from that region. Due to the existence of action noise and complicated barriers in our tasks, it is reassuring to see that the MDP results confirmed all the null effects predicted by the desire theory, suggesting the significant results from humans were not due to unexpected side effects of the task setup. From a theoretical perspective, algorithms following reward maximization have increasingly become tools for understanding human mental states through their actions in both adults and developmental studies (for a review, see Jara-Ettinger, 2019). The discrepancy between the MDP and human results systematically demonstrated in our study suggest that future studies should be cautious when using MDP as an approximation of human planning, especially when multiple resources of rewards are presented as conflicting desires.

7.2. Intentional commitment: A bug or a feature?

Our study showed that the human mind regulates conflicting desires by intention with a commitment to a fixed future, thus deviating from an MDP model that only acts to maximize expected utility. One straightforward interpretation is that these results show a "bug" in the human mind, adding another example to the long list of human irrational decisions — they could not even follow the shortest path to a goal on a simple 2D map without barriers, as in Experiments 1 and 3. They cling to prior inertia and resist re-planning, even when the environmental changes have made their intentions suboptimal.

Alternatively, intentional commitment may be viewed as a "feature" of the mind that serves important functions. It should be noted that planning in the real world can be extremely complex. Such problems involve huge action and state spaces with a long-time horizon—how would you find the optimal solution to achieve a successful and happy life? As nicely summarized in a recent book comparing human and machine intelligence, planning in a human lifetime involves approximately 20 trillion decisions (Russell, 2019). No model, human or MDPs, can claim to find the optimal solution in this vast space. Ultimately, any model would have to impose certain constraints to simplify the process of solving problems at the cost of being suboptimal in some edge cases. In this way, intentional commitment may serve as a constraint for two purposes: (a) computational efficiency, by allowing the model to make simplified decisions, and (b) facilitating coordination, by providing a predictable future to support one's coordination both with future selves and with other people. These functions are not necessarily mutually exclusive, and both may contribute to the value of intentional commitment as a feature of the human mind. They may jointly explain the various behavioral signatures observed in this study.

7.2.1. Computationally bounded rationality

Human beings are resource-bounded agents. In order to make rational decisions in real-time, they must use their computational resources wisely within the constraints of time, energy, and memory (Simon, 1955, 1990; Bratman, 1987). This idea, known as bounded rationality, has been adopted in psychology to explain why humans sometimes deviate from optimal decision-making, particularly in the presence of cognitive biases identified in behavioral economics (Dasgupta et al., 2019; Gershman, 2021). More recently, computational modeling studies have also employed this principle to develop models of efficient, human-like planning that take into account the costs of computation (Lieder & Griffiths, 2019; Ho et al., 2022; Callaway et al., 2022).

This view can explain the "goal perseverance" in Experiments 1 and 2. Intentional commitment allows humans to form a plan with a course of actions leading to a fixed-future. This reduces the demand for real-time online computations, as humans only need to adhere to the plan at each step. As a result, switching to a new goal can increase the computational costs in two ways. First, the switch itself may involve cognitive control that consumes mental effort (for a review, see Shenhav et al., 2017). Second, the pursuit of a new goal requires abandonment of the previous plan and the formation of a new plan involving additional computations. Therefore, to avoid these two additional computations, humans tend to refuse to re-plan, even if the original plan becomes sub-optimal in terms of the length of the path. The same explanation applies to the "temporal leap" effect in Experiment 3. If a partial plan for a destination has already been formed, humans tend to resist re-planning even when the presence of a new destination makes the original destination sub-optimal. In addition, by planning ahead with a chain of intentions, humans can reduce deliberation during online decision-making, as they can simply follow the plan.

The computationally bounded rationality (CBR for short) hypothesis may also explain the “self-binding” effect in Experiment 2. One can reasonably assume that intention can reduce the computational load by focusing on a fixed-future, rather than evaluating the expectations of multiple futures. Therefore, taking the fixed-future path is less mentally demanding because it involves planning for only one destination, whereas the open-ended path requires humans to always evaluate their actions with respect to two destinations, which is more mentally demanding. By taking the fixed future path, humans take advantage of the environment, making one destination objectively suboptimal on that path, so that humans are no longer mentally distracted by it.

However, the CBR explanation of intention also faces challenges. First, while some of our maps involve carefully arranged barriers, others are just plain 2D grids without any barriers. Intuitively, planning and replanning in these simple maps is not computationally demanding. Therefore, the results from these simple maps may not provide the most convincing evidence to support the CBR account of intentional commitment. Parallel to our work, a recent study also found a “goal persistence” effect that is similar to the “goal perseverance” reported here (Chu & Schulz, 2022). In this study, participants were presented with cartoons depicting two moral goals (e.g., helping a hungry or lost kitten) and were asked to choose one. An explicit aim of this task is to minimize cognitive load, both for the plan and goal switch. Despite the simplicity of the task, both adults and children (aged 4-6 years) tended to stick to their originally chosen goal, even when a later-present cost associated with it was higher than that of the alternative. This suggests that even in tasks where the cognitive load is negligible, intentional commitment can still occur. To further reveal the role of computational cost in intentional commitment, it would be interesting to explore how commitment varies as a function of the computational demands of the task in future studies.

In addition, certain results in our study are difficult to explain by the CBR. In particular, the "commitment avoidance" effect in Experiment 2b suggests that when humans did not have enough time to deliberate, they tended to prefer the open-ended path to keep their options open. This runs counter to the CBR explanation that fixed-future is more computationally effective. Moreover, arguably, deciding whether a path is fixed or open-ended also involves considerable computation — participants need to evaluate and compare the expected utility of each destination along each path. It is unclear how the benefit of taking the fixed future path can outweigh the cognitive cost of comparing these two paths in the first place. It may be more computationally efficient if one just randomly chooses a path instead of carefully comparing them. Taken together, these findings suggest that intentional commitment may serve other purposes beyond CBR.

