Attributed Intelligence

by

Marta Kryven

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Computer Science

Waterloo, Ontario, Canada, 2018

Examining Committee Membership.

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Jacob Feldman Professor, Department of Psychology, Rutgers University
Supervisor	William Cowan Associate Professor, Cheriton School of Computer Science, University of Waterloo
Internal Member	Peter Van Beek Professor, Cheriton School of Computer Science, University of Waterloo
Internal Member	Daniel Vogel Associate Professor, Cheriton School of Computer Science, University of Waterloo
Internal-external Member	Wayne Oldford Professor, Statistics and Actuarial Science, University of Waterloo

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Human beings quickly and confidently attribute more or less intelligence to one another. What is meant by intelligence when they do so? And what are the surface features of human behaviour that determine their judgements? Because the judges of success or failure in the quest for 'artificial intelligence' will be human, the answers to such questions are an essential part of cognitive science. This thesis studies such questions in the context of a maze world, complex enough to require non-trivial answers, and simple enough to analyse the answers in term of decision-making algorithms.

According to Theory-theory, humans comprehend the actions of themselves and of others in terms of beliefs, desires and goals, following rational principles of utility. If so, attributing intelligence may result from an evaluation the agent's efficiency – how closely its behaviour approximates the expected rational course of action. Alternatively, attributed intelligence could result from observing outcomes: billionaires and presidents are, by definition, intelligent. I applied Bayesian models of planning under uncertainty to data from five behavioural experiments. The results show that while most humans attribute intelligence to efficiency, a minority attributes intelligence to outcome.

Understanding of differences in attributed intelligence comes from a study how people plan. Most participants can optimally plan 1-5 decisions in advance. Individually they vary in sensitivity to decision value and in planning depth. Comparing planning performance and attributed intelligence shows that observers' ability to attribute intelligence depends on their ability to plan. People attribute intelligence to efficiency in proportion to their planning ability. The less skilled planners are more likely to attribute intelligence to outcome.

Moreover, model-based metrics of planning performance correlate with independent measures of cognitive performance, such as the Cognitive Reflection Test and pupil size. Eyetracking analysis of spatial planning in real-time shows that participants who score highly on independent measures of cognitive ability also plan further ahead. Taken together, these results converge on a theory of attributed intelligence as an evaluation of how efficiently an agent plans, such that depends on the observer's cognitive abilities to carry out the evaluation.

Acknowledgements

I would like to thank all the people who made this thesis possible.

My adviser at Waterloo, Bill Cowan, for his open-minded interest in human nature, reading of the many drafts and dedication to editing and statistics.

My co-adviser at MIT, Josh Tenenbaum, for his wonderful insights, wise guidance and academic support.

Tomer Ullman, for his careful attention to experimental protocols, reasonable rhetoric and stimulating discussion.

Jacob Feldman, for his valuable feedback, many helpful suggestions and an incredible external examiner report.

My examining committee, whose comments made this thesis so much better: Peter Van Beek, Daniel Vogel and Wayne Oldford.

Mehrdad Jazayeri and his lab for their kind help with eye-tracking setup.

Tobi Gerstenberg, for his extraordinary good humour and eye-tracking skills.

Jan Drugowitsch, for his thoughtful advice on Drift Diffusion Models.

Daniel Estandian, Kellsey Allen and John McCoy for being wonderful friends on my couch-surfing exploits.

My friends and colleagues for their curiosity, inspiration and engaging questions: Chris Baker, Kevin Smith, Jessica Hamrick, Falk Lieder, John Widler, Sam Gershman, Ignazio Saez and Max Kleiman-Weiner.

The organisers of CBMM Summer Course in Woods Hole in 2015, where this project started.

Minjung Kim, for her friendship and advice.

Liam Levy, for energy, incredible optimism and contact juggling magic.

My brother, for the presentation remote, which I take everywhere.

My family.

The many human subjects who participated in my experiments.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

> To read without joy is stupid.
>
> John Williams

The NASA engineers building the Voyageur spacecraft had a problem: the spacecraft would leave the solar system forever, and the engineers wanted unknown aliens who might encounter it to attribute intelligence to its builders. Their solution, an assembly of sounds, images and text, is highly unlikely to evoke any associations in the cognition of truly alien creatures. But even if the other beings do not share our senses, might they nevertheless see intelligence in the mere intent of sending a message?

In common-sense psychology attributing mental states to others seems intuitive. However, a large body of artificial intelligence (AI) research shows that it is by no means trivial. Most humans easily infer each others' thoughts and feelings, as a practical evolutionary advantage for any social being [18]. At the same time, the information processing problem associated with social perception has proved immensely difficult for AI. Unlike many aspects of human problem-solving that have been successfully abstracted in AI (e.g. playing computer games or traffic scheduling), social perception appears to require complex, abstract and overlapping representations.

Formally defining intelligence is hard because our intuitive understanding of it has many sides. Clearly, some behaviours appear more intelligent than others. Rote memorisation is not as intelligent as the ability to adapt to new situations and solve new problems. It is intelligent to infer laws of nature by observation. It is intelligent to compose music. An organism that seems intelligent at first strikes us as unintelligent once we discover that it is inflexible. A spider-web may seem a magnificent feat of architecture until we realise that it is produced by a genetically-scripted behaviour without awareness or an explicit plan of deceiving the prey.

In general, humans associate intelligence with reasoning, creativity and problem-solving. However, the abstract mental properties of intelligence can be interpreted in different ways. People may see intelligence as a disjunction of abilities [42, 80, 22] or disagree about the behaviours expected of an intelligent agent [139]. Impulsive or silly behaviours, such as gambling, are sometimes evaluated as intelligent if they lead to a favourable outcome [38].

Understanding how humans attribute intelligence has valuable applications. For example, making software act intelligently in human terms produces a more trusted AI. Marketing intelligent-looking applications is easier than marketing uninterpretable results [67, 120, 145, 130]. Moreover, a theory of attributed intelligence has important social implications. Understanding the psychological mechanism by which people make value judgements about others could explain why some people see rich as deserving of wealth and poor as undeserving, and inform an effective social policy for promoting equal opportunities. Furthermore, although we expect intelligent individuals to have successful careers, intelligence quotient or test scores do not predict career outcomes as well as social skills [84, 50]. Understanding the mechanism by which humans arrive at accurate insights about others may inform more accurate formal assessments of intelligence itself.

This thesis proposes a methodology for investigating planning and plan evaluation in an experimental task that combines behavioural metrics, psychophysics and eyetracking. Intuitive evaluations of intelligence are modelled formally by probabilistic computations conditioned on observed behaviour, based on a Partially Observable Markov Decision Process (POMDP) [62, 5]. The model is compared to empirical results. This thesis proposes and tests the **efficiency hypothesis** of attributed intelligence: that people attribute intelligence in proportion to their rational evaluation of how efficiently the agent maximises its utility over time. The results show that participants attribute intelligence to planning efficiency in proportion of their own planning ability. The better people can approximate optimal planning, the more likely they are to attribute intelligence to efficiency. Poor planners often evaluate short-term outcomes as a heuristic. Moreover, both planning and attributing intelligence varies with independent metrics of cognitive performance, suggesting that both depend on a common cognitive resource.

The structure of the next chapters is the following. Chapter 2 reviews the history of social perception and mental state attribution to motivate a study of attributed intelligence. Chapter 3 describes a Bayesian computational framework of spatial planning in a probabilistic environment based on a theory of Partially Observable Markov Decision Process (POMDP). The framework is used throughout the thesis as a formal model of planning under uncertainty. Chapter 4 applies the framework to interpret human attributions of intelligence, showing that the majority of observers (about two thirds of the tested sample) attributed intelligence to the efficiency of the observed behaviour. A smaller group, however, attributed intelligence to outcome. Chapter 5 measures human planning in a naturalistic maze task, showing that the variability between individuals can be quantified as a variability in sensitivity to decision value. Chapter 6 describes experimental results showing that an individual's problem-solving ability predicts attributed intelligence. Chapter 7 describes an empirical eye-tracking study of planning, prompting further refinements of the model. Chapter 8 summarises the contributions of this thesis and directions of future research.

Chapter 2

Background

> One notable way of misunderstanding the actions of people in the past is to over-rationalise them.
>
> Bernard Williams

This chapter begins by a review of the history of social attribution in Section 2.1, leading up to the view agents are assumed to act intentionally and be rational. Section 2.2 discusses related studies of mental state attribution from an information-processing perspective and provides background for understanding the modelling methodology in this thesis.

2.1 A Brief History of Social Perception

Psychological science originates in the 18-th century empiricist approach to the nature of knowledge. A predominant view at the time, most notably held by Decartes, stated that experience merely awakens innate ideas. Empiricists, however, argued that knowledge is learned from experience. For example, Hume argued that causality is learned from experiencing conjunctions: if A and B always follow each other people expect the co-occurrence to continue and say that A causes B [57]. Infant experiments reveal that

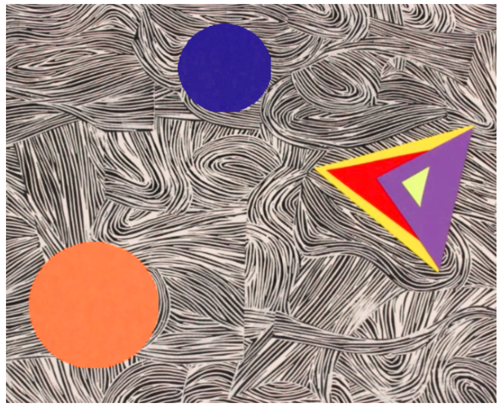


Figure 2.1: An artist interpretation of a Heider and Simmel cartoon. Image credit: Liz Little (modified with permission).

attribution of causality indeed emerges gradually, although its rudiments appear as early as the first six month of life [107, 113].

Empiricism made perception an acceptable object of experimental study, a measurable phenomenon comparable to the movement of physical bodies. Furthermore, Darwin's study of universal emotions (happiness, sadness, anger, fear, surprise, and disgust) showed that emotional behaviour predates the emergence of humanity and can be explained by physical causes, without necessitating a non-physical mind [24].

A second trend foreshadowing psychological science was phenomenology, a style of philosophy associated with Husserl in the late 19th and the early 20th centuries. Phenomenology understands mental states as having an intention, reflecting, but not identical to, the objective reality. Phenomenology clearly distinguished between a *phenomenon* (a mental representation leading to a qualitative experience) and the object being experienced. Empiricism, and later phenomenology, provided a foundation for the scientific study of the human mind.

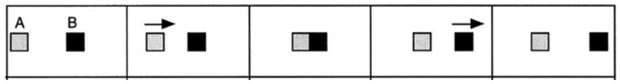


Figure 2.2: A launching display.

2.1.1 Attribution theory

In 1940s Fritz Heider and Marianne Simmel made a short film of moving geometric shapes [53]. It showed a larger triangle moving in the presence of two smaller circles (a single frame is shown in Figure 2.1). Most viewers see the circle as a nice character who avoids the angry triangle. People describe the film by attributing thoughts and emotions to the figures, saying things such as: *'The triangle is bullying the smaller guy' 'The other circle wants to help his friend but is scared'*, or *'The big triangle is angry'*. Remarkably, all viewers in the nonclinical population attribute mental states to the cartoon shapes, with nobody saying that they saw a film about moving shapes [1].

This simple experiment shows that viewers automatically engage in elaborate social inference and attribute mental states even to inanimate objects. Animated displays such as this reveal the building blocks of social perception and are still used to study social attribution today [138, 95, 96, 58]. But how is social attribution learned and why does it emerge?

Behaviourists claim that organisms respond to rewarding stimuli. If one perceives a moving triangle as aggressive, it must be because they learned to avoid angular fast-moving bodies in the past [148]. In contrast, an information-processing approach interprets perception as reverse-engineered computation: an observer infers the agent's motivation (e.g pursuing a moving target, searching, helping another) by reconstructing the information-processing problem faced by the agent [6, 9, 58, 138].

To understand how inferences of emotions, goals, and beliefs, such as occur in Heider and Simmel experiment, build on each other we will consider the order in which they emerge during development. The earliest inference that infants make is differentiating between agents (physical bodies that generate behaviour) and objects (bodies that move only when a force is applied to them), which occurs at 6 month of age [127].

Perception of causal interactions is studied using launching displays, as shown in Figure 2.2: Object (A) moves until it is adjacent to object (B), at which point (A) stops moving and (B) starts moving [83]. Such displays create a subjective impression of (A) causing (B)'s motion. In adult viewers the strength of the causal perception is mediated by three perceptual parameters: (1) relative speeds of objects, (2) speed-mass interactions, and (3) spatio-temporal delays between events. The perception of launching interactions as causal disappears completely if the delay between collision and launch is 500ms or more [109]. In contrast, a body that moves from rest, without an obvious physical cause is seen as animate. Adult viewers attribute animacy automatically, based on changes of velocity, angular direction, or alignment between the principal axis of the object and its direction of motion [132, 116].

Attributions of causality and agency are straightforward enough to be modelled by a mapping of sensory cues to precepts. However, this does not mean that simple sensory cues, such as change of velocity, are all there is to recognising agency. A leaf blown in the wind seems alive only until the observer identifies it as a leaf. Seeing moving bodies as possessing intent may be a practical trick for predicting and anticipating physical motion and behaviour [82, 18]. The use of intentional metaphors is often apparent in language, when people say things such as: 'The ball wants to go into the left hole'.

Following the understanding of agents and objects, six-to-ten-month-old babies can infer physical goals and distinguish between helpful and malicious behaviour [138]. At 14 months a baby expects an agent to act efficiently in accord with their goals, such as running straight to the ball that they intend to pick up [43]. Two year old children understand and value competence, preferring to play with competent agents, who can perform an action on the first attempt, even if the agent is antisocial [60].

Reasoning about beliefs and knowledge of others, usually referred to as Theory of Mind (ToM) is mentally demanding. ToM emerges gradually after 3-4 years of age, at which time children can clearly articulate that some beliefs are mistaken [114, 124] and often break down in mental health disorders [125, 1]. In primates the evidence of social reasoning, such as forming alliances and deception [146] is observed with frequency proportional to the size of the neocortex [20].

2.1.2 Attributing Emotions

> (-0.88,1), (-0.5,0) Yorick (0,1) I knew him, Horatio:
> a fellow of (1,1) (0.74, 0.69), of (0, 0.8) (0.85, 0.1) (0.68, 0.2): he hath borne me on his back a thousand times; and now, how (-1,1) in my imagination it is!
> *Alas, poor Yorick! I knew him, Horatio: a fellow of infinite jest, of most excellent fancy: he hath borne me on his back a thousand times; and now, how abhorred in my imagination it is!*
>
> Hamlet, Act 5

Skilled puppeteers effortlessly convey elation and melancholy, pride and resentment, vice and generosity by manipulating inanimate objects. The illusions of complex thoughts and feelings arise automatically, despite the audience's full knowledge that puppets are inanimate. Neurophysiological and psychophysical evidence shows emotion processing is fast. For example, an angry face elicits response in the amygdala within 40 milliseconds of exposure [118]. People generally spot angry faces faster than happy faces [36], presumably because avoiding angry con-specifics helped our ancestors to survive. Moreover, emotional expressions, both happy and sad, influence consumer decisions and judgements of value, even if the exposure to the emotion cue is too brief to reach awareness [144]. The effect of emotional expression on consumption can be reversed, however, if the customer consciously notices the cue [106].

Attributing basic emotions to facial expressions happens before cognitive reflection. There are two basic models of categorising emotions that produce significant correlations with corresponding brain activity: a combination of five or six categories of basic facial expressions [72] (the same ones that were studied by Darwin [24]), the *circumplex* model, described by a two-dimensional space of Valence and Arousal [121, 98], or a combination of these two [100]. These models do little justice to the richness of human sensibilities. (The epigraph rewrites a sentence from Hamlet by replacing every sentiment by a number corresponding to valence, dominance and arousal).

An alternative view is that our ability to identify others emotions is better described by complex inference about others' feeling and thoughts based on contexts and relationships. For example, such inference may involve distinguishing between existential frustration and outward hostility, thoughtless generosity and political investment [93]. The approach, known as *appraisal theory*, aims to characterise emotions in terms of interpretations of the value of the events people encounter based on an intuitive causal theory [30]. The dimensions of emotion appraisal are thus causal: is the event caused by self or other? is the event expected? what is the subjective value of the outcome? was the outcome intended? how likely was the outcome given the subject's abilities?, and so on [115]. Appraisal theory explains the representational spaces in medial prefrontal cortex (MPFC) as well as intuitive judgements about situations that cause different emotions better than 5-6 emotion dimensions or the circumplex model [122].

People also have prior expectation of an agent's ability to experience various cognitive states and feelings. For example elephants are judged as far more capable of feeling love, embarrassment or rejection than goats, robots or beetles. Moreover, people attribute complex cognitive states (e.g. sympathy, melancholy, intelligence) less consistently and with higher variance than sensory states (e.g. feeling cold, feeling hungry) [139]. It is unclear whether the variance in mental state attribution comes from a disagreement about definitions, or from a difficulty recognising such states. The strength of prior expectations could be measured by the amount of observation it takes to change the observer's inference, however, so far this has not been measured experimentally.

2.1.3 Attributing Goals and Beliefs

Any sequence of behaviour can have a multitude of causes. Thus, reliably inferring mental states requires reducing the hypotheses space by common-sense assumptions, such as assuming that others are rational [48, 28]. For example, it is easy to interpret running by a likely cause, such as catching a train. It is harder to come up with unlikely explanations, such as that the person has spotted an undesirable acquaintance. Both children and adults more readily attribute goals to agents who act efficiently [9, 43, 96].

This common-sense assumption of rationality is known as Dennett's *rationality principle*, when applied to interpreting physical actions, and as *Grice's maxims* [48] when applied to interpreting communication. The view is that people expect agents act efficiently to achieve its goals, given the agents' knowledge and abilities. Assuming rationality allows observers to infer hidden goals and anticipate behaviour. In a collaborative environment the agent might maximise the utility of a partner, rather than its own, and such actions are still seen as rational [34, 138].

Mental state attribution may occur by several theoretical mechanisms. Theory-theory assumes that organisms have a theory of how rational beings behave independent of one's own decision-making mechanisms, although used to rationalise one's actions [102]. An alternative account, Simulation Theory, states that organisms use themselves to understand others, and interpret behaviour by performing a mental simulation of the actions they observe [41]. Both theories have their strengths and weaknesses.

Simulation Theory is often applied to explain physical inference, such as deciding how much force it takes to move an object (but see [112]). When used as a model of social attribution Simulation Theory assumes that observers use their own decision-making mechanisms to anticipate behaviour, and so predictions should be biased by the observers preferences. For example, people find it difficult to choose lunch for a friend whose dietary preferences contradict their own [128]. However, according to Simulation Theory people are expected to experience empathy as an effortless consequence of identifying the other's emotional state, which has not been observed. Automatic emotion contagion occurs only with basic emotions, most notably fear and disgust [47]. In contrast, whenever empathy involves cognitive appraisal people generally find it effortful and actively avoid it [21]. Thus, it is unlikey that Simulation Theory can explain attributing mental states.

According to Theory-theory observers use a folk theory of rational behaviour and over-attribute rationality when interpreting others' actions and their own [112]. Indeed, people generally believe themselves and others to be more rational than people actually are. For example, people have little awareness that their choices are influenced by recency, primacy and availability heuristics [137], by the order of presenting evidence [117] or by social pressure [14]. Human expectations of how they and others reason resemble normative theories of logic and utility more closely than does actual behaviour [45, 88, 137].

2.2 Computational Modelling

Mental state attribution is fast for primitive states (agency, anger, excitement), but slow for casual inference (embarrassment, pride, amusement). In a series of thought experiments Valentino Braitenberg proposed that primitive robotic devices built with only sensors and motors can induce an intuitive sense of agency, memory and intent [17]. For example, a vehicle with two light-powered motors connected to sensors on the same sides of the vehicle will flee the light, creating an impression that it does not like the light. A vehicle with motors connected to the opposite sides will approach and appear to like the light (Figure 2.3). To a human observer such vehicles might appear as possessing intent.

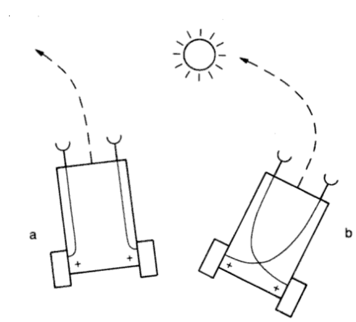


Figure 2.3: Left: Braitenber's vehicle. Right: A moth attracted to light

Braitenberg hypothesised that a combination of sensors, motors and simple generative rules selected through an evolutionary process could develop fitness qualities which humans identify as intelligent. Agents that are better adapted to their environment achieve better evolutionary success and may appear to be more rational. Given a suitable environment, evolutionary process is expected to gradually approach an approximately optimal behavioural strategy.

Pantelis and Feldman tested Braitenberg's paradigm in a software simulation of reproducing autonomous agents competing for resources. They found that evolved agents were indeed perceived as more intelligent [96]. However, on average evolved agents also achieved better outcomes compared to nonevolved agents. After accounting for outcome, the agent's evolved status was still a significant, but a weak predictor of perceived intelligence. Thus, on average participants likely assessed how well the agent was doing, although some adjusted their assessment by taking the agent's strategy into account.

In addition, participants attributed more intelligence to agents who maintained a consistent goal state, copied the actions of their adversaries and avoided fighting. Participants were better able to infer mental states of evolved agents compared to nonevolved agents, suggesting that people are better at interpreting agents that are more rational [96]. Thus, attributing intelligence likely depends on the observer's ability to interpret the agent's behaviour as rational, infer the agent's goal states and evaluate how well the agent's intention is carried out. However, it is unclear whether attributions of intelligence are sensitive to the level of perceived rationality and whether intelligence can be attributed based on the agent's actions alone when the outcome is not seen, or when the outcome is negative despite the agents best rational actions.

One way of modelling mental state inference formally combines Bayesian planning with an expected utility calculus. Such models formalise the rationality principle by defining a rational action as one that maximises the agent's expected utility [62, 13], taking into account uncertainty over observations, action outcomes and knowledge of the environment.

Of special interest is a class of computational ToM theory models encoding goal-based agents and their interactions. According to the rationality principle, beliefs and desires lead to rational actions that bring into being a goal state of the world. Likewise, inference over mental states conditioned on behaviour finds a combination of desires and beliefs that would have generated the observed behaviour. Goal-based rational agents are conveniently implemented by a combination of Bayesian forward planning and backward filtering algorithms, together called the Bayesian theory of mind (BToM) [6, 9, 58, 138, 7, 5, 68].

Since formal approaches to planning, such as MDP or a POMDP are designed to solve practical problems, one might suppose that they encode the principles of human folk theory of rationality, formalising the rationality principle. BToM reasons as follows. Any solution to a planning problem is a sequential process in discrete time. At each time step the agent's state is described by beliefs, observations, affordances, rewards and goals. At each step the agent chooses from its set of actions the action expected to maximise its pay-off,

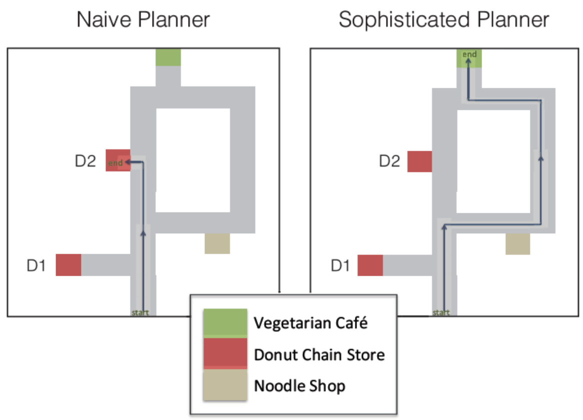


Figure 2.4: Image from [32] showing a 'self-conscious' Bayesian agent planning a route to avoid an unhealthy temptation.

within a time horizon. An observer who knows some components of the planning process can sharpen the distribution over the unknowns, treating behaviour as a random variable dependent on hidden beliefs and goals.

A simple BToM models goal inference, which occurs when the observer sees an agent approach an object and infers that the object is the agent's goal. An observer who only sees the agent take the same direct path to the goal can still predict the agent's actions correctly in a new environment by knowing that the agent goes to the goal by the shortest route. A Bayesian model based on an MDP can describe such reasoning [8] and make predictions that match those of humans [43]. Likewise, BToM models can infer changing or multiple goals [9]. Multi-agent models can encode interactions between agents. For example, maximising the other's reward is seen as helping, while minimising the other's pay-off is seen as hindering [138].

BToM models also capture behaviour that appears as self-awareness, preferences or beliefs [58, 32]. An example in Figure 2.4 shows a BToM agent planning a route to get lunch. A 'naive' BToM agent on the left has no knowledge of its own reward function and can not predict its future decisions. The 'naive' agent plans the shortest route toward the green cafe which it prefers, but upon passing a doughnut chain, it is tempted to satisfy its hunger immediately and settles for the lesser option. A 'sophisticated' BToM agent plans a longer route, which guarantees that it reaches its favourite goal and gets a higher pay off in the long run. BToM methodology is a convenient modelling tool to interpret attributed intelligence as inference over the agent's efficiency, which I will call the *efficiency hypothesis of intelligence*, discussed in detail in Chapter 3.

BToM methodology has important limitations. The most prominent arise from objections to utility calculus per se. Deciding between work and skiing, an apple and an orange, or a new xylophone and a retro bicycle, assumes that each object and experience is reduced to a value on a common utility scale. In practice, eliciting value assignments is hard and the results are often inconsistent.

This is a reasonable objection. One possible response is that all models address reality only at some level of abstraction. The mechanism by which decision values are obtained is external to the model. A vale of an experience can be estimated by a proxy of behavioural correlates of decision-making, such as neural activation in brain areas related to value processing (anterior cingulate (ACC), dorso-lateral prefrontal cortex (dlPFC), the orbitofrontal cortex (OFC) [101, 27, 19]). Psychophysical measurements, such as reaction times and choice probabilities are another common estimate of the hidden decision value [94, 101, 129, 70, 56]. In economics, the approach to estimating the utilities of goods through patterns of consumption has a long history as revealed preferences [111]. It remains to be seen whether behavioural metrics produce consistent value estimates. Since behavioural metrics result from aggregating many measurements during a choice trial, to the extent that inconsistency in human behaviour is driven by decision noise, random effects should cancel out resulting in a more consistent estimate of decision value. However, inconsistency in human behaviour can be also caused by a systematic error due to heuristics and biases, leading to the second objection.

The second objection is that naive utility calculus is more often explained by heuristics and biases than by normative utility calculations. The perceived probability of a hypothesis is higher when the hypothesis is described as a disjunction of typical examples [35, 135] but lower when it is defined by atypical examples [123, 49]. People are more likely to accept a claim if a straw-man argument is presented as an alternative [33, 143]. Numeric estimates (e.g. estimating an age of a tree or a distance between cities) can be primed by irrelevant numbers prompted at the start [136].

Indeed, individual decisions are often inefficient, sub-optimal and seemingly irrational. This, however, does not rule out applications of utility theory to understanding social attribution. When attributing intelligence to an agent we may expect it to behave more rationally than we would ourselves. Even one's own accidental behaviour can be rationalised in retrospect. Moreover, utility-theory-based models might explain individual decisions better if resource-rationality is taken into account. Thus, when making a decision agents are not only maximising an observable monetary pay-off and minimising an observable monetary cost. In most real-life situations agents also minimise decision cost reflected by decision time and cognitive effort. Taking time and effort into account, most human decisions can be explained as rational [73].

The third objection is that the BToM models work well only on small hypothesis sets. Most real-world tasks include large actions spaces and infinite planning horizons, making POMDP-based planning intractable. The response to this objection is to consider that people usually prune the hypothesis space and evaluate only a few salient option while dismissing multitudes of uninteresting ones. The many models of hypotheses sampling, such as sampling by utility of outcomes [76], importance sampling [89], or Markov Chain sampling [25] provide possible solutions to how the hypothesis space may be pruned by discarding the less relevant options [25]. Given an efficient way of pruning of the available action space as well as a way of managing planning horizons, observers could still engage in Bayesian planning in such simplified scenarios.

Moreover, the utility calculus can fail in multi-agent scenarios. For example, an agent who wants to maximise average utility will be driven to eliminate every agent whose utility is lower than average. An agent who seeks to maximise total utility will be indifferent between multitudes of others with minimal non-zero utility and a small group of high-utility others. Although humans, and other social animals, care about the utility of others, such inferences appear clearly nonsensical.

This reasonable objection is not yet successfully resolved and remains an important question for future research. Indeed, it makes no sense to talk about aggregate utility so as to treat two half-persons as equivalent to a whole individual. However, it is possible to describe utilities of social agents as dependent on each other. For example, for agents A and B: $U(A) = R(A) + \beta U(B)$ and $U(B) = R(B) + \alpha U(A)$, where $R(A), R(B)$ are reward functions. By arithmetic this reduces to $U(A) = (R(A) + \beta R(B))/(1 - \alpha\beta)$. Collective utility modelling captures common intuitions that people seek to share their positive experience to feel better. It remains to see whether behaviours resulting from combining an appraisal theory of emotions with methods for approximating group utility correlate with measures produced by psychophysical or neurophysiological measurements.

This thesis concerns itself with models of intelligence attribution in problem-solving scenarios involving one agent. BToM modelling has proven successful in capturing the quantitative and qualitative intuitive-psychology judgements of adults and children in similar goal-directed scenarios [6, 5, 59, 138, 7]. Attributed intelligence is a new question that has not yet been addressed. It has important implications for building better AI and understanding practical social issues. The next chapter addresses specific applications of the BToM methodology to attributed intelligence and serves as a reference to support the experimental methodology used in the rest of this thesis.

Chapter 3

The Planning Framework

> Nature has placed mankind under the governance of two sovereign masters, pain and pleasure. It is for them alone to point out what we ought to do, as well as to determine what we shall do.
>
> Jeremy Bentham, *The Principles of Morals and Legislation*, 1780

This chapter describes a formal model of how people plan and how they interpret the plans made by others. People associate intelligence with problem-solving. Intelligence may thus be attributed on the basis of an evaluation of the agent's problem-solving process. The chapter begins by discussing a naturalistic model environment in which intelligent behaviour can be observed. Next it discusses a formal definition of planning and of inference mechanisms used to evaluate planning in the model environment based on a Partially Observable Markov Decision Processes (POMDP) [62]. The goal of the modelling is to define how observers can identify an agent as efficient and to distinguish attributions of intelligence to efficiency from attributing intelligence to outcome.

3.1 The Problem of Planning Under Uncertainty

Consider a simple example of planning under uncertainty. You are shopping at a mall when you discover that you are missing something: a backpack, a glove, a scarf. Where should you look? You remember when you last had the missing item and reconstruct the places you visited since. You consider the order in which to visit those places, so as to minimise the time and effort spent searching. You will likely consider the effort of going from one place to another and the probability of finding the lost item in each place.

This humble example captures the essence of most planning: a mostly known environment, critical unknown information, which must be gained by costly actions and a goal that terminates the execution of the plan. Even the first action requires some prior planning. In a simple case all planning may be completed prior to the first action. In more complex cases the plan is likely to evolve as more information is gained. Such planning under uncertainty is central to many natural behaviours, from hunting-and-gathering to managing investments. The rewards are sparse, the uncertainties are many, and to get by, organisms must maintain distant goals while reasoning many steps in advance.

To study planning in natural environments we abstract the planning problem as a model-world consisting of mazes with corridors and rooms. Consider the example of lost keys. The layout of the maze corresponds to the spaces in which the keys are likely to be. Some rooms are far and remote, so that visiting them is costly. Some rooms are close. Some rooms are big, highly likely to contain the lost object, and some are small and unlikely to be of interest. And while looking for the lost object, one intends to traverse the maze in a way that minimises the expected searching time.

The planning problem is a search for a hidden goal in a world with progressively disclosed information. A schematic example of the problem as seen from an overhead view is shown in Figure 3.1. An agent looks for a target in a maze-like environment. It is familiar with the layout of the world and knows that the goal exists, but does not know where it is. The agent can move into or within rooms, but not through walls and must inspect each of the spaces in turn to find the goal. The goal may be in any of the locations not yet seen by the agent. In the example in Figure 3.1 the solid arrow shows the optimal route, which reveals the whole space in the fewest steps, the dashed arrow shows a suboptimal path.

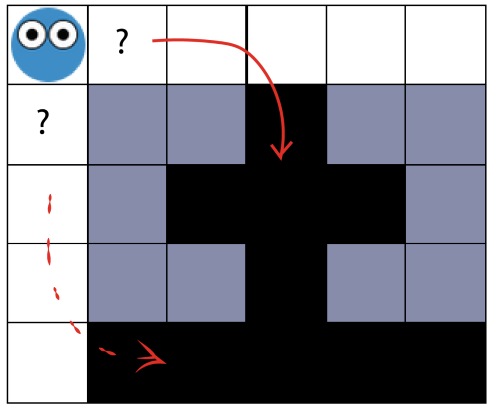


Figure 3.1: Dark cells indicate that the agent has not yet seen the area. Known areas are coloured as walls (grey) or floor (white). The goal is behind one of the black squares.

Assuming that the agent wants to get to the goal quickly, it wants to reveal unseen cells while minimising the length of its path. So, the agent balances the expected rewards, measured in the number of cells revealed, against the costs, measured as the number of moves, in effect maximising the expected utility of the path [87]. Depending on the agent's location, moving can result in revealing new cells and updating its beliefs about the world. To minimise one's expected path the agent must reason about the possible outcomes of observations encountered along the way.

POMDP-based planning captures sequential decision-making such as this, with costs, rewards and observation confidence, all of which vary over time. POMDPs can admit a variety of belief models, observation models, reward functions, cost functions and discount rate models resulting in a multitude of routine decision-making scenarios. For example, an agent who steeply discounts future rewards seeks fast gratification. An agent with a noisy observation model seems to doubt its senses, since it needs to sample each observation several times.

To evaluate the agent's reasoning an observer infers the agent's mental state (desires, goals and beliefs) and infers a planning procedure consistent with the observed behaviour. Following the logic of BToM, an observer attributes intelligence to the agent by assuming

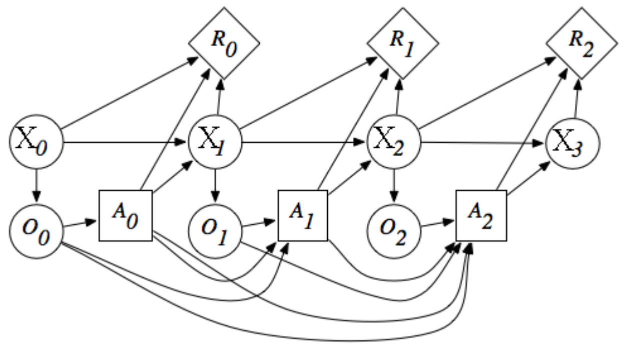


Figure 3.2: (a) POMDP is a sequential process described by beliefs X_t, observations O_t, rewards R_t and actions A_t at time step t. Arrows indicate a causal relationship.

that the agent is using a planning procedure to achieve its goal while minimising cost. Based on this assumption, the observer is likely to conclude that agents who adopt a more efficient planning procedure also look more intelligent.

3.2 The POMDP Model

To define a POMDP model we need to formally describe the agent and the world, which is a maze like the one in Figure 3.1.

