ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Perspectives on Psychological Science
1–14

DOI: 10.1177/17456916231196766
www.psychologicalscience.org/PPS
Sage

Building Human-Like Artificial Agents: A General Cognitive Algorithm for Emulating Human Decision-Making in Dynamic Environments

Cleotilde Gonzalez

Dynamic Decision Making Laboratory, Social and Decision Sciences Department, Carnegie Mellon University

Abstract

One of the early goals of artificial intelligence (AI) was to create algorithms that exhibited behavior indistinguishable from human behavior (i.e., human-like behavior). Today, AI has diverged, often aiming to excel in tasks inspired by human capabilities and outperform humans, rather than replicating human cogntion and action. In this paper, I explore the overarching question of whether computational algorithms have achieved this initial goal of AI. I focus on dynamic decision-making, approaching the question from the perspective of computational cognitive science. I present a general cognitive algorithm that intends to emulate human decision-making in dynamic environments, as defined in instance-based learning theory (IBLT). I use the cognitive steps proposed in IBLT to organize and discuss current evidence that supports some of the human-likeness of the decision-making mechanisms. I also highlight the significant gaps in research that are required to improve current models and to create higher fidelity in computational algorithms to represent human decision processes. I conclude with concrete steps toward advancing the construction of algorithms that exhibit human-like behavior with the ultimate goal of supporting human dynamic decision-making.

Keywords

AI algorithm, decisions from experience, choice, memory, learning, instance-based learning theory

One of the initial goals of artificial intelligence (AI) was to build algorithms[1] capable of replicating human behavior (i.e., *human-like behavior*, as outlined by Lake et al., 2017; Simon, 1983; Turing, 1950). However, in contemporary AI, the focus often shifts away from replicating human behavior and deepening our understanding of cognition. Instead, the emphasis lies in using human abilities as an inspiration to develop algorithms that can perform, predict and classify events more accurately and rapidly than humans can (i.e., *unlike* human behavior), thus, diverging from the goal of advancing our understanding of cognition.

Although the contemporary AI approach will continue to be very valuable, I argue that in unstable and complex real-world decision-making situations (i.e., *dynamic decision-making*, or DDM), we need to focus on building learning algorithms that simulate human cognitive processes so that we can support humans in making decisions in real-world tasks (Gonzalez et al., 2003, 2017). Here, I aim to discuss (a) the ways in which current computational decision-making algorithms may make human-like decisions and (b) how current computational models may be lacking precision and development of human cognition. Many aspects of human decision-making processes in dynamic environments are unknown or imprecise, particularly as they relate to DDM (Gonzalez et al., 2017). In this paper I reflect on some of the unknowns of DDM and on some of the AI achievements with regard to building agents that exhibit human-like decision-making in dynamic tasks.

I focus on a cognitive-science perspective of DDM, in which human decision-making in dynamic environments is conceived as a closed-loop learning process and information processing (Gonzalez, 2017),[2] and I highlight some of the gaps that researchers need to

Corresponding Author:
Cleotilde Gonzalez, Dynamic Decision Making Laboratory, Social and Decision Sciences Department, Carnegie Mellon University
Email: coty@cmu.edu

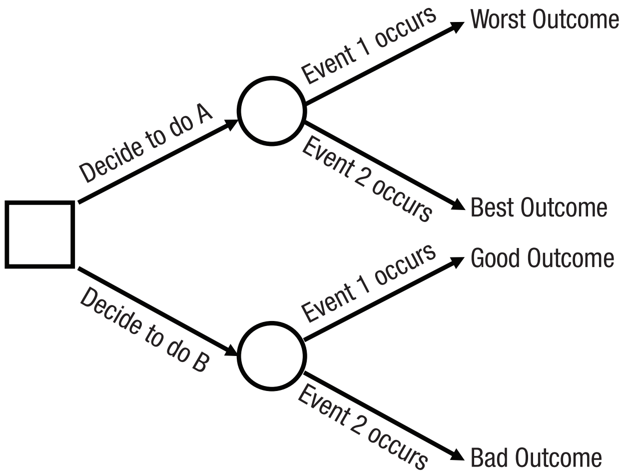

Fig. 1. Open-loop linear process of choice. This is the dominant model of human judgment and choice. A choice between A and B is made and an outcome is obtained depending on the occurrence of a probabilistic event.

consider when building human-like algorithms for DDM. I present a general cognitive algorithm that intends to emulate human decision-making in dynamic environments and reach a level of cognitive precision and systematic calculation to be carried out automatically. This algorithm, defined in instance-based learning theory (IBLT; Gonzalez et al., 2003), is precise in some aspects, but it remains vague and requires formal mathematical formulations in other relevant aspects of cognition. While introducing the various steps of the general IBLT algorithm, I consider accumulated evidence that supports the human-likeness of IBLT models and the gaps and steps required to advance the algorithmic representations of this process.