7.2.2. Internalized ToM as an inner eye

Another important function of intentional commitment could be that it makes human actions more predictable. Therefore, it is easier for an intention-agent to arrange long-term plans by coordinating with its future self as well as coordinating with other agents (Bratman, 1987). Unlike many artificial intelligence agents that are often designed to solve specific tasks, humans frequently encounter tasks that are embedded within complex networks. Navigating between these intertwined tasks provides a high degree of freedom; however, excessive flexibility can pose a significant challenge when it comes to coordination. Imagine our ancestors hunting mammoths together; they would not stand a chance unless they all committed to one of the prey simultaneously and persistently. Any flexibility in that commitment might place the entire group in peril. For these reasons, intention and commitment has been studied primarily in the social context across various fields, including philosophy (e.g., Gilbert, 2013), developmental psychology (e.g., Tomasello et al., 2005; Siposova et al., 2018), moral psychology (e.g., Lombrozo, 2009; Cushman, 2015) and computational modeling (e.g., Kleiman-Weiner et al., 2016; Tang et al., 2022). In the social cooperative context, one not only needs to form an intention with commitment but also to demonstrate that intention to others (Goffman, 1959; Dragan et al., 2013; Shafto et al., 2014; Ho et al., 2016).

Importantly, the cognitive capacity developed for cooperation is not limited to situations where people interact with others. Instead, it can be internalized through socialization and manifest in individual actions (Vygotsky, 1930/1978). Thus, even when acting alone, without external evaluation from others, humans may still reflect on their own actions with this internalized socialization as their “inner eye” (Humphrey, 1986/2002). In collaborative settings, demonstrating one’s intentions is as crucial as understanding others’ intentions (Tomasello, 2010). To gain an evolutionary advantage, it is essential to turn ToM inwards to make one's own mind more readable from an objective perspective. This involves online monitoring of the consequences of one's own decision-making process, such as what mental states one's own actions will convey in the eyes of others. In this sense, ToM is being internalized as a type of metacognition (Flavell, 1979; Bandura, 1989). Indeed, it has been argued that metacognition originates from monitoring the minds of others and is built on internalized ToM (Gopnik, 1993; Carruthers, 2009; Heyes et al., 2020). Our results support this internalized ToM hypothesis: even in an individual task, human actions concern not only their efficiency, but also how their intentions will be recognized by ToM as an internalized third-party observer. We now interpret our results from this perspective.

The “goal perseverance” effect in Experiments 1 and 2 suggested that humans maintained a clear intention by not changing the destination during the course of their actions. This was evidenced by the humans' higher consistency between the destination first revealed by the agent and the final destination they reached. This consistency can support a third-party observer's prediction of an agent's intention once it is revealed. Moreover, a follow-up study has further supported this interpretation by showing that the "goal perseverance" effect becomes stronger when an actual third-party observer watches the participant doing the task or when the participant performing the task in parallel with another participant in the same map without any interactions (Cheng et al., 2022). Similarly, in Experiment 3, by forming partial plans for the distant future, humans made their actions more predictable in the sense that a third-party observer could already predict humans' future destinations in advance.

The most delicate results on how human planning interacts with ToM come from Experiments 2a and 2b, in which we used carefully designed crossroads to probe the mental process of choosing different paths. At the crossroad, humans were forced to choose between two paths that were equally efficient in terms of expected utilities, but different in the eyes of a third-party observer. The fixed-future path allows for a clear interpretation of intention, whereas the open-ended path delays such a clear interpretation. In addition, the preference of paths may not be fixed but may change over time, as the formation of intention is a deliberative process that takes time and effort (Harman, 1986). The relationship between deliberation, commitment, and their consequences is nicely captured by the ancient Chinese philosopher Laozi, who states that "those who make promises lightly should seldom be believed." This statement implies that commitments have a binding force that constrains human behavior; therefore, one should not make commitments hastily without proper deliberation. Consistent with this perspective, Experiment 2a showed that humans tended to demonstrate intentional commitment only after they had taken sufficient steps to form an intention. In Experiment 2b, we not only replicated this "self-binding" effect at a late stage of navigation, but also zoomed in on the early stage of navigation, which showed a significant "commitment avoidance" effect, in which humans preferred the open-ended path when the crossroad was approached immediately. This flip-flop of preference cannot be explained by the desire theory, which should be indifferent to the difference between the fixed and open-ended paths, as confirmed by the MDP results. Instead, human results indicated that humans did commit, but they did not do so lightly, as Laozi suggested. These results again are consistent with the metacognitive perspective that when humans have not made a decision, they do not just "sit back," as Searle has put it, and let the conflicting desires lead to random choices. Instead, they deliberately kept their choices open until they made up their minds.

Taken together, the internalized ToM offers a perspective that provides a coherent interpretation of the various effects overed here. We show that, from a third-party perspective, recognizing intention from human trajectories is quite different from recognizing intention from the trajectories of a purely desire-driven agent. The behavioral signatures showing how and when humans commit are consistent with the perspective that one function of intention is to make human actions more predictable and interpretable, thereby supporting coordination.

7.3. Toward a computational model of intention

Our results demonstrate that intention is a distinctive mental state in human decision-making. But how does an intention emerge from conflicting desires precisely? There are different approaches to modeling the intention formation process computationally, ranging from simple to complex. The simple, perhaps naive, way is to implement intention as breaking the tie of conflicting desires by randomly choosing one and then committing to it. This approach can certainly model the "goal perseverance" effect (in Experiments 1 and 2) to some extent, since this model would only pursue one designated destination until it has been reached. However, it will not be able to explain the “self-binding” effect, as the open-ended path does not hinder the reach of any destinations. More critically, this random choice assumption may trivialize the question of how an agent can choose randomly in the first place, and fails to connect to some in-depth discussion of how intentions are formed (Harman, 1986; Furstenberg et al., 2015).