The model world is described by discrete time, $0 \leq t \leq T$, and a grid of cells, $W = \{w(i,j)\}\ 0 \leq i \leq width,\ 0 \leq j \leq height, \quad w(i,j) \in \{wall, empty, goal\}$. A unique cell contains a goal: $\exists(i_g, j_g) \rightarrow w(i_g, j_g) = goal$.

The agent knows its location at time t, L_t, and acts based on its belief about the world, which is a set of probabilities $X_t = \{P(\mathbf{x_s})_t\}$. Here $\{\mathbf{x_s}\}$ is the set of all possible world states with the goal in cell s. X_0 is the set of the agent's initial beliefs.

The generic POMDP process shown in Figure 3.2 is a sequence of belief state transitions. The model has no objective knowledge of the world W and makes decisions based on subjective beliefs X. Each belief state, X_t, encodes the agent's knowledge of the world at time t. A transition to the next belief state occurs as a consequence of performing an action A_t and collecting new observations O_t. Actions are chosen based on their expected rewards R_t.

3.2.1 Initial Beliefs

For simplicity, assume that the location of the walls is initially known to the agent. Since a wall can not be the target location, X_0 encodes every wall cell as 0: $\forall i, j,\ w(i,j) = wall \rightarrow X_0(i,j) = 0$. Non-wall grid cells are initialised with positive probabilities, corresponding to the probability that the cell contains the goal. Thus, assuming that every cell is equally likely to contain the goal, X_0 is initialised with a uniform prior: $\forall i, j,\ w(i,j) \neq wall \rightarrow X_0(i,j) = 1/n$, where n is the number of non-wall cells in the world.

Each element of X_t, which is a set encoding beliefs about the goal location, represents the probability that a specific cell contains the goal. For example, for the 6x5 grid-world shown in Figure 3.1 the initial belief X_0 is initialised as follows.

Initial belief, X_0, assuming an unbiased agent:

0.05	0.05	0.05	0.05	0.05	0.05
0.05	0	0	0.05	0	0
0.05	0	0.05	0.05	0.05	0
0.05	0	0	0.05	0	0
0.05	0.05	0.05	0.05	0.05	0.05

3.2.2 Belief Updating

The beliefs are updated based on new observations. The agent can see cells adjacent to its location that are not occluded by walls. Formally, a cell s is visible if the four rays cast from each of the corners of L_t to the corresponding corners of cell s do not intersect walls. The observations are a grid of probabilities $O_t = P(W|X_t, L_t)$, such that for every visible cell (i_v, j_v) $0 \leq O_t(i_v, j_v) \leq 1$ and for every invisible cell (i_i, j_i) $O_t(i_i, j_i) = 1$. For simplicity we restrict the model to deterministic observations setting $\forall i_v, j_v \rightarrow O_t(i_v, j_v) = 0$. (Setting $0 < O_t(i_v, j_v) < 1$ could encode observation uncertainty.)

The agent moves one grid cell at a time, choosing among deterministic actions $A(L_t) \in \{N, S, W, E\}$. Actions moving into a wall are forbidden. In the general case the action model can be probabilistic, encoded by a set of probabilities $P(A_t|X_t)$, so that taking an action has probabilistic results. For example, given non-deterministic actions it is possible that the agent chooses to move West, but the result is a move to the South.

At each step t the agent updates its beliefs by observations using standard Bayesian updating, $X_{t+1} = X_t O_t$ [1]. The distribution of beliefs narrows during the course of the trial as the agent collects more observations.

Consider belief updating in the above example. The agent enters the world at the top-left location $(0, 0)$. The first observation received by the agent reveals the cells to the East (right) and to the South (down) visible from its location, represented by the first row and column of the belief matrix. None of the revealed cells contain the target, and so the set of beliefs X_1 becomes:

Beliefs after the first observation is made, X_1, assuming an unbiased agent:

0	0	0	0	0	0
0	0	0	0.1	0	0
0	0	0.1	0.1	0.1	0
0	0	0	0.1	0	0
0	0.1	0.1	0.1	0.1	0.1

The agent chooses between going South or East based on its belief X_1, as described in the next subsection. Since the value of visiting a cell is proportional to the probability of the cell being the exit, the value of going East should be higher, since more cells can be accessed by a shorter route.

3.2.3 Decision Value

After updating observations the agent chooses an action a from a set A_t based on its value function $Q(a, L_t, X_t)$ calculated from a Bellman Equation for POMDP [51]. In

[1] Here the notation $X_t O_t$ defines element-wise multiplication of the two sets, followed by a normalisation

general terms, the value function is equal to a sum of an immediate reward from performing action a and an expectation of future rewards in states accessible after that action. For a comprehensive review of methods used to model and solve such equations please see [51]. In a general case the outcomes of actions are probabilistic and actions have different costs. The model in this thesis in a special case with deterministic actions and a constant cost:

$$Q(a, L_t, X_t) = X_t(L_t + a)\rho(X_t, L_t + a) + \gamma \sum_{X_{t+1}} \max_{a_i \in A(L_t+a)} \{Q(a_i, L_t + a, X_{t+1})\}, \quad (3.1)$$

where $0 < \gamma < 1$ is a cost factor, affecting the value for future rewards, ρ is the **expected reward** of visiting location $L_t + a$ and $X_t(L_t + a)$ is the probability that the goal is in the cell $L_t + a$. The first term is the value of the immediate reward of performing action a given current beliefs. The second term describes the expected value of future rewards. Given a constant action cost, the expected reward of entering a goal state after moving one cell $\rho(X_t, L_t + a) = k$, where $k > 0$ is a constant. The cost rate γ encodes the requirement of arriving at the goal state sooner.

An optimal agent always chooses the action with the highest value. More generally, the agent chooses an action with a probability proportional to the action value by applying a reward function $R : X_t \times A_t \mapsto \mathbb{R}$, which effectively translates action values to probabilities. The reward function is defined as:

$$R(Q_t(a_i)) = \frac{exp(Q_t(a_i)/\tau)}{\sum_j exp(Q_t(a_j)/\tau)} \quad (3.2)$$

where τ is a *softmax* parameter controlling decision noise described in detail in the next subsection.

3.2.4 Deviations from optimal policy

The schematic model in Figure 3.2 describes a generic agent that acts optimally. However, in reality deviations from optimal policy are common. Such deviations can be intentional, for example as a way to control uncertainty [133] or unintentional, resulting for example from constraints on the available cognitive resources [44].

For this reason, the planning framework considers three causes of deviation from the optimal policy: **guessing** (biased initial knowledge about the world), **decision noise** and **forgetting**. Initial knowledge encodes prior belief about the goal's location. In terms of belief bias, an unbiased agent has uniformly distributed priors and a guessing agent has non-uniformly distributed initial beliefs. Decision noise is captured by τ, a parameter controlling the strength of mapping between action values and the probability of taking an action. Forgetting regresses the agent's beliefs toward the mean, so that the agent's belief about previously observed locations gradually decay. While this model does not fully capture the complexities of human planning, it provides a computational ideal-observer benchmark to test against experimental data.

Softmax decision noise

Softmax decision noise causes deviations from optimal policy, with the probability of deviation inversely proportional to the relative value of the available actions. Decision noise in effect treats values as probabilities by normalising values in a cell to add up to 1. The degree to which decisions deviate from optimality is controlled by τ. As $lim_{\tau \to 0}$ the agent deterministically chooses the action with the highest value, and as $lim_{\tau \to \infty}$ the agent acts at random. Intermediate values of τ result in choosing actions probabilistically, with the probability of choosing an action increasing with its expected value.

The choice of τ depends on the range of values $Q(a, L_t, X_t)$ generated by the model, which depend on the degree of uncertainty over the goal location. The implementation used in this thesis produces action values $0 \leq Q(a, L_t, X_t) \leq 1$, where 1 is the value of entering the goal state and intermediate values between 0 and 1 reflect the estimated distance to the goal. Importantly, although model values fall in the same range as probabilities, the model values are not probabilities since the sum of values of all actions in a cell is not required to add up to 1.

To choose a good range of τ for model values in this range we run a simulation with value distributions of length 4 sampled from $(0, 1)$ so that: $v_{i,1} + v_{i,2} + v_{i,3} + v_{i,4} = 1$,

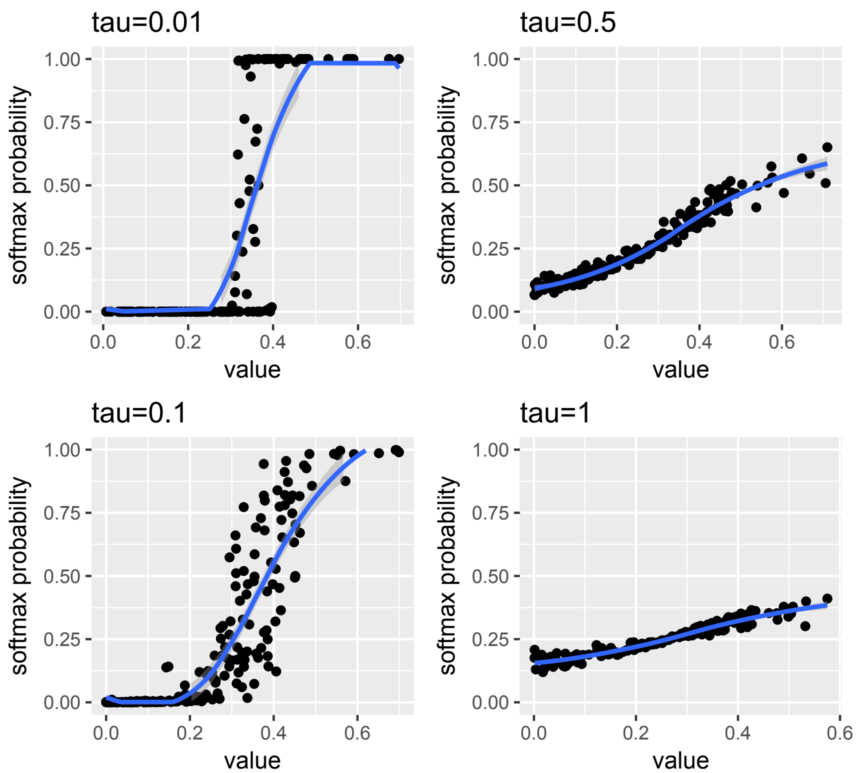


Figure 3.3: A simulation of mapping probability values sampled from $(0, 1)$ to softmax probabilities.

$i \in (1, 50)$, sampling values as if they were probabilities. To each set of probabilities v_{j_i} corresponds a set of softmax probabilities pr_{j_i}. Figure 3.3) shows softmax probability plotted against value, showing that $\tau = 1$ results in a practically equal probability of choosing any value, $\tau = 0.001$ results in choosing optimally, and $\tau \in [0.01, 0.1]$ produces intermediate probabilities.

Guessing

Biased initial knowledge about the world simulates guessing: an agent with biased initial knowledge appears to guess the target location. In contrast, a non-biased agent assumes a uniformly distributed probability of each unseen location containing the goal. In the example discussed above, a biased agent might start with a hunch that the goal is at the bottom-left:

Example of an initial belief, X_0*, assumed by a biased agent. One cell is considered highly likely to contain the goal –* 0.81 *in contract to* 0.01 *for all other cells:*

0.01	0.01	0.01	0.01	0.01	0.01
0.01	0	0	0.01	0	0
0.01	0	0.01	0.01	0.01	0
0.01	0	0	0.01	0	0
0.01	0.81	0.01	0.01	0.01	0.01

Behaviours produced by decision noise and by biased priors often resemble one another. For example, going down in the example shown in Figure 3.1 may result either from a hunch that the goal is in the bottom-left corner, as captured by a biased belief or from choosing the option with a lesser value due to decision noise.

Forgetting

A forgetting parameter $0 \leq f \leq 1$ regresses the agent's beliefs toward the mean after each step so that the agent gradually forgets whether a previously observed cell is empty and must re-check already visited locations.

3.2.5 Inferring a planning procedure from observed behaviour

The planning framework can be applied to infer parameters that generated a sequence of actions by calculating the probability of the agent's path given the parameters (equation 3.4 and equation 3.3). So, the observer infers the most likely combination of parameters driving planning, and uses it to predict how well the agent will perform in the long run.

The observer who assumes that the deviations from optimal behaviour arise from decision noise will perform an inference that takes only τ into account, the *SOFTMAX* model:

$$P(\tau \mid W, path) \propto \prod_{L_i, a_i} P(\tau \mid W, L_i, a_i) \tag{3.3}$$

A slightly more sophisticated observer infers the causes of sub-optimal behaviours by varying the values of three parameters: **forgetting, initial beliefs** and **softmax decision noise**, the *Model3* model.

$$P(\tau, f, X_0 \mid W, path) \propto \prod_{L_i, a_i} P(\tau, f, X_0 \mid W, L_i, a_i) \tag{3.4}$$

In theory, the observer could keep in mind many possible causes of the agent's behaviour. However, evaluating efficiency in this way can be computationally costly. The observer may need to consider many counter-factual agents and environments to come to a conclusion about the fitness of this particular agent in relation to hypothetical others. An observer may thus have empirical heuristics such as: 'Agents with decision noise will generally do better than agents who forget' or 'An agent with one source of noise is better than an agent with two'. Even so, evaluating the agent still incurs the computational cost of inferring its parameters.

Attributed Intelligence

This thesis tests two hypothesis of attributed intelligence. The **outcome hypothesis** supposes that participants attribute intelligence based on the agent's outcome. The **efficiency hypothesis** suggests that participants attribute intelligence based on the agent's efficiency. To differentiate between attributing intelligence to outcome and to efficiency, we need a formal definition of efficiency and of outcome.

The agent's **efficiency** is defined as a measure of how well the agents' planning procedure maximises reward and minimises cost. To evaluate the agent's efficiency an observer must understand the costs and rewards facing the agent and how to plan under uncertainty to maximise the probability of success. The agent's efficiency can be measured by various model-based metrics, such as: (1) a relative rankings of agents with different sets of parameters, (2) the magnitude of the inferred decision noise and (3) the fraction of optimal moves made by the agent.

Likewise, the agent's **outcome** is defined as a metric of how the agent does in a single trial. The agent's outcome can be measured by perceptual metrics, such as: (1) the number of moves to achieve the goal or (2) the number of times a cell is revisited [2]. Outcomes differ with transient aspects of the environment. While inefficient planning will occasionally produce low cost lucky outcomes, an efficient optimal planning is expected to maximise the expected reward overall. In effect, the outcome is a random variable, which varies trial to trial, and efficiency is the expectation of outcome.

Implementation

There are many ways to implement POMDP planning. A common approach to solving POMDP planning completely is n-step lookahead [51], which entails simulating a decision tree with nodes every time an action can be taken or an observation is made, so as to achieve the highest average payoff given all possible future outcomes. Approaches that are common with solving Markov Decision Process (MDP), such as Value Iteration, [110] can be used to plan the trajectory after the goal location is known, for example if there is only one dark cell remaining. MDP planning can be also used to simulate guessing, however such an interpretation of guessing is not modelled in this thesis.

Animated videos were generated in a free Java programming environment called Processing 3.0, which is convenient for generating graphics. The modelling, inference and video generation described in this manuscript were run using a Java simulation of equations 3.1 and 3.2. The implementation evaluates each action (N, W, S, E) at the agent's location while keeping track of the visited cells, so that each cell is accounted for only once. In case of the earlier example, the order in which the value of each cell will be counted is shown in Figure 3.4. The agent is located as $(0, 0)$ and can go East or South.

[2] For example, during pilots some participants pointed out that agents who backtrack seem less intelligent.

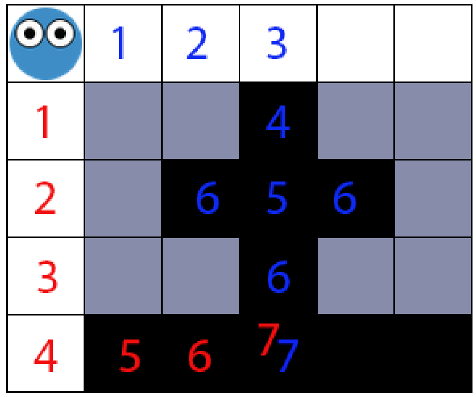


Figure 3.4: The order in which the value of cells will be accounted for.

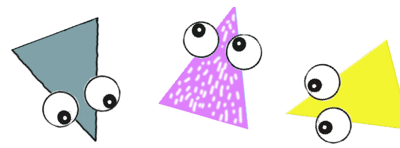
In the example above, assuming that the agent is optimal with a belief X_1 as described above and $\gamma = 0.7$, the value of going East for the steps shown in Figure 3.4 is:

$\gamma^0 \times 0 + \gamma^1 \times 0 + \gamma^2 \times 0 + \gamma^3 \times 0.1 + \gamma^4 \times 0.1 + \gamma^5 \times 0.3 + \gamma^6 \times 0.1 + ... = 0.343 \times 0.1 + 0.24 \times 0.1 + 0.16 \times 0.3 + 0.11 \times 0.01 + ...$ The full value of going East is 0.148 and the full value of going South is 0.093.

The inference implementation also tracks the frequency with which the agent's trajectory includes zero-valued actions. An example of a zero-valued action is entering empty rooms. Humans taking zero-valued actions are most likely doing so unintentionally, likely due to inattention. The frequency of such zero-value moves can be a proxy metric of attention of human participants. A metric of attention is essential for interpreting the results of experiments conducted online.

In summary, this chapter describes a formal model of agents in a partially observable environment and of an observer evaluating the intelligence of such agents. The model represents deviations from optimal panning, as generated by decision noise, biased beliefs and forgetting. The model can be used to generate examples of optimal and sub-optimal planning behaviours and compare such model-generated normative solutions to how people actually plan. An observer who attributes intelligence to an agent's efficiency is represented by a Bayesian inference over the agent's planning procedure. An observer who attributes intelligence to outcome is represented by model-free metrics such as the length of the agent's path.

The following chapters apply this model to explaining how humans plan and to how they evaluate plans made by others. Chapter 4 describes an experimental study in which people attribute intelligence to computer-generated agents controlled by the model. While the experiment supports the efficiency hypothesis of intelligence, it also revels that a sub-population attributes intelligence to outcome. Chapter 5 is an experimental study of how people actually plan, in which human plans are analysed both empirically in terms of psychophysical measurements, and by comparison to models. The results show that the plans averaged across individuals, taking a majority vote at every step, are indeed model-optimal. However, about a half of the individual solutions deviate from the optimal path. Chapter 6 looks for correlations between individuals' own planning abilities and their attributions of intelligence and shows that more proficient planners are generally more likely to attribute intelligence to efficiency. Chapter 7 describes an eye-tracking experiment that elucidates the nature of mental processing required by planning tasks.



Chapter 4

Attributed Intelligence: Outcome and Efficiency

People expect others to behave efficiently [28, 48, 43, 5] and describe behaviour as stupid if it disagrees with their expectations of efficiency [2]. However, reality is often different: movements slip, attention wanders, memories change, beliefs are false and decisions are distorted by biases [137]. How do people make sense of inefficient agents? According to the **efficiency hypothesis**, attributed intelligence reflects the observer's inference of the agent's efficiency. At the same time, people may see lucky others as deserving [71] and assume that agents are responsible for their outcomes regardless of external circumstances [38]. According to the **outcome hypothesis** attributed intelligence reflects the observers appraisal of the agent's outcome.

Consider an intelligent agent exploring a 2-D maze in search of a treat. The agent is familiar with the layout of the maze and knows there is a reward in the maze, but does not know where the reward is. The agent's view is obstructed by walls, so that at any time only part of the maze is visible. The reward is equally likely to be in any location. The agent incurs a cost when it moves, reflecting the physical effort of motion. Figure 4.1 shows several examples of such mazes, seen from an overhead view. In an environment such as this the agent's outcome can be measured by the path-length to the goal or by inverse of the number of times the agent revisits previous locations. In the two experiments described below participants view animated movies of agents and evaluate their intelligence.

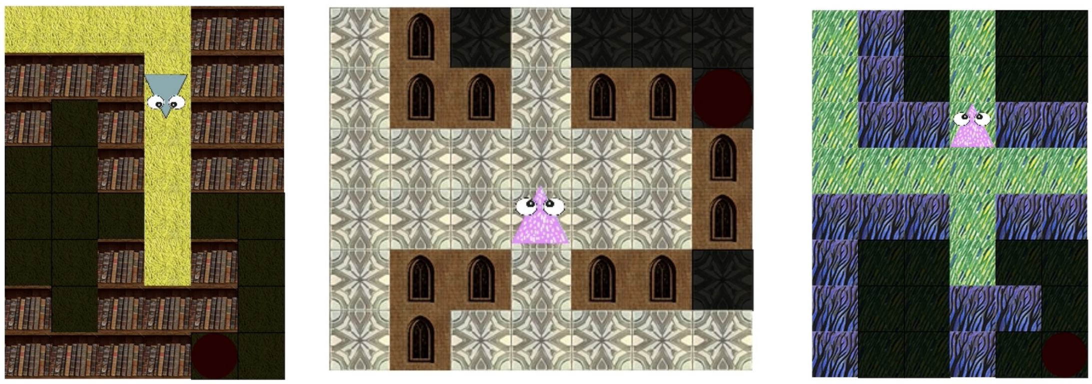


Figure 4.1: Examples of maze environments shown to participants. Mazes are from a bird's eye view. Dark cells indicate that the agent has not yet seen the area, while cells previously seen are patterned as walls or floor. The agent can move through empty areas, but not through walls. The goal (a red circle) can be anywhere.

In our maze example an efficient agent may visit rooms in a systematic order that ensures seeing the whole maze in the fewest steps. In contrast, an inefficient agent may visit rooms in an inefficient order and seem the less intelligent of the two. An agent who moves at random until it stumbles upon the reward might appear as outright unintelligent. Regardless of the agent's planning procedure, given an uncertain environment and the probabilistic nature of rewards, the agent is expected to encounter a variety of outcomes. It might be **lucky** and find the goal in the first place it looks or it might be **unlucky** and search the maze exhaustively.

An observer seeing a lucky agent might wonder whether the agent's trajectory is justified and attribute intelligence only if the agent's planning is efficient. In contrast, the observer might decide that if the agent achieved its goal quickly, then it must have done something right, and so must be intelligent. Likewise, seeing an efficient, but an unlucky, agent the observer might reason that the agent would have achieved a better outcome under most counterfactual scenarios, and judge the agent as intelligent. In contrast, an observer who attributes intelligence to outcome might decide that the agent is not intelligent, since it

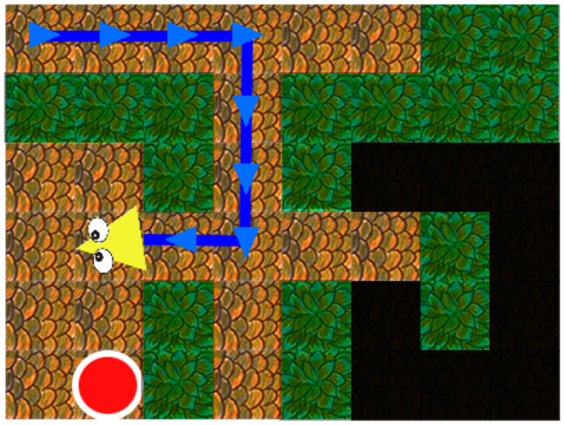


Figure 4.2: An example of an *optimal-lucky* condition. The agent makes an (optimal) decision to go into the room on the left (reader's perspective). Incidentally, it finds the goal quickly.

took a long path to the goal. Alternatively, it is possible that attributed intelligence is explained neither by outcome nor efficiency, in which case we should see no relationship between the agent's planning or outcome and the intelligence attributed to the agent.

The first experiment is an exploratory study run with a small group of participants in the lab to validate the experimental procedure and test the two hypotheses. The results support the efficiency hypothesis. The second experiment applies the experimental procedure of the first experiment to a larger group of participants recruited online and shows that intelligence attribution varies among participants. While two-thirds of the sample attributed intelligence to efficiency, the remainder attributed intelligence to outcome.

The experiments received ethics clearance from a University of Waterloo Research Ethics Committee and from the MIT Ethics Review Board. Full instructions are available in the Appendix.

4.1 Experiment 1

If the efficiency hypothesis is correct, then there should be no difference in ratings of optimal agents, regardless of their outcome, and most of the variation in attributed intelligence is explained by model-based metrics of agent's efficiency (defined below). If the outcome hypothesis is correct, then lucky agents should be rated higher than the unlucky ones, regardless of their planning policy, and most of the variation in attributed intelligence is explained by the agent's outcome, measured as the length of path in **steps** and the frequency of **revisits** of the same cells. To test the two hypotheses, participants rate the intelligence of various optimal and sub-optimal agents generated by the POMDP framework. Both optimal and sub-optimal agents incurred a variety of lucky (low cost) and unlucky (high-cost) outcomes.

The agents are sampled from six POMDP planners generated by varying the values of the three parameters in *Model3* (forgetting, initial beliefs and decision noise). Each planner is assigned an **optimality rank**, an ordinal measurement reflecting its expected performance as measured by simulating planning trajectories on a set of procedurally generated mazes [4]. In addition, the *random* planning model showed an agent that appeared to move at random, but constrained to reach the goal in the number of steps not greater than the average number of steps taken by the other agents. The full list of planning behaviours along with their optimality ranks is shown in Table 4.1.

4.1.1 Participants

Twelve participants were recruited from a University of Waterloo participant pool, 4 females and 8 males, median age 27.5. The participants were volunteers and their small group size is due to this experiment being an initial study to test the experimental design.

4.1.2 Method

Participants are told that they would watch movies of different mice looking for a treat, and instructed that each dark square was equally likely to hide the treat. Participants

Rank	Label
1	optimal (lucky or unlucky)
2	softmax
3	lucky-guess
4	softmax-guess
5	softmax-forgetful
6	softmax-forgetful-guess
7	random

Table 4.1: Agent policies generated varying the three parameters according to *Model3*. The conditions are ranked according to their performance across different mazes. For example, the lucky-guess agent is rank 3, and a random agent is rank 7, indicating that on average a lucky-guess agent would find the goal faster than a random agent, over a set of different mazes.

are told that a mouse wants to get to the treat as quickly as possible. The appearance of the mouse, the layout of the maze, the textures of the maze, and the POMDP parameters were varied on each trial. At any time the cells yet unseen by the mouse were portrayed as dark, and cells previously seen portrayed as patterned (see Figure 4.2). In every movie, the mouse eventually found the goal.

Participants are asked to watch each movie at least once, and to rate the intelligence of each mouse using a Likert scale ranging from 1 (least intelligent) to 5 (Most intelligent). Full instructions are available in the Appendix, and the set of stimuli is available at http://www.cgl.uwaterloo.ca/~mkryven/attributionStimuliLab.zip. Each participant read the instructions on the computer screen, and viewed four familiarisation examples followed by the stimuli in two blocks. After viewing each movie, participants marked their rating in a provided answer sheet.

4.1.3 Stimuli

Thirty animated movies of an agent (described as a 'mouse') were shown on a high-resolution LCD display in two blocks, using Psychtoolbox [16]. The movies were presented in two blocks to compare the early and the late trials, to see if ratings change over time. After viewing each movie a participant marked their rating in a provided answer sheet. The movies were generated by solving POMDP planners on 9 mazes with each of the planning models described above. Each movie shows a triangular agent moving through a maze at a uniform speed. The agent orients in the direction of its motion, so that in the corridors the agent moves straight ahead, and upon arriving at intersections it makes turn between -90 and 90 degrees until it is facing in the chosen direction. Each movie was labelled according to its agent condition. A 'lucky' or 'unlucky' label reflects the outcome cost. The location of the goal was counterbalanced so that in one half of the mazes the mouse could find it equally quickly by optimal planning or by pure luck. In the other half of the mazes, the goal was placed so that an optimal mouse had to search exhaustively, while a mouse making a lucky guess could finish in fewer steps.

In summary, there were 8 possible movie conditions: *optimal-lucky, optimal-unlucky, softmax, lucky-guess, softmax-guess, softmax-forgetful, softmax-guess-forgetful* and *random*. Each of the conditions was shown four times. The *random* condition occurred twice, once in each block. Both *random* movies showed agents getting to the goal in fewer than 20 steps, matching the average length of movies in non-random conditions.

4.1.4 Results

Most of the variation in the ratings of each trial can be explained by the agent's condition, with more efficient agents rated as more intelligent. Analysis of Variance (ANOVA) of *rating* [1] against $(condition, steps \times revisits, block)$ shows a significant main effect of condition, $(F(7, 18) = 67.8453, p < .0001)$, with mean ratings of each condition are shown in Figure 4.3). There was also a small main effect of block $(F(1, 17) = 7.6421, p = 0.013)$, with early trials rated higher than later trails $(d = 0.24, p = p = .013)$. The effects of steps and revisits are not significant $(p = .3)$. The adjusted R^2 of the regression model is .9423.

According to Tukey HSD post-hoc test, the difference between the ratings of *optimal-lucky* and *optimal-unlucky* agents was not significant $(p = 1)$, in support of the efficiency

[1]Mean ratings of each trial calculated between participants.

hypothesis. At the same time, there was a significant difference between the *optimal-lucky* and the *lucky-guess* conditions ($p = .005$, $d = .79$), meaning optimal agents were rated higher then suboptimal lucky ones. The *softmax* agents are rated as more intelligent than *softmax-forgetful* ($p < .0001$, $d = 1.1$) and the *softmax-forgetful* higher than *softmax-guess-forgetful* ($p = .001$, $d = 0.9$). The *lucky-guess* agents are rated higher than *softmax-forgetful* ($p = .002$, $d = 0.88$) condition. The difference between *lucky-guess* and *softmax* conditions was not significant ($p = .92$). In summary, the participants attributed the highest intelligence to optimal agents, and judged forgetful agents as less intelligent than the noisy agents. Random-looking agents were seen as least intelligent of all, and are not included in the subsequent analysis. Thus, the results support the efficiency hypothesis over the outcome hypothesis, meaning people are able to tell when an agent was being lucky, and seeing rationally planning agents as more intelligent than agents who got to the goal quickly, but did so in a way that indicated sub-optimal planning.

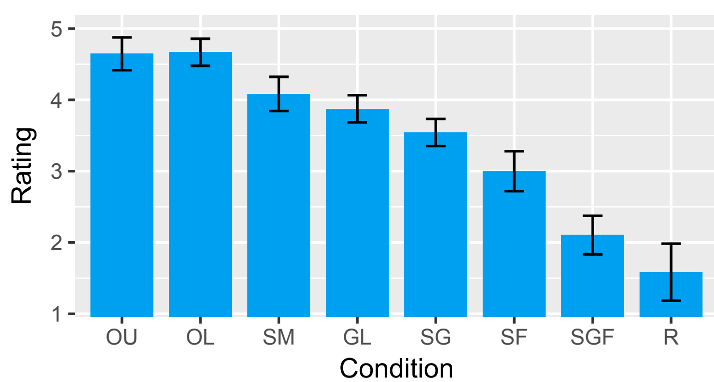


Figure 4.3: Ratings by condition *OL-optimal-lucky*, *OU-optimal-unlucky*, *SM-softmax*, *GL-lucky-guess*, *SG-softmax-guess*, *SF-softmax-forgetful*, *SGF-softmax-guess-forgetful*, *R-random*. Error bars indicate 95% confidence intervals.

A a significant multiple linear regression predicts attributed intelligence based on optimality rank and block, $F(2, 25) = 117.4, p < .0001$ with an adjusted $R^2 = .896$. The predicted rating is equal to $4.961 - 0.462(rank) - 0.179(block)$, where block is coded as 1 or 2 and rank is given in Table 4.1. Optimality was a significant predictor or rating ($p < .0001$), but not block ($p = 0.11$).

However, if the inference of efficiency is not available to the observer, could revisits and steps be used as a heuristic for attributing intelligence? ANOVA of *rank* against ($revisits \times steps$) shows a significant effect of steps ($F(1, 24) = 9.3630, p = .005$) and of revisits ($F(1, 24) = 7.4435, p = 0.01$), thus indicating the number of steps and revisits could be used as a proxy to optimality. For example, forgetting leads to a longer route and more revisits. A multiple linear regression predicting ratings based on revisits, steps and block is significant, $F(4, 23) = 7.152, p < .0001$ with an adjusted $R^2 = .476$. Predicted rating is equal to $4.27 + 0.07(steps) - 0.29(revisits) - 0.5(block)$. However, neither steps ($p = .2$), revisits ($p = .1$) nor block ($p = .07$) are significant predictors.

In summary, model-based optimality metrics explain some of the variance in human attributions of intelligence better than perceptual outcome cues. However, model-based optimality metrics also accurately predict the absence of a significant difference between the ratings of lucky or unlucky optimal agents, in support of the efficiency hypothesis.

4.1.5 Discussion

The results of Experiment 1 validate the experiment design and support the efficiency hypothesis. Participants appear to attribute intelligence in proportion to the agent's efficiency, which measures how closely the agent approximates the optimal path. However, the planning problems viewed by participants are simple, and it is unclear how intelligence may be attributed if the participants themselves cannot efficiently approximate an optimal trajectory in the agent's task.

Although in practice people often revisit places while searching, forgetful agents in the task were consistently judged as unintelligent. Given that the revealed areas of the maze remain visible to the observer, the observer probably assumes that they are also visible to the agent. In addition, forgetful agents make comparatively more revisits and take longer paths, which introduces a dependency between planning accuracy and outcome. For this reason, forgetful agents are not included in the subsequent experiment. Excluding forgetful agents decreases the discriminating power of the optimality rank inferred according to *Model3*. Other graded metrics of efficiency, such as **decision noise** inferred by fitting *SOFTMAX*, and the **fraction of optimal steps** are more useful in that regard.

In addition, in this experiment the agent always finds the goal. However, in many real-life situations observers attribute intelligence having seen only brief examples of behaviour, without witnessing an outcome. For example, someone admitting a mistake is more intelligent than a person insisting on being right; having a savings account is more intelligent than buying a cell phone insurance. Such inferences occur because the observer compares the statistical expectation of the consequences of the agent's actions to that of possible alternatives. To address this limitation the next experiment includes examples of partial plans, in which the movie stops part-way, before the outcome is seen.

4.2 Experiment 2

The second experiment tests the efficiency hypothesis and the outcome hypotheses with a more diverse selection of online participants. It includes examples of incomplete solutions, in which the movie stops after the agent chooses one of the rooms, but before the reward is found. Forgetful agents are excluded, which leaves four POMDP planners with varying initial beliefs and decision noise: *optimal, softmax, lucky-guess* and *softmax-guess*. A *random* model represents an agent that appears to move at random, though constrained to reach the goal in the number of steps not greater than the average number of steps taken by the other agents.

Complete trials are assigned three model-based efficiency metrics: an **optimality rank** and **decision noise** τ_i inferred by the SOFTMAX model, and the **fraction of optimal steps**, $optstep_i$, where i indexes the trial. Agents' efficiency is assumed to be proportional to the fraction of their optimal steps, and inversely proportional to decision noise. In addition, complete examples are assigned two outcome metrics: the number of **steps** and **revisits**. Incomplete trials are labelled as *optimal* or *suboptimal* since they are too short to generate an accurate inference of model-based metrics. The efficiency and outcome metrics are independent, the number of steps and τ_i have no measurable correlation, Spearman $r = -.02, p = .94$.