A Cognitive Perspective of Dynamic Decision-Making

Human decision-making is a complex, high-level cognitive process that is clearly different from other cognitive processes in at least two ways: It builds on other elements of our cognitive system such as perception, memory, and attention, and beyond the judgment process traditionally characterized in the decision-making literature, decision-making is uniquely identified by its essential component, the *process of choice* (Payne et al., 1993).

Choice is the act of selecting among alternatives, whether they are present at the same time or they develop sequentially over time. The process of choice is highly influenced by the cognitive steps that occur before a choice is made (e.g., perception, recognition, and judgment) as well as those that occur with the execution of the action selected and after a result is observed (e.g., feedback and learning). Notably, the judgment process, which precedes choice, involves evaluating the merits and determining the preferences for different alternatives. These two processes, judgment and choice, have been the core of the study of behavioral decision-making for many decades (Einhorn & Hogarth, 1981; W. M. Goldstein & Hogarth, 1997; Hastie, 2001).

The dominant model of judgment and choice is an *open-loop linear process*, illustrated in Figure 1. In this template, explicit- and simultaneous-choice options are represented as "branches" of a decision tree, and there is no learning or feedback loop. Uncertainty in the environment are events in the world described by probabilities or likelihoods, and the consequences of choice are a subjective evaluation of some expected value (i.e., subjective expected utility, or SEU; Hastie, 2001). However, this principle of maximization as a description of human choice has been criticized since its inception because of its discrepancies with the actual *bounded-rationality* decision process (Simon, 1955, 1957). Prospect theory (Kahneman & Tversky, 1979) was proposed as a description of human decision-making under risk, and it prevails as the most accepted general descriptive model of risky choice. Furthermore, it has been associated with an impressive list of biases from optimal decisions (Kahneman et al., 1982; Tversky & Kahneman, 1974). Similarly, ecologically rational mechanisms have been investigated not as a source of human error but as a way to highlight how decisions depend on each environment. The concept of ecological rationality aims to demonstrate how heuristics may succeed or fail in particular situations (Gigerenzer et al., 2011; Gigerenzer & Todd, 1999).

Beyond the prevailing views of judgment and choice summarized above, making decisions under uncertainty in changing environments requires consideration of additional aspects of our cognitive system such as recognition, feedback, and learning from experience. The study of how these cognitive processes influence our decisions has been uncommon in the literature of decision sciences (Einhorn & Hogarth, 1981; Gonzalez, 2017). Simon (1955) highlighted the importance of learning in the choice process. However, decision theory in the 1950s and 1960s addressed learning only in gambles in which choices are independent such as the study of probability learning in the prediction of the occurrence of two mutually exclusive events (Estes, 1964, 1976). Although simple, these learning experiments were connected to initial research in DDM (Edwards, 1962).

DDM may be conceptualized as a control process: a *closed-loop learning process*, illustrated in Figure 2, in which decisions are made sequentially and are influenced by the result of previous decisions (i.e., experience), by external events in the environment, and by constraints such as time limitations and complexity of the task (Brehmer, 1992; Gonzalez, 2017; Gonzalez et al., 2017).

DDM is largely determined by the characteristics of dynamic environments and the constraints of human cognition. Dynamic environments demand a sequence of interdependent decisions made under changes that are endogenous (caused by the decision maker's own past decisions), exogenous (by factors beyond the decision maker's control), or both (Edwards, 1962). In dynamic systems, the state at time *t* depends on the state of the system and decisions made in the past, generating loops that may be self-reinforcing or self-diminishing over time (Gonzalez et al., 2005). These characteristics make dynamic systems very complex, not only in terms of the number of elements of the system (i.e., structural complexity) but also in terms of the relationships and interdependencies among these elements over time (i.e. dynamic complexity; Gonzalez et al., 2017).

Simon (1983) noted that human learning is terribly slow. Indeed, research in DDM has shown that making decisions in dynamic environments can be very challenging for humans. For example, people do not always improve their decisions with practice in a task (Brehmer, 1980), and their performance may remain suboptimal even with full and immediate feedback, unlimited time, and high-performance incentives (Diehl & Sterman, 1995; Sterman, 1994). People are generally poor at handling systems with long feedback delays (Brehmer, 1992; Sterman, 1989), and they have difficulty learning in situations involving environmental constraints, such as workload and time pressure (Gonzalez, 2004, 2005; Kerstholt & Raaijmakers, 1997). Thus, building artificial agents that emulate human decision-making in dynamic environments is not equivalent to building agents that behave "