7.3.1. Modeling intention for saving computational resources

Another approach to modeling intention lies within the CBR framework. Recent studies have shown that human planning can be modeled as the rational use of bounded computational resources (Ho et al., 2022; Callaway et al., 2022). As discussed earlier, CBR can provide explanations for many of the effects observed here and will very likely play an important role in future work on the computational model of intention. However, there are theoretical challenges in understanding intentions from a purely computational-resource perspective. First, to explain the gradually emerging self-binding effect (Experiments 2a and 2b), CBR must assume that the computational cost of following the fixed-future and open-ended paths vary as a function of the steps-to-crossroad. Such a shift in computational resources has not been indicated in existing literature on CBR. Second, the CBR framework highlights the quantitative difference between intention and desire in terms of computational cost. By contrast, psycholinguistic studies have identified several qualitative differences between intention and desire in terms of semantics (Malle & Knobe, 2001; Perugini & Bagozzi, 2004). For example, while desire can be fulfilled in many different ways, the intention must be satisfied by the right outcome achieved in the right way (Schult, 2002). It is unclear how to derive these qualitative differences just from the quantitative difference in the computational cost. Third, forming an intention requires deliberation which takes time and effort (Harman, 1986). After that, maintaining the intention also requires continuous devotion of effort, without which, the intention will fail due to a weakness of will (Holton, 1999). Arguably, these deliberate processes consume humans' limited computational resources. It is unclear how the additional computational burden of intention can be justified in a framework that emphasizes computational efficiency. These open questions need to be addressed in future studies. It is very likely that a cognitive theory of intention and its prediction of behavior will ultimately be consistent with CBR. However, we argue that in order to arrive at such a theory, it is important to bring in perspectives beyond CBR to uncover the mental representations involved in human planning with intentional commitment. Revealing these representations is necessary for quantifying the computational cost of manipulating them in the first place.

7.3.2. Modeling Intention as Internalized ToM

Alternatively, following the internalized ToM framework discussed earlier, intention can be modeled as a distinctive mental state that functions as a form of metacognition. This modeling approach can be built on the existing Bayesian modeling of ToM. Bayesian ToM assumes that mental states, including intentions, are unknown and have to be inferred from the observed actions using Bayes' rule. To turn this Bayesian ToM inwards as an "inner eye" (Humphrey, 1986/2002) for analyzing one's own actions, one has to assume that humans do not know their own intentions, and have to infer them by observing their own actions. While this intention-after-action assumption sounds counterintuitive, it has been supported by neuroscientific evidence showing that the neural signal to initiate an action occurs *before* humans are aware of the decision to take that action (Libet, 1985; Soon et al., 2008). These findings lead to the theory that the decision itself may not cause the action, but rather severs as a post hoc explanation of one's own action (Wegner & Wheatley, 1999; Wegner, 2002; Pockett et al., 2006; for a recent review, see Pacherie, 2011). Neuropsychological evidence is also consistent with the philosophical arguments that humans may constantly confabulate a reason for their own actions, treating them as if they are governed by a clear-cut intention rather than a complex interplay of various candidates (Dennett, 1996, p. 128).

Critically, the internalized ToM we propose here assumes that while intention formation may involve post hoc explanation, intention itself is not merely a post hoc explanation or an epiphenomenal by-product (for neuroscience evidence, see Haggard, 2005). Our behavioral results demonstrate that intentions effectively guide participants through a course of actions with commitment, even if they may be absent at the beginning of sequential actions. To integrate Bayesian ToM, neuroscientific evidence of intention-after-action, and the behavioral signatures of intentional commitment in the current study, we propose a bootstrapping process between intention and action. In the beginning, as neuroscience suggested, humans act spontaneously without a clear intention. Next, an online internalized ToM observes these actions and explains them with a probabilistic distribution of possible intentions. This probabilistic inference of intention is not epiphenomenal, but will have a real impact on future actions, with the more likely intention having a larger chance in deciding the very next action. This new action will then be observed by the internalized ToM, which will reinforce the intention that best explains it. This positive feedback loop allows the model to converge quickly on a single clear-cut intention that drives future actions. Using ToM for bootstrapping an intention has been successfully applied in cooperation contexts, where a group of agents initially do not know their joint plan, but must use ToM to infer it by observing the group actions (Shum et al., 2019, Wu et al., 2019; Tang et al., 2020; 2022). This bootstrapping process has also been implemented in a recent study to model the "self-binding" effect in the individual task reported in this study (Cheng et al., 2022). It is interesting to note that similar bootstrapping processes can be applied to both cooperative and individual tasks. This is consistent with the view that internalized ToM as a metacognitive process may originate from the ability developed for social coordination (Gopnik, 1993; Carruthers, 2009; Heyes et al., 2020).

7.4. Intention as an exercise of free will?

Finally, there are still certain aspects of intention cannot be adequately captured by computational resources or an internalized ToM perspective. Intention has been considered a component of free will, which distinguishes human rational actions from those of animals (Mele, 2009). While animals may just let their basic desires drive their actions, human desires are insufficient to generate most of the meaningful actions in a civilized society. As Searle (2003) noted, “I can not just sit back and watch the action unfold in the same way as I do when I sit back and watch the action unfold on a movie screen.” Similarly, the authors of this paper can testify that they would never be able to finish this paper if they just sat back and let their desires drive their actions. This phenomenon has been interpreted as the gap between desires and actions, as desires are casually insufficient to generate actions. This gap involves a presumption of "free will," which is the conscious experience that we are acting freely, rather than being forced to act because of some unstoppable desires. When intention fills the gap between desire and action, it is an exercise of free will. In this sense, capturing the free will aspect of intention through mathematical formulation, if plausible at all, remains elusive and is likely to remain a long-term challenge.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Acknowledgments

This research was funded by National Natural Science Foundation of China Grants 32071044 and 31871096. The authors also wish to acknowledge anonymous reviewers for helpful comments on previous drafts.

Data availability

All data, code, and materials are available at the Open Science Framework (https://osf.io/k5e69/).