If participants attribute intelligence based on efficiency, they should rate optimal agents higher, even if the outcome is not shown, and most of the variation in intelligence attributed to agents in complete trials should be explained by the agent's efficiency according to model-based metrics. If participants instead attribute intelligence based on outcome, there should be no significant difference between their ratings of optimal and sub-optimal incomplete trials, and most of the variation in intelligence attributed to complete trials should be explained by the agent's outcome. It is also possible that participants attribute intelligence based neither on efficiency nor on outcome.

4.2.1 Participants

Thirty-two participants were recruited via Amazon Mechanical Turk, restricted to US participants, 2 were discarded for failing to answer questions. The analysis thus included 30 participants (11 females, median age 34).

4.2.2 Stimuli

Forty animated movies were generated by 5 POMDP planners, including random, on 8 different mazes. The appearance of the agent, the layout, textures, the orientation of each maze, the location of the goal, and the specific planning model used to generate each movie are varied trial to trial. Movies are labelled according to the agent's planning model, and as 'lucky' or 'unlucky' according to the agent's outcome, resulting in 8 conditions: *optimal-lucky, optimal-lucky, softmax, lucky-guess, softmax-guess, optimal-partial, suboptimal-partial* and *random*. The list of conditions is shown in Table 4.2.

Each of the *optimal-unlucky, lucky-guess, softmax* and *softmax-guess* conditions occurred 4 times and *optimal-lucky* occurred three times. Two movies generated by the *random* model constrained to 20 steps to match the average length of movies in non-random conditions occurred two times as an attention check. Half of the trials were incomplete. Of the incomplete movies, 12 showed a suboptimal decision (*suboptimal-partial* condition) and 7 showed an optimal decision (*optimal-partial* condition).

Rank	Label
1	optimal-(lucky or unlucky)
2	softmax
3	lucky-guess
4	softmax-guess
5	random
	suboptimal-partial
	optimal-partial

Table 4.2: List of conditions along with their optimality rank.

4.2.3 Method

Participants read the instructions on a computer screen in a web browser, and viewed 4 familiarisation examples followed by 40 movies randomly sorted into two blocks to compare early and late trials to see if ratings change over time. After viewing each movie, participants selected a rating from a Likert scale between 1 (least intelligent) to 5 (most intelligent). At the end of the survey participants were asked: *How did you make your decision?*. Full instructions are available in the Appendix, and the set of stimuli is available at http://www.cgl.uwaterloo.ca/~mkryven/attributionStimuliTurk.zip.

4.2.4 Results

The results are analyzed separately for complete and incomplete trials. For complete trials, repeated measures ANOVA of *rating* [2] against $(condition, steps \times revisits, block)$ shows a significant main effect of condition, $(F(5, 11) = 53.1001, p < .0001)$. There were no significant effects of steps $(F(1, 11) = 1.8901, p = .196)$, revisits $(F(1, 11) = 0.7212, p = .4)$ or block $p =, 7$. The adjusted R^2 of the regression model is .929. According to Tukey HSD post-hoc test the difference between *optimal˙lucky* and *optimal˙unlucky* agents is not significant, $d = 0.47$, $p = .219$, as well as the difference between *lucky˙guess* and *softmax* agents is $d = 0.57$, $p = .796$. As in the first experiment, the *random* condition was judged

[2] These are mean ratings of each trial calculated between participants.

as least intelligent ($p < .0001$). For partial trials, Welch Two Sample t-test indicates that the ratings of optimal trials ($M = 3.49, SD = 0.253$) are not significantly different from the suboptimal trials ($M = 3.3, SD = 0.274$), $t(13.583) = 1.4726, p = .163$. The mean ratings of each condition are shown in Figure 4.4). In contrast to the first experiment, there is not enough evidence to support the efficiency hypothesis. However, the result of ratings of incomplete trials partly support the outcome hypothesis.

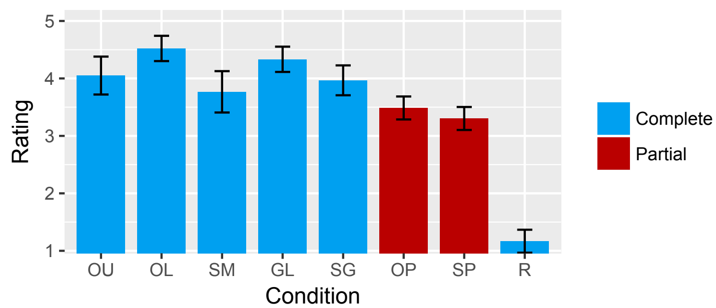


Figure 4.4: Comparing ratings between the two groups *OL-optimal-lucky*, *OU-optimal-unlucky*, *SM-softmax*, *GL-lucky-guess*, *SG-softmax-guess*, *OP-optimal-part*, *SP-suboptimal-part*, *R-random*. Error bars indicate 95% confidence intervals.

Two Groups of Participants

The online participants in the second experiment might have preferred lucky agents. Alternatively, they may have used more than one way of attributing intelligence. To find out, we analyse their verbal responses. Two independent raters coded the participant answers to the question *'How did you make your decision?'* into two groups: **outcome** and **efficiency**. Comments that fall into the **efficiency** group talk about looking for efficient planning strategies that can minimise the agent's cost over time, which corresponds to our definition of efficiency. Comments that fall into the **outcome** group talk about the length of path or the number of revisits on the current trial. For example, a response was coded as **efficiency** if it said: *'Based on if the mouse checked every nook and cranny.'* and as **outcome** if they said *'Based on how long it took for the rat to find the treat'*.

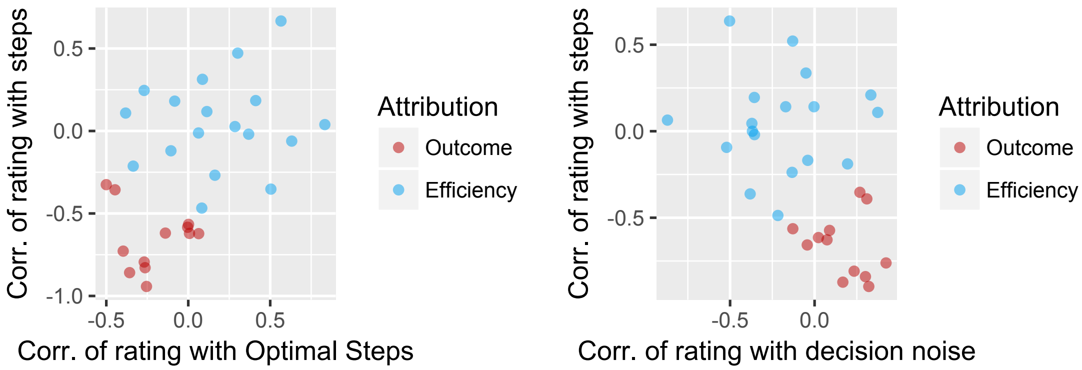


Figure 4.5: Spearman rank correlations of individual ratings with path length and with model-based metrics coloured according to self-report as Outcome or Efficiency. Each data-point corresponds to a participant.

The raters agreed on 27 out of 30 participants, coding 8 of them as **outcome** and 19 as **efficiency**. The remaining 3 were assigned to the outcome group after a discussion, since their comments were hard to interpret (for example *Depends on if the mouse made logical decisions with no backtracking.*). The results that follow are not affected by the choice of method for handling these 3 exceptions.

Another way to cluster participants is to plot a 2-dimensional space of Spearman rank correlations of individual ratings with model-based and with outcome metrics. The model-based dimension is represented by calculating the correlations of individual ratings with $optstep_i$ or with τ_i. The outcome dimension is represented by correlations of individual's ratings with the length of each trial. The plots shown in Figure 4.5 situate participants in the space of the two correlations. Each point represents a participant, coloured by the individual's self-report as 'Outcome' or 'Efficiency'

Unsupervised clustering on the correlation space using a Gaussian Mixture Model (GMM) [12] identifies two clusters of participants as more likely than three clusters or one cluster (BIC of a one-component model -19.5418, two-component -30.0407 and three-component -24.8234). Two-component GMM also coincides with the two groups of participants independently coded as 'Outcome' or 'Efficiency' based on their free-form response as to how they made their judgements. The 8 participants independently agreed on as

outcome by both raters (using verbal measures) were also identified as belonging to the outcome cluster by the two-component GMM. So, participants in the two groups attributed intelligence in opposite ways: either based on the agent's efficiency, or based on the agent's outcome.

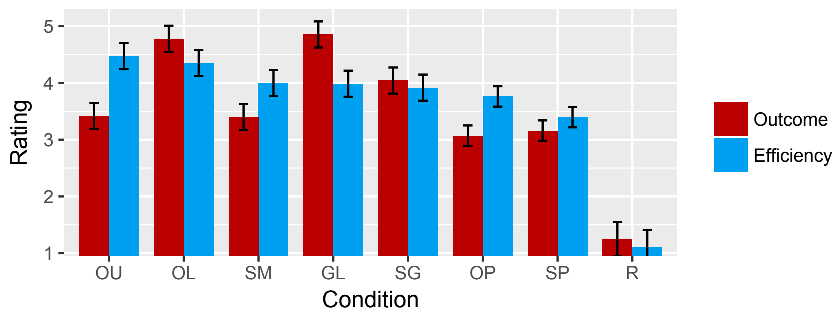


Figure 4.6: Comparing ratings between the two groups *OL-optimal-lucky*, *OU-optimal-unlucky*, *SM-softmax*, *GL-lucky-guess*, *SG-softmax-guess*, *OP-optimal-part*, *SP-suboptimal-part*, *R-random*. Error bars indicate 95% confidence intervals.

Next, we analyse the responses of the **efficiency** and **outcome** groups. The mean ratings of each condition by group of participants are shown in Figure 4.6. For complete trials, repeated measures ANOVA of *rating* against (*condition* $\times$ *group*) shows significant effects of condition ($F(5, 30) = 62.5010, p < .0001$) and a significant interaction between condition and group ($F(5, 30) = 8.6635, p < .0001$). The adjusted R^2 of the regression model is .894.

According to Tukey HSD the **outcome** but not the **efficiency** group rated lucky examples higher in terms of intelligence. The difference between the **outcome** group's ratings of *optimal˙lucky* and *optimal˙unlucky* agents is significant ($d = 1.48, p = .0001$) as well as between ratings of *guess˙lucky* and *softmax* agents (difference $1.55, p < .0001$). The differences between the **efficiency** group's ratings of the lucky and unlucky agents are not significant. The difference between ratings of *optimal˙lucky* and *optimal˙unlucky* agents is $p = 1$, and between ratings of *guess˙lucky* and *softmax* agents is $p = 1$. In summary, the outcome, but not the efficiency participants preferred lucky agents regardless of their planning policy.

Next, we analyse the ratings of optimal and suboptimal (excluding the *random* agents) agents by both group. For the efficiency group, Welch Two Sample t-test indicates that the ratings of optimal trials ($M = 4.4, SD = 0.19$) are significantly higher from the suboptimal trials ($M = 3.99, SD = 1.1$), $t(16.128) = 3.8936, p = .001$. In contrast, for the outcome group the ratings of optimal ($M = 3.95, SD = 1.25$) and suboptimal ($M = 4.1, SD = 0.8$) agents are not significantly different $t(12.601) = 0.31079, p = .7$. So, the efficiency, but not the outcome participants prefer efficient agents regardless of their outcome.

For ratings of partial trials by the outcome group, Welch Two Sample t-test indicates that the ratings of optimal trials ($M = 3, SD = 0.645$) are not significantly different from the suboptimal trials ($M = 3.1, SD = 0.487$), $t(10.039) = 0.47357, p = .6$. In contrast, ratings of partial trials by the efficiency group were higher for optimal ($M = 3.7, SD = 0.14$) than for suboptimal trials ($M = 3.39, SD = 0.24$) trials, $t(16.982) = 3.9689, p = .001$. So, there is no evidence that participants who self-identify as attributing intelligence to outcome rate optimal and suboptimal agents differently if the outcome is not observed, which provides conclusive evidence in favour of attributing intelligence to outcome. In contrast, those who self-identify as attributing intelligence to efficiency attribute more intelligence to optimal decisions, even when they cannot see the outcome.

A linear regression was calculated to predict intelligence rating attributed by the efficiency group to full trials based on optimality rank (excluding the random agents). A significant regression equation was found, $F(1, 17) = 9.305, p = .007$ with an adjusted $R^2 = .3157$. Predicted rating is equal to $4.339 - 0.153(rank)$, where rank is given in Table 4.2. Optimality is a significant prefictor of rating ($p = .005$). Intelligence attributed by the efficiency group may be also explained by τ_i. Calculating a polynomial regression to predict ratings of full trials based on τ_i results in a significant equation $F(2, 16) = 8.243, p = .003$, with an adjusted $R^2 = .4459$. Predicted rating is equal to $4.14474 - 0.64333(\tau_i) + 0.66787(\tau_i^2)$. Both τ ($p = .01$) and τ^2 ($p = .048$) are significant predictors of rating. In contrast, a linear regression of rating against $steps \times revisits$ is not significant, $F(3, 15) = 0.1585, p = .923$.

A linear regression of intelligence attributed by the outcome group to full trials based on optimality rank (excluding the random agents) is not significant, $F(1, 17) = 1.118, p = .3$ as well as the polynomial regression of rating based on τ_i, $F(2, 16) = 1.116, p = .4$. However, a significant linear regression equation was found predicting ratings of the outcome group against $steps \times revisits$, ($F(3, 15) = 32.99, p < .0001$, Adjusted $R^2 = .842$). Predicted rating is equal to $6.336 - 0.133(steps) - 0.36(revisits)$. Only steps ($p = .002$) but not revisits ($p = .08$) were a significant predictors. In summary, the ratings of participants in the **efficiency** group can be predicted by the model-based metrics, supporting the efficiency hypothesis. However, the ratings of participants in the **outcome** group are better explained by outcome. [3]

4.2.5 Discussion

Converging evidence from model-based statistical analysis and verbal responses reveals participants used two different methods for attributing intelligence: attributing intelligence to efficiency, and attributing intelligence to outcome (Figure 4.6). While the majority (19 participants) rated more efficient solutions as more intelligent, a minority relied instead on outcome. Attributing intelligence to outcome was not observed in the first experiment. However, MTurk participants are likely to exhibit a wider variety of behaviours.

How should we explain attributing intelligence to outcome? Participants in the outcome group may not understand the agent's task and fall back on counting the number of steps. To address this limitation subsequent experiments described in this thesis include an instruction quiz. The responses of participants who fail to answer the quiz correctly are not analyzed.

In addition, participants may attribute intelligence based on prior experience with other humans, and find it difficult to interpret model-generated behaviours. If so, participants could apply the outcome heuristic to behaviours that they have difficulty interpreting and attribute high intelligence to any behaviour they can rationalise as opposed to just the optimal strategies. For example, an observer might think: 'I see why someone would search that way, so this agent is intelligent.' or 'That does not make sense.'. If so, then

[3] The relationship between the ratings of each group and model-based or outcome metrics can be also shown by calculating simple Pearson correlations.

participants should rate all behaviours generated by other humans equally. Conversely, observers who attribute intelligence to efficiency should rate trajectories in proportion to their model-based efficiency metrics. To address this limitation in Chapter 6, participants evaluate behaviours generated by other humans.

Moreover, the efficiency hypothesis implicitly assumed that people can generate the optimal strategy. But what if the optimal strategy is not available to the observer? According to Simulation Theory participants could use themselves as a model for evaluating others, in which case people who plan sub-optimally should attribute high intelligence to other sub-optimal planners (e.g. If I play lottery I should think that playing lottery is smart.) as opposed to just the lucky sub-optimal trajectories. In contrast, participants who specifically attribute intelligence to outcome, should rate only lucky sub-optimal agents as intelligent. At the same time, even participants who attribute intelligence to efficiency may judge others might depend on the basis od their own planning skills. Young children, for example, attribute goals to reaching actions only if the action is known to the child: an experimenter pulling on the tablecloth to get a duck makes sense to an infant only after the infant acquires the pulling action [103]. If participants' judgements of intelligence are limited by their own planning ability, participants in the efficiency group should attribute more intelligence to efficiency in proportion to their own planning ability. The experiment described in Chapter 6 tests this hypothesis and compares the individuals' planning to their attributions of intelligence to others.

Chapter 5

An Empirical and a Model-Based Study of Planning Under Uncertainty

Misce stultitiam conciliis brevem

Horace

Might the variability in individual planning explain the variability in attributing intelligence? The efficiency hypothesis implicitly assumes that participants are able to generate the optimal trajectory given simple mazes. But how likely are participants to plan efficiently? On one hand participants' ability at attributing intelligence to efficiency might critically depend on their planning skills. If so, participants who attribute intelligence to outcome might lack the ability to produce an efficient solution themselves and fall back on outcome as a heuristic. On the other hand, participants could use themselves as a reference for predicting how others plan. If so, participants who plan sub-optimaly should evaluate other sub-optimal planners highly, not just the lucky ones. The experiment described in this chapter measures human planning using a maze planning task, similar to the agent's task in Chapter 4, with a difference that participants control the agent searching the maze. The results inform an empirical model of human planning and illustrate variability in participants' ability to approximate optimal planning.

The POMDP-based framework described in Chapter 3 formalises rational planning and can be used to measure planing efficiency (e.g. by inferring the level of decision noise that can explain a particular trajectory). However, in a general case, solving a planning procedure by a POMDP fully is computationally intractable [62]. In practice people are likely to use heuristics to balance computation cost against expected utility [74, 44]. Most formal models of searching in natural environments deal with this complexity by interpreting search as step-by-step information sampling [78, 77, 85, 86], in effect planning one step in advance. For example, models of searching hypothesis spaces in active learning [78] and Bayesian Ideal Observer models [85] reward the agent for maximising the information gained (IG) one step at a time. In addition, novelty preference models, common in robotics, directly give the agent small rewards for taking novel actions [15, 11, 26]. At the same time, it appears counter-factual reasoning that seems to accompany natural behaviour is not represented by IG-based models, making them an unlikely candidate for explaining how humans actually search.

Previous work on exploratory search shows that mental and spatial search engages a shared cognitive mechanism: for example priming strategies of spatial foraging affects how humans subsequently search for words in memory [55]. This suggests that there may be a shared cognitive resource that handles planning under uncertainty as well, so as to enable knowledge transfer across domains. However, so far there has been little work on decision-making models that could explain human planning under uncertainty and make falsifiable quantitative predictions of behaviour.

An empirical model of human planning should be informed by quantitative empirical evidence. In general terms, such a model should predict the observed *distribution of decisions*, explain the *variability between individuals* and the *relative decision difficulty* reflected by decision times of human participants. The next Section outlines the experimental methodology, using which we empirically calculate the three metrics. The result of fitting the POMDP-based framework to human trajectories shows that participants rarely plan optimally, however the probabilities of taking each action are proportional to model-based estimates of action value. Model-free RT distribution analysis shows that participants subdivide planning problem into sub-problems with nodes at observations and plan more than one observation in advance.

5.1 Method

How might a rational agent form a plan in a maze planning problem? Optimal planning algorithms in machine learning compute a policy that prescribes a decision for each state before moving. Once the policy is computed, the algorithm simply rolls out the policy, matching pre-computed behaviour to the states it encounters [110]. If participants employ a *pre-computed policy*, they spend a long time thinking during the starting state, with only small differences in reaction times through the rest of the trial, caused by searching for the goal at observation locations.

In contrast, current formal models of search under uncertainty assume decisions to be made step-by-step, which in the context of the maze task means planning one step at a time in every cell, or planning each time new information is gained (one observation at a time). If participants make decisions *one step at a time* then the time spent in a cell should depend only on the number of exits leading out of the cell and on visual search. If participants consider decisions *one observation at a time*, then the time spent in a cell should depend only on whether this cell reveals new observations, and the number of exits, but not on subsequent observations.

Moreover, participants might balance the computational costs associated with planning by *planning several observations ahead* and by *evaluating a subset of actions* in each state. If participants evaluate all available options, then the mental effort measured as decision time should increase with the number of exits. If participants decide between at most two options, then the mental effort should not depend on the number of exits from a cell. There also may be individual differences between participants, with some planning further in advance or subdividing the space differently.

Each planning decision, such as deciding which way to go in a maze, requires a way of estimating the values of the available actions. For example, a participant might think: 'I should go North then South because the northern room is bigger'. Participants might estimate the values of actions by visually scanning the maze for a route that lets them see the most space quickly, or engage in mental arithmetic balancing steps against cells revealed. The hidden value estimates made by participants can be measured indirectly through empirical probability of taking an action, since we assume that participants prefer more rewarding actions. And when modelling the planning procedure formally, an accurate planning model should produce objective value estimates proportional to the *empirical action values* of human participants.

In addition, the relationship between empirical action value and decision time might be informative about the nature of computations driving decisions. According to Drift-Diffusion Models (DDM) of decision-making, actions with higher value should be chosen faster [129, 29]. So, assuming that action likelihood measures the hidden action value estimate, highly likely actions should elicit faster RT. Moreover, according to DDM the choice is easy when the difference between action values is large, since a larger difference in values translates to a steeper 'drift' [129, 29]. So, RT should be proportional to the differences between the empirical probabilities of actions in a state.

Next, we turn to collecting empirical data: trajectories of human participants, empirical likelihoods of actions in each state and reaction times during the task. Using empirical evidence allows us to test and falsify the above assumptions.

5.1.1 Procedure

Participants are instructed to find a hidden goal location ('exit') in a series of mazes, by controlling an agent inside a maze. Each maze consists of a grid of cells, where each cell is either a space or a wall. The agent can move one grid cell at a time: N, W, S, E and has a view of the maze limited by walls. Each of the non-wall cells can contain the exit. Spaces that are obstructed by walls are not visible to the agent. Participants must navigate the maze until the exit is found, which may be not until all spaces are seen. Upon reaching the exit the trial automatically terminates.

All walls are visible at the start of the trial, so participants initially know the layout of the rooms, but not the location of the goal, which is marked as a bright red circle once visible. Participants are instructed that each of the dark cells is **equally likely** to hide the 'exit', and that they should find it in as few steps as possible. The full set of instructions is listed in the Appendix.

After reading the instructions, participants complete three practice mazes and answer instruction-comprehension questions. Participants who fail to answer the instruction-comprehension questions correctly proceed with the experiment, but their responses are discarded. The participants' reaction time (RT) (the time to move from one cell to another) and trajectory is recorded on each trial. At the end of the experiment participants are asked how they made their decisions.

5.1.2 Participants

A total of 120 participants were recruited via Amazon Mechanical Turk, restricted to US residents, 46 female and 74 male, mean age 33, SD=10.13. Half of the participants did the experiment in the Bonus condition, in which the top 20% of participants received a bonus for finishing with the lowest total step cost. The other half in the No Bonus condition received no bonus. Bonus was intended to motivate attention, as is common practice with online participants [79].

The exclusion procedure involved an instruction quiz and an automated check for multiple answers coming from the same IP address. Ten participants were excluded for failing to answer quiz questions and five more because they were generated by one person with multiple MTurk accounts [1]. Experiments received ethics clearance from a University of Waterloo Research Ethics Committee and from an MIT Ethics Review Board. The full experimental procedure and the set of stimuli used in this and subsequent experiments can be downloaded from `http://www.cgl.uwaterloo.ca/mkryven/solving/int.php`

5.1.3 Stimuli

The test stimuli are the 12 mazes, shown in Figure 5.6.

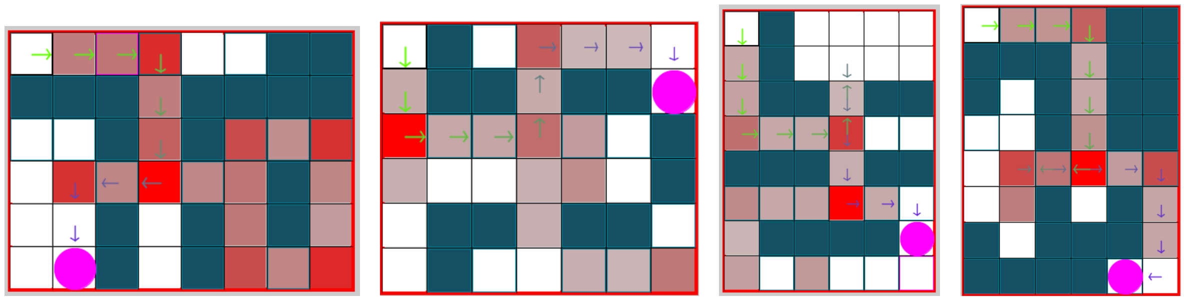


Figure 5.1: Examples of mean reaction time (RT) heat-maps in each of the visited maze locations. Note that the heat-maps are approximate, since the median RT is calculated independently of whether the cell is visited for the first time or revisited, however RT are shorter upon revisits. Longer reaction times are mapped to a more saturated red hue.

5.2 Results

5.2.1 Model-Free Analysis

Planning Depth and Planning Breadth

RT shorter than 200ms were removed, since they are likely accidental [75], removing 0.25% of responses. Another 0.25% of responses longer than 10 seconds are also removed. We assume that RT is a sum of a normally distributed non-decision time, such as time to perform a click, and decision time. The decision time varies with decision difficulty, while the non-decision time depends only on participant due to factors such as age, gender other participant-specific effects [37]. Informally, median RT appears to be longer in certain maze locations, as shown on Figure 5.1.

If the participant effect on non-decision time is significant, then the data needs to be normalised to reduce the participant effect. To test for participant effects on non-decision time we select a subset of RT in cells in which no observations were revealed,

[1] It is possible that two different MTurk workers use the same IP address. However, the person also gave the same answers to demographics each time he repeated the experiment and took fewer steps on each re-run.

rt_{NoObs}. From this subset we recursively remove outliers more than two SD from the mean [119] to maximise the accuracy of ANOVA [141]. ANOVA of rt_{NoObs} against (*bonus*, $agegroup, gender$, *participant*) shows significant effects of all factors, where age group is coded by dividing participants into two equal groups (younger and older) by a median split. The effect of bonus is $F(1, 6781) = 56.718, p < .0001$, of age $F(1, 6781) = 241.468, p < .0001$, of gender $F = (1, 6781) = 105.079, p < .0001$ and of participant $F(85, 6781) = 46.186, p < .0001$. According to Tukey HSD the difference between the No Bonus and Bonus condition is $(d = 11ms., p < .0001)$, between older and younger participants $(d = 22ms., p < .0001)$ and between females and males $(d = 15ms., p < .0001)$. Mean of the median participant's RT in no-observations states is 397ms. SD=59ms.

Assuming that the participant effect is linear and additive, it can be reduced by subtracting each participant's median RT in no-observation cells from their data. Repeating the same ANOVA on the normalised data shows significant, but smaller, effects of gender $(F(1, 6626) = 15.8064, p < .0001)$, age $(F(1, 6626) = 5.9892, p = .01)$ and participant $(F(85, 6626) = 4.2449, p < .0001)$ and no effects of bonus $(p = .9)$. After normalisation mean of the median participant's RT in no-observations states is 0ms, SD=18ms. Although normalisation significantly reduces the participant effect, the actual effect turned out to be non-linear, so residual participant effects remain in the normalised data.

Decision times might depend on observations, the visible area of the maze and the agent's location in the maze. To test for differences in mental processing between maze locations, the steps in each path were coded based on the function of the location in the path, as shown in Figure 5.2. *Observation states* are states labelled 'D', 'O' and 'G'. An 'O' state is associated with decision-free observations, such as looking around a corner. In a 'G' state the goal first becomes visible and all that remains is to approach it. A 'D' state requires a decision regarding which way to go (e.g. an intersection). An 'X' state is the starting state, when the participants first sees the maze. The differences between RT distributions in each state reveals different in demands on mental processing arising during the task.

Since the RT distributions are clearly not Gaussian (see Figure 5.2), t-test validity is limited. Neither is the RT distribution log-normal, according to Kolmogorov-Smirnov Test $(D = 0.71567, p < .0001)$. Median RTs in each state are as following: In 'N' states $Mdn =$

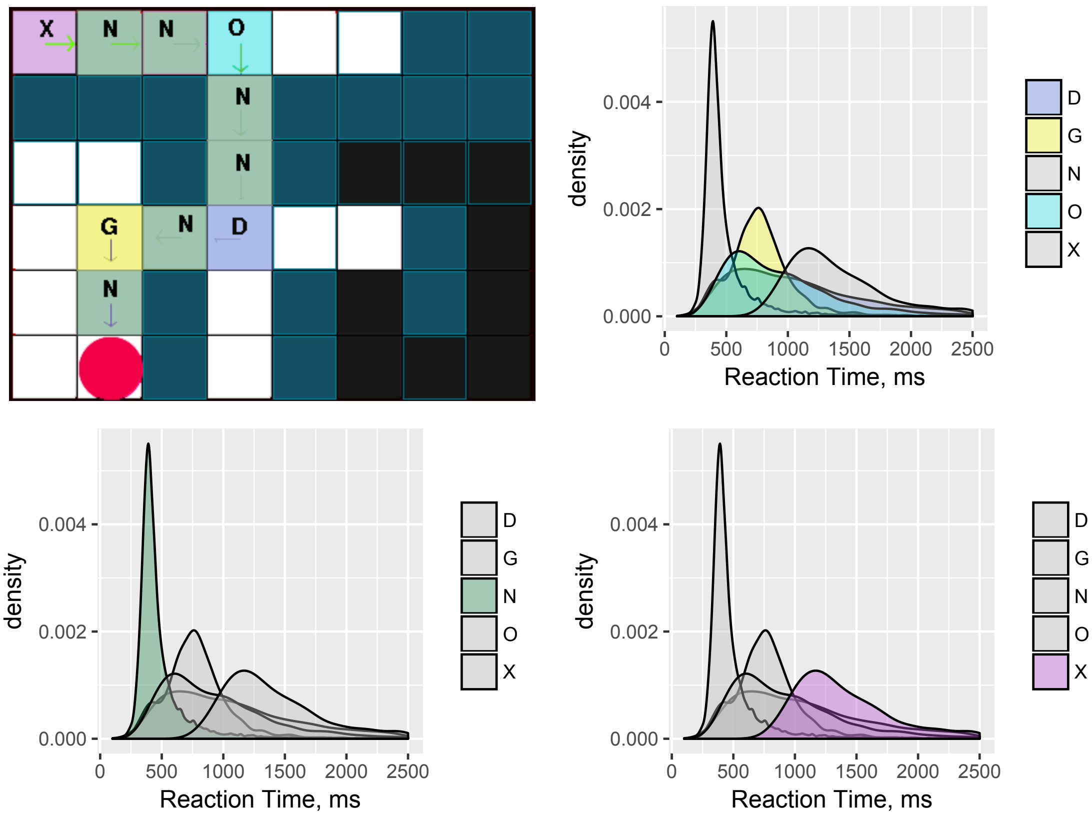


Figure 5.2: Left: Maze location codes: **X** - the starting state , **N** - non-observation states, **O** - observation, no planing required, **D** - decision state, **G** - the goal is observed. Right: distribution density of RT mean-subtracted by participant in each type of cells.

427ms, in 'G' states Mdn = 765ms, in 'O' states Mdn = 850ms, in 'D' states Mdn = 1036ms, and in X states Mdn = 1426ms The significance of the differences between RT distributions between states is demonstrated by MannWhitney tests, which do not assume normality. Four pairwise Mann-Whitney tests indicate the following differences in RT: in 'G' states RT was greater compared to 'N' states ($d = 308ms, W = 1500600, p < .0001$); in 'O' states RT was greater compared to 'G' states ($d = 95ms, W = 946320, p < .0001$); In 'D' states RT was greater than in 'O' states ($d = 161ms, W = 1574000, p < .0001$); In 'X' states RT was greater than in 'D' states ($d = 450ms, W = 903000, p < .0001$). This result rejects the hypothesis that participants plan their actions *one step at a time*.

To calculate the cost of visual search for the goal, we select RT in observation states ('O' and 'D') and remove outliers greater than 2 SD away from the mean. For the remaining data-set ($Mdn = 868ms, sd = 389$) a significant regression equation was found, $F(1, 3561) = 9.046, p = .003$ with an adjusted $R^2 = .002$. Predicted RT is equal to $871ms + 12.6ms \times n$, where n is the number of revealed cells. Thus, the difference between RT in 'O' and 'N' states is too large to be explained by visual search for the target, and thus must be due to planning the next steps.

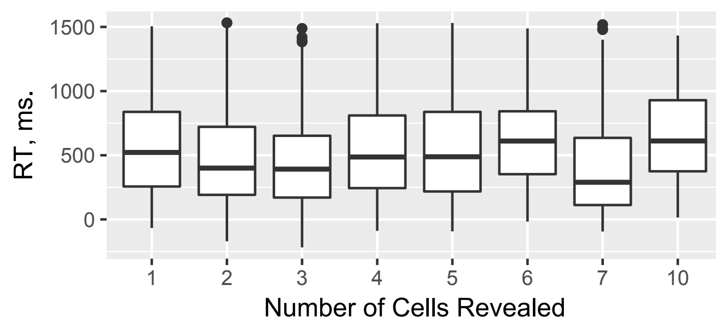


Figure 5.3: Mean RT by number of revealed cells.

In addition, we expect visual search in 'G' states to be easier than in 'O' states. In 'O' states participants process the entire revealed space because the exit is not seen. In 'G' states search terminates early once the exit is noticed. The observed difference between RT in 'O' and 'G' states is too large to arise from differences in visual processing, and so must reflect the differences in planning costs. Thus, *pre-computed policy* hypothesis is rejected, since the differences between RT in observation and non-observations states are too large to be explained by visual search for the goal.

In summary, RT analysis shows that participants make some of the decisions as they go along and mostly plan at observation locations, treated as decision nodes. While it is evident that participants plan their path from one observation to the next, they might plan more than one observation in advance. Participants may also prune unpromising decision options and evaluate only the a subset of alternative trajectories.

A 'D' state can have at most four exits (N,E,W,S), of which participants may evaluate

Figure 5.4: Number of available exits from 'D' states that are used.

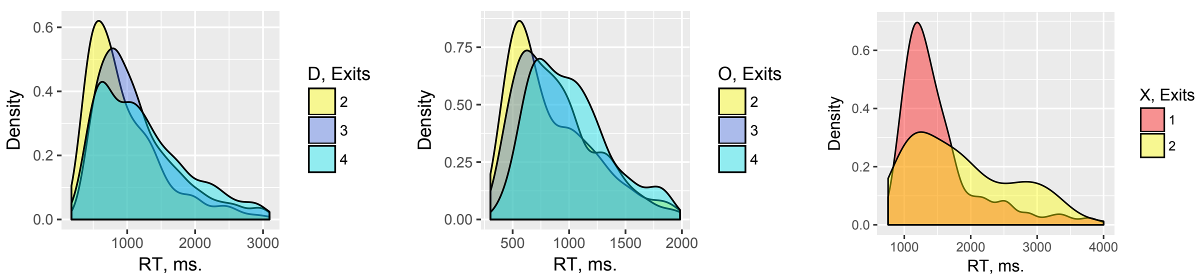


Figure 5.5: RT tends to increases with the number of exits.

only a subset. Participants might restrict their choices to the two most promising actions or use only exits that do not go backwards. For example, Figure 5.4 shows that participants rarely use all of the exits available in a cell, possibly pruning unpromising alternatives. But even if certain exits from a state are never used in practice, participants might be still considering every option, as may be evident from analysing their RT.