Appendix A

MDP model

We employed the Markov decision process (MDP) as an implementation of the desire theory following the MEU principle, in which desires are defined as the reward function, and an agent acts to maximize its expected long-term future rewards. The definition of an MDP includes a state space S, an action space A, a transition function, $T(s,a)$; a utility function, $R(s,a)$. The solution of an MDP is an optimal policy π, which takes s as input, and outputs a probabilistic distribution of actions given s, $P_\pi(a|s)$. An agent acts by sampling an action from this distribution. The above definition and the solution of an MDP do not involve a formulation of intention.

State Space

The agent's state was its location, defined as a tuple with 2D coordinates $(X\_coordinate, Y\_coordinate)$. In Experiments 1 and 3, the size of the state space was 225, including each cell in a 15×15 map. In Experiments 2a and 2b, the cells occupied by a barrier were excluded from the state space.

Action Space

In all experiments, the agent can travel one cell in one of four directions: $a \in \{(0,1),(1,0),(0,-1),(-1,0)\}$.

Transition function

The transition function takes state s and action a as input and outputs $p(s')$, a probability distribution of the next state s'. In all experiments, the two destinations were set as the termination states. In Experiment 1, the agent could reach only four nearby states. The agent moved to the cell in the direction of its action with a probability of 9/10, with the other nearby cells evenly splitting the rest of the probability of 1/10 (transition noise). In Experiments 2a and 2b, the noise could push the agent to one of the eight nearby cells. With a probability of 14/15, the agent moved to the cell in the direction of its action, with the other seven nearby cells evenly splitting the remaining probability of 1/15. Experiment 3 was identical to Experiment 1, except that the transition noise was 0.

Reward function

The reward function takes state s and action a as the input, and outputs a scalar as the short-term reward. In all the experiments, the reward had two components: 30 for reaching any of the destinations and, -1/30 for every movement on the map.

Solving MDP

The optimal policy of the MDP was solved by value iteration using the Bellman optimality equation (Bellman, 1957) of the value function V. This was an iterative bootstrapping process.

At time *t*+1, the new value function V_{t+1} is derived from the value function V_t. The optimal value function V^*can be found when this iterative process converges:

$$V^*(S) = \max_a \left[R(s,a) + \gamma \sum_{s \in \mathbb{S}} p(s' \mid s,a) V^*(s') \right] \tag{1}$$

Where γ is the discount factor. In all the MDP simulations, γ was fixed to 0.9, which is a commonly used value.

Policy

The optimal policy can be derived from the optimal value function V^* in two steps:

First, we derived the optimal action-value Q^* from V^*:

$$Q^*(s,a) = R(s,a) + \gamma \sum_{s \in \mathbb{S}} p(s' \mid s,a) V^*(s') \tag{2}$$

Then, we derived a Boltzmann policy as proportional to $Q^*(s,a)$, indicating the probability of taking an action a given its state s,

$$P_\pi(\, a \mid s \,) \propto \exp\big(\beta Q^*(s,a)\big) \tag{3}$$

The Boltzmann policy took β as a rationality parameter. When $\beta \rightarrow 0$, the agent acts in a more random way; when $\beta \rightarrow \infty$, the agent chooses the action greedily based on the optimal Q-value. Here we chose β = 2.5 following previous studies that modeling human ToM with MDP (Baker et al., 2009, 2017). With this value, the action selection will be dominated by the maximum $Q(s, a)$, but still deviates from it with a small probability, to capture the fact that human decision-making is not entirely rational.

Bayesian theory of mind (BToM)

We used BToM (Baker et al., 2009) to infer the agent's destination over time. As there were only two destinations, we only needed to plot the posterior of the destination actually reached by the agent, denoted as $P_{reached}$. The posterior of the destination not reached was always 1- $P_{reached}$. Given an agent's trajectory (the state-action pair up until $T \geq t$), the posterior of the agent's destination was proportional to the product of the action likelihood and prior probability of the destination:

$$P_\beta(destination|action_{1:T},\ state_{1:T}) \propto \prod_{t=1}^{T} P_\beta(action_t \mid destination, state_t) * P(destination) \quad (4)$$

The action likelihood function $P_\beta(action_t \mid destination, state_t)$ was derived from an MDP policy, which was similar to the policy in Equation (3), except it only considered one destination as its goal. The initial $P_0(destination)$ was set to 0.5 for both potential destinations.

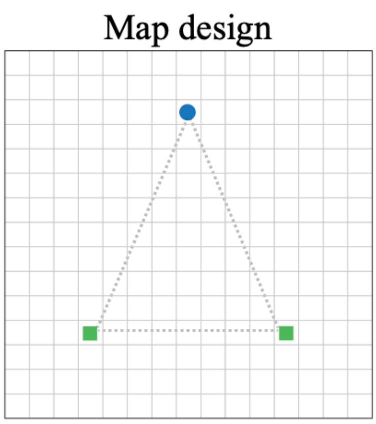
Appendix B

Experiment 1

The two destinations were aligned vertically or horizontally, with Manhattan distances varying from [4, 6, 8]. The Manhattan distance between the agent's starting positions and the two destinations were identical, achieved by letting the three entities form an invisible isosceles triangle with the agent placed along the perpendicular bisector of the invisible line connecting the two destinations as the base of the triangle (see Figure S1). The length of the perpendicular bisector varied from [6, 7, 8]. There were nine combinations of the base length and height of the triangles, causing the distance between the agent's starting position and two destinations to vary from 8 to 12, with a mean of 10.

Figure S1

Map construction in Experiment 1



Note. The agent (blue circle) and the two destinations (green squares) formed an invisible isosceles triangle, depicted by the dashed lines, which were invisible during experimental displays. The line connecting these two destinations serves as the base of the triangle. The agent was placed on the perpendicular bisector of the isosceles triangle. Each map in Experiment 1 was randomly assigned one of the invisible isosceles triangles, whose size, location, and orientation were randomized across trials.

The map for each trial was generated as follows. First, the order of the nine triangles was shuffled, with each triangle assigned to one of the first nine trials. The triangle in the 10^{th} trial was randomly sampled. Second, the triangle for each trial was randomly placed on the map with the constraint that the vertices were not on border grids. Third, the triangle was rotated by an angle that was randomly sampled from [0, 90, 180, 270] degrees.