To test whether RT depends on the number of exits, observation and non-observation states are analysed separately. First, we consider 'D' states after removing outliers greater than 2 SD away from the mean ($Mdn = 1001ms, SD = 606$) linear regression of RT against exits is significant ($F(1, 2349) = 46.89, p < .0001$) with an adjusted $R^2 = .019$. Predicted RT is equal to $733ms + 127ms \times e$, where e is the number of exits. For 'O' states after removing outliers greater than 2 SD away from the mean ($Mdn = 818ms, SD = 371$)

linear regression of RT against exits is significant, ($F(1, 1493) = 42.07, p < .0001$) with an adjusted $R^2 = .0268$. Predicted RT is equal to $687ms + 76ms \times e$. For 'X' states after removing outliers greater than 2 SD away from the mean ($Mdn = 1358ms, SD = 651$) linear regression of RT against exits is significant, ($F(1, 473) = 18.15, p < .0001$) with an adjusted $R^2 = .0349$. Predicted RT is equal to $1187ms + 336ms \times e$. Distribution densities of the data used in the above analysis are shown on Figure 5.5. In contrast, in non-observation states, excluding dead-ends (states with only one exit) and outliers greater than 2 SD away from the mean, linear regression of RT against exits is not significant ($p = .14$).

In summary, in observation states RT increases with the number of exits, suggesting that on average participants *evaluate all actions* available in a state. In contrast, no such relationship is evident in non-observation states, and so non-observation states are rarely used for planning. Most real-life planning problems involve more than four possible actions in a state, in which case it is unlikely that people evaluate all actions; the result described above is specific to our planning problem and more research in needed to understand how planning breadth scales in the general case.

Moreover, since the RT distributions in observation and non-observation states overlap, it is likely that observation states are also used to roll-out pre-planned actions, suggesting that participants plan *more than one observation in advance*. In addition, there may be differences between individuals not apparent in the averaged data.

One Observation Model

To examine the hypothesis that participants plan one observation in advance, we define a heuristic that considers costs and rewards one observation node at a time. Assume decision state S has at most four exits: N,W,S,E. Exit i, $i \in N, W, S, E$ reveals n_i squares after s_i steps. The One Observation Model (OOM) chooses the direction with the highest value $n_i - ks_i$, where k is an unknown cost coefficient. Choosing $2 < k < 3$ fits 75% of the individual decisions made by participants. Seven out of 12 solutions produced by the fitted OOM agree with the majority solutions shown on figure 5.6. However, the remaining 5 OOM solutions shown on Figure 5.7 capture neither the majority solution, nor the most

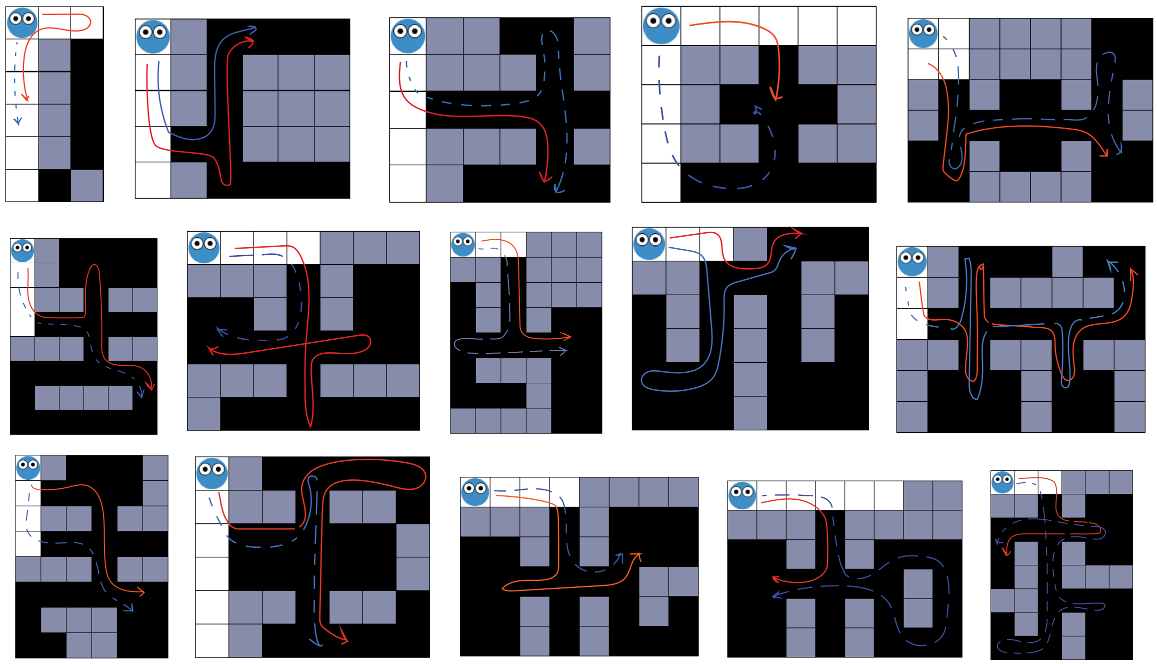


Figure 5.6: Examples of paths taken by participants. The solid line shows the average path. The dashed arrow shows the second most common alternative. The first three images are practice mazes.

common alternative. Thus, OOM can not explain participant's trajectories, which rejects the hypothesis of *planning one observation at a time.* [2]

Observation Decision-Tree Model

Participants conceptualise mazes in terms of observations rather than steps and plan their actions several observations in advance. When first seeing the maze participants might identify the observation states, approach the first observation point, at which point they must reason about uncertainty. Participants might either complete the entire plan while

[2] An alternative one-step heuristic with action value equal to n_i/ks_i predicts that the chosen actions are better than the best alternative, $n_c/s_c > n_a/s_a$. Since cost must be positive, $k > 0$. However, only 62% of decisions made by participants satisfy this constraint.

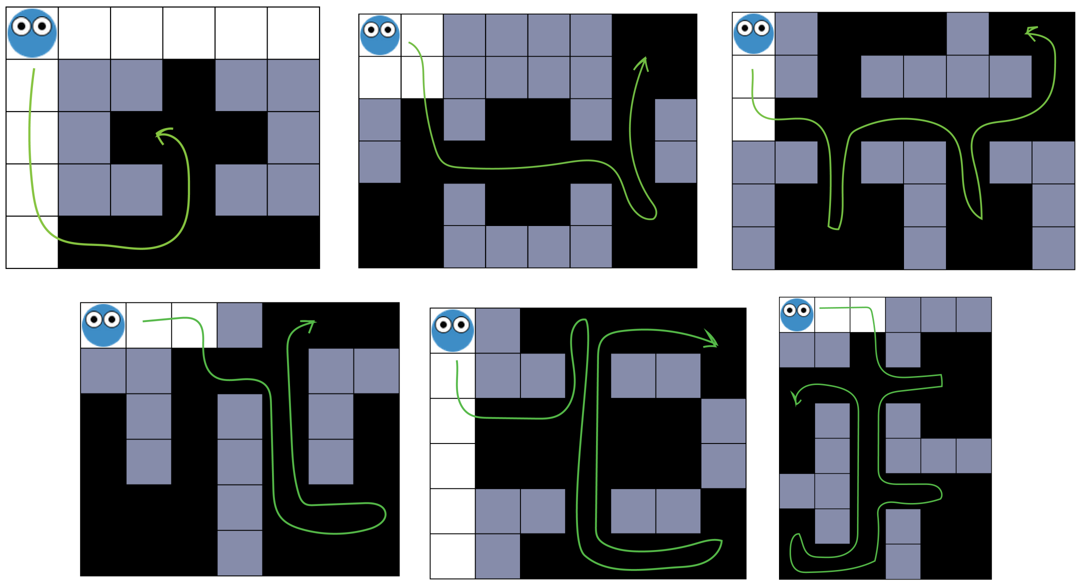


Figure 5.7: Solutions of the OOM heuristic that are different from the POMDP-based solution.

at the first observation state, plan a limited number observations ahead, or revise their plan at each observation.

Participants are likely to think of each observation as associated with a cost, how far away is it, and a reward, how many cells does it reveal. Such reasoning can be formalised by a decision tree with decision-nodes at observation states. Each observation state is associated with a cost in steps s_i and a number of revealed cells n_i as a reward, where i indexes observation states. The full decision tree of a maze includes all possible trajectories such that eventually lead to revealing the entire space. Each node is associated with a reward, which is a function of the path length from the root until the node and the revealed cells. The optimal trajectory goes from the root of the full decision tree to the leaf with the highest reward.

If all planning is completed at the first observation state, then RT at the first observation state should be long, while in subsequent states RT should be not different from 'G' states. In contrast, if participants do some planning at each observation, then RT in 'D' states may decrease with each subsequent observation, but should be always longer than in 'G'

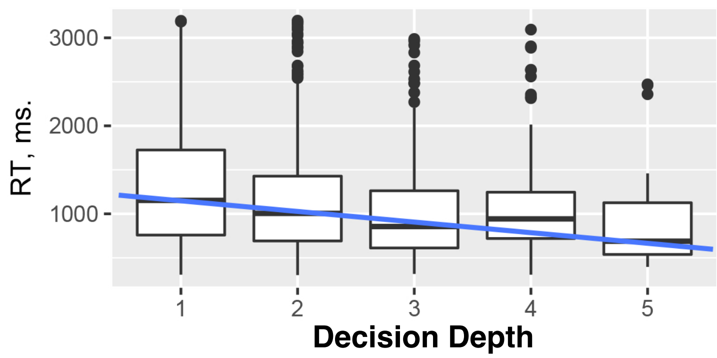


Figure 5.8: Mean RT by decision number.

states.

To test whether most planning is completed at the first observation state we label each decision with its position in the path. The first decision state in a path is labelled "1", the second decision state is labelled "2" and so on. The mean RT by decision number after removing outliers 2 SD away from the median are shown on Figure 5.8. A significant linear regression of RT against decision number was found, $F(1, 920) = 62.87, p < .0001$ with adjusted $R^2 = .0312$. Predicted RT is equal to $1384ms - 108ms \times d$, where d is decision number. According to Mann-Whitney tests, there is a significant difference between RT in 'D' state at the fourth decision and 'G' states ($d = 183ms, W = 55223, p < .0001$), but not at the fifth decision ($p = .5$).

In summary, planning effort increases with the depth of the planning sub-tree that remains to be traversed. With each subsequent observation the difficulty of the problem decreases, requiring lass mental effort. However, mental effort spent in D states is nearly always greater than in 'G' states, suggesting that participants update their plan every time new evidence is received. Possibly, updating leads to a greater confidence over the planning policy, or participants might plan only up to a limited observation depth and update the plan as one goes along.

One way to implement an observation decision tree is the N-Observation Model (NOM), which applies OOM recursively. NOM builds a maze decision tree and chooses the leaf with the highest value of $\sum_{i=1}^{i=N} n_i - k \sum_{i=1}^{i=N} s_i$, where k is an unknown cost coefficient, n_i is the number of revealed cells at each node and s_i is the number of steps from one node to the next. NOM is equivalent to minimising the length of path that reveals the entire maze and predicts the majority trajectories shown in figure 5.6, however, deviations from the majority trajectory remain to be explained.

The Empirical Action Probability

Assume a state S is visited by N subjects and $\{n_w, n_e, n_n, n_s\}$ are the number of subjects taking each of the actions $\{W, E, N, S\}$ respectively. The variability of choices in each state is described by *empirical action probability*, calculated as:

$$Pr_i = \frac{n_i}{N}, \quad i \in \{E, W, N, S\} \tag{5.1}$$

The *average path* taken by participants in each maze, defined as the sequence of most likely actions in each cell, is shown in Figure 5.6 by a solid red line. The average path is also equivalent to the path taken by the majority of participants. The most common alternative solution is the dashed blue line shown for comparison.

Assuming that empirical action probability measures the hidden action value, we test two predictions of DDM: that highly likely actions elicit faster RT and that RT is proportional to the differences between action values. The differences between action values can be measured as *entropy over action probabilities*:

$$E(S) = -\sum_{I \in \{E,W,N,S\}} Pr(I) log(Pr(I)) \tag{5.2}$$

As shown in Figure 5.9, highly likely actions are indeed taken faster (Pearson $r = -.28, p < .0001$), and high entropy states elicit longer RT (Pearson $r = .42, p < .0001$), in agreements with DDM predictions.

The next section discusses model-based analysis of the individual trajectories and the average solution in the light of the formal POMDP framework described in Chapter 3.

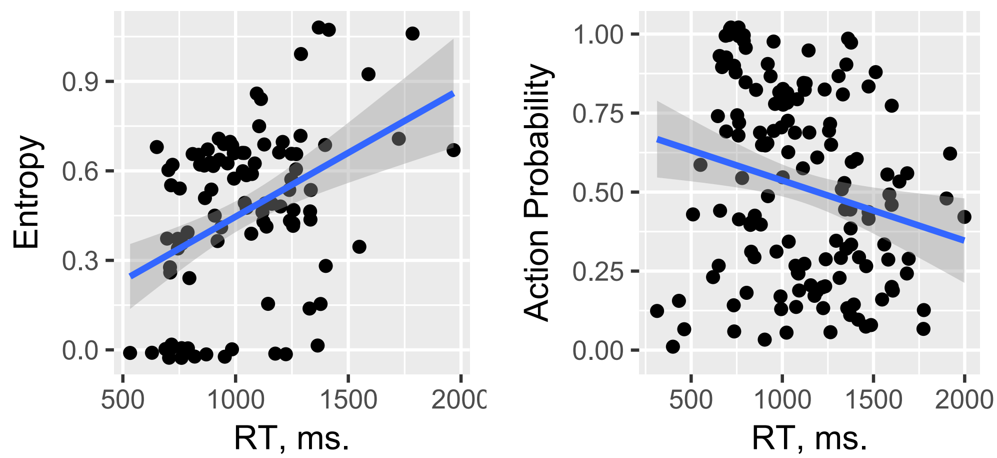


Figure 5.9: Higher entropy states elicit longer RT. RT is faster for highly likely actions. Bands indicate 95% confidence intervals of the linear model.

5.2.2 POMDP-Based Inference

By fitting the POMDP-based model as described in Chapter 3, each solution is assigned model-based planning accuracy metrics: decision noise, *SOFTMAX* $\tau_{p,j}$, and fraction of optimal steps $optstep_{p,j}$. Here p refers to participant and j is the trial. A solution (p, j) labelled as *optimal* if $\tau_{p,j} = 0$, *softmax* if $\tau_{p,j} > 0$, and *softmax0* if $\tau_{p,j} > 0$ and the trial included zero-valued actions. Importantly, participants who produce POMDP-optimal solutions do not necessarily implement an internal POMDP, since there may be other methods that give the same result.

For each participant we calculate model-based metrics of **planning accuracy**: the **fraction of optimal steps** taken by the participant, $optstep_p$, and the mean **decision noise** inferred from the participant's solutions, τ_p. On average individuals took 84% optimal steps (standard deviation 5.9%) and made 75% optimal decisions at decision points ('D' states) (standard deviation 8.6%). There was no evidence of a difference between the fractions of optimal steps $t = -0.6204, p = .5$ of participants in the Bonus and the No Bonus conditions.

Table 5.1: Model-based inference over individual trials.

	optimal	softmax	softmax0
Bonus	45%	48%	7%
No Bonus	45%	39%	16%

The average paths taken by participants, shown in Figure 5.6, are identical to the optimal solution. Individual solutions are optimal on 45% of the trials (Table 5.1). The most common non-optimal solutions taken by participants are shown by the blue dashed lines in Figure 5.6. Non-optimal decisions may minimise cognitive effort, with some participants more inclined to save effort than others. Alternatively, participants may make non-optimal decisions intentionally, because of mistaken assumptions that the experimenter is trying to trick them.

5.2.3 Variability Between Individuals

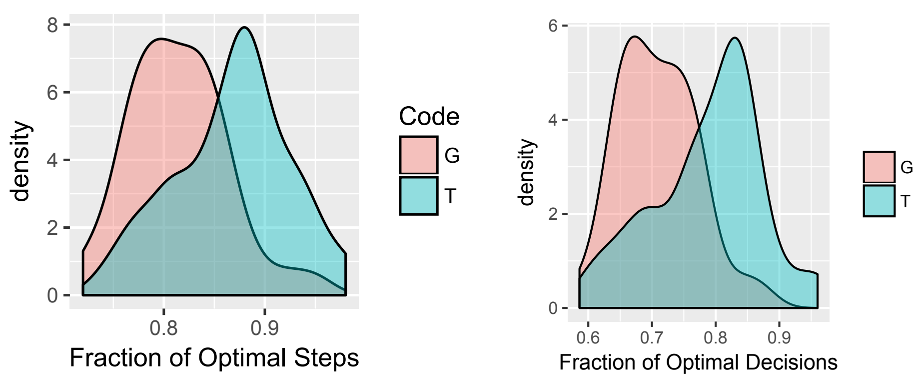


Figure 5.10: Participants self-describing as **thinking** (right) were more optimal than those self-describing as **guessing** (left)

Participants' self-reported reasons for their decisions offer insight into their decision-making. Inspecting the answers, it appears that some participants preferred words referring to intuition, feelings and guessing, suggesting that they generated their solutions by a more intuitive approach. In contrast, other participants self-reported as generating solutions by cognitive effort and deliberate optimisation. Two independent raters coded all responses into two categories: *thinking* and *guessing*. The raters agreed on 90% of participants, coding 45 of them as **thinking** and 37 as **guessing**. For example, a response was coded as **thinking** if it read: *'I tried to maximise the number of squares revealed per step.'* and as **guessing** if it read *'I followed my gut*. The remainder were discarded from the analysis. The results that follow are not affected by the choice of method for handling these participants. According to a Welch two sample t-test, participants who self-described as **thinking** on average made more optimal steps ($M = 0.87, SD = 0.06$) ($t(79.978) = 5.5898, p < .0001$) than those who self-described as **guessing** ($M = 0.81, SD = 0.05$). Likewise, considering only 'D' states, participants who self-described as **thinking** were more optimal ($M = 0.79, SD = 0.084$) ($t(78.668) = 4.8318, p < .0001$) than those who self-described as **guessing** ($M = 0.71, SD = 0.06$) (Figure 5.10).

5.2.4 Guessing

What do participants mean by 'guessing'? One way to interpret 'guessing' is deciding by chance, making decisions indistinguishable from choosing at random. Alternatively, guessing can mean approximating decision value by choosing actions with the likelihood proportional to their reward, as modelled by a softmax rewards (section 3.2.3). It is also possible that some of the participants who self-describe as guessing decide at chance, while others approximate decision value.

To identify participants who decide at chance, each individual's decisions were fitted to the Bernoulli distribution. The Bernoulli distribution describes binary choices, so each decision in a 'D' state was coded as '1' if the person chose the optimal action (computed by the POMDP model described above) and '0' otherwise. If a participant was as likely to make the optimal move as any other move, then the participant is considered deciding at chance. Such a test provides an upper bound estimate of guessing at chance, since most 'D' states have more than two exits. As a result, 15 participants were identified as deciding at chance, 9 of whom were previously coded as 'guessing' and 6 as 'thinking'.

One might expect that participants who guess might well make their decisions by a

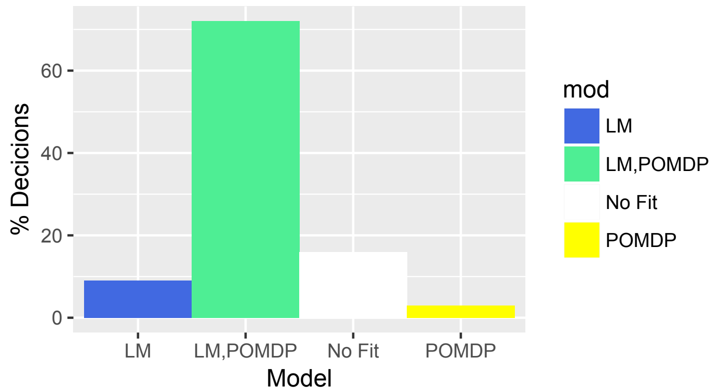


Figure 5.11: Fractions of decisions explained only by optimal POMDP, only by OOM, by both models and by neither model.

heuristic, compared to those who plan. However, participants who self-reported as 'thinking' are better fitted by a OOM model than those who were 'guessing', $t = 4.2777, p < .0001$. On average 'thinking' participants made 82% of OOM-fit decisions, and 'guessing' made 74%. Thus, in the context of maze planning, guessing can be explained neither by deciding at chance, nor by limiting one's planning horizon to one observation at a time.

Comparing the OOM and the optimal POMDP as shown on Figure 5.11 shows that the two models agree in most cases. However, OOM can not explain full trajectories as well as the POMDP model, suggesting that participants plan more than one observation node in advance.

Action Values and Action Likelihoods

Next, we test the assumption that the model-based action values are proportional to empirical action likelihoods. Figure 5.12. [3] shows that in 'D' states model-based action values and action likelihoods are indeed correlated, Pearson $r = .35, p < .0001$.

The difference between the two best actions and the likelihood of choosing the better action are correlated as well (Figure 5.13), Pearson $r = -.23, p = .002$, so participants are

[3]The model-based values depend on γ, here $\gamma = 0.6$.

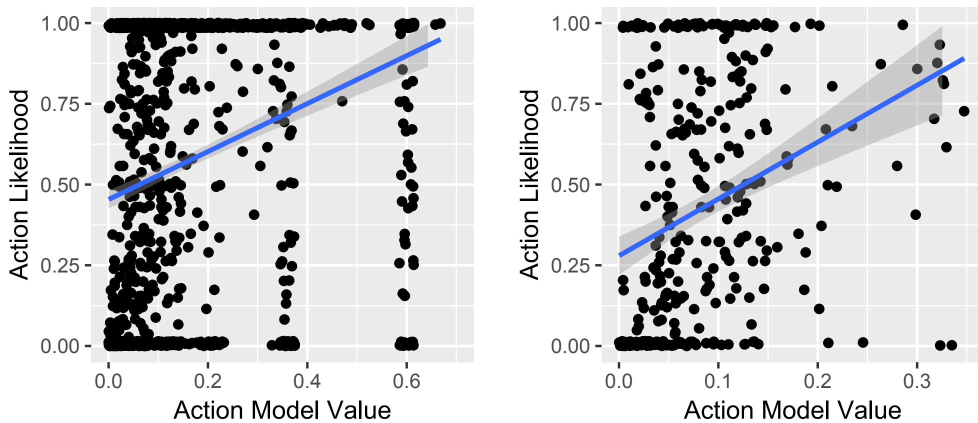


Figure 5.12: The relationship between model-based action values and empirical action likelihoods. Left: all states. Right: only 'D' states.

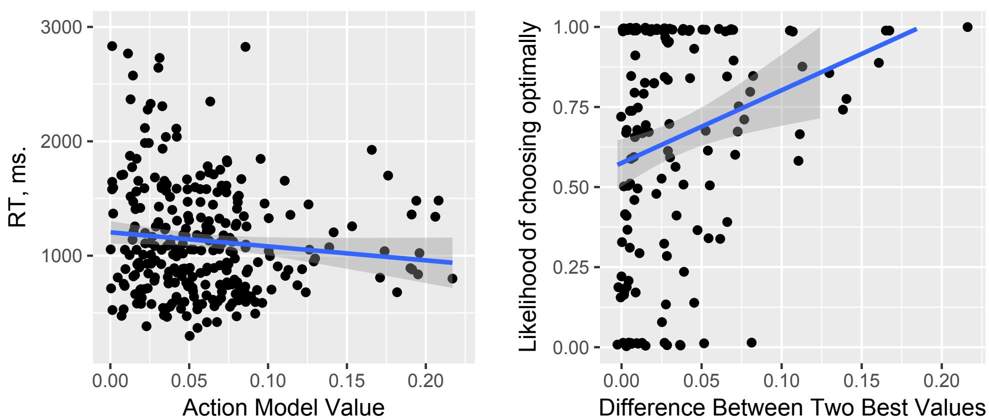


Figure 5.13: Left: action values plotted against empirical RT. Right: the difference between the two highest model-based action values and likelihood of choosing the better action.

more likely to choose optimally when the difference is action values is higher. In addition, actions with a higher value are chosen faster (Figure 5.13, left), Pearson $r = .13, p < .0001$, however, there is no evidence of a relationship between the difference in values and RT, $p = .9$.

The softmax function described in Chapter 3 select actions with likelihood propor-

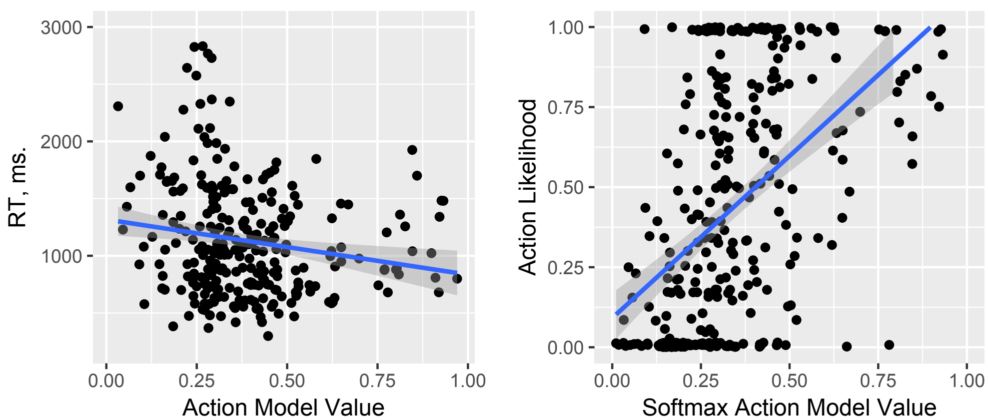


Figure 5.14: Left: Softmax of the model-based action values plotted against RT. Right: Softmax of the model-based action values is highly correlated with action likelihood.

tional to their value by transforming values as probabilities. The parameter τ controls the degree of values normalisation. For example, for $\tau = 0.1$ $softmax(0.1, 0.5, 0.1, 0.0) = (0.018, 0.958, 0.018, 0.006)$, and for $\tau = 0.05$ $softmax(0.1, 0.5, 0.1, 0.0) = (0.0003, 0.9993, 0.0003, 0)$. The softmax model fits the data quite well, with the highest Pearson correlation $r = .83, p < .0001$ for $\tau = 0.01$ (Figure 5.14, right). Actions with a higher softmax value are chosen faster (Figure 5.14, left), Pearson $r = .23, p < .0001$ and lower levels of decision noise are fitted more often than higher levels (Figure 5.15).

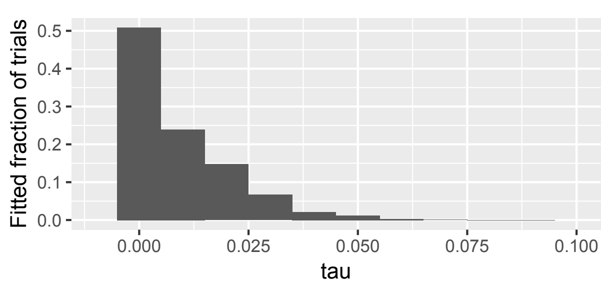


Figure 5.15: Frequency of different levels of τ fitted to individual solutions. Optimal solutions correspond to $\tau = 0$.

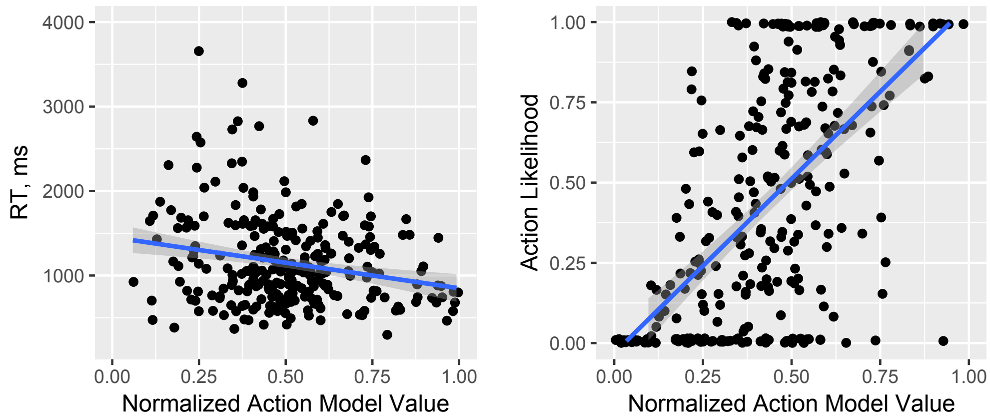


Figure 5.16: Left: Normalised model-based action values plotted against RT. Right: Normalised model-based action values are highly correlated with action likelihood.

A simpler alternative to softmax is to normalise all actions in a cell. Such a normalising transformation may be considered a linear version of softmax:

$$\{v_w^{'} = \frac{v_w}{\sum v_i}, v_e^{'} = \frac{v_e}{\sum v_i}, v_n^{'} = \frac{v_n}{\sum v_i}, v_s^{'} = \frac{v_s}{\sum v_i}\}. \tag{5.3}$$

The normalised values explain the action likelihoods and RT equally well, Pearson $r = .83, p < .0001$ (Figure 5.16, right) and with RT is $r = .23, p < .0001$ (Figure 5.16, left). In conclusion, empirical action likelihoods can be explained by a softmax mapping or by a normalisation of model-based values, so that the likelihood of selecting an action is indeed proportional to the action rewards.

5.3 Discussion

The research described in this chapter gathers empirical evidence to describe an empirical model of how people plan. Fitting the POMDP-based framework to empirical evidence shows that the optimal POMDP-based model rarely predicts individual planning. However, the model is useful for defining model-based metrics of planning efficiency, which measure how closely individuals approximate the optimal plan. According to model-based metrics participants who self-report as thinkers more closely approximate optimal planning than participants who guess. However, self-attributed descriptions are only approximately accurate, since efficiency metrics of the '*thinking* and *guessing* groups overlap. On one hand people may believe their actions to be more rational then they are. At the same time, people may self-describe as 'guessing', simply because they have no conscious awareness of their thought process, and not because their actions are independent of value.

Empirical analysis of reaction times shows that participants divide planning into subproblems tied to observations. A planning problem requires the participant to plan sequences of steps leading from one observation to the next as well as a sequence of decisions that specify the order in which observation cells are visited. On average participants evaluate all of the at most 4 of the available exits from an observation state. The average trajectory produced by participants coincides with the optimal trajectory generated by the optimal POMDP planner. However, this result does not mean that given arbitrary mazes the average path taken by participants will always be optimal - only that given simple mazes such as these, the average path that humans take requires reasoning several observations in advance. Based on the current results, accurately quantifying the limits of planning breadth and depth will require more complex mazes.

Whenever trajectories generated by participants are not optimal, such trajectories can be explained by a softmax mapping of model-based action values to empirical action probabilities. Assuming that empirical action probabilities reflect the average hidden value estimates made by participants, human value estimates approximate action values generated by the POMDP framework. In addition, human choices are also likely to be affected by decision difficulty, reflected by the depth and breadth of the planning planning problem. The relationship between the strength of mapping of action values to probabilities and the difficulty of the planing problem remains to be explored.

Participants could also implicitly interpret the planning task in a social context, thinking of the experimenter as an adversary who is hiding the exit. If so, participants might think that the adversary would place the exit in the last cell an optimal agent would look. To test if participants expect to be tricked, we must reason about 5 types of planners:

- L0: Non-recursive optimal agents with uniform initial beliefs.
- L0: Non-recursive softmax agents with uniform initial beliefs.
- L1: Recursive optimal agents with one level of recursion. An L1 agents assume that the exit is hidden in the last place that an L0 agent will look.
- L2: Recursive optimal agents with two levels of recursion. An L2 agent believes that the exit is hidden in the last place L1 or L0 will look.
- L3: Recursive optimal agent with two levels of recursion. An L3 agent believes that the goal is hidden in the last place an L0, L1 or L2 will look.

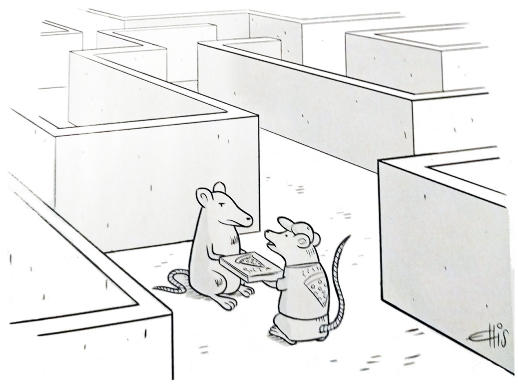
Recursion is costly, so for participants who use recursive reasoning the number of trials fitted by L1, n_{L1} should be greater than the number of trials fitted by L2, n_{L1} and $n_{L2} > n_{L3}$
.

The choice and reaction time data collected in this experiment can be also interpreted as a Drift-Diffusion Model (DDM) [129, 94]. A DDM interprets decision-making as a process of accumulating evidence for each choice over time, until enough evidence for one option is reached. When applied to value-based decision-making, such models assume that the decision is the result of integrating value computations [129]. DDMs predict that the choices are sensitive to relative differences in values, but not to absolute value.

DDMs take longer to decide between options relatively close in value, which is indeed observed in our results. RT and the difference in model-based values is anticorrelated, Pearson $r = -.2, p < .0001$. A second prediction of DDMs is that the likelihood of choosing the lesser option is higher when choosing between options close in value, which is evident in Figure 5.13. Fitting a DDM entails fitting two parameters to each choice: the evidence accumulation rate, μ, and the diffusion coefficient, σ. The parameters probably differ between individuals and between thinking and guessing groups.

The results described in this chapter provide a set of human solutions, which can be used to study attributed intelligence. The construct validity of the intelligence attribution task would be stronger if participants rated actual humans rather than model-generated behaviours. At the same time, the described maze-planning task can be used to measure individual planning efficiency as a proxy for individual planning ability. The next chapter combines the planning task and the intelligence attribution task described in Chapter 4 to look for an effect of planning ability on attributing intelligence to others.

Chapter 6



"Sorry I'm late. It took me forever to find this place."

Planning Ability Predicts Attributed Intelligence

Chapter 4 shows that observers attribute intelligence by evaluating an agent's efficiency or their outcome. At the same time, Chapter 5 shows that participants describe their planning as 'thinking' or 'guessing' in a way that agrees with model-based metrics of planning accuracy and sensitivity to decision value. These results suggest that observers' attributions of intelligence might critically depend on their planning ability. Those who find planning easy may easily evaluate the efficiency of the viewed trajectories and attribute intelligence to efficiency. Conversely, less skilled participants may simply rely on agents' outcomes.

This chapter compares the results of the intelligence attribution task and the planning task within participants. The planning task elicits metrics of **planning accuracy** based on self-report and model-based inference. The intelligence attribution task elicits metrics of **sensitivity to efficiency**, measured as the difference between the participants' ratings of optimal and suboptimal trials and of **sensitivity to outcome**, measured as a difference between the ratings of lucky and unlucky trials, as well as verbal-self reports. If attributed intelligence depends on planning ability, then the variation in sensitivity to efficiency should be explained by planning accuracy. However, if attributed intelligence is independent of planning, then metrics of planning accuracy and sensitivity to efficiency should be independent of each other.