Experiment 2a

Design of critical-crossroad maps

To create a critical-crossroad map, certain barriers were purposely placed so that (i) an agent could not reach any destination without encountering a critical crossroad; (ii) if there was more than one critical crossroad, the length of the shortest path to each critical crossroad was always identical; and (iii) this length was manipulated according to the steps-to-crossroad condition introduced in the main text. To increase the variety of maps, the critical crossroads were created either by one barrier only or by two barriers aligned vertically or horizontally. For example, for the 6-steps-to-crossroad condition, seven barriers were required to create two one-barrier critical crossroads that participants could reach in six steps (see Figure 5A).

Figure S2

Sample maps with different steps-to-crossroad in Experiment 2a

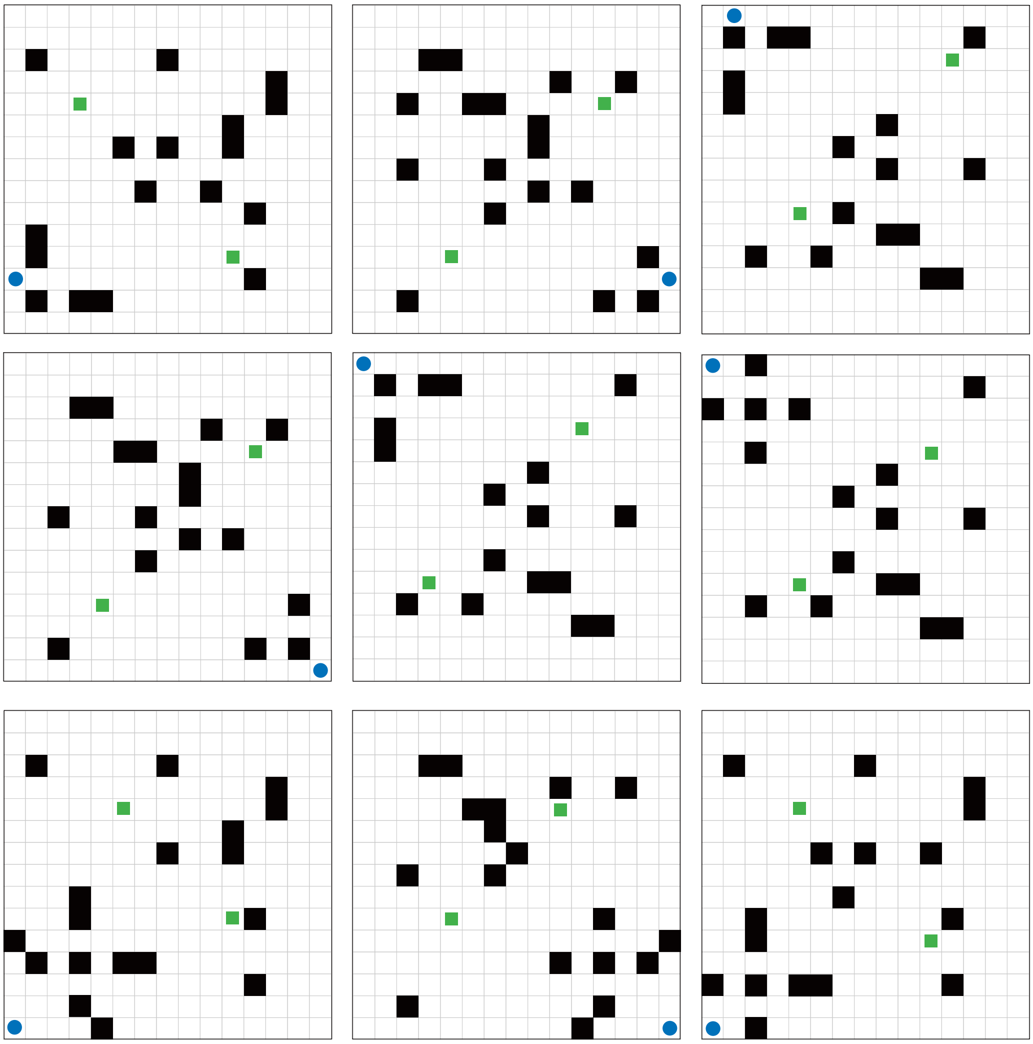
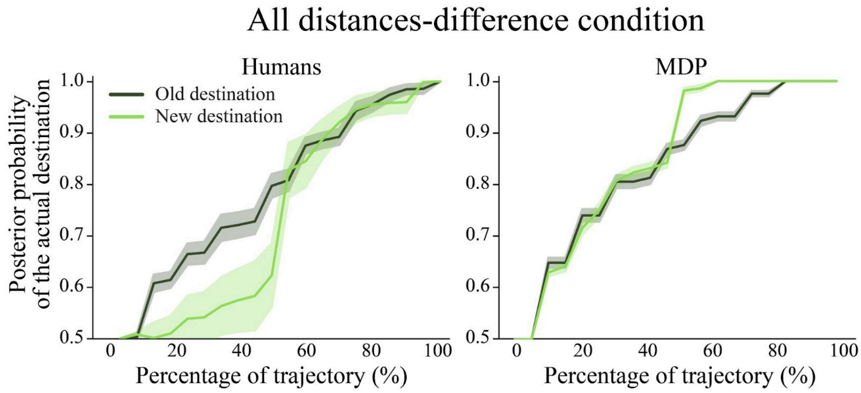


Figure S3

Faster BToM inference of the old destination in human trajectories



Note. The posterior of the BToM inference of the final destination reached by an agent over the time course of the trajectories from all distance-difference conditions. The error shading reflects 95% confidence intervals. As the length of the trajectories between the old and new destinations varied across different conditions, it was difficult to directly compare the speed of BToM inference. Nevertheless, a generally faster BToM inference for the old destination was still observed only in the human trajectories.

References

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin, 82*(4), 463–496. https://doi.org/10.1037/h0076860

Ajzen, I. (1985). *From intentions to actions: A theory of planned behavior* (pp. 11-39). Springer Berlin Heidelberg.