The order in which participants complete the planning and the attribution tasks is significant. If participants complete the attribution task before the planning task, they may be biased by the exit locations they encounter, and repeat some of the trajectories they see. Viewing sub-optimal examples may also bias observers to guess. At the same time, engaging in planning can improve participants' planning ability, and bias their intelligence attribution toward efficiency. In the experiment described in this chapter all participants completed the planning task followed by the attribution task. This design has a potential weakness, in the difficulty of reaching a strong conclusion regarding the effect of planning practice on intelligence attribution. However, it has the advantage of excluding retrospective effects of prior familiarity with the agent's task on planning.

6.1 Participants

Participants were recruited via Amazon Mechanical Turk restricted to US residents, out of 110 participants 104 completed the experiment. An attention check included an instruction comprehension quiz, an automated check for multiple answers coming from the same IP address, and the number of zero-valued steps. 10 participants were excluded for failing the attention check: three did not answer the instruction comprehension question, and the others made zero-valued steps on half or more of the trials. Of the remaining 94 participants 41 were female and 53 male, mean age was 34, SD=9.7.

6.2 Method

Participants completed two tasks, always in the same order. Participants first completed a planning task, in which they searched for a goal in a series of mazes, using the procedure described in Chapter 5. Participants next completed an intelligence attribution task, in which they watched and rated replayed solutions of other participants, who have previously completed the planning task. The intelligence attribution task follows the same procedure as used in Chapter 4. The stimuli were presented in a web-browser using a Java Script interface developed in our lab, and hosted on a University of Waterloo web-server.

6.3 Planning Task

Participants look for a hidden goal location ('exit', marked when visible as a bright red circle) in a series of mazes, by controlling an agent using a mouse. The agent moves one grid square at a time: N, W, S or E and has a 180 degree view of the maze limited by walls. The maze is initially dark, but is uncovered as the agent moves along, so participants initially know the layout of the rooms and location of the barrier walls, but not where the goal is. Participants are instructed that each of the dark squares is equally likely to hide the 'exit', and that they should find it in as few steps as possible. All participants received a performance-based bonus of up to $1 for finishing all mazes while minimising step cost. To receive the maximal bonus participants need to achieve a step cost within 5% of the optimal solution. The bottom 10% of participants receive a zero bonus.

6.3.1 Stimuli

After reading the instructions, participants complete three practice mazes and answer an instruction comprehension quiz. Participants who answer the quiz incorrectly proceed with the experiment, but their responses are discarded. The test stimuli are 12 more mazes, which are shown in Figure 6.1. Black cells indicate unseen areas. Empty cells that were previously revealed are shown white. The location of the walls is always visible, shown as grey. The agent can move through empty areas, but not through walls. The starting location is always assigned to the top-left corner and the exit is hidden behind one of the black squares. At the end of the experiment participants are asked how they made their decisions. The reaction times (times to make a move) and path are recorded for each trial. Instructions are listed in the Appendix. The full experimental procedure is available at http://cgl.uwaterloo.ca/mkryven/attribute/int_exp.php.

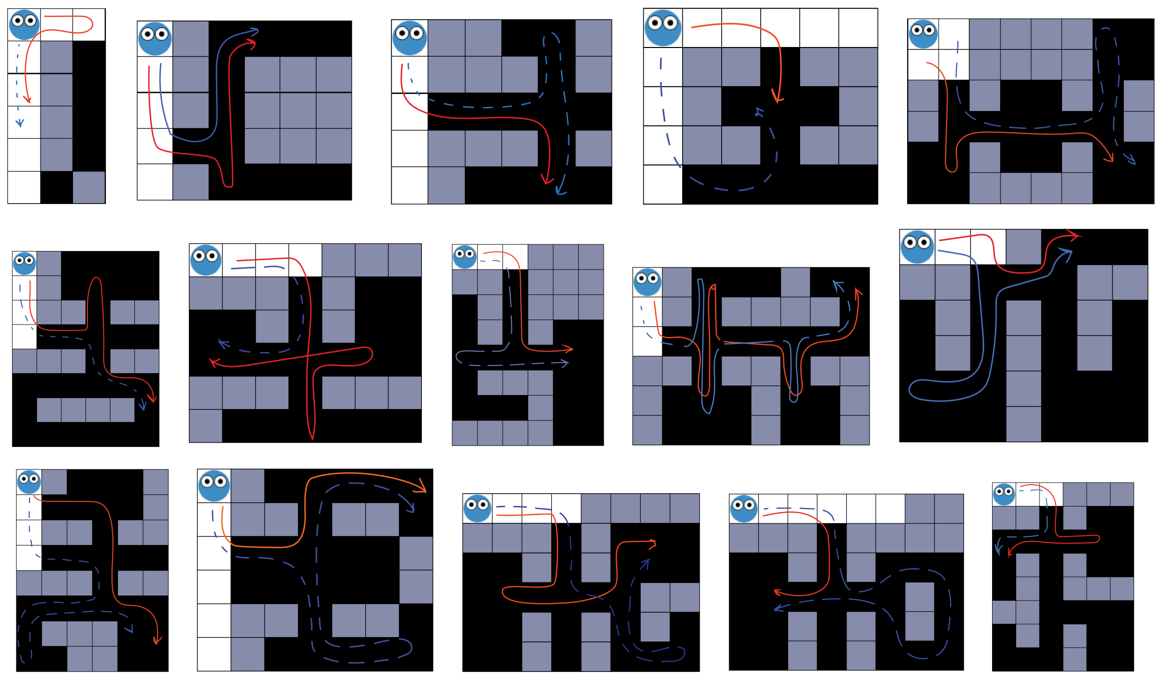


Figure 6.1: Paths taken by participants. The solid line shows the optimal path, which is also the most common solution. The dashed arrow shows the second most common path, which is non-optimal. The first three images are practice mazes.

6.3.2 Results

Each path is analysed using the *SOFTMAX* model described in Chapter 3 and assigned model-based planning accuracy metrics: decision noise, *SOFTMAX* $\tau_{p,j}$, and fraction of optimal steps $optstep_{p,j}$. Here p refers to participant and j is the trial. A solution (p, j) labelled as *optimal* if $\tau_{p,j} = 0$ (51% of the trials), *softmax*, if $\tau_{p,j} > 0$ (46% of the trials) and *softmax0* if $\tau_{p,j} > 0$ and the trial included zero-valued actions (3% of the trials). All trajectories were consistent with $\tau \in [0, 0.1]$, as shown in Figure 5.15, with lower levels of decision noise fitted more frequently than higher levels. The majority path is identical to the optimal solution, shown in Figure 6.2.

Two independent raters coded participant's responses as **thinking** or **guessing**. For example, a response is categorised as **thinking** if the participant said: *'I tried to maximise*

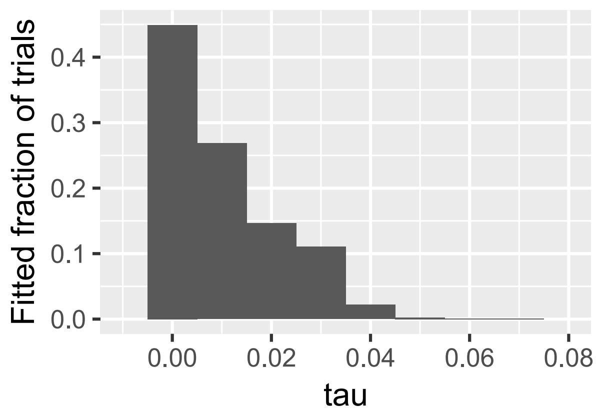


Figure 6.2: Frequency of τ fitted to individual trajectories. Optimal trajectories correspond to $\tau = 0$, lower levels are fitted more often than higher levels.

the number of squares revealed per step.' and as **guessing** if they said *'I followed my gut'*. The raters agreed on 92% of participants (86 out of 94), coding 52 of them as **thinking** and 34 as **guessing**. The remaining 8 were excluded from the subsequent analysis.

Each participant is assigned three planning accuracy metrics: their verbal-self report code, the fraction of optimal steps they took ($optstep_p$), and the mean decision noise ($SOFTMAX\ \ \tau_p$), fitted to the individual's solutions. Here p indexes the participant. The mean fraction of optimal steps is 86.3% ($SD = 4.8\%$). According to Welch Two Sample t-test participants who self-reported as **thinking** were on average closer to optimal ($t(81.405) = 2.8292, p = .006$), taking optimal steps 87% of the time. The **guessing** group took optimal steps 84% of the time. So, the two ways in which participants self-describe their planning, as more a deliberate or more intuitive, are associated with generating trajectories that differ according to model-based metrics.

6.3.3 Discussion

In summary, we find that the optimal POMDP model predicts the majority solution. That is, in a given trajectory, the action taken by the majority of the participants is the same action that an optimal POMDP would take. This does not mean that given arbitrary mazes the majority solution will always be optimal, only that given simple mazes such as these participants were able to produce an optimal plan. At the same time, suboptimal trajectories are explained by solving a POMDP with a *softmax* decision noise, so that the likelihood of choosing an action is proportional to its model-based value. Participants' self-reports, 'thinking' or 'guessing', correspond to differences in planning efficiency, measured by model-based metrics.

In the two replications of the planning task (the first one described in Chapter 5) participants self-report as 'thinking' and 'guessing'. At the same time, participants have previously self-reported as attributing intelligence to 'outcome' or to 'efficiency', suggesting that there might be a correspondence between solving mazes and attributing intelligence. The next section replicates the intelligence attribution task and compares individual planning accuracy to intelligence attribution.

6.4 Intelligence Attribution Task

Participants rated the intelligence of maze-solving agents, using the procedure described in Chapter 4. The movies are generated by replaying representative solutions (including optimal, suboptimal and pseudo-random solutions with lucky and unlucky outcomes) of previous participants. Half of the viewed solutions are optimal, matching the optimality statistics typical of human solutions. As before, the experiment includes incomplete trajectories to detect participants who attribute intelligence to outcome. Participants who attribute intelligence to outcome should rate all incomplete trajectories equally, regardless of the agent's efficiency, while those attributing intelligence to efficiency should rate incomplete optimal trajectories higher. The ratings of complete trials are used to measure the participant's sensitivity to efficiency and outcome.

The movies show a variety of representative behaviours and outcomes, corresponding to one of 8 conditions: *optimal-lucky*, *optimal-unlucky*, *optimal-fair*, *softmax-unlucky*, *softmax-lucky*, *softmax0*, *optimal-partial* and *suboptimal-partial*. Incomplete examples are labelled *optimal-partial* or *suboptimal-partial*. A *lucky, fair* or *unlucky* label reflects the agent's

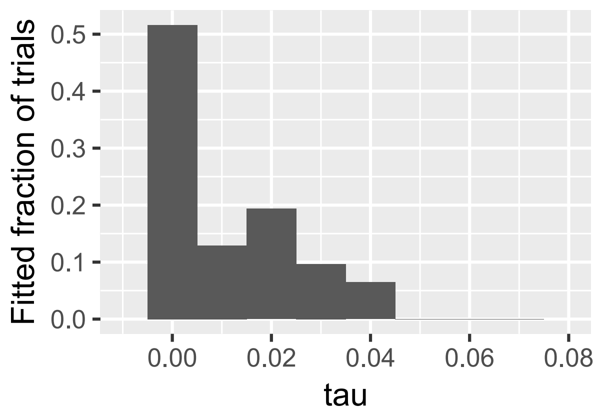


Figure 6.3: Frequency of different levels of τ fitted to movies. Half of the solutions were optimal, with $\tau = 0$.

outcome measured in number of steps to the goal. Agents who encounter the exit in the first room they visit are labelled as *lucky*; agents who search the maze exhaustively are labelled as *unlucky*; other agents are labelled as *fair*. So, a *softmax-lucky* agent finds the goal quickly by sheer luck, analogous to the *lucky-guess* condition in the second experiment. The *softmax-unlucky* agent corresponds to the *softmax* or *softmax-guess* conditions in the second experiment. The *softmax0* condition is similar to the *random* condition in the Experiments 1 and 2 and captures lapses of attention (e.g. re-entering empty rooms) and typos. We hypothesised that unlike pseudo-random behaviours generated by a model, such human behaviours may be rationalised. For example, someone visiting every square in a maze may be looking for hidden trap-doors. Each movie is additionally labelled with an inferred fraction of optimal steps, $optstep_i$ and $SOFTMAX$ τ_i, where i indexes the movie. The frequencies of fitted τ_i are shown in Figure 6.3. [1]

[1]Given that participants have no prior knowledge of the maze and can always see the already visited areas the *optimality ranks* inferred by *Model3* do not apply to the current experiment design.

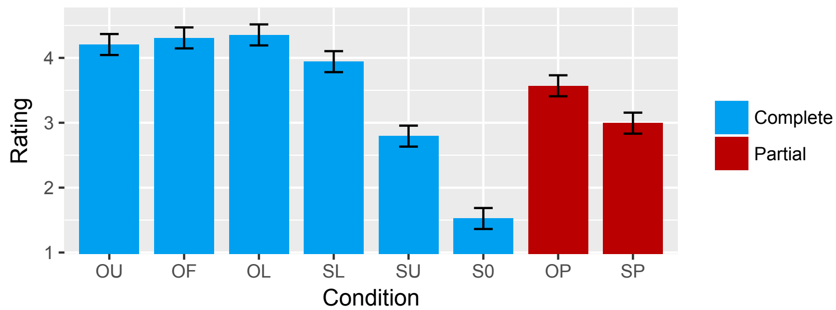


Figure 6.4: Average ratings for conditions *OL-optimal-lucky*, *OF-optimal-fair*, *OU-optimal-unlucky*, *SL-softmax-lucky*, *SU-softmax-unlucky*, *S0-softmax0*, *OP-optimal partial*, *SP-suboptimal partial*. Error bars indicate 95% confidence intervals.

6.4.1 Stimuli

Participants read the instructions on a computer screen in a web browser, and viewed four familiarisation examples followed by the 28 test movies. Each condition occurred four times in a randomised order. After viewing each movie, participants rated the intelligence of the agent selecting a rating from a Likert scale between 1 (less intelligent) and 5 (more intelligent) by choosing a radio button. At the end of the survey participants were asked: *How did you make your decision?* Full instructions are listed in the Appendix. The experimental procedure is available online at http://www.cgl.uwaterloo.ca/mkryven/attribute/int_exp.php

6.4.2 Results

ANOVA of *rating* against *condition* shows a significant main effect of condition, ($F(7, 24) = 73.45, p < .0001$). The adjusted R^2 of the regression model is .942. Mean ratings for each condition are shown in Figure 6.4. The difference between ratings of *optimal* conditions was not significant. According to Tukey HSD the difference between ratings of *optimal-unlucky* and *optimal-lucky* conditions was not significant ($d = 0.15, p = .9$) as well as the difference between *optimal-lucky* and *optimal-fair* ($d = 0.04, p = .99$) and *optimal-unlucky*

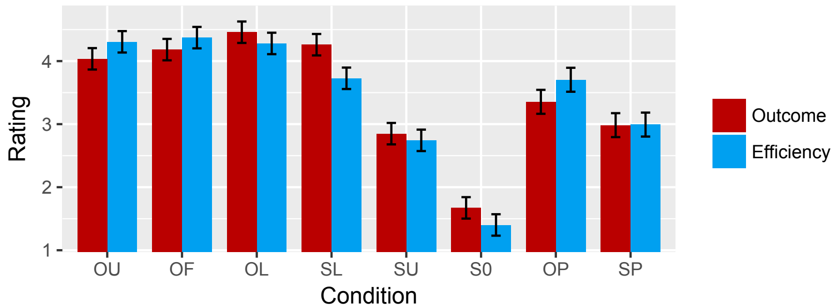


Figure 6.5: Comparing intelligence attribution of **outcome** and **efficiency** groups. *OL-optimal-lucky*, *OF-optimal-fair*, *OU-optimal-unlucky*, *SL-softmax lucky*, *SU-softmax unlucky*, *OP-optimal-partial*, *SP-suboptimal-partial*, *S0-softmax0.*

and *optimal-fair* ($d = 0.1, p = .99$) conditions. So, participants rated all optimal trials highly regardless of outcome.

Incomplete optimal trials were rated higher than incomplete suboptimal trials ($d = 0.59, p = .028$). So incomplete trials were rated according to their efficiency, rejecting the outcome hypothesis [2]. At the same time, suboptimal decisions with lucky consequences are rated differently than unlucky ones. The difference between *optimal-unlucky* and *softmax-unlucky* is significant ($d = 1.41, p < .0001$), showing that among unlucky agents optimal agents are preferred. However there is no evidence of a difference between ratings of *optimal-lucky* and *softmax-lucky* ($d = .41, p = 0.2$) trials.

Two independent judges coded participant's responses to *'How did you make your decision?'* as 'outcome' or 'efficiency'. The raters agreed on 93% of participants, coding 33 as **outcome** and 55 as **efficiency** . The remaining 6 gave mixed answers, such as: *I tried to base it on if I would do the same thing and how few of steps were taking* and were assigned to a group after a discussion. For complete trials, repeated measures ANOVA of *rating* against ($condition \times group$) reveals a significant effect of *condition* ($F(5, 36) = 173.9816, p < .0001$) and a significant interaction between *group* and *condition* ($F(5, 36) = 3.0124, p = .02$).

[2]The reader will remember that according to the outcome hypothesis intelligence is attributed to the agent's outcome. Thus, the attribution can not be made if the outcome is not observed.

Group	Number of responses	Optimal Steps
Thinking-Efficiency	41	$m = .88, sd = .05$
Guessing-Outcome	21	$m = .84, sd = .03$
Guessing-Efficiency	13	$m = .86, sd = .05$
Thinking-Outcome	11	$m = .86, sd = .04$

Table 6.1: Planning and Intelligence Attribution

Mean ratings are shown in Figure 6.5.

Both groups rated optimal trials highly, regardless of outcome. According to Tukey HSD, the difference between **outcome** group's ratings of *optimal-lucky* and *optimal-unlucky* agents is not significant ($p = .4$). Likewise, there is no significant difference between **efficiency** group's ratings of *optimal-lucky* and *optimal-unlucky* agents ($p = 1$). However, there is no evidence that the **outcome** group rated lucky suboptimal and optimal agents differently. While both groups rated optimal agents highly regardless of outcome, only the efficiency group preferred efficient agents to lucky but inefficient ones.

In contrast to the results of the second experiment, both groups rated incomplete optimal trials higher than the suboptimal ones. For ratings of partial trials by the outcome group, a Welch Two Sample t-test indicates that the ratings of optimal trials ($M = 3.35, SD = 0.19$) are significantly different from the suboptimal trials ($M = 2.98, SD = 0.04$), ($t(3.2535) = 3.7789, p = .028$). Likewise, ratings of partial trials by the efficiency group were higher for optimal ($M = 3.7, SD = 0.41$) than for suboptimal trials ($M = 2.99, SD = 0.3$) trials, ($t(5.5407) = 2.8127, p = .033$).

Comparing verbal-self reports in the planning and attribution tasks shows that responses are correlated (Kendall's rank $\tau = 0.4, p = .0002$). Participants who self-report as **thinking** are more likely to self-report as **efficiency**. A third of participants also gave inconsistent self-reports, such as **thinking-outcome** and **guessing-efficiency**, suggesting that participants do not use themselves as a model to evaluate others.

For each participant we calculate their **sensitivity to efficiency** defined as the difference between the participant's average ratings of optimal and suboptimal trials and

sensitivity to outcome defined as the difference between participant's average rating of lucky and unlucky trials. The sensitivities to outcome and efficiency measured in this way are independent of self-report and anti-correlated, Pearson ($r = -0.27, p = .0086$). Figure 6.6 shows the relationship between individual sensitivities to outcome and efficiency for participants in each group. Thus, although better efficiency is on average expected to result in better outcomes, whenever the two dimensions can be dissociated they appear to contribute to attributed intelligence in opposite ways. As participants become more sensitive to efficiency they also become less sensitive to outcome.

Two factor ANOVA of sensitivity to efficiency against self-report of $planning \times attribution$ shows significant effects of attribution ($F(1, 82) = 7.9467, p = .006$), but no significant effect of planning ($p = .2$) and no interactions between factors ($p = .3$). Tukey HSD identifies significant differences between the **thinking-efficiency** and the **guessing-outcome** participants, ($d = 0.49, p = .03$), as well as between **efficiency** and **outcome** groups ($d = 0.41, p = .006$). The data on which ANOVA was calculated is shown on Figure 6.6. So, the **efficiency** group attribute more intelligence to efficiency compared to the **outcome** group.

Two factor ANOVA of sensitivity to outcome against self-report of $planning \times attribution$ shows significant effects of attribution ($F(1, 82) = 10.4198, p = .002$) and of planning ($F(1, 82) = 8.1147, p = .006$), with no interactions between factors ($p = .14$). Tukey HSD shows that **guessing-outcome** participants are more sensitive to outcome compared to **thinking-efficiency** participants, ($d = 0.49, p = .001$) and **guessing-efficiency** are more sensitive to outcome compared to **thinking-efficiency** participants, ($d = 0.45, p = .01$). There is also a significant differences between **guessing** and **thinking** groups ($d = 0.26, p = .01$) and between **outcome** and **efficiency** groups ($d = 0.33, p = .001$). The data used to calculate ANOVA is shown on Figure 6.6. So, self-reporting as **guessing** or as *outcome* is correlated with attributing more intelligence to outcome.

The relationship between planning and attributed intelligence is confirmed by considering the correlation between planning efficiency, measured as the fraction of optimal steps during planning, and sensitivity to efficiency. The sensitivity to efficiency and $optstep_p$ are positively correlated ($r = .54, p < .0001$), showing that participants attributed intelligence to efficiency in proportion to their own planning skills.

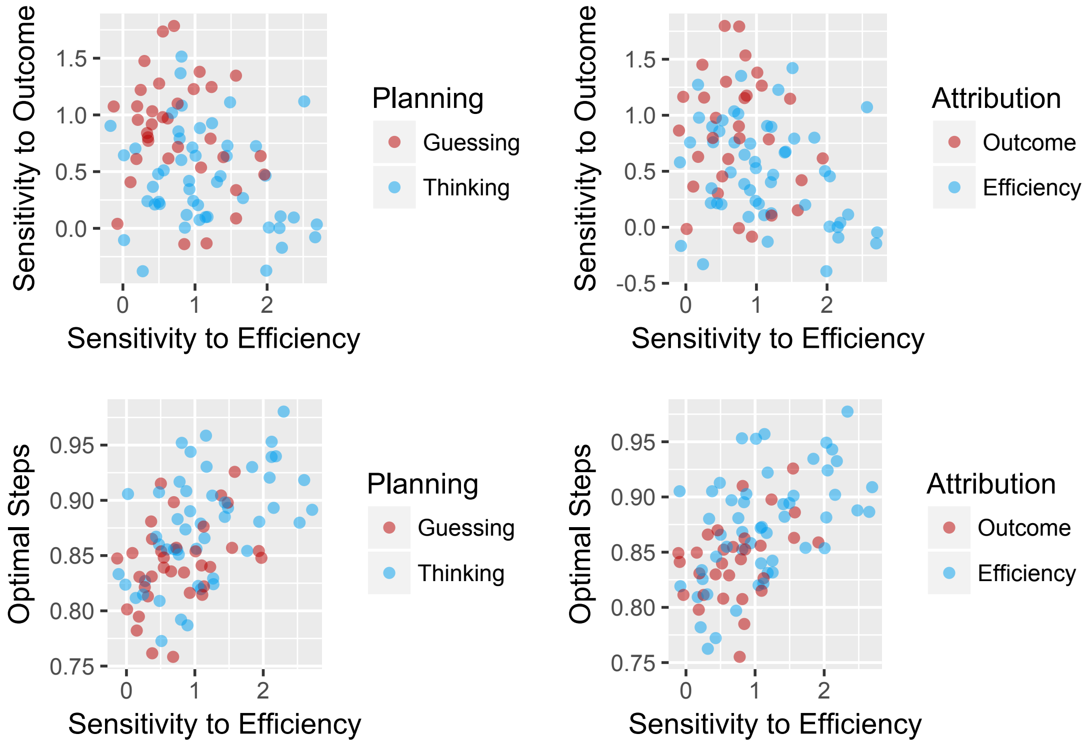


Figure 6.6: Plots show the relationships between self-report and model-based metrics. Individual **sensitivity to efficiency**, defined as the difference between the participant's average ratings of optimal and suboptimal trials and **sensitivity to outcome**, defined as the difference between participant's average rating of lucky and unlucky trials.

Two factor ANOVA of $optstep_p$ against self-report of *planning* $\times$ *attribution* shows a significant effect of attribution ($F(1, 82) = 6.8677, p = .01$), but not of planning ($p = .07$). According to Tukey HSD, the *efficiency* participants are more likely to make optimal moves compared to the *outcome* participants ($d = 0.03, p = .01$), and the *thinking-efficiency* participants are more likely to move optimally compared to the *guessing-outcome* group ($d = 0.04, p = .01$). (Table 6.1). Two factor ANOVA of $optstep_p$ against sensitivity to efficiency and sensitivity to outcome shows a significant effect of efficiency ($F(1, 83) = 34.2377, p < .0001$) but not of outcome ($p = .3$). The data used to calculate the ANOVA is shown on Figure 6.6.

In addition, according to Welch Two Sample t-test participants who self-reported as **efficiency** were on average more optimal ($t(3.2069) = 75.81, p = .002$), taking optimal steps 87% of the time. The **outcome** group took optimal steps 84% of the time. Lastly, participants are not using themselves as a model for evaluating others, since the outcome group, despite being on average more likely to plan sub-optimally, rated only the lucky sub-optimal planners highly.

Together, these results show that planning skills affect attributed intelligence. While participants in the efficiency group attempt to evaluate the proximity of the observed behaviour to the optimal strategy, people are able attribute intelligence to efficiency in proportion to their planning skills.

6.4.3 Discussion

The results show people attribute intelligence to efficiency in proportion to their planning skills. Attributing intelligence to outcome is correlated with poor planning. However, the correspondence between verbal self-reports as **efficiency** or **outcome** and as **thinking** or **guessing** is approximate, verbal self-reports may be a weak predictor of behaviour. On one hand, some participants who were coded as **thinking-outcome** might have incorrectly rationalised their behaviour. On the other hand, participants coded as **guessing-efficiency** might have had difficulty planning, but improved their planning skills by the time they completed the planning task.

In agreement with Experiment 2 in Chapter 4, about a third of participants self-identify as attributing intelligence to outcome. However, this time the **outcome** group rated all optimal agents highly and rated incomplete optimal trials higher than the suboptimal ones, suggesting that planning practice may affect intelligence attribution by improving one's planning skills. To validate this hypothesis a controlled experiment needs to compare participants who practice planning before attributing intelligence, to a control group who complete an irrelevant task. Notably, even after planning practice, **outcome** participants still attribute less intelligence to efficiency than the **efficiency** ones. Participants might learn to recognise optimal solutions before being able to identify mistakes in inefficient ones. The **outcome** group may also lack empathy, which is primed by the planning task.

However, poor planning and lack of empathy might have a common cause since both are cognitively demanding [21]. The cognitive cost of empathy added to the cost of planning might be prohibitive to observers with limited mental resources. If planning ability depends on a general cognitive resource, then variation in both planning accuracy and attributed intelligence may be explained by an independent metric of general intelligence.

The efficiency hypothesis explains human reasoning in simple goal-directed scenarios. However, many real-life situations are complex, the observer's mental resources are limited and so intelligence attribution may be noisy. In complex scenarios people might use a mixture of outcome and efficiency approaches. Indeed, comparing a participant's sensitivity to outcome and to efficiency shows that participants fall along a continuum rather than falling into separate groups. Since intelligent agents are more likely to achieve good outcomes, the outcome heuristic is expected to do better than chance. So, in real-life scenarios implicit attributions of intelligence to outcome may be made automatically but revised upon reflection. For example, stereotype-consistent judgments are exacerbated during fast responses, but are rarely reported if participants are given time to think [10]. One way to investigate this possibility is to collect evaluations of intelligence in real time, such as asking participants to give continuous evaluations by controlling a joystick.

Furthermore, even if the observer is skilled, the agent's problem might be too difficult for the observer. In such a case, do people immediately fall back on attributing intelligence to outcome, or do they evaluate the agent planning partly, for example, by following its reasoning up to a certain depth? Given that the accuracy of Bayesian inference increases with an increasing number of observations, the observer might choose to evaluate a subset of decisions instead of evaluating the whole plan and come to a conclusion once a sufficient confidence is reached.

Lastly, attributing intelligence to lucky agents may be an adaptive automatic response. When one's resources are limited by a cognitive load, people perceive lucky persons as more likeable [92] and young children tend to evaluate lucky others as nice [91, 90]. Liking lucky others is rational, to the extent that it promotes favourable social connections. [3]. To differentiate between liking lucky others and attributing intelligence to outcome it would

[3] A comment by a participant in Experiment 3: 'Generally I attributed intelligence to their efficiency, but also we all like lucky winners!'

be necessary to separately ask participants about their liking of the agent and about the agent's intelligence. One measure of liking is to instruct participants to share bonus with other participants. Moreover, if liking lucky agents is adaptive, then attributing intelligence to outcome should disappear if the agent appears mean.

Importantly, the efficiency hypothesis implicitly assumes that the observer understands the agent's abilities and constraints and has the cognitive resources needed to evaluate the agent's planning. The intelligence attribution task is designed to make inference of the agent's mental states easy by explicitly informing participants about the agent's goal and its view of the maze. Under such conditions, competent planners interpret the agent's intelligence as planning efficiency. One important difference between the stimuli used for intelligence attribution and the actual human planning is that participants' speed varies during a planning trial, while in the intelligence attribution movies the agents move at a uniform speed. Varying the agent's speed during the pilots was confusing to the participants (see Section 4.2.5), however pausing for 500ms at every observation cell is not what people actually do. Showing the participants' planning in real-time might convey a more accurate idea of thinking.

In summary, most participants attribute intelligence to goal-directed agents by evaluating the efficiency of the agent's planning, as suggested by the rationality principle. To the extent that the skills needed to accurately evaluate the agent's planning are not available to the observer, attributing intelligence to outcome may be a rational alternative. Both methods optimise the use of the resources available to the observer.

Chapter 7

Eyetracking

Chapter 5 shows that participants differ in planning ability. Those who self-report as 'thinking' are more optimal according to model-based metrics. Reaction time (RT) analysis shows that participants divide maze problems into planning sub-problems with decision nodes at observation states, and that mental effort spent in a maze location increases with the number of available actions. However, RT are insufficient to make conclusions about the planning depth and breadth of individual participants. A replication of the maze-solving experiment with eye-tracking may offer additional evidence in that regard. Moreover, it is unclear whether verbal self-report and model-based metrics measure performance specific to the planning task, or to a more general cognitive ability.

Searching for a goal in a maze, participants are likely to think visually and trace hypothetical paths to find the best route. Although eye-movements generally do not follow problem-solving algorithms as a one-to-one mapping between fixations and cognitive processing, the stochastic distributions of fixations inside a maze are expected to reveal elements of the problem that are maintained in working memory in the current stage of the task [31]. In particular, the distance from the agent to fixation, **fixation depth**, might reflect the depth of participant's planning. **Fixation depth** can be measured as the distance from the agent's location to fixation in three ways: (1) a distance in a straight line; (2) the number of steps from the agent to reach the fixated location; (3) the number of observation nodes between the agent and the fixated location. At the same time, the frequency with which different types of cells (dark, white, wall) are fixated reflects the distribution of participant's attention within the maze. So, optimal planning might involve fixating a higher proportion of dark cells and looking further ahead and eye-movements while collecting information about the environment might predict the optimality of subsequent planning.

In addition, if attributed intelligence and planning depend on a common cognitive resource, there may be a correspondence between planning/plan evaluation and numeric intelligence, as measured by a version of the Cognitive Reflection Test (CRT) [39]. The CRT consists of three problems that measure a person's numerical ability, rational thinking and inhibition of incorrect impulsive response. It is a relatively simple task correlated with more time-consuming established measures of general intelligence, such as intelligence quotient and rational thinking [131, 140]. If planning accuracy depends on general cognitive ability then the variation in model-based metrics (decision noise inferred by the *SOFTMAX model* and the fraction of optimal steps) can be explained by the participant's CRT score. In addition, pupil size correlates with cognitive performance [134], working memory [63] and expected utility of the task [61]. So, simple pupil dilation measured during eye-tracking provides a proxy metric of a participant's cognitive performance.

The experiment described in this chapter replicates the maze-solving experiment with eye-tracking to test two hypotheses. First, planning accuracy may correlate with specific patterns of eye-movements and fixation depth. Second, model-based metrics of planning accuracy reflect a general cognitive ability. The next section reviews related work on eye-tracking during planning, followed by the experimental method in Section 7.2 and the results of the planning task in Section 7.3. Section 7.4 describes the results of the intelligence attribution task and comparing model-based metrics of planning and intelligence attribution.

7.1 Related Work

Previous research on planning in a fully observable space using eye-tracking shows that participants differ in how far ahead they plan, and proposes a classification of eye-movements into two categories: *motor guidance* and *exploration* [149]. *Motor guidance* fixations immediately precede motor actions (e.g. guiding the hand moving the cursor). *Exploration* fixations look ahead, collecting information about the environment, while the cursor is stationary, their signature is the divergence of eye and the cursor. While *motor guidance* fixations are highly stereotypical, *exploration* fixations are idiosyncratic. Some participants explore extensively before moving, while others alternate between exploration and guidance [149]. In addition, saccadic eye movements are often formalised by Information Gain models [104] as selecting locations most relevant to the spatio-temporal demands of the task [52, 105]. Early sequences of fixations when first seeing an environment are often formalised as discovering its structure [104, 105].

So, when first seeing a maze a participant studies the environment, the rooms and the walls in a maze, which engages a specific eye-movement pattern. Through the rest of the trial participants usually generate a combination of motor guidance and planning fixations. While navigating previously observed parts of the maze (fixations landing on white cells) the eye-movements are expected to be similar to fixations described in [149]. In contrast, fixations landing on black cells might relate to *decision* planning and reasoning about probabilities.

At the same time, in the intelligence attribution task, participants who evaluate the agent's efficiency might do something like predicting the agent's behaviour and then comparing it to observed actions. According to Event Segmentation Theory, humans routinely make short-term predictions of trajectories of moving objects or agents and segment the stream of perceptual experience into events when predictions fail [147, 103]. Might participants' predictions of the agent's behaviour influence how they evaluate its intelligence?

Such predictions can be measured indirectly using eye-movement analysis, since action perception is accompanied by patterns of eye-movements that alternate between fixating the supposed goal and the agent. Action interpretation is known to be idiosyncratic. Healthy adults consistently segment streams of visual information into events with intra-individual agreement higher than the inter-individual agreement [126], suggesting that translating perception to events may recruit the observer's planning mechanisms [103]. On one hand, participants might limit themselves to certain predictions, only expecting the agent to approach the next observation. On the other hand, people could also look ahead of the agent along the path they anticipate it to take.

Moreover, the agents in the intelligence attribution task move through the maze in a way that elicits smooth pursuit eye-movements. Such eye-movements are automatically initiated by sensory motion cues and can not be performed at will. However, smooth pursuit eye-movements vary between tasks, in a way that results from an interaction between low-level cues (e.g. motion) and high-level cognitive processes (goals, beliefs). Thus, the observer's beliefs about the agent's mental states might influence smooth pursuit eye-movements as well [66].

7.2 Method

The experiment used the methodology described in Chapter 6 while tracking eye-movements. Eye movements were recorded using the EyeLink 1000 Plus Desktop Mount with a head-free upgrade, sampling at 1000 Hz from the left eye. The eye-tracker measures projections of gaze onto the screen and pupil size in eye-tracker pixel units. The pupil size measured by the device depends on the eye-tracker calibration, which optimises tracking for each participant. So, pupil size measurements indicate relative effort within participants but not between participants.