Ajzen, I., Czasch, C., & Flood, M. G. (2009). From intentions to behavior: Implementation intention, commitment, and conscientiousness. *Journal of Applied Social Psychology, 39*(6), 1356–1372. https://doi.org/10.1111/j.1559-1816.2009.00485.x

Ariely, D., & Wertenbroch, K. (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science, 13*(3), 219–224. https://doi.org/10.1111/1467-9280.00441

Audi, R. (1974). Intending. *The Journal of Philosophy*, *70*(13), 387–403.

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, *1*(4), 1–10.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition, 113*(3), 329–349. https://doi.org/10.1016/j.cognition.2009.07.005

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist, 44*(9), 1175–1184. https://doi.org/10.1037/0003-066X.44.9.1175

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52,* 1–26. https://doi.org/10.1146/annurev.psych.52.1.1

Bellman, R. (1957). *Dynamic programming*. Princeton, NJ: Princeton University Press.

Brand, M. (1984). *Intending and acting: Toward a naturalized action theory*. The MIT Press.

Bratman, M. (1987). *Intention, plans, and practical reason.* Harvard University Press.

Bryan, G., Karlan, D., & Nelson, S. (2010). Commitment devices. *Annual Review of Economics*, 2(1), 671-698.

Callaway, F., van Opheusden, B., Gul, S., Das, P., Krueger, P. M., Griffiths, T. L., & Lieder, F. (2022). Rational use of cognitive resources in human planning. *Nature Human Behaviour*, 6(8), 1112-1125.

Carruthers, P. (2009). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences, 32*(2), 121–138. https://doi.org/10.1017/S0140525X09000545

Cheng, S., Zhao, M., Tang, N., Zhao, Y., Zhou J., Shen, M., & Gao T. (2021). *Intention beyond desire: Spontaneous intentional commitment regulates conflicting desires [Data and code]*. https://osf.io/k5e69/

Cheng, S., Zhao, M., Zhu, J., Zhou J., Shen, M., & Gao T. (2022, July 1). Intentional commitment as goal perseverance in human planning [Workshop on Social Intelligence in Humans and Robots]. *The annual meeting of the Robotics: Science and Systems 2022 (RSS 2022)*. 2022. https://social-intelligence-human-ai.github.io/

Cheng, S., Zhao, M., Zhu, J., Zhou, J., Shen, M., & Gao, T. (2022). Intentional commitment through an internalized theory of mind: Acting in the eyes of an imagined observer. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 44, No. 44).

Chisholm, R. (1966). Freedom and Action. In K. Lehrer (Eds), *Freedom and Determinism*. Random House.

Choe, K. S., Keil, F. C., & Bloom, P. (2005). Children's understanding of the Ulysses conflict. *Developmental Science, 8*(5), 387–392. https://doi.org/10.1111/j.1467-7687.2005.00426.x

Chu, J., & Schulz, L. (2022). " Because I want to": Valuing goals for their own sake. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 44, No. 44).

Cohen, P. R., & Levesque, H. J. (1990). Intention is choice with commitment. *Artificial Intelligence*, *42*(2–3), 213–261.

Cushman, F. (2015). Deconstructing intent to reconstruct morality. *Current Opinion in Psychology*, *6*, 97-103.

Dasgupta, I., Schulz, E., Tenenbaum, J. B., & Gershman, S. J. (2020). A theory of learning to infer. *Psychological Review, 127*(3), 412–441. https://doi.org/10.1037/rev0000178

Davidson, D. (1963). Actions, reasons, and causes. *The Journal of Philosophy*, *60*(23), 685–700.

Dennett, D. C. (1987). *The intentional stance.* The MIT Press.

Dennett, D. C. (1996). *Kinds of minds: Toward an understanding of consciousness*. Basic Books.

Dragan, A. D., Lee, K. C., & Srinivasa, S. S. (2013, March). Legibility and predictability of robot motion. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 301-308). IEEE.

Elster, J. (1987). *The multiple self*. Cambridge University Press.

Elster, J. (2000). *Ulysses unbound: Studies in rationality, precommitment, and constraints*. Cambridge University Press.

Fawzi, A., Balog, M., Huang, A., Hubert, T., Romera-Paredes, B., Barekatain, M., ... & Kohli, P. (2022). Discovering faster matrix multiplication algorithms with reinforcement learning. *Nature*, 610(7930), 47-53.

Fishbach, A., & Shah, J. Y. (2006). Self-control in action: Implicit dispositions toward goals and away from temptations. *Journal of Personality and Social Psychology, 90*(5), 820–832. https://doi.org/10.1037/0022-3514.90.5.820

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Fodor, J. A. (1992). A theory of the child's theory of mind. *Cognition*.

Freud, S., & Jones, E. (Ed.). (1922). *Beyond the pleasure principle.* (C. J. M. Hubback, Trans.). The International Psycho-Analytical Press. https://doi.org/10.1037/11189-000

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences, 7*(7), 287–292. https://doi.org/10.1016/S1364-6613(03)00128-1

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature, 415*(6873), 755. https://doi.org/10.1038/415755a

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56*(2), 165–193. https://doi.org/10.1016/0010-0277(95)00661-H

Gershman, S. (2021). *What makes us smart: The computational logic of human cognition*. Princeton University Press.

Gilbert, M. (2013). *Joint commitment: How we make the social world*. Oxford University Press.

Goffman, E. (1959). *The presentation of self in everyday life.* Doubleday.

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist, 54*(7), 493–503. https://doi.org/10.1037/0003-066X.54.7.493

Gollwitzer, P. M., & Brandstätter, V. (1997). Implementation intentions and effective goal pursuit. *Journal of Personality and Social Psychology, 73*(1), 186–199. https://doi.org/10.1037/0022-3514.73.1.186

Gollwitzer, P. M., & Keller, L. (2016). Mindset theory. *Encyclopedia of personality and individual differences*, 1-8.

Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. In M. P. Zanna (Ed.), *Advances in experimental social psychology,* Vol. 38, pp. 69–119). Elsevier Academic Press. https://doi.org/10.1016/S0065-2601(06)38002-1

Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences, 16*(1), 1–14, 29–113. https://doi.org/10.1017/S0140525X00028636

Haggard, P. (2005). Conscious intention and motor cognition. *Trends in Cognitive Sciences, 9*(6), 290–295. https://doi.org/10.1016/j.tics.2005.04.012

Harman, G. (1986). *Change in view: Principles of reasoning.* The MIT Press.

Heyes, C., Bang, D., Shea, N., Frith, C. D., & Fleming, S. M. (2020). Knowing ourselves together: The cultural origins of metacognition. *Trends in cognitive sciences*, *24*(5), 349-362. https://doi.org/10.1016/j.tics.2020.02.007

Ho, M. K., Abel, D., Correa, C. G., Littman, M. L., Cohen, J. D., & Griffiths, T. L. (2022). People construct simplified mental representations to plan. *Nature, 606*(7912), 129–136. https://doi.org/10.1038/s41586-022-04743-9

Ho, M. K., Cushman, F., Littman, M. L., & Austerweil, J. L. (2021). Communication in action: Planning and interpreting communicative demonstrations. *Journal of Experimental Psychology: General, 150*(11), 2246–2272. https://doi.org/10.1037/xge0001035

Ho, M. K., Littman, M. L., MacGlashan, J., Cushman, F., & Austerweil, J. L. (2016). Showing versus doing: Teaching by demonstration. *Advances in Neural Information Processing Systems 29*.

Holton, R. (1999). Intention and weakness of will. *The journal of philosophy*, 96(5), 241-262.

Humphrey, N. (2002). *The inner eye*. Oxford University Press on Demand. (Original work published in 1986)

Jara-Ettinger, J. (2019). Theory of mind as inverse reinforcement learning. *Current Opinion in Behavioral Sciences*, *29*, 105-110.

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. Trends in Cognitive Sciences, 20(8), 589–604. https://doi.org/10.1016/j.tics.2016.05.011

Jara-Ettinger, J., Schulz, L. E., & Tenenbaum, J. B. (2020). The Naïve Utility Calculus as a unified, quantitative framework for action understanding. *Cognitive Psychology, 123,* Article 101334. https://doi.org/10.1016/j.cogpsych.2020.101334

Jiang, K., Stacy, S., Chan, A., Wei, C., Rossano, F., Zhu, Y., & Gao, T. (2021). Individual vs. Joint Perception: a Pragmatic Model of Pointing as Smithian Helping. *Proceedings of the Annual Meeting of the Cognitive Science Society.* https://doi.org/10.48550/arXiv.2106.02003

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583-589.

Kahneman, D., & Tversky, A. (1982). The psychology of preferences. *Scientific American, 246*(1), 160–173. https://doi.org/10.1038/scientificamerican0182-160

Kleiman-Weiner, M., Gerstenberg, T., Levine, S., & Tenenbaum, J. B. (2015). Inference of Intention and Permissibility in Moral Decision Making. In *Proceedings of the Annual Meeting of the Cognitive Science Society.*

Kleiman-Weiner, M., Ho, M. K., Austerweil, J. L., Littman, M. L., & Tenenbaum, J. B. (2016). Coordinate to cooperate or compete: abstract goals and joint intentions in social interaction. *Proceedings of the 38th Annual Conference of the Cognitive Science Society.*

Kryven, M., Ullman, T. D., Cowan, W., & Tenenbaum, J. B. (2021). Plans or Outcomes: How do we attribute intelligence to others? *Cognitive Science*, 45(9). https://doi.org/10.1111/cogs.13041

Leonard, J. A., Lee, Y., & Schulz, L. E. (2017). Infants make more attempts to achieve a goal when they see adults persist. *Science*, 357(6357), 1290-1294. https://doi.org/10.1126/science.aan2317

Lewin, K. (1931). Environmental forces in child behavior and development. In C. Murchison (Ed.), *A handbook of child psychology* (pp. 94–127). Clark University Press. https://doi.org/10.1037/13524-004

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences, 8*(4), 529–566. https://doi.org/10.1017/S0140525X00044903

Lieder, F., & Griffiths, T. L. (2019). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences, 43,* Article e1. https://doi.org/10.1017/S0140525X1900061X

Liu, S., & Spelke, E. S. (2017). Six-month-old infants expect agents to minimize the cost of their actions. *Cognition*, 160, 35-42.

Liu, S., Ullman, T. D., Tenenbaum, J. B., & Spelke, E. S. (2017). Ten-month-old infants infer the value of goals from the costs of actions. *Science*, 358(6366), 1038-1041.

Lombrozo, T. (2009). The role of moral commitments in moral judgment. *Cognitive Science*, *33*(2), 273-286. https://doi.org/10.1111/j.1551-6709.2009.01013.x

Malle, B. F., & Knobe, J. (2001). The distinction between desire and intention: A folk-conceptual analysis. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 45–67). The MIT Press.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of neuroscience methods*, 164(1), 177-190.

Mele, A. R. (2003). *Motivation and agency*. Oxford University Press.

Mele, A. R. (2009). *Effective intentions: The power of conscious will.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195384260.001.0001

Miller, N. E. (1944). Experimental studies of conflict. In J. M. Hunt, *Personality and the behavior disorders* (pp. 431–465). Ronald Press.

Mischel, W., Ebbesen, E. B., & Raskoff Zeiss, A. (1972). Cognitive and attentional mechanisms in delay of gratification. *Journal of Personality and Social Psychology, 21*(2), 204–218. https://doi.org/10.1037/h0032198

Moskowitz, G. B., & Grant, H. (Eds.). (2009). *The psychology of goals*. Guilford press.