Participants first complete a calibration trial comprising the usual Eyelink 9-point calibration routine followed by a custom pupil diameter calibration while looking at black, grey, and white rectangle subtending 15 degrees of visual angle. Each rectangle is displayed for 4 seconds, during which time the participant's pupil size is recorded. Following calibration participants answer a mental arithmetic question '*How much is 14 times 15?*' to measure pupil dilation during mental effort. The pupil size is recorded for 4 seconds starting at 4 seconds after the question is presented. The question is presented on a grey screen.

Next, participants complete the planning task followed by the intelligence attribution task. Upon completing the eye-tracking session, participants provide a free-form answer to two questions: *How did you make your decisions while solving mazes?* and *How did you* *make your decisions while judging others?* by writing the answer on a provided answer sheet. Lastly, participants complete an untimed CRT.

The validity of the original CRT is disputed owing to the popularity of the test [23]. A common solution is providing a different but analogous set of problems [23]. In our version of CRT participants answer the following questions:

- A slime mould doubles in size every 2 hours. One gram of slime mould can fill a container in 8 hours. How long does it take to fill half of the same container?
- It takes 10 people 10 hours to knit 10 scarfs. How long does it take 100 people to knit 100 scarfs?
- A coffee and a sandwich cost $12. A sandwich costs $10 more than a coffee. How much does a coffee cost?

During the experiment the experimenter was nearby, ready to handle technical problems. [1] Stimuli were presented on a 21.5-inch Apple LCD Display (Display width: 18.7 inches (47.498cm), display height 10.5 inches (26.67cm)), 1024×768 resolution, which is the lowest possible resolution [2], 60 Hz refresh rate. The stimuli were presented using Psychtoolbox [16]. The display was viewed from a distance of 60 cm.

7.2.1 Participants

Thirty-five participants, median age 27, standard deviation 12, 18 females and 17 males were tested with approval of an MIT University Institutional Review Board and a University of Waterloo ethics committee. Four were excluded because of technical problems. Participants were recruited from the MIT Psychology subject pool, which includes a mix of MIT students and members of the general public, who sign up to participate via an online system and are paid $12 for a 30 minute session. All had normal or corrected to normal vision.

[1] This is to make sure that the eye-tracker is recording and has not disconnected or crashed.

[2] Pixel size 0.46mm $\times$ 0.34mm

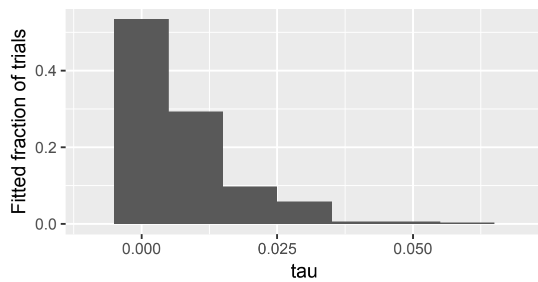


Figure 7.1: The distribution of decision noise inferred from solutions produced by participants according to the *SOFTMAX* model.

7.3 Planning task

7.3.1 Stimuli

Seventeen mazes including 5 practice mazes and 12 test mazes are presented to each participant centred on the screen, subtending between 5 and 8 cells. Each maze cell is 80 pixels tall and 80 pixels wide (36mm $\times$ 28mm, 1.7×1.3 degrees of visual angle). The start location is always assigned to the top left corner of the maze. Participants are instructed to look for a hidden 'exit', marked when visible as a bright red square in a series of mazes, by controlling an agent using arrow keys on a keyboard. The agent moves one grid square at a time: N, W, S, E and has a 180 degree view of the maze limited by walls. The maze is initially dark, but is uncovered as the agent moves along, so participants initially know the layout of the rooms and location of the barrier walls, but not where the goal is. Participants are instructed that each of the dark squares is equally likely to hide the 'exit', and that they should get to it in as few steps as possible. To incentivise attention at the end of the experiment participants see how many steps they have taken compared to others before them. [3] During the trial current step count is displayed above the maze. Once the exit is reached, the participant immediately starts the next maze.

[3] This data is available based on pilot and previous runs of the experiment

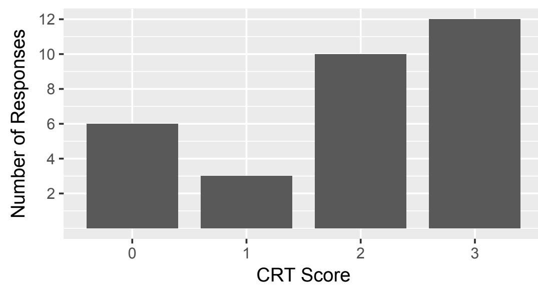


Figure 7.2: The distribution of CRT scores across the tested participant sample.

7.3.2 Results

Model-Based Inference

For each participant we calculate model-based metrics of planning performance, the fraction of optimal steps taken by the participant, $optstep_p$, and the mean decision noise inferred from the participant's solutions by $SOFTMAX$, τ_p, where p is the participant index. In addition, relative pupil size is calculated as a ratio on the mean pupil size while viewing the grey screen during calibration to the maximal size of the pupil during mental arithmetic.

Of all solutions generated by participants 47% were optimal ($\tau = 0$) , another 47% had decision noise ($\tau > 0$) and another 6% included zero-utility actions, most likely due to a difficulty using keyboard with the eye-tracking setup. Participants took 87% optimal steps on average, standard deviation 6%. As shown in Figure 7.1, among the suboptimal solutions, solutions with lower decision noise were more frequent. Of the 32 participants 6 self-reported as solving mazes by 'guessing', 4 of whom also self-reported as attributing intelligence to outcome. Another 25 self-reported as solving mazes by 'thinking'. The distribution of CRT scores in Figure 7.2 shows that about a third of participants answered all questions correctly, a third answered two questions, and another third answered one or fewer questions.

Next, we test the relationship between the model-based metrics and the independent metrics of cognitive performance (the CRT score and pupil relative size during mental

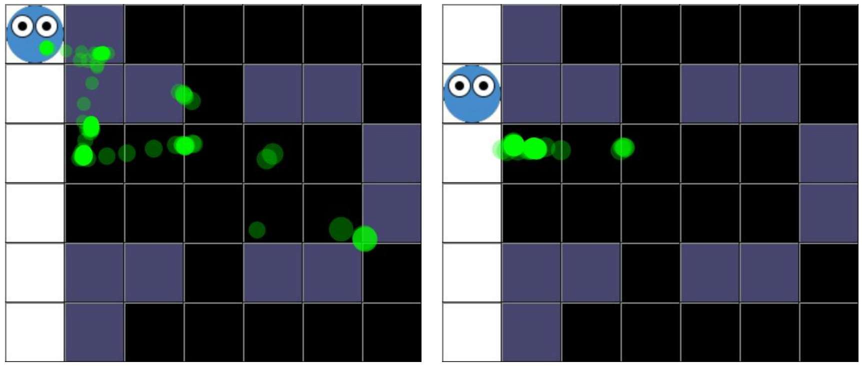


Figure 7.3: Example 1. Left: The start of the trial shows the participant scanning the layout of the maze. Right: After taking a step the participant looks at the cells to be revealed next.

effort). ANOVA of CRT score against $optstep_p$ and τ_p shows a significant effect of τ_p ($F = 7.5079, p = .01$) but not of $optstep_p$ ($F = 0.1109, p = 0.7$). So, the participants' CRT scores predicts their decision noise. ANOVA of the relative pupil size vs. $optstep_p$ and τ_p shows a significant effect of $optstep_p$ ($F = 6.8387, p = .01442$), but not of τ_p ($F = 3.7832, p = .06$). So, the variation in the participants' relative pupil size predicts planning planning optimality. Pearson correlation between the relative pupil size and $optstep_p$ is $r = .53, p = .002$ and between CRT score and τ_p $r = -.45, p = .009$. In summary, the variation in planning accuracy can be explained by the variation in cognitive performance measured by CRT and relative pupil size.

Eye-movement Analysis

All recorded eye-movements can be viewed frame-by-frame here: http://www.cgl.uwaterloo.ca/mkryven/comic_all_subjects_1.pdf. Real-time replays with superimposed fixations can be downloaded from: http://www.cgl.uwaterloo.ca/mkryven/replayMovies.zip. Each green dot corresponds to a recorded of gaze projection sampled at 500Hz. A saturated green dot indicates several superimposed samples, a fixation. Transparent dots are artefacts, such as blinks or saccades. The size of each dot is proportional to pupil size.

A pre-processing script [4] interprets the data recorded by the eye-tracker as fixations. Each fixation is labelled by the coordinates of the maze cell where it falls. If a fixation drifts across several maze cells then it is recorded as two fixations, which keeps track of all fixated maze cells. Each fixation inside the maze is labelled as falling on a black, white or wall cell and assigned the median pupil size, the distance in pixels from the agent and the distance in steps from the agent. Next, we informally describe the patterns of eye-movements that occur during a trial. Two trial examples are shown in Figures 7.3-7.6 and Figures 7.7 - 7.8. Example 1 in Figures 7.3 - 7.6 shows an optimal solution. Example 2, Figures 7.7 - 7.8, shows a non-optimal path.

[4] I wrote the script because I needed extra processing on top of the standard eyetracking fixation-processing algorithm.

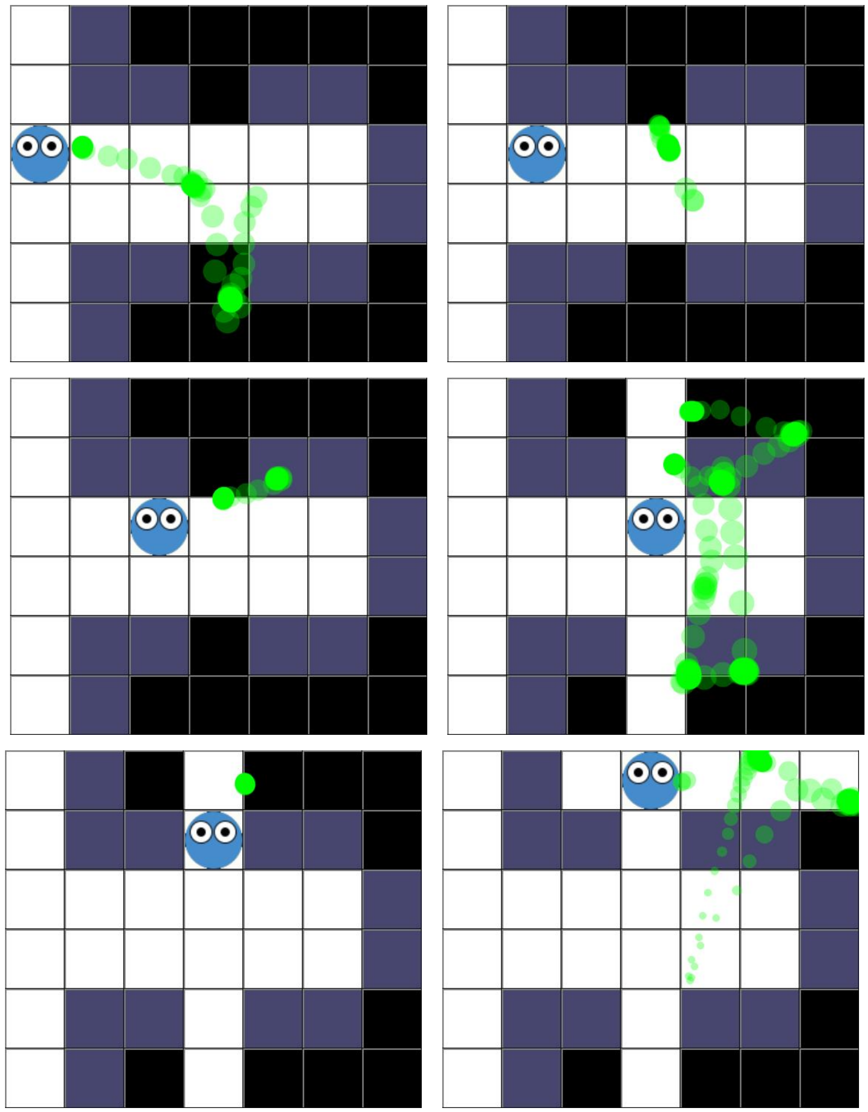


Figure 7.4: Example 1, continued. The participant alternates between looking at the immediate next steps (motor guidance) and looking ahead at unrevealed areas (decision planning).

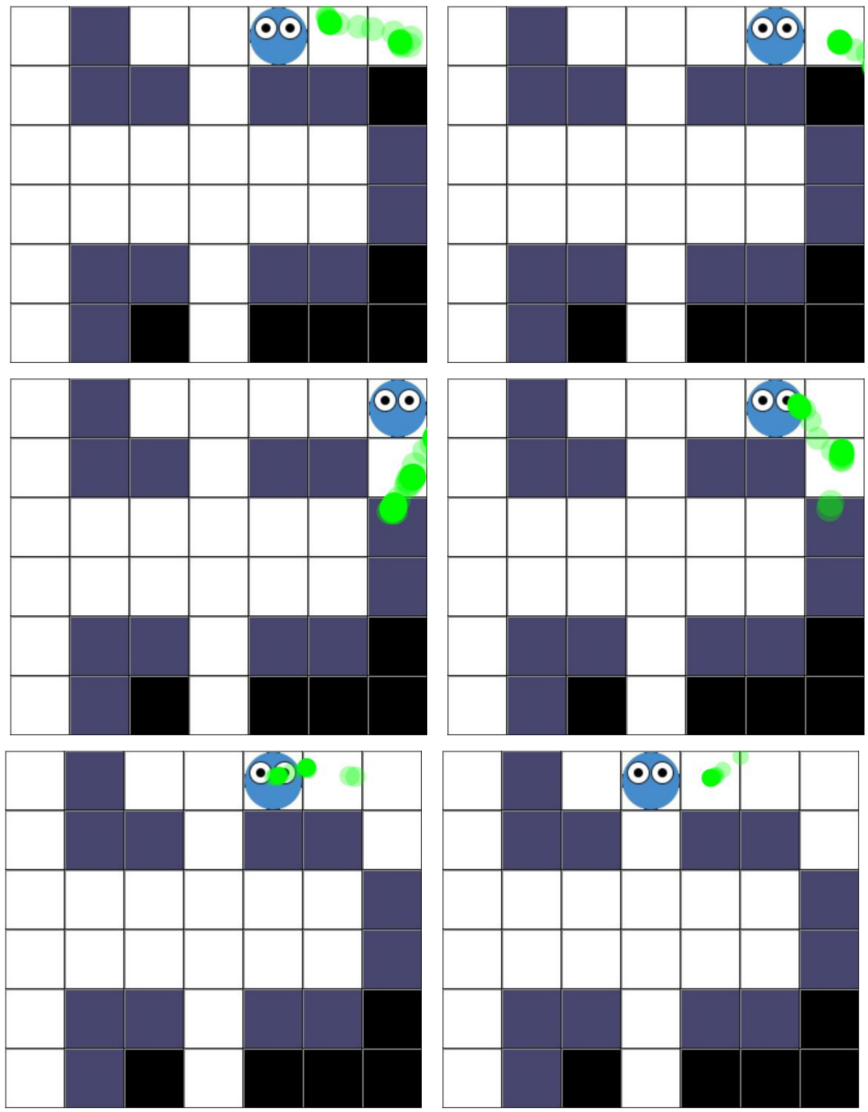


Figure 7.5: Example 1, continued. The participant looks at white cells near the agent.

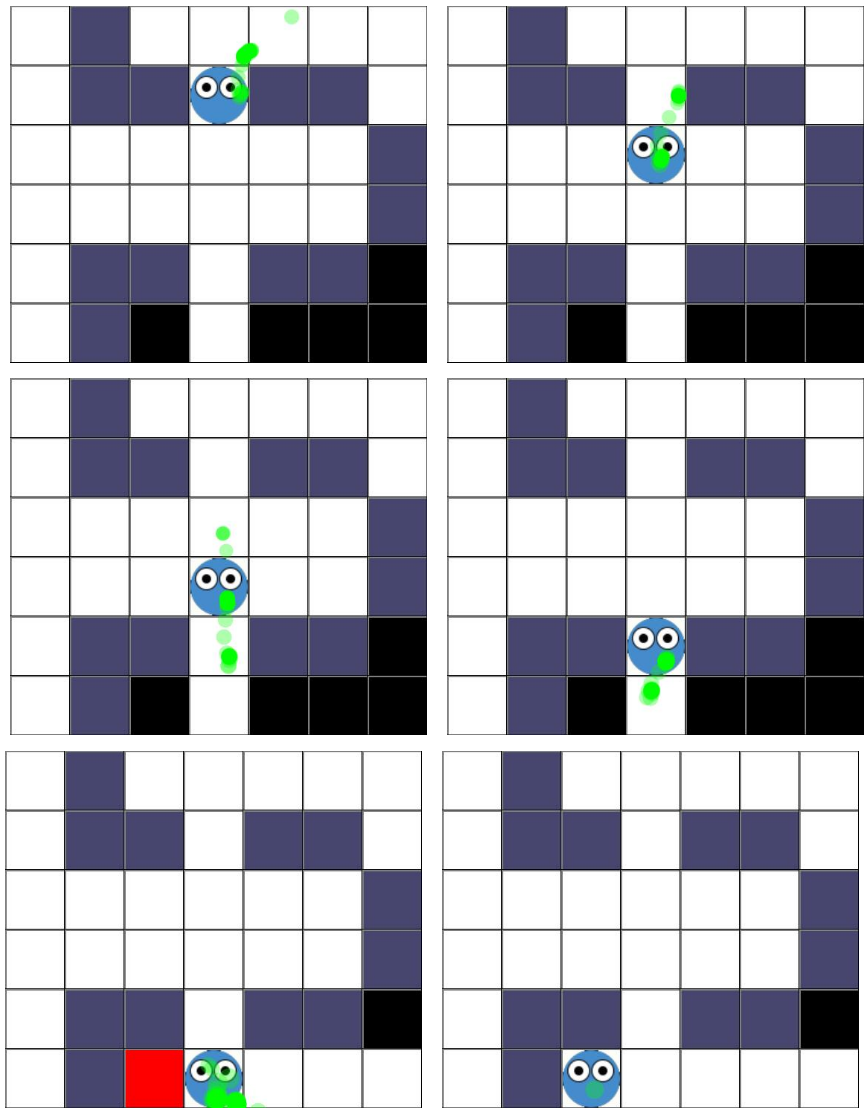


Figure 7.6: Example 1, continued. The participant looks on white cells immediately ahead, similar to *motor guidance* fixations described in [149], or at the agent.

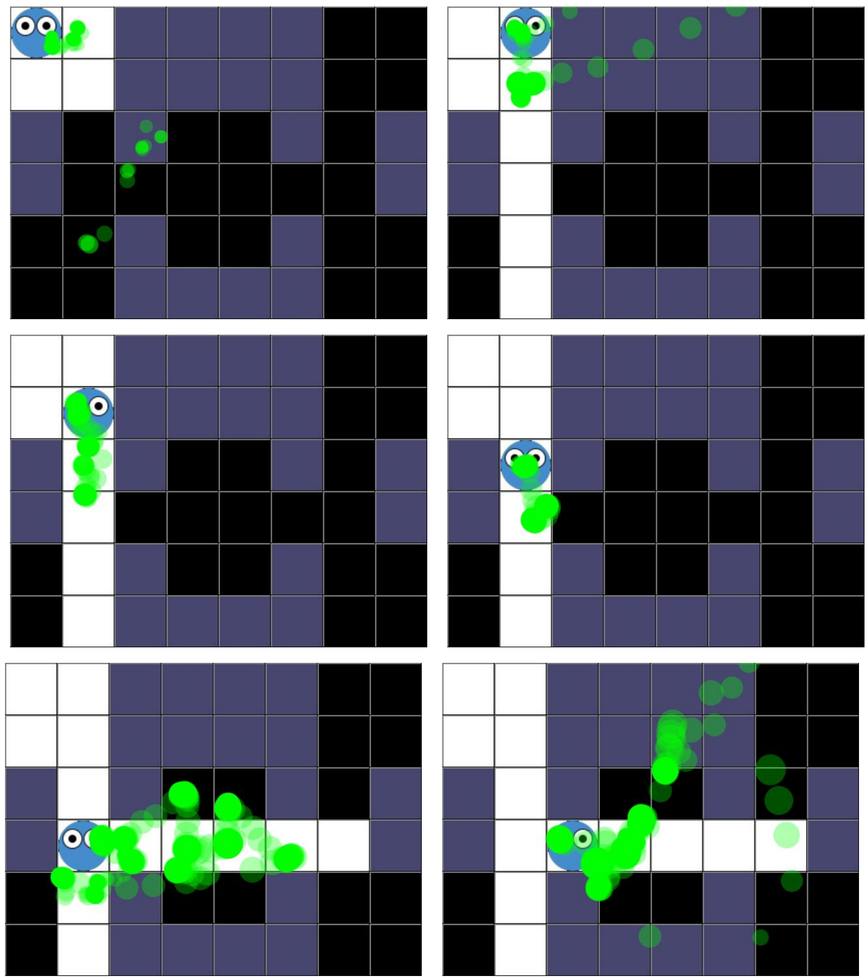


Figure 7.7: Example 2. After scanning the maze the participant approaches the first decision point. Fixations occurring at this stage land on white cells, suggesting motor guidance.

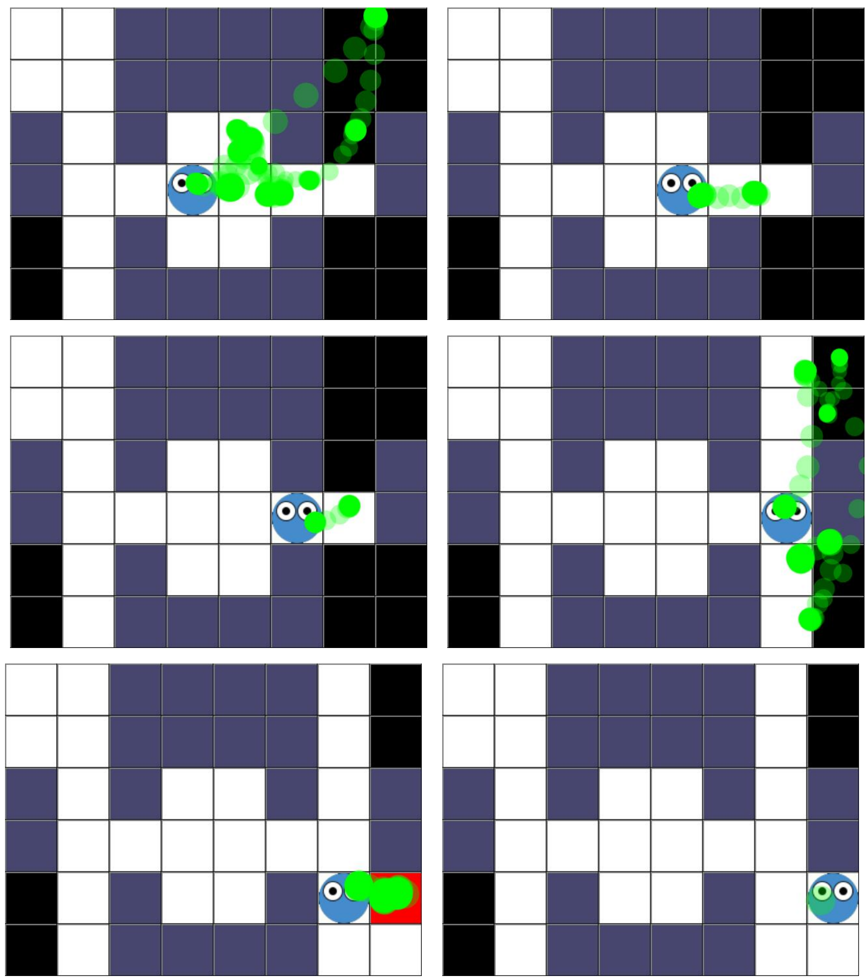


Figure 7.8: Example 2, continued. At the first decision point the participant fixates the cells associated with two alternative choices (going right and going down). More fixations on black and white cells are made at observation locations, while moving between observations is accompanied by making relatively few fixations on white cells.

Figures 7.3-7.8 show eye-movements leading the future actions. While traversing already revealed areas eye-movements fall on the immediate next steps. When approaching unrevealed areas (black cells), eye-movements alternate between possible directions, presumably assessing the value of each choice. The correspondence between eye-movement states and RT states discussed in Chapter 5 is not one-to-one. Rather, there appear to be three distinct patterns of eye-movements.

- **Scanning Maze Layout**: In the starting cell participants scan the maze to discover its structure.
- **Guidance**: Looking at already revealed (white) cells close to the agent.
- **Decision Planning**: Looking at black cells associated with different possible trajectories.

To test whether planning depends on eye-movement patterns, we calculate frequency of looking at black, white, wall or agent cells in a maze (($D_{black}, D_{white}, D_{wall}, D_{agent}$) for each participant. For example, frequency of looking at black cells D_{black} is the total duration of fixations on black cells divided by the total duration of all fixations inside the maze. Median distance from agent to fixation measures how far ahead participants look at each type of cell (($pd_{black}, pd_{wall}, pd_{white}$). Moreover, since the distance in pixels to fixation ignores the maze structure of corridors and rooms, distances in steps to fixation ($sd_{black}, sd_{white}, sd_{wall}$) are also calculated, as the number of steps to reach the fixated location, as an alternative metric of looking ahead.

ANOVA of $optstep_p$ against ($D_{black}, D_{wall}, D_{agent}$) shows a significant effect of D_{wall} ($F = 5.6522, p = .024$), but not of other variables. The Pearson correlation of $optstep_p$ and D_{wall} is $r = -.39, p = .03$, showing that better planners are less likely to fixate walls. ANOVA of $optstep_p$ against (($pd_{black}, pd_{wall}, pd_{white}$) shows a significant effect of pd_{black}, ($F = 4.0457, p = .05$) but not of other variables. The Pearson correlation of $optstep_p$ and pd_{black}, $r = .35, p = .05$ shows that the fraction of optimal steps increases with median distance to fixated black cells (Figure 7.9). ANOVA of $optstep_p$ against $sd_{black}, sd_{white}, sd_{wall}$ shows no significant effects. In summary, more optimal planners fixate

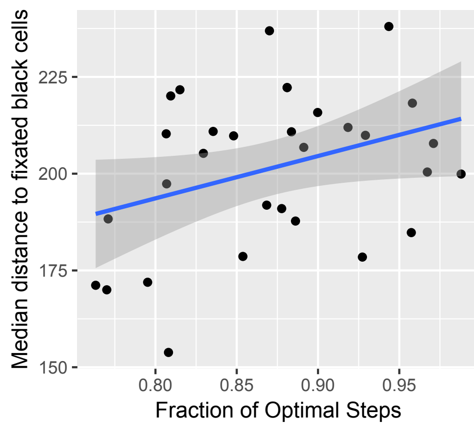


Figure 7.9: The median distance to fixated black cells increases with fraction of optimal steps. Bands represent 95% confidence intervals for means of the linear model.

black cells further away, with distance to fixation measured as a straight line between the fixation and the agent. However, there is no evidence that more optimal planners look further ahead in terms of distance measured in steps. One possible explanation of this effect is that the distance in steps is not available to the visual system as the eye-movements are planned. Even without this information, the visual system plans eye-movements further away under conditions when the observers have more cognitive resources at their disposal and can process the sensory evidence collected by such eye-movements (but see page 110 later in this Chapter).

Next, we test whether participants fixation patterns vary with CRT score by calculating the Spearman rank correlations of CRT score with each of the above metrics. Spearman rank correlation of CRT and D_{black} is $r = .49, p = .006$, of CRT and D_{wall} is $r = -.65, p < .0001$, of CRT and D_{white} is $r = -.36, p < .05$ and of CRT and D_{agent} is $r = .46, p < .01$. Spearman rank correlations of CRT and median distance to fixation are significant for all three types of cells: $r = .56, p = .001$ for black, $r = .47p = .009$ for white and $r = .5p = .005$ for walls. Spearman rank correlations of CRT and median distance to fixation measured in steps are significant for fixations on white cells ($r = .47, p = .009$) and walls ($r = .54, p = .002$) but not black cells ($r = .11p = .5639$). In summary, participants with a higher CRT score preferentially fixate black cells and the agent, and spend less time looking on white cells and walls. They also look further ahead, with distance to fixation measured as a straight line between the fixation and the agent. However, there is no evidence that distance in steps to fixation depends on the participants' CRT score. While participants with higher CRT score look more steps ahead during motor planning, most looking at white cells is accomplished in a straight line.

Taking only the fixation data while **Scanning Maze Layout** we calculate the frequency of looking at black, white, wall or agent cells $((D_{S,black}, D_{S,wall}, D_{S,agent})$, as well as the median distances to fixation in pixels $((pd_{S,black}, pd_{S,wall}, pd_{S,white})$. ANOVA of fraction of optimal steps against $(D_{S,black}, D_{S,wall}, D_{S,agent}$ shows no significant effects. Likewise, ANOVA of optimal step fraction against $(pd_{S,black}, pd_{S,wall}, pd_{S,white}$ shows no effects. Thus, there is no evidence that decision-making optimality depends on the type of fixations participants make at the start of the trial.

Decision-Tree Model

As discussed in Chapter 5, participants may conceptualise mazes in terms of observations rather than steps, with most mental effort occurring after new cells are revealed. So, participants might think of each observation as associated with a cost, how far away is it, and a reward, how many cells does it reveal. Such reasoning can be formalised by a decision tree with decision-nodes at observation states. An illustration of the tree-solving paradigm is shown in Figure 7.10. Each fixation falls at a certain *observation depth* in the decision sub-tree, measured as the number of observations that occur until the fixated cell becomes visible. Each observation state is associated with a cost in steps s_i and a number of revealed cells n_i as a reward, where i indexes observation states. The full decision tree of a maze includes all possible trajectories through the decision tree, such that eventually lead to revealing the entire maze. The optimal trajectory goes from the root of the full decision tree to the leaf with the highest reward. The decision-tree paradigm can be also used to part-solve a maze, up to a certain depth. So, a solution of depth d finds the optimal path within a sub-tree of depth d. So, when arriving at an observation, a participant may either follow a pre-calculated plan, or compute a new plan up to a certain *planning depth*

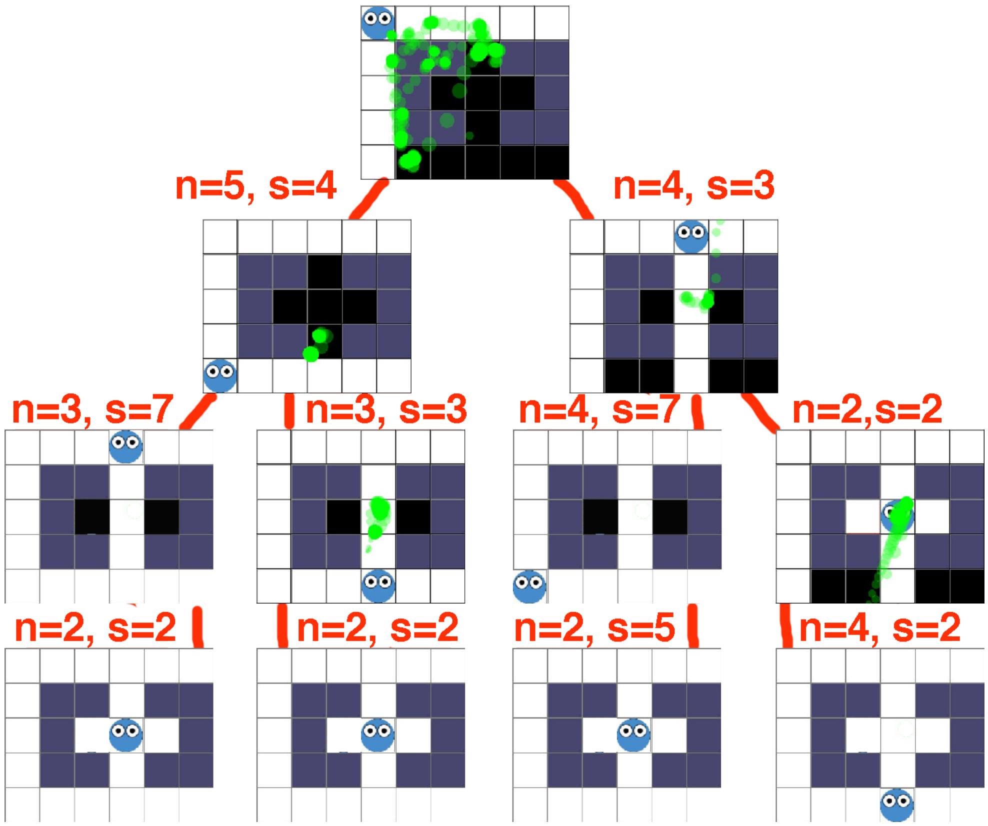


Figure 7.10: A simple maze represented by decision-tree. Examples of eye-movements are shown in each of the states visited by participants.

for the decision sub-tree rooted at the current location. The *planning depth* is defined as the depth of the sub-tree used to compute the plan.

To solve a maze-planning problem using the decision-tree model we use the formal framework described in Chapter 3, but with states corresponding to observation nodes instead of cells. And when given a path through the tree, the same framework can be used to infer the *planning depth* of the path at each observation node, by looking for a sequence of sub-trees within which the corresponding path fragments are optimal.

Next, we analyse the distribution of fixations within a decision tree to measure planning depth in terms of observations and test for a relationship between participant's eye-

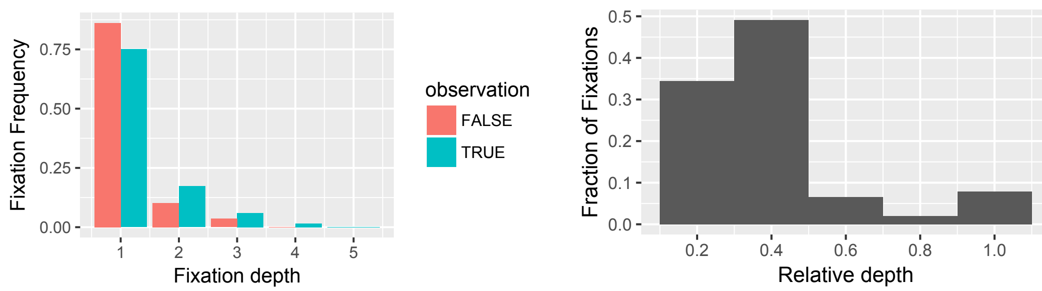


Figure 7.11: Fixation depth at observation and no-observation locations while looking at black cells. Most fixations on black cells fall one observation node in advance. However, the decision tree becomes more shallow as a trial progresses and so the depth of fixations on each step relative to the maximal possible depth are shown on the right.

movements and model-based decision value. Inspection of eye-movements at decision states reveals that eye-movements are often biased toward the chosen branch of the decision tree. So, participants preferentially look into the branch they are about to visit. However, fixations on white cells are explained by motor planning. So, the relationship between eye-movements and choice is analysed for fixations on black cells, which likely accompanies reasoning about the likelihoods of finding the exit.