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin, 126*(2), 247–259. https://doi.org/10.1037/0033-2909.126.2.247

Pacherie, E. (2006). Toward a Dynamic Theory of Intentions. In S. Pockett, W. P. Banks, & S. Gallagher (Eds.), *Does consciousness cause behavior?* (pp. 145–167). MIT Press. https://doi.org/10.7551/mitpress/9780262162371.003.0009

Pacherie, E. (2011). Self-agency. In S. Gallagher (Ed.), *The Oxford handbook of the self* (pp. 442–464). Oxford University Press.

Perugini, M., & Bagozzi, R. P. (2004). The distinction between desires and intentions. *European Journal of Social Psychology, 34*(1), 69–84. https://doi.org/10.1002/ejsp.186

Pockett, S. (2006). The Neuroscience of Movement. In S. Pockett, W. P. Banks, & S. Gallagher (Eds.), *Does consciousness cause behavior?* (pp. 9–24). MIT Press.

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Penguin.

Russell, S. J., & Norvig, P. (1995). *Artificial intelligence*: A modern approach. Prentice Hall.

Schelling, T. C. (1980). T*he strategy of conflict: with a new preface by the author*. Harvard University Press.

Schelling, T. C. (1984). *Choice and consequence*. Harvard University Press.

Schopenhauer, A. (1969). *The world as will and representation*. Dover Publications. (Original work published 1818)

Schult, C. A. (2002). Children's understanding of the distinction between intentions and desires. *Child Development, 73*(6), 1727–1747. https://doi.org/10.1111/1467-8624.t01-1-00502

Searle, J. (2003). *Rationality in action*. The MIT Press.

Searle, J., & Willis, S. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge University Press.

Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology, 71,* 55–89. https://doi.org/10.1016/j.cogpsych.2013.12.004

Shah, J. Y. (2005). The Automatic Pursuit and Management of Goals. *Current Directions in Psychological Science, 14*(1), 10–13. https://doi.org/10.1111/j.0963-7214.2005.00325.x

Shah, J. Y., Friedman, R., & Kruglanski, A. W. (2002). Forgetting all else: On the antecedents and consequences of goal shielding. *Journal of Personality and Social Psychology, 83*(6), 1261–1280. https://doi.org/10.1037/0022-3514.83.6.1261

Sheeran, P., Webb, T. L., & Gollwitzer, P. M. (2005). The interplay between goal intentions and implementation intentions. *Personality and social psychology bulletin*, 31(1), 87-98.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual review of neuroscience*, 40, 99-124.

Shu, T., Peng, Y., Zhu, S. C., & Lu, H. (2021). A unified psychological space for human perception of physical and social events. *Cognitive Psychology*, 128, 101398. https://doi.org/10.1016/j.cogpsych.2021.101398

Shum, M., Kleiman-Weiner, M., Littman, M. L., & Tenenbaum, J. B. (2019, July). Theory of minds: Understanding behavior in groups through inverse planning. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 6163-6170).

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., & others. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484–489.

Silver, D., Singh, S., Precup, D., & Sutton, R. S. (2021). Reward is enough. *Artificial Intelligence*, 299, 103535.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2013). Life after P-Hacking. *Meeting of the Society for Personality and Social Psychology*. https://doi.org/10.2139/ssrn.2205186

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69,* 99–118. https://doi.org/10.2307/1884852

Simon, H. A. (1990). Bounded rationality. *Utility and probability*, 15-18.

Siposova, B., Tomasello, M., & Carpenter, M. (2018). Communicative eye contact signals a commitment to cooperate for young children. *Cognition*, *179*, 192-201.

Soon, C. S., Brass, M., Heinze, H. J., & Haynes, J. D. (2008). Unconscious determinants of free decisions in the human brain. *Nature neuroscience*, 11(5), 543-545.

Steele, Katie & H. Orri Stefánsson. (2020). Decision Theory. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy.* Retrieved from https://plato.stanford.edu/archives/win2020/entries/decision-theory.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643–662. https://doi.org/10.1037/h0054651

Strotz, R. H. (1955). Myopia and inconsistency in dynamic utility maximization. *The Review of Economic Studies*, *23*(3), 165-180.

Sutton, R. S., & Barto, A. G. (1998). *Introduction to reinforcement learning*. The MIT Press.

Tang, N., Gong, S., Zhao, M., Gu, C., Zhou, J., Shen, M., & Gao, T. (2022). Exploring an imagined "we" in human collective hunting: Joint commitment within shared intentionality. In *Proceedings of the annual meeting of the cognitive science society*.

Tang, N., Stacy, S., Zhao, M., Marquez, G., & Gao, T. (2020). Bootstrapping an Imagined We for Cooperation. In *Proceedings of the annual meeting of the cognitive science society.*

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, 1(1), 39-60.

Tomasello, M. (2010). *Origins of human communication*. MIT press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28*(5), 675–735. https://doi.org/10.1017/S0140525X05000129

Tversky, A., & Shafir, E. (1992). Choice under conflict: The dynamics of deferred decision. *Psychological Science, 3*(6), 358–361. https://doi.org/10.1111/j.1467-9280.1992.tb00047.x

Vinyals, O., Babuschkin, I., Czarnecki, W. M., Mathieu, M., Dudzik, A., Chung, J., ... & Silver, D. (2019). Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 575(7782), 350-354.

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior.* Princeton University Press.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (Cole, M., John-Steiner, V., Scribner, S. & Souberman, E., Eds.). Harvard University Press. (Original work [ca. 1930-1934])

Wegner, D. M. (2002). *The illusion of conscious will.* MIT Press.

Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist*, 54(7), 480-492. https://doi.org/10.1037/0003-066X.54.7.480

Wellman, H. M. (1992). *The child's theory of mind.* The MIT Press.

Wellman, H. M. (2014). *Making minds: How theory of mind develops.* Oxford University Press.

Wu, S. A., Wang, R. E., Evans, J. A., Tenenbaum, J. B., Parkes, D. C., & Kleiman-Weiner, M. (2021). Too Many Cooks: Bayesian Inference for Coordinating Multi-Agent Collaboration. *Topics in Cognitive Science*, 13(2), 414-432.