Participants may fixate black cells associated with a decision even before they arrive at a decision state, possibly reasoning about the upcoming choice. Figure 7.11 shows that participants look ahead into the decision sub-tree at both decision states ('D') and neutral states ('N') while approaching an observation or a decision. So, planning is only approximately linked to observations. However, fixation depth is easier to analyse if only decision states are considered ('D' cells described in Chapter 5), without integrating fixations on tree branch throughout the trial, although this limits the discriminatory power of the analysis.

In each decision state visited by participants we calculate the mean duration in milliseconds of fixating on black cells in each branch of the corresponding decision sub-tree. This duration is correlated with the model-based action value (Pearson $r = .36, p < .0001$) and with its empirical probability (Pearson $r = .24, p = .002.$). ANOVA of empirical prob-

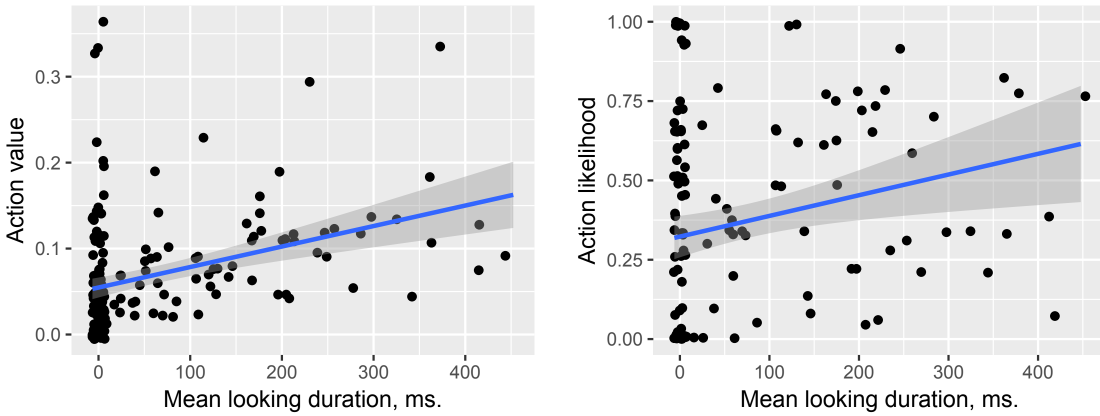


Figure 7.12: Left: Model-based action value plotted against the mean looking duration on black cells in the corresponding tree branch. Right: Action likelihood at decision states plotted against the mean looking duration on black cells in the corresponding tree branch. Bands represent 95% confidence intervals for means of the linear model.

ability against fixation duration on black and on white cells shows a significant effect of black ($F = 10.166, p = .002$), but not of white cells ($F = 3.172, p = .08$).

In agreement with Chapter 5, model-based action values predict the empirical probability of the actions, Pearson $r = .59, p < .0001$. The Pearson correlation of the softmax of action value ($\tau = .04$) and empirical probability is $r = .93, p < .0001$. So, participants fixate tree branches in proportion to the associated decision value. Actions, which are never taken are still fixated (Figure 7.12, right) and so fixation duration may be a better correlate of an action is being evaluated than the empirical probability of the action.

Next, we test for a link between planning accuracy and fixation depth measured in observations. When passing a decision point, a participant might look at several consecutive observation depths. Thus, we consider the maximal depth fixated on each step. Accurate planners might look deeper into the decision tree. In addition, they also may finish their planning early. So, fixations on black cells early in the trial might be associated with planning, wheres fixations on black cells later on might indicate decision retrieval.

Each visited decision state is associated with a **decision depth**, equal to the depth

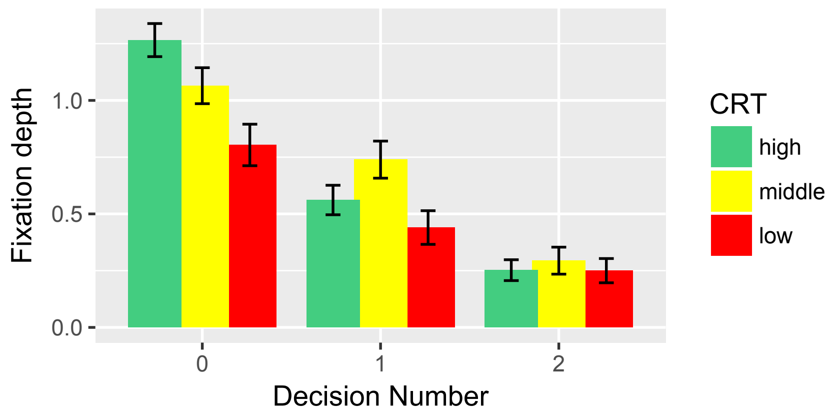


Figure 7.13: The mean fixation depth through the trial by participants with high, middle and low crt scores.

of the decision sub-tree rooted at the agent's current location. ANOVA of fixation depth against sub-tree depth and CRT shows a significant effect of CRT ($F = 8.3268, p = .0003$), of sub-tree depth ($F = 98.5822, p < .0001$), and a significant interaction between factors ($F = 3.8062, p = .004$). All participants fixate deeper at the start of the trial. However, participants with high CRT scores look deeper compared to the lower scoring participants at the first decision occurring in the trial (Figure 7.13). According to TukeyHSD, at the first decision the difference between fixation depth of high and low scoring participants is significant ($.46, p < .0001$).

In addition, ANOVA of $optstep_p$ against fixation depth and sub-tree depth shows a small effect of fixation depth ($F = 4.8254, p = .02$), but not of sub-tree depth and no interactions between factors. Spearman correlation between $optstep_p$ and fixation depth is $r = .09, p = .009$. ANOVA of τ_p against fixation depth and sub-tree depth shows a small effect of fixation depth ($F = 6.1525, p = .01$), but not of sub-tree depth and no there are interactions between factors. Spearman correlation between decision and fixation depth is $r = -.12, p < .0001$

In summary, participants with high CRT scores look deeper into the decision tree early in the trail. In addition, optimal planning is weakly correlated with fixation depth, suggesting that fixating deeper into the decision tree is associated with more accurate planning.

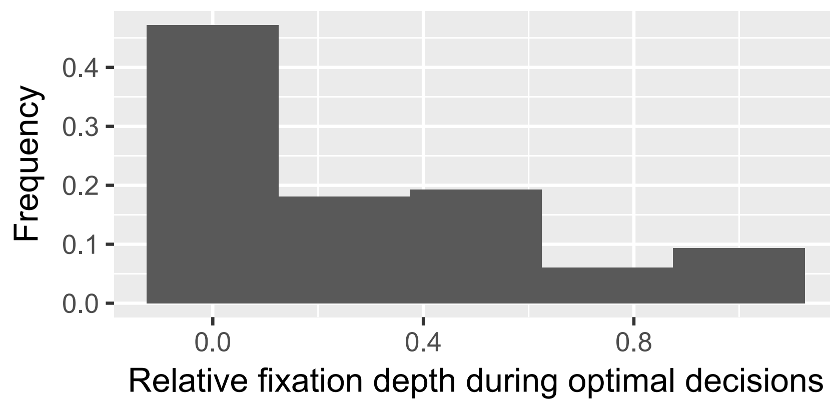


Figure 7.14: The relative depth of fixations during optimal decisions.

Importantly, participants may see deeper into the decision tree than they fixate because smaller rooms may be visible without fixating them directly [5]. Figure 7.14 shows the depth of fixations relative to the maximal possible fixation depth during optimal decisions. To make an optimal decision participants must have a plan equivalent to an optimal policy at the full depth of the decision sub-tree. Because participants rarely fixate at full depth, either the decision policy has been prepared in advance, or the participants can see deeper than they look.

By fitting the decision-tree model to paths we calculate the *planning depth* of each decision made by participants. Figure 7.15 (left) shows the planning depth of each decision plotted against its fixation depth. Planning and fixation depth are positively correlated, Spearman $r = .22, p < .0001$, so fixating deeper into the decision tree results in deeper planning. Figure 7.15 (left) shows that sometimes participants fixate deep into the decision tree, and make decisions that cannot be explained by a limited planning depth (decision at planning depth 0). Thus, limited planning depth is not the only reason why the POMDP model and participants decisions disagree. Alternative reward models, such as softmax decision noise or a variable discount rate may be useful in that regard.

However, fixations reflect planning depth only approximately. Fixation depth is mea-

[5]Recall that a cell subtends 1.7×1.3 degrees of visual angle, and so participants can see approximately the 8 cells surrounding the fixation quite clearly.

sured as the shortest path, measured as the number of observations to the fixated location. Planning depth is measured as the depth of the optimal planning sub-tree. So, maximal planning depth of a sub-tree is often greater than maximal possible fixation depth. In the example in Figure 7.10 maximal possible fixation depth at the root of the tree is two. However the depth of the planning tree is 4. Plotting the maximal possible fixation depth against actual measured fixation depth as shown in Figure 7.15 (right), shows that the depth of fixation is higher in deeper decision trees, Pearson $r = .37, p < .0001$. As discussed earlier, the actual fixation depth likely differs from the maximal possible fixation depth because participants likely see deeper into the decision tree than the depth of fixation. An extended model of fixation depth is needed to control for this effect. If, after controlling for the perceived fixation depth the disparity between planning depth and fixation depth remains, that would suggest that participants are rolling out a plan made several observations in advance. Alternatively, participants might re-plan their path within the smaller sub-tree every time they come upon a new observation. A participant might reason that since the complexity of the planning problem decreases after each observation, each solution regeneration improves their plan.

Discussion

The analysis of eye-movements during planning clarifies the earlier RT analysis. Eye-movements at simple observation states (labelled 'O' in Chapter 5) usually accompany *motor guidance*. In contrast, eye-movements at decision states (labelled 'D' in Chapter 5) include *decision planning* fixations that alternate between possible paths, as well as *motor guidance* fixations approaching the chosen direction. Eye-movements also illustrate that decision planning can be spread-out across several steps, rather than occurring exclusively in decision states.

Eye-movement analysis supports the earlier conclusion that participants sub-divide planning by observations. Reasoning about mazes in terms of observations is conveniently formalised by a decision-tree solver, which treats paths between observations as costs and observations as rewards. Inverting the decision-tree solver can be used to infer the depth of the optimal planning sub-tree starting at the agent's current location, thus estimating the

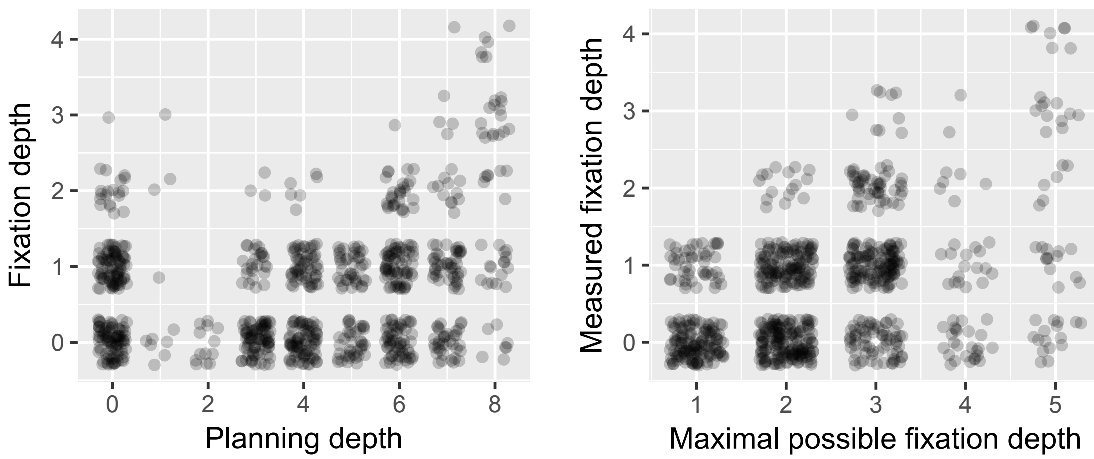


Figure 7.15: Left: Inferred planning depth against fixation depth. Participants usually plan deeper than they fixate. Right: Maximal possible depth of fixation in the current sub-tree against actual fixation depth. The dot locations are jittered for readability.

depth at which each decision can be explained by optimal planning. The depth of tree nodes fixated by *decision planning* eye-movements is positively correlated with planning depth inferred by the decision-tree model, suggesting that eye-movements are used to evaluate available decision options.

Moreover, the duration of looking into a tree branch is proportional to its objective decision value. So, eye-movements can be interpreted as sampling the value of each choice. Participants with higher CRT scores look further ahead and are better at directing their attention to the relevant aspects of the maze (the black cells and the cell where the agent is located), which may be explained by a general action-planning ability. In addition, the depth at which maze decision trees are fixated varies with CRT, so that participants with higher CRT scores look deeper into possible counter-factual outcomes of observations they might encounter. Analysing the planning depth by inference over the decision-tree solver reveals a positive relationship between looking deeper into the decision tree and more accurate planning. However, eye-movements measure planning depth only approximately. Decision-tree based inference shows that participants are usually able to plan deeper than they look and might provide a more accurate metric of planning depth.

Several other questions will be addressed in future research. What do good planners do better than poor ones? A larger working memory capacity might allow participants to evaluate more alternatives. Better planners may select options to evaluate more effectively. Moreover, quality of planning might depend on fixation sequence. For example, the order in which tree branches are fixated may contribute to primacy or recency effects. In addition, eye-movements during planning are accompanied by series of pupil dilations. The extent of pupil dilation may be associated with the sensitivity to decision value, and the baseline pupil size when approaching decisions might be predictive of the tendency to update one's plans or exploit a planning policy. If so, then baseline pupil size might predict planning depth according to decision-tree inference.

More accurate planning might depend on higher sensitivity to decision value sampled by eye-movements. If so, sampling decision value can be modelled by a drift-diffusion model, where each fixation contributes a fragment of evidence for a decision option. So, a higher sensitivity to decision value might entail a steeper drift or a lower noise in the drift-diffusion model. Moreover, if eye-movements sample decision value, then priming observers to fixate at particular locations by increasing its sensory salience can potentially influence decision quality.

7.4 The Intelligence Attribution Task

The intelligence attribution task follows the procedure described in Chapter 6. Half of the solutions were optimal, matching the optimality statistics typical of human solutions. Others showed examples of deviations from optimal policy.

7.4.1 Stimuli

Participants read the instructions on a computer screen and viewed 3 familiarisation examples followed by 26 test movies, 8 of which showed incomplete trajectories, stopping before the exit was found. Stimuli were generated by replaying solutions of participants in the planning task descried in Chapter 5.

The movies were labelled according to inverse planning inference as *softmax* or *optimal*. The complete examples were also labelled as *lucky* or *unlucky* according to the length of path taken by the agent. In summary, there were 6 movie conditions: *softmax-unlucky, softmax˙lucky, optimal˙unlucky, optimal˙lucky, softmax0, subpotimal-partial* and*optimal˙part*. Each of the conditions occurred 4 times. The *softmax0* condition occurred two times since such trajectories are infrequent among human solutions. Each movie was assigned a fraction of optimal steps in the movie, $optstep_i$ and the inferred decision noise τ_i. After viewing each movie, participants rated the intelligence of each solution by selecting a rating from a Likert scale between 1 (less intelligent) to 5 (more intelligent). Full instructions are listed in the Appendix.

7.4.2 Results

The results replicate the main findings described in Chapter 6, followed by a comparison of the model-based metrics of intelligence attribution to relative pupil size and the CRT score. For each participant we calculate metrics of individual sensitivity to efficiency and to luck. Sensitivity to efficiency is the difference between the participant's average ratings of optimal and suboptimal trials. Sensitivity to luck is the difference between participant's average rating of lucky and unlucky trials.

ANOVA of *rating* for $(participant, condition)$ shows a main effect of *participant* ($p < .0001, F = 3.6873$) and of condition ($p < .0001, F = 93.5942$). The adjusted R-squared of the regression model is $r^2 = .4336$. The mean ratings of each condition are shown in Figure 7.16). According to Tukey HSD the difference between the optimal lucky and unlucky conditions is not significant, $-0.3, p = .17$, however there is a significant difference between the ratings of lucky and unlucky *softmax* conditions $-0.56, p = .0002$. There is also a significant difference between ratings of incomplete conditions, with optimal incomplete conditions rated higher than suboptimal ones $-0.89, p < .0001$.

Participants' metrics of planning accuracy predict their sensitivity to efficiency. ANOVA

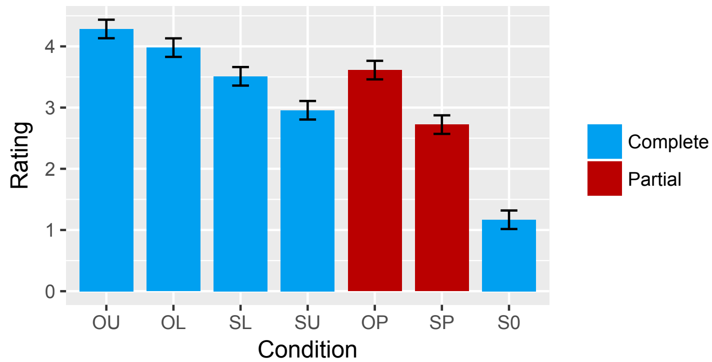


Figure 7.16: Comparing ratings between the two groups *OL-optimal-lucky*, *OU-optimal-unlucky*, *SL-softmax-lucky*, *SU-softmax-unlucky*, *S0-softmax0*, *OP-optimal-partial*, *SP-suboptimal-partial*. Error bars indicate Standard Error of the Mean.

of $optstep_p$ against sensitivity to efficiency and sensitivity to luck shows a marginally significant effect of sensitivity to efficiency ($F = 4.499, p = .043$) but not of sensitivity to luck ($F = 0.0331, p = .86$). Pearson correlation of sensitivity to efficiency and $optstep_p$ is $r = .37, p = .04$. In addition, regressing τ_p against sensitivity to luck is significant ($F = 6.1909p = .01$), adjusted R-squared .14. The Pearson correlation between sensitivity to luck and τ_p is $r = .41, p = .02$.

Analysing complete trials separately, ANOVA of *rating* for ($participant, condition, steps, revisits$) shows a main effect of participant ($p = .002, F = 1.9804$), of condition ($p < .0001, F = 137.0509$), steps ($p < .0001, F = 50.9197$) and an interaction between steps and revisits ($p = .01, F = 6.0016$). Earlier the effects of steps and revisits in rating were explained by the presence of the outcome group. This time only 5 participants self-reported as 'outcome', (4 of whom also self-reported as 'guessing'), which is insufficient to analyse the outcome and strategy groups separately. However, there may be differences in individual intelligence attributions based on the participant's CRT score.

Indeed, the ANOVA of *rating* against ($participant, CRT, condition$) shows main effects of participant, ($F = 3.9168, p < 0.0001$), of condition ($99.41958, p < 0.0001$) and a significant interaction between CRT and condition ($F = 4.8062, p < 0.0001$). The mean ratings are shown in Figure 7.16). According to Tukey HSD participants with low CRT

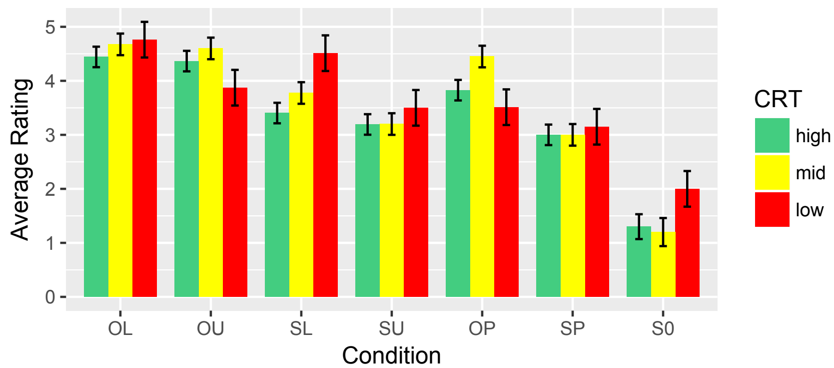


Figure 7.17: Comparing ratings between high, middle and low scoring participants *OL-optimal-lucky*, *OU-optimal-unlucky*, *SL-softmax-lucky*, *SU-softmax-unlucky*, *S0-softmax0*, *OP-optimal-partial*, *SP-suboptimal-partial*. Error bars indicate Standard Error of the Mean.

scores (those scoring 0 and 1) rated lucky examples higher. At the same time, there is no evidence that participants with high and middle CRT scores attribute intelligence to outcome. The difference between *optimal-lucky* and *optimal-unlucky* conditions is significant, ($d = 0.89, p = .01$). However, the difference is not significant for high ($d = 0.077, p = 1$) and middle scoring ($d = 0.075, p = 1$) participants. Likewise, difference between the *softmax-lucky* and *softmax-unlucky* conditions is significant for participants with low CRT scores ($d = 1.02, p = .01$), but not significant for high ($d = 0.2, p = 99$) and middle scoring ($d = 0.57, p = 46$) participants. So, only participants with low CRT scores attribute intelligence to outcome.

Participants with low CRT scores show no evidence of difference between ratings of optimal and suboptimal incomplete trials ($d = 0.36, p = .99$). At the same time, the difference is significant for middle ($d = 1.45, p < .0001$) and high-scoring participants ($d = 0.83, p = .002$). Therefore, there is no evidence that participants with low CRT scores attribute intelligence to efficiency. In contrast, high and middle scoring participants attribute intelligence to efficiency even if the outcome is not seen.

Testing whether participants' relative pupil size is linked to their attributions of intelligence, ANOVA of relative pupil size against sensitivity to efficiency, sensitivity to luck and fraction of optimal steps shows a significant effect of sensitivity to efficiency

($F = 5.2731, p = .03$) and of optimal step fraction ($F = 6.8725, p = .01$), but not of sensitivity to luck ($F = 0.9040, p = 0.4$). Pearson correlation of relative pupil size and sensitivity to efficiency is $r = .37, p = .046$. Thus, relative pupil size during mental effort predicts participants' attribution of intelligence to efficiency.

The results agree with the observation described in Chapter 6 that planning ability predicts attributed intelligence. The variation in attributed intelligence is also explained by independent metrics of cognitive performance, pupil size and CRT score. So, participants who have more cognitive resources at their disposal are more likely to attribute intelligence to efficient plans.

7.5 Discussion

The variations in planning ability can be explained by independent metrics of cognitive ability (CRT score and pupil size). So, participants' evaluations of plans made by others reflect a subjective judgement of cognitive ability, rather than a metric of task-specific planning. Available cognitive resources, efficient planning and attributing intelligence to efficiency go together. Participants who score highly on CRT are both more likely to plan efficiently and to attribute intelligence to efficiency, so that having more cognitive resources available for planning also correlates with skill at evaluating plans made by others.

Possibly, people engage a shared planning mechanism when attributing intelligence and when making plans. Such a mechanism must be able to work with arbitrary perspectives, affordances and beliefs. However, this does not mean that, as suggested by Simulation Theory, people automatically imagines themselves in the agent's place. Rather, learning to plan in a variety of domains likely leads to skill and knowledge transfer among unrelated planning activities. The extent to which planning ability is shared across domains needs clarification by future research.

Eye-movement analysis suggests that people likely represent maze problems as graphs consisting of observation nodes and paths between them, than use a step-based representation. Both step-based and observation-based models predict the average trajectories generated by participants. However, the decision-tree model also requires fewer computations and explains the differences in mental effort (RT) and eye-movements throughout the trial.

Planning frameworks such as these offer a variety of ways to formalise costs and rewards. The calculations in this thesis adopt a simple model of discounted utility, which discounts the rewards exponentially with distance. The downside of discounted utility models is that the values they produce depend on the discount rate. So, changing the discount rate affects the solutions predicted by the model. The discount rate used in this thesis was fitted by simulation of procedurally-generated grid-world mazes. It was validated in practice by comparing its predictions to solutions generated by human participants. Nevertheless, the discounted rate reward model might be more appropriate for reasoning about long-term investment rather a short path through a maze. Alternative ways to calculate rewards need to be considered in future work. The decision-tree model is particularly well-suited for reasoning about maze exploration as information gain, where the participant's goal is to minimise the entropy (e.g. Shannon entropy) over one's beliefs while minimising cost. Reward calculations could also be based on prospect theory [64] or might extend the One Observation model described in section 5.2.

Regardless of the reward function used by the model, comparing empirical action likelihoods to model-based action values shows that participants reason about rewards within a certain precision and scale. According to the softmax model of rewards, which fits well the empirical likelihoods of actions during the planning task, participants compare the values of actions locally, relative to other actions in the same state. So, empirical probability does not depend on values of actions in earlier states. For example, the softmax model treats equally deciding between one cent and two cents on one trial and between one dollar and two dollars on a subsequent trial. However, people should be more disposed to undertake risk when choosing between low rewards than when choosing between high rewards. The sensitivity to decision value might depend on the magnitudes and the outcomes of choices on previous trials. There must be better models of how decision values are scaled taking the absolute action value and the choice history into account.

Chapter 8

General Discussion

Intelligence is difficult to define explicitly. Yet people readily make attributions of intelligence and the lack of it - 'We know it when we see it'. But how do we go from seeing to knowing, from observing external behaviour to evaluating the reasoning process of someone else? The efficiency hypothesis explains intelligence attribution in the context of goal-directed agents whose distribution of rewards and mental states are known to the observer. It asserts that observers attribute intelligence to a capacity for reasoning about probabilities, as opposed to certain quantities. If so, hypothetical aliens encountering the NASA spacecraft could reason that its architects are intelligent, because sending a message requires the sender to consider the (minuscule) probability of encountering other sentient species in outer space.

Intelligence attributed to efficiency reflects the statistical expectation of outcome, on the assumption that the distribution of rewards in the environment is known and stable. However, in many real-life situations the observer's knowledge of the reward distribution is uncertain. Attributing intelligence to outcome, in contrast, requires only the general assumption that intelligent agents are more likely to achieve good outcomes. Such a heuristic, although imprecise, is informative in the absence of other inference. The outcome heuristic is often employed in marketing uninterpretable algorithms: 'We should trust the program because it produces reasonable classification performance' although may lead to prejudice and biases (billionaires are intelligent by definition).

Both attributing intelligence to outcome and poor planning, are correlated with poor performance on the Cognitive Reflection Test (CRT). This suggests that the 'outcome' participants may have difficulty reasoning about probabilities. [1] However, such participants could also attribute intelligence to a different cognitive ability. Participants whose numerical reasoning is weak might compensate by learning reward probabilities implicitly, and likewise by assuming that lucky agents have successfully learned the implicit distributions of rewards. The observation that participants who do a task in a certain way rate others highly if they seem to do the task in the same way resembles *homophily* - the phenomenon of preferring others with similar physical or acquired characteristics (e.g. social class, beliefs, education) [69, 81]. However, homophily based on cognitive style remains little researched.

The results described in Chapter 6 show that all participants exhibit some sensitivity to outcome and efficiency when evaluating others. On one hand, participants might compensate for the lack of confidence in their evaluations of efficiency by attributing intelligence to *a combination of efficiency and outcome*. On the other hand, attributing intelligence to both outcome and efficiency may reflect an implicit assumption that intelligent agents expect the environment to change. In a stable environment assuming that the reward distributions are constant is an advantage. However, in a changing environment, such as most real-world situations, intelligent agents continuously update their knowledge of how the rewards are distributed. [2] Thus, the variance in attributions of intelligence might reflect interpreting deviations from the optimal policy as uncertainty over the distribution of rewards.

Taking a broader perspective, goal-directed agents are only the tip of the iceberg of intelligent behaviour. The efficiency hypotheses assumes that intelligent behaviour must be intentional; an observer attributing intelligence infers the agent's goal, recognises that the agent approaches it in an efficient way, and evaluates the agent's planning procedure. However, agents may appear to behave intelligently while unaware of the goal. For example,

[1] A participant in the planning task: *To me this task is mostly a matter of luck, since you have no way of knowing where the exit is.* Upon reading that the exit location is not deterministic the participant decided that the task is entirely outside of his control.

[2] A participant's comment after solving all mazes optimally: *'I was so greedy it is embarrassing.'*

creative activity, such as composing music, requires intelligence, however often proceeds by gradual elaboration, without a well-defined end goal. In practice, people attribute intelligence to unintentional or exploratory agents, such as artists. Possibly, but not necessarily, observers might do so by attributing a specific goal to an exploratory agent even if the agent itself is not aware of it. But might observers also evaluate actions as intelligent without necessitating goal-based inference? More research is needed to test this possibility.

In addition, intelligence is often attributed to evidence of social awareness. For example, many people see octopuses as intelligent because octopuses seem to be aware of humans as social agents: they hide when watched, extend tentacles to make contact with divers and break equipment, apparently, to sabotage experiments [46]. This suggests that attributed intelligence could result from assuming that an octopus has a Theory of Mind (ToM), since not all social agents are seen as intelligent – ants are certainly not. The relationship between the degree of attributed intelligence and the agent's ability to apply ToM to social interactions remains to be explored.

There may be a practical evolutionary reason for humans to have an 'intelligence detector' to identify intelligent others that are worth modelling on. If people attribute high intelligence to optimal social models, then the agents seen as the most intelligent would be such that generate the optimal information gain for the observer. For example, a teenager might attribute more intelligence to a classmate than to a parent. People may see other members of the same profession as the most intelligent, and members of other professions as less so. In summary, 'intelligent' might be intuitively defined as 'better than me in something I want to do'. [3].

But if one needs an intelligence detector to identify who to model on to become intelligent, how does one start? Children prefer lucky agents [90] and those who can perform a sporting action on the first attempt [60], in effect favouring agents with good outcomes. However, physical fitness and luck are necessary for success independently of and in addition to intelligence. Simply modelling on agents who are fit and lucky would not produce efficient learning per se. A practical intelligence detector should thus reason backwards,

[3] Participant's comment: *'Generally when I saw that they did something that I would have done I rated them as 3. Whenever I saw them do something that I haven't thought of that was good, I gave them a 5'.*

rationalising behaviour that the observer can not spontaneously produce, but can nevertheless recognise as efficient. [4]

Lastly, the results of the planning task show that participants plan their actions when new observations are received, making plans several observations in advance. The POMDP-based model used in this thesis describes planning in a general case. However, given that participants limit their planning horizon, there may be simpler models that explain human behaviour. For example, A* type models [110] can minimise the path through a maze by going to a node with a minimal total path length plus expected path length to the goal. Comparison of reaction-time predictions made by different planning models remains a topic of future research. The choice of the planning model used to explain solutions does not alter the conclusion presented in this thesis that people plan several observations in advance and attribute intelligence to evidence of efficient planning.

In conclusion, common sense suggests that intelligence attribution is based on perceptual evaluation of behaviour according to which more complex behaviour or better outcomes seem more intelligent. However, the results presented in this thesis reveal a different story: attributed intelligence depends on reasoning about the mental states of other agents. People expect intelligent organisms to reason probabilistically about uncertain environments and attribute intelligence in proportion to the quality of probabilistic reasoning ascribed to the agent. This research takes a step toward understanding how people interpret intelligent behaviour in a variety of domains. The next chapter summarises the findings of this thesis and describes possible directions of future work.

[4] A participant in Chapter 7: *I was systematically visiting rooms left to right and top to bottom. Then I saw someone going into bigger room first, then into a smaller room and I thought, that's interesting, why would they do that? That was smart! Now I would like to go back and play again.*

Chapter 9

Conclusions and Future Work

The main contributions of this thesis are:

- An experimental method for evaluating planning and attributed intelligence in naturalistic environments;
- An empirical study of planning under uncertainty;
- A theory of attributed intelligence.

9.1 Summary

Chapter 4 measures empirical, common-sense attributions of intelligence to planning under uncertainty during exploratory search. The proposed *efficiency hypothesis* explains intelligence attributions made by the majority of participants as resulting from a subjective evaluation of the agent's planning. However, a minority of participants attributes intelligence to outcome, how quickly the agent completes a particular trial.

Thus, understanding how humans plan is essential to understanding how actions are interpreted as intelligent or not. Chapter 5 describes an experimental planning task to quantitatively study planning in natural environments. The proposed methodology uses non-verbal representations of costs (path length) and rewards (revealed cells) in a form that people can easily understand regardless of age and education. The decision times, decisions and eye-movements that occur during planning reflect the subjective value of the underlying planning decisions. Analysing planning behaviour as probabilistic Bayesian computations shows that on average participants produce optimal plans. At the same time, individuals choose non-optimal actions with likelihoods proportional to the expected utility of the choice.

Chapter 6 shows that individuals' attributions of intelligence depend on their planning ability. Competent planners attribute intelligence to efficiency, while less skilled ones attribute intelligence to outcome. The results suggest that planning practice leads to attending to efficiency when attributing intelligence, possibly as a result of improved planning skills. Chapter 7 shows that eye-movements correlate with the hidden process of estimating decision value. In addition, both planning ability and attributing intelligence to efficiency vary with metrics of cognitive ability. So, both planning and attributing intelligence probably rely on a shared cognitive resource, such that is measured by Cognitive Reflection Test and pupil dilation. Participants who have more cognitive resources at their disposal are more adept at attributing intelligence to efficiency and at making efficient plans and so attributing intelligence to outcome is due to limited cognitive resources.

9.2 Future work

Efficient modelling hypothesis of intelligence. An experiment that tests the *efficient modelling* hypothesis of attributed intelligence might involve three tasks. First, an incentivised planning task assesses participants' planning. Second, an intelligence attribution task assesses participants' evaluations of others. Finally, participants complete an incentivised planning task to measure learning. Incentivising the planning task motivates participants to learn from others during the intelligence attribution task. If the *efficient modelling* hypothesis is correct, we expect to see more intelligence attributed to efficiency under such circumstances, since individual outcomes are less relevant to learning planning strategies.

Intelligence attribution and planning in development. The intuitive experimental methodology described in this thesis is engaging, easy to learn and suitable for children and adults. So far, little is known about planning and attributing intelligence during development. Although children recognise and value competence [60], attributing intelligence to efficiency requires complex mental computations. In contrast, attributing intelligence to outcome may be available early. So, children might transition from judging by outcome to judging by efficiency as their Theory of Mind (ToM) and cognitive abilities develop. Young children might attribute intelligence differently than adults. For example, children might rate efficient agents as intelligent so far as the problems solved by the agents are useful for learning, so that the efficient modelling hypothesis, attributing intelligence to agents who are efficient models, might hold for children, but not for adults.

Implicit assumption about the environment. Giving participants no explicit instructions about the distribution of rewards or about the stability of the world, the planning task can be used to elicit implicit assumptions held by participants. When exploring the reward distribution participants are more likely to converge on the true distribution if it is close to their implicit assumptions. Participants might also be influenced by the outcome of the previous trial, a phenomenon commonly described as 'win stay, lose switch'. Assuming an unstable environment should result in a higher likelihood of switching after an unlucky outcome. To simplify the analysis, participants may be instructed that the experiment may have been designed by the good, the bad, or the random experimenter, priming three possible hypotheses about the distribution of rewards.

Shared resource. Furthermore, electrophysiological correlates of planning and of intelligence attribution may clarify the nature and time-course of the common resource recruited by the two tasks. If people attribute intelligence by comparing a reference solution to the agent's actions, then the functional neuroanatomy involved in planning should be recruited prior to observing the agent's decisions in intelligence attribution task. Observing unexpected actions that differ from the observer's predictions might correlate with higher functional activation [65]. So, attributed intelligence might correlate with two factors: (1) is the action expected? and (2) is the action optimal? The observer's prediction may be sub-optimal, but if observing an optimal action leads to learning, attributed intelligence should be high. Alternatively, participants might attribute high intelligence to expected actions , regardless of planning policy. Moreover, participants who attribute intelligence to outcome might recruit different functional networks (possibly associated with emotions rather than planning) compared to participants who attribute intelligence to efficiency.

Planning under uncertainty in abstract domains. The planning task developed in this thesis raises several questions of interest to future research. Planning under uncertainty is central to most natural behaviour, from hunting-and-gathering to managing investments. The rewards are usually sparse, the uncertainties are many, and to get by, organisms must maintain distant goals and reason in advance. Previous work on exploratory searching shows that mental and spatial search engages a shared cognitive mechanism: priming strategies of spatial foraging affects how humans subsequently search for words in memory [55]. This suggests that there may be a shared cognitive resource that handles planning under uncertainty as well, so as to enable knowledge transfer across domains. If all planning under uncertainty evokes an abstract mechanism of probabilistic reasoning, then priming a particular style of planning in one domain should influence how people subsequently make their choices in other domains. For example, priming attention to costs in the planning task should make people spend frugally while shopping. In contrast, instructing participants to reveal as much space as possible should make people spend more.

Models of value-based reasoning. The model of planning supported by the experiments in this thesis puts the evaluation of action values at the centre of planning. The softmax reward function produces a human-like mapping of action values to probabilities. However, softmax rewards ignore the magnitude of choices faced on earlier trials. In contrast, humans remember the choices they faced, retrieve the reasoning they used and reason about action values globally within a certain time-window. More recent values may be weighted more strongly, or weighted with respect to reference points, as suggested by prospect theory. This concern may be partly addressed by Drift-Diffusion (DDM) models, in which action values are fitted to to eye-movements, RT and electrophysiological measurements [29, 129]. Model-based metrics of value appraisal can be further compared with established metrics of impulsivity [142, 97], reward appraisal [99] and personality [40]. Reward appraisal is known to differ between individuals, e.g. diminished reward appraisal is common in depression [3, 108, 54]. The methodology described in this thesis can be used to measure value processing and characterise the mechanisms by which differences between individuals arise.

The common principle motivating all biological cognition is the need to make sense of uncertain evidence. Perception, from the brightness of sunlight to fleeting emotions of an acquaintance, arises from a process of inference combining expectations and sensory stimuli. Human behaviour may seem irrational at times, simply because most cognition is unconscious and because first impressions are often wrong. However, a careful application of computational modelling, psychophysics and behavioural experiments reveals a system precisely balancing computation costs and quality of inference while adapting to uncertain environments. Humans intuitively identifying efficiency as the basis of intelligence likely generalise this principle, that efficiency is to intelligence what transporting entropy across a boundary is to life.

References

[1] Frances Abell, Frances Happe, and Uta Frith. Do triangles play tricks? attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development*, 15(1):1–16, 2000.

[2] Balazs Aczel, Bence Palfi, and Zoltan Kekecs. What is stupid? people's conception of unintelligent behavior. *Intelligence*, 53:51–58, 2015.

[3] Nicholas B Allen, John Trinder, and Catherine Brennan. Affective startle modulation in clinical depression: preliminary findings. *Biological psychiatry*, 46(4):542–550, 1999.

[4] Daniel Ashlock, Colin Lee, and Cameron McGuinness. Search-based procedural generation of maze-like levels. *IEEE Transactions on Computational Intelligence and AI in Games*, 3(3):260–273, 2011.

[5] Chris L Baker, Julian Jara-Ettinger, Rebecca Saxe, and Joshua B Tenenbaum. Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, 1:0064, 2017.

[6] Chris L Baker, Rebecca R Saxe, and Joshua B Tenenbaum. Bayesian theory of mind: Modeling joint belief-desire attribution. In *Proceedings of the Thirty-Second Annual Conference of the Cognitive Science Society*, pages 2469–2474, 2011.

[7] Chris L Baker and Joshua B Tenenbaum. Modeling human plan recognition using bayesian theory of mind. *Plan, activity, and intent recognition: Theory and practice*, pages 177–204, 2014.

[8] Chris L Baker, Joshua B Tenenbaum, and Rebecca Saxe. Bayesian models of human action understanding. *Advances in neural information processing systems*, 18:99, 2006.

[9] Chris L Baker, Joshua B Tenenbaum, and Rebecca R Saxe. Goal inference as inverse planning. In *Proceedings of the 29th annual meeting of the cognitive science society*, 2007.

[10] Mahzarin R Banaji and Curtis D Hardin. Automatic stereotyping. *Psychological science*, 7(3):136–141, 1996.

[11] Marc G Bellemare, Sriram Srinivasan, Georg Ostrovski, Tom Schaul, David Saxton, and Remi Munos. Unifying count-based exploration and intrinsic motivation. In *Proceedings of Neural Information Processing 2016*, 2016.

[12] Tatiana Benaglia, Didier Chauveau, David Hunter, and Derek Young. mixtools: An r package for analyzing finite mixture models. *Journal of Statistical Software*, 32(6):1–29, 2009.

[13] Jeremy Bentham. The principles of morals and legislation. *Amherst, NY: Prometheus Books*, 1988.

[14] Rod Bond and Peter B Smith. Culture and conformity: A meta-analysis of studies using asch's (1952b, 1956) line judgment task. *Psychological bulletin*, 119(1):111, 1996.

[15] Ronen I Brafman and Moshe Tennenholtz. R-max-a general polynomial time algorithm for near-optimal reinforcement learning. *Journal of Machine Learning Research*, 3(Oct):213–231, 2002.

[16] David H Brainard. The psychophysics toolbox. *Spatial vision*, 10:433–436, 1997.

[17] Valentino Braitenberg. *Vehicles: Experiments in synthetic psychology.* MIT press, 1986.

[18] Martin Brüne and Ute Brüne-Cohrs. Theory of mind. evolution, ontogeny, brain mechanisms and psychopathology. *Neuroscience & Biobehavioral Reviews*, 30(4):437–455, 2006.

[19] Bingni W Brunton, Matthew M Botvinick, and Carlos D Brody. Rats and humans can optimally accumulate evidence for decision-making. *Science*, 340(6128):95–98, 2013.

[20] Richard W Byrne and Nadia Corp. Neocortex size predicts deception rate in primates. *Proceedings of the Royal Society B: Biological Sciences*, 271(1549):1693, 2004.

[21] Daryl Cameron, Cendri Hutcherson, Amanda Ferguson, Julian Scheffer, and Michael Inzlicht. Empathy is a choice: People are empathy misers because they are cognitive misers. *Rotman School of Management Working Paper No. 2887903*, 2016.

[22] John B Carroll. *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press, 1993.

[23] Jesse Chandler, Pam Mueller, and Gabriele Paolacci. Nonnaïveté among amazon mechanical turk workers: Consequences and solutions for behavioral researchers. *Behavior research methods*, 46(1):112–130, 2014.

[24] Charles Darwin. *The expression of the emotions in man and animals*. John Murray, London, 1872.

[25] Ishita Dasgupta, Eric Schulz, and Samuel J Gershman. Where do hypotheses come from? *Cognitive Psychology*, 96:1–25, 2017.

[26] Peter Dayan and Terrence J Sejnowski. Exploration bonuses and dual control. *Machine Learning*, 25(1):5–22, 1996.

[27] Benedetto De Martino, Stephen M Fleming, Neil Garrett, and Raymond J Dolan. Confidence in value-based choice. *Nature neuroscience*, 16(1):105–110, 2013.

[28] Daniel Clement Dennett. *The intentional stance*. MIT press, 1989.

[29] Jan Drugowitsch, Rubén Moreno-Bote, Anne K Churchland, Michael N Shadlen, and Alexandre Pouget. The cost of accumulating evidence in perceptual decision making. *Journal of Neuroscience*, 32(11):3612–3628, 2012.

[30] Phoebe C Ellsworth. Appraisal theory: Old and new questions. *Emotion Review*, 5(2):125–131, 2013.

[31] Julie Epelboim and Patrick Suppes. A model of eye movements and visual working memory during problem solving in geometry. *Vision research*, 41(12):1561–1574, 2001.

[32] Owain Evans, Andreas Stuhlmüller, and Noah D Goodman. Learning the preferences of ignorant, inconsistent agents. *arXiv preprint arXiv:1512.05832*, 2015.

[33] Philip M Fernbach, Adam Darlow, and Steven A Sloman. When good evidence goes bad: The weak evidence effect in judgment and decision-making. *Cognition*, 119(3):459–467, 2011.

[34] J Fisac, Chang Liu, Jessica B Hamrick, Shankar Sastry, J Karl Hedrick, Thomas L Griffiths, and Anca D Dragan. Generating plans that predict themselves. In *Workshop on the Algorithmic Foundations of Robotics (WAFR)*, 2016.

[35] Craig R Fox and Amos Tversky. A belief-based account of decision under uncertainty. *Management science*, 44(7):879–895, 1998.

[36] Elaine Fox, Victoria Lester, Riccardo Russo, RJ Bowles, Alessio Pichler, and Kevin Dutton. Facial expressions of emotion: Are angry faces detected more efficiently? *Cognition & emotion*, 14(1):61–92, 2000.

[37] James L Fozard, Max Vercruyssen, Sara L Reynolds, PA Hancock, and Reginald E Quilter. Age differences and changes in reaction time: the baltimore longitudinal study of aging. *Journal of gerontology*, 49(4):P179–P189, 1994.

[38] Robert H Frank. *Success and luck: Good fortune and the myth of meritocracy*. Princeton University Press, 2016.

[39] Shane Frederick. Cognitive reflection and decision making. *The Journal of Economic Perspectives*, 19(4):25–42, 2005.

[40] Adrian Furnham and Jeremy D Thorne. Need for cognition. *Journal of Individual Differences*, 2013.

[41] Vittorio Gallese, Christian Keysers, and Giacomo Rizzolatti. A unifying view of the basis of social cognition. *Trends in cognitive sciences*, 8(9):396–403, 2004.

[42] Howard Gardner. *Frames of mind: The theory of multiple intelligences*. Basic books, 2011.

[43] György Gergely and Gergely Csibra. Teleological reasoning in infancy: The naıve theory of rational action. *Trends in cognitive sciences*, 7(7):287–292, 2003.

[44] Samuel J Gershman, Eric J Horvitz, and Joshua B Tenenbaum. Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, 349(6245):273–278, 2015.

[45] Thomas Gilovich. *How we know what isn't so*. Simon and Schuster, 2008.

[46] Peter Godfrey-Smith. *Other Minds: The Octopus and the Evolution of Intelligent Life*. HarperCollins UK, 2017.

[47] Alvin I Goldman and Chandra Sekhar Sripada. Simulationist models of face-based emotion recognition. *Cognition*, 94(3):193–213, 2005.

[48] H Paul Grice, Peter Cole, Jerry Morgan, et al. Logic and conversation. *1975*, pages 41–58, 1975.

[49] C Hadjichristidis, JM Stibel, SA Sloman, DE Over, and RJ Stevenson. Opening pandoras box: Selective unpacking and superadditivity. In *Proceedings of the European society for the study of cognitive systems 16th annual workshop*, 1999.

[50] Joshua D Hall, Anna B OConnell, and Jeanette G Cook. Predictors of student productivity in biomedical graduate school applications. *PloS One*, 12(1):e0169121, 2017.

[51] Milos Hauskrecht. Value-function approximations for partially observable markov decision processes. *Journal of Artificial Intelligence Research*, pages 33–94, 2000.

[52] Mary Hayhoe and Dana Ballard. Eye movements in natural behavior. *Trends in cognitive sciences*, 9(4):188–194, 2005.

[53] Fritz Heider and Marianne Simmel. An experimental study of apparent behavior. *The American Journal of Psychology*, pages 243–259, 1944.

[54] Jeffrey B Henriques and Richard J Davidson. Decreased responsiveness to reward in depression. *Cognition & Emotion*, 14(5):711–724, 2000.

[55] Thomas T Hills, Peter M Todd, and Robert L Goldstone. Search in external and internal spaces: Evidence for generalized cognitive search processes. *Psychological Science*, 19(8):802–808, 2008.

[56] Ming Hsu, Meghana Bhatt, Ralph Adolphs, Daniel Tranel, and Colin F Camerer. Neural systems responding to degrees of uncertainty in human decision-making. *Science*, 310(5754):1680–1683, 2005.

[57] David Hume. *A treatise of human nature.* Clarendon Press, Oxford, 2077.

[58] Julian Jara-Ettinger, Chris L Baker, and Joshua B Tenenbaum. Learning what is where from social observations. In *Proceedings of the Thirty-Fourth Annual Conference of the Cognitive Science Society*, pages 515–520, 2012.

[59] Julian Jara-Ettinger, Hyowon Gweon, Laura E Schulz, and Joshua B Tenenbaum. The naïve utility calculus: computational principles underlying commonsense psychology. *Trends in cognitive sciences*, 20(8):589–604, 2016.

[60] Julian Jara-Ettinger, Joshua B Tenenbaum, and Laura E Schulz. Not so innocent. toddlers? inferences about costs and culpability. *Psychological science*, 26(5):633–640, 2015.

[61] Marieke Jepma and Sander Nieuwenhuis. Pupil diameter predicts changes in the exploration–exploitation trade-off: Evidence for the adaptive gain theory. *Journal of cognitive neuroscience*, 23(7):1587–1596, 2011.

[62] Leslie Pack Kaelbling, Michael L Littman, and Anthony R Cassandra. Planning and acting in partially observable stochastic domains. *Artificial intelligence*, 101(1):99–134, 1998.

[63] Daniel Kahneman. *Attention and effort*, volume 1063. Prentice-Hall Englewood Cliffs, NJ, 1973.

[64] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the econometric society*, pages 263–291, 1979.

[65] Jorie Koster-Hale and Rebecca Saxe. Theory of mind: a neural prediction problem. *Neuron*, 79(5):836–848, 2013.

[66] Eileen Kowler, Cordelia D Aitkin, Nicholas M Ross, Elio M Santos, and Min Zhao. Davida teller award lecture 2013: the importance of prediction and anticipation in the control of smooth pursuit eye movements. *Journal of vision*, 14(5):10–10, 2014.

[67] Viktoriya Krakovna and Finale Doshi-Velez. Increasing the interpretability of recurrent neural networks using hidden markov models. *arXiv preprint arXiv:1606.05320*, 2016.

[68] Marta Kryven, Tomer Ullman, William Cowan, and Joshua B Tenenbaum. Outcome or strategy? a bayesian model of intelligence attribution. In *Proceedings of the Thirty-Eighths Annual Conference of the Cognitive Science Society*, 2016.

[69] Paul F Lazarsfeld, Robert K Merton, et al. Friendship as a social process: A substantive and methodological analysis. *Freedom and control in modern society*, 18(1):18–66, 1954.

[70] Maël Lebreton, Raphaëlle Abitbol, Jean Daunizeau, and Mathias Pessiglione. Automatic integration of confidence in the brain valuation signal. *Nature neuroscience*, 18(8):1159–1167, 2015.

[71] Melvin Lerner. *The Belief in a Just World; a Fundamental Delusion*. New york: Plenum Press, 1980.

[72] Robert W Levenson. Basic emotion questions. *Emotion Review*, 3(4):379–386, 2011.

[73] Falk Lieder, Ming Hsu, and Thomas L Griffiths. The high availability of extreme events serves resource-rational decision-making. In *Proceedings of the Cognitive Science Society*, volume 36, 2014.

[74] Falk Lieder, Dillon Plunkett, Jessica B Hamrick, Stuart J Russell, Nicholas Hay, and Thomas Griffiths. Algorithm selection by rational metareasoning as a model of human strategy selection. In *Advances in neural information processing systems*, pages 2870–2878, 2014.

[75] R Duncan Luce. *Response times: Their role in inferring elementary mental organization*. Oxford University Press on Demand, 1986.

[76] Elliot A Ludvig, Christopher R Madan, and Marcia L Spetch. Extreme outcomes sway risky decisions from experience. *Journal of Behavioral Decision Making*, 27(2):146–156, 2014.

[77] Douglas B Markant and Todd M Gureckis. A preference for the unpredictable over the informative during self-directed learning. In *Proceedings of the 36th annual conference of the Cognitive Science Society*, pages 958–963, 2014.

[78] Douglas B Markant, Burr Settles, and Todd M Gureckis. Self-directed learning favors local, rather than global, uncertainty. *Cognitive science*, 40(1):100–120, 2016.

[79] Winter Mason and Siddharth Suri. Conducting behavioral research on amazons mechanical turk. *Behavior research methods*, 44(1):1–23, 2012.

[80] Kevin S McGrew. The cattell-horn-carroll theory of cognitive abilities: Past, present, and future. *D. P. Flanagan, J. L. Genshaf, P. L. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests*, page 136182, 2005.

[81] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.

[82] Douglas L Medin, Salino G García, et al. Grounding principles for inferring agency: Two cultural perspectives. *Cognitive Psychology*, 95:50–78, 2017.

[83] Albert Michotte. *The perception of causality.* Basic Books, 1963.

[84] Liane Moneta-Koehler, Abigail M Brown, Kimberly A Petrie, Brent J Evans, and Roger Chalkley. The limitations of the gre in predicting success in biomedical graduate school. *PloS one*, 12(1):e0166742, 2017.

[85] Jiri Najemnik and Wilson S Geisler. Optimal eye movement strategies in visual search. *Nature*, 434(7031):387–391, 2005.

[86] Jonathan D Nelson, Craig RM McKenzie, Garrison W Cottrell, and Terrence J Sejnowski. Experience matters information acquisition optimizes probability gain. *Psychological science*, 21(7):960–969, 2010.

[87] Jv Neumann, Oskar Morgenstern, and HW Kuhn. *Theory of games and economic behavior*. Princeton University Press, Commemorative Ed edition (1 May 2007), 1953.

[88] Richard E Nisbett and Timothy D Wilson. Telling more than we can know: Verbal reports on mental processes. *Psychological review*, 84(3):231, 1977.

[89] Man-Suk Oh and James O Berger. Adaptive importance sampling in monte carlo integration. *Journal of Statistical Computation and Simulation*, 41(3-4):143–168, 1992.

[90] Kristina R Olson, Mahzarin R Banaji, Carol S Dweck, and Elizabeth S Spelke. Children's biased evaluations of lucky versus unlucky people and their social groups. *Psychological Science*, 17(10):845–846, 2006.

[91] Kristina R Olson, Yarrow Dunham, Carol S Dweck, Elizabeth S Spelke, and Mahzarin R Banaji. Judgments of the lucky across development and culture. *Journal of Personality and Social Psychology*, 94(5):757, 2008.

[92] Kristina R Olson, Andrea S Heberlein, Elizabeth Kensinger, Christopher Burrows, Carol S Dweck, Elizabeth S Spelke, and Mahzarin R Banaji. The role of forgetting in undermining good intentions. *PLoS ONE*, 2013.

[93] Andrew Ortony, Gerald L Clore, and Allan Collins. *The cognitive structure of emotions*. Cambridge university press, 1990.

[94] John Palmer, Alexander C Huk, and Michael N Shadlen. The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of vision*, 5(5):1–1, 2005.

[95] Peter C Pantelis, Chris L Baker, Steven A Cholewiak, Kevin Sanik, Ari Weinstein, Chia-Chien Wu, Joshua B Tenenbaum, and Jacob Feldman. Inferring the intentional states of autonomous virtual agents. *Cognition*, 130(3):360–379, 2014.

[96] Peter C Pantelis, Timothy Gerstner, Kevin Sanik, Ari Weinstein, Steven A Cholewiak, Gaurav Kharkwal, Chia-Chien Wu, and Jacob Feldman. Agency and rationality: Adopting the intentional stance toward evolved virtual agents. *Decision*, 3(1):40, 2016.

[97] Jim H Patton, Matthew S Stanford, et al. Factor structure of the barratt impulsiveness scale. *Journal of clinical psychology*, 51(6):768–774, 1995.

[98] Marius V Peelen, Anthony P Atkinson, and Patrik Vuilleumier. Supramodal representations of perceived emotions in the human brain. *Journal of Neuroscience*, 30(30):10127–10134, 2010.

[99] Diego A Pizzagalli, Dan Iosifescu, Lindsay A Hallett, Kyle G Ratner, and Maurizio Fava. Reduced hedonic capacity in major depressive disorder: evidence from a probabilistic reward task. *Journal of psychiatric research*, 43(1):76–87, 2008.

[100] Jonathan Posner, James A Russell, Andrew Gerber, Daniel Gorman, Tiziano Colibazzi, Shan Yu, Zhishun Wang, Alayar Kangarlu, Hongtu Zhu, and Bradley S Peterson. The neurophysiological bases of emotion: An fmri study of the affective circumplex using emotion-denoting words. *Human brain mapping*, 30(3):883–895, 2009.

[101] Alexandre Pouget, Jan Drugowitsch, and Adam Kepecs. Confidence and certainty: distinct probabilistic quantities for different goals. *Nature neuroscience*, 19(3):366, 2016.

[102] David Premack and Guy Woodruff. Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, 1(04):515–526, 1978.

[103] Gabriel A Radvansky and Jeffrey M Zacks. *Event cognition*. Oxford University Press, 2014.

[104] Raghu Raj, Wilson S Geisler, Alan C Bovik, and Robert A Frazor. Contrast statistics for foveated visual systems: Fixation selection by minimizing contrast entropy. *JOSA A*, 22(10):2039–2049, 2005.

[105] Laura Walker Renninger, Preeti Verghese, and James Coughlan. Where to look next? eye movements reduce local uncertainty. *Journal of Vision*, 7(3):6–6, 2007.

[106] Bruce Rind and Prashant Bordia. Effect on restaurant tipping of male and female servers drawing a happy, smiling face on the backs of customers' checks. *Journal of Applied Social Psychology*, 26(3):218–225, 1996.

[107] Philippe Rochat, Rachel Morgan, and Malinda Carpenter. Young infants' sensitivity to movement information specifying social causality. *Cognitive development*, 12(4):537–561, 1997.

[108] Jonathan Rottenberg, Karen L Kasch, James J Gross, and Ian H Gotlib. Sadness and amusement reactivity differentially predict concurrent and prospective functioning in major depressive disorder. *Emotion*, 2(2):135, 2002.

[109] Sverker Runeson, Peter Juslin, and Henrik Olsson. Visual perception of dynamic properties: Cue heuristics versus direct-perceptual competence. *Psychological review*, 107(3):525, 2000.

[110] Stuart Russell, Peter Norvig, and Artificial Intelligence. Artificial inteligence, a modern approach. *Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs*, 25:27, 1995.

[111] Paul A Samuelson. A note on the pure theory of consumer's behaviour. *Economica*, 5(17):61–71, 1938.

[112] Rebecca Saxe. Against simulation: the argument from error. *Trends in cognitive sciences*, 9(4):174–179, 2005.

[113] Rebecca Saxe and Susan Carey. The perception of causality in infancy. *Acta psychologica*, 123(1):144–165, 2006.

[114] Rebecca Saxe, Susan Carey, and Nancy Kanwisher. Understanding other minds: linking developmental psychology and functional neuroimaging. *Annu. Rev. Psychol.*, 55:87–124, 2004.

[115] Klaus R Scherer and Ben Meuleman. Human emotion experiences can be predicted on theoretical grounds: evidence from verbal labeling. *PloS one*, 8(3):e58166, 2013.

[116] Brian J Scholl and Patrice D Tremoulet. Perceptual causality and animacy. *Trends in cognitive sciences*, 4(8):299–309, 2000.

[117] Eric Schwitzgebel and Fiery Cushman. Expertise in moral reasoning? order effects on moral judgment in professional philosophers and non-philosophers. *Mind & Language*, 27(2):135–153, 2012.

[118] Philippe G Schyns and Aude Oliva. Dr. angry and mr. smile: When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, 69(3):243–265, 1999.

[119] Mark Van Selst and Pierre Jolicoeur. A solution to the effect of sample size on outlier elimination. *The quarterly journal of experimental psychology*, 47(3):631–650, 1994.

[120] N Siddharth, Brooks Paige, Alban Desmaison, Van de Meent, Frank Wood, Noah D Goodman, Pushmeet Kohli, Philip HS Torr, et al. Inducing interpretable representations with variational autoencoders. *arXiv preprint arXiv:1611.07492*, 2016.

[121] Amy E Skerry and Rebecca Saxe. A common neural code for perceived and inferred emotion. *Journal of Neuroscience*, 34(48):15997–16008, 2014.

[122] Amy E Skerry and Rebecca Saxe. Neural representations of emotion are organized around abstract event features. *Current biology*, 25(15):1945–1954, 2015.

[123] Steven Sloman, Yuval Rottenstreich, Edward Wisniewski, Constantinos Hadjichristidis, and Craig R Fox. Typical versus atypical unpacking and superadditive probability judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(3):573, 2004.

[124] Beate Sodian. The development of deception in young children. *British Journal of Developmental Psychology*, 9(1):173–188, 1991.

[125] Beate Sodian and Uta Frith. Deception and sabotage in autistic, retarded and normal children. *Journal of child psychology and psychiatry*, 33(3):591–605, 1992.

[126] Nicole K Speer, Khena M Swallow, and Jeffery M Zacks. Activation of human motion processing areas during event perception. *Cognitive, Affective, & Behavioral Neuroscience*, 3(4):335–345, 2003.

[127] ES Spelke, AT Phillips, AL Woodward, D Sperber, D Premack, and A Premack. Infants' knowledge of object motion and human action. *Causal cognition: A multidisciplinary debate*, 1995.

[128] Nicolette Sullivan, Cendri Hutcherson, Alison Harris, and Antonio Rangel. Dietary self-control is related to the speed with which attributes of healthfulness and tastiness are processed. *Psychological science*, 26(2):122–134, 2015.

[129] Satohiro Tajima, Jan Drugowitsch, and Alexandre Pouget. Optimal policy for value-based decision-making. *Nature communications*, 7, 2016.

[130] Hui Fen Tan, Giles Hooker, and Martin T Wells. Tree space prototypes: Another look at making tree ensembles interpretable. *arXiv preprint arXiv:1611.07115*, 2016.

[131] Maggie E Toplak, Richard F West, and Keith E Stanovich. The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory & cognition*, 39(7):1275, 2011.

[132] Patrice D Tremoulet and Jacob Feldman. Perception of animacy from the motion of a single object. *PERCEPTION-LONDON-*, 29(8):943–952, 2000.

[133] Konstantinos Tsetsos, Rani Moran, James Moreland, Nick Chater, Marius Usher, and Christopher Summerfield. Economic irrationality is optimal during noisy decision making. *Proceedings of the National Academy of Sciences*, 113(11):3102–3107, 2016.

[134] Jason S Tsukahara, Tyler L Harrison, and Randall W Engle. The relationship between baseline pupil size and intelligence. *Cognitive psychology*, 91:109–123, 2016.

[135] Amos Tversky and Daniel Kahneman. Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2):207–232, 1973.

[136] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. *science*, 185(4157):1124–1131, 1974.

[137] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. In *Utility, probability, and human decision making*, pages 141–162. Springer, 1975.

[138] Tomer Ullman, Chris Baker, Owen Macindoe, Owain Evans, Noah Goodman, and Joshua B Tenenbaum. Help or hinder: Bayesian models of social goal inference. In *Advances in neural information processing systems*, pages 1874–1882, 2009.

[139] Kara Weisman, Carol S Dweck, and Ellen M Markman. Varieties of experience: A new look at folk philosophy of mind. In *Proceedings of the Thirty-Eighths Annual Conference of the Cognitive Science Society*, 2016.

[140] Matthew Welsh, Nicholas Burns, and Paul H Delfabbro. The cognitive reflection test: how much more than numerical ability? In *CogSci*, 2013.

[141] Robert Whelan. Effective analysis of reaction time data. *The Psychological Record*, 58(3):475, 2008.

[142] Stephen P Whiteside, Donald R Lynam, Joshua D Miller, and Sarah K Reynolds. Validation of the upps impulsive behaviour scale: a four-factor model of impulsivity. *European Journal of Personality*, 19(7):559–574, 2005.

[143] Paul D Windschitl and John R Chambers. The dud-alternative effect in likelihood judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(1):198, 2004.

[144] Piotr Winkielman, Kent C Berridge, and Julia L Wilbarger. Unconscious affective reactions to masked happy versus angry faces influence consumption behavior and judgments of value. *Personality and Social Psychology Bulletin*, 31(1):121–135, 2005.

[145] Scott Wisdom, Thomas Powers, James Pitton, and Les Atlas. Interpretable recurrent neural networks using sequential sparse recovery. *arXiv preprint arXiv:1611.07252*, 2016.

[146] Guy Woodruff and David Premack. Intentional communication in the chimpanzee: The development of deception. *Cognition*, 7(4):333–362, 1979.

[147] Jeffrey M Zacks, Nicole K Speer, Khena M Swallow, Todd S Braver, and Jeremy R Reynolds. Event perception: a mind-brain perspective. *Psychological bulletin*, 133(2):273, 2007.

[148] Jamil Zaki, Seth Kallman, G Elliott Wimmer, Kevin Ochsner, and Daphna Shohamy. Social cognition as reinforcement learning: Feedback modulates emotion inference. *Journal of cognitive neuroscience*, 2016.

[149] Min Zhao and Andre G Marquez. *Understanding Humans' Strategies in Maze Solving.* Doctoral dissertation, Psychology Department, Rutgers University, 2013.

APPENDICES

Instructions, Chapter 5, Escape a maze

The following instructions were displayed in a computer monitor using a browser. The experiment used our own interface, with code written in HTML 5, JavaScript and PHP.

Welcome to our study!
To participate in this study you need a desktop computer or a laptop, not a mobile device.
In this study you will move in a maze and try to find the exit in fewest moves possible.
There are 3 practice mazes and 12 test mazes.
The study takes approximately 8 minutes.
The top 20 % of participants who find the exit in fewest moves receive a 10 % bonus!
Thanks for participating!

Participants in the No Bonus condition saw the same set of instructions, but without the line about the bonus.

Instructions, Chapter 4, Experiment 1

The following instructions were displayed on a computer monitor before starting the experiment using Psychtoolbox [16].

Thank you for participating in our experiment!
You will see a mouse in a maze. The mouse is hungry. The mouse is familiar with the space, but does not know where the food is and wants to get to the food.
The food can be anywhere.
Rate how intelligent is the mouse on the scale 5 (more intelligent) .... 1 (less intelligent).
Please write your rating in the provided answer sheet.
We will start with 4 familiarization examples.

After completing the 4 practice trials the participant saw a message:

If you have questions about the experiment please ask them now.
Otherwise press any key to continue.

The experimenter was in an adjacent room to answer questions at that may arise this point. Most participants did not have questions.

Instructions, Chapter 4, Experiment 2

The following welcome instructions were displayed in a computer monitor using a browser via an interface created in *Survey Gizmo TM.*

Welcome to our study!
In this study you will see 40 videos of a mouse trying to get to a target.
For each movie, you will rate the intelligence of the mouse.
The study takes approximately 18 minutes.
Thanks for participating!

The welcome instructions were followed by an informed consent agreement and a form collecting age and gender. After accepting the informed consent, participants saw task-specific instructions:

Instructions, please read carefully! In this study you will see a many hungry mice looking for food in a maze. The food can be anywhere in the maze. Each movie shows a different mouse. Each mouse is familiar with the maze it is running through, and knows where the walls are. In the beginning, a mouse DOES NOT KNOW WHERE THE FOOD IS. Areas not yet seen by a mouse are initially dark, and become uncovered as the mouse goes along. You will rate how intelligent each mouse is, on a scale of 1 (less intelligent) to 5 (more intelligent). Each movie takes approximately 10 seconds. It is important to watch each movie at least once before submitting a rating. On the next page are 4 example movies as a warm-up, after that each movie will be displayed on a separate page. When you are ready, please continue to the next page.

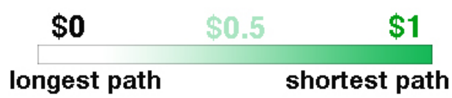
Instructions, Chapter 6, Experiment 1

The following instructions were displayed in a computer monitor using a browser. The experiment used our own interface, with code written in HTML 5, JavaScript and PHP.

Welcome to our study!
To participate in this study you need a desktop computer or a laptop, not a mobile device.
This study has two parts.
In the first part you will look for an exit in a maze.
In the second part you will view and evaluate the solution of other participants.
The study takes approximately 20 minutes.
Thanks for participating!

The welcome instructions were followed by an informed consent agreement and a form collecting age and gender. After accepting the informed consent, participants saw instructions specific to the maze-solving task:

PLEASE READ THE FOLLOWING INSTRUCTIONS CAREFULLY.
Your task is to exit the maze by reaching the red square in as few steps as possible.
You can move one square at a time,
by clicking on the white squares near your character.
Blue squares are walls. You cannot see through the walls,
so the squares you cannot see yet are black.
The exit is equally like to be hidden behind any of the black squares.
In the end of the experiment we will add all
steps you took and show you how you did compared to previous results.
You have an opportunity to earn a bonus for completing the mazes in fewer steps.



After completing practice examples participants answered instructions quiz.

Great, you have finished practice!
Please answer the instructions quiz below to proceed with the experiment.

	Question 1: My task is to ..
o	visit every square in the maze.
o	finish in as little time as possible.
o	solve the mazes in as few steps as possible.
o	click as fast as possible.
	Question 2: My bonus will be bigger if I ...
o	finish the mazes in less time.
o	finish the mazes in fewer steps.
o	am lucky at guessing.
o	The bonus will be given at random.

Additionally, a short reminder message was displayed during the experiment on the top of the screen.

Find the exit in as few steps as possible. The exit is equally like to be hidden behind any of the black squares.

After finishing the maze-solving task, participants were asked to provide free form answers:

Thank you for completing the first half of the experiment! Please answer the question below to move on to the second half. How did you make your decisions about which way to go? _______________________________- This is optional, but we would love to hear any comments that you may have. _______________________________-

Before beginning the intelligence attribution task participants read the following instructions:

PLEASE READ THE FOLLOWING INSTRUCTIONS CAREFULLY.
Your task is to evaluate another person's solution on the scale
1 (less intelligent) to 5 (more intelligent).
The person wants to get to the exit in fewest moves and knows
that the exit is equally likely to be behind any of the black squares.

In some videos you will see only a part of the path.

Please take a moment to review each maze before viewing the solution,
then press ‘ ¿¿ Play/Pause’ to start the video.
Each video shows a different person.
It is important to view each video at least once.

Instructions, Chapter 7, Planning Task

Welcome to our study!
IT IS IMPORTANT TO READ THE INSTRUCTIONS CAREFULLY
Part 1: Decision-making
Escape the maze in AS FEW MOVES AS POSSIBLE by reaching the exit (the red square).
Blue squares are walls. You cannot see through the walls. The squares you cannot see yet are black.
The exit is EQUALLY LIKELY to be behind ANY of the black squares.
In the end you will see how many steps you took compared to other people.
There are 5 practice mazes and 12 test mazes.
Press SPACE to start, then use keyboard arrow keys to move.

Instructions, Chapter 7, Intelligence Attribution Task

Part 2: Judging Others

You will see 26 solutions of other people to puzzles like the ones you just solved.

Each video shows a different person.

Please rate each solution between 1 (less intelligent) and 5 (more intelligent) using the mouse.

Some videos will show only a part of the path.

Take a moment to study each maze, then press SPACE to view the solution.

(Press SPACE to start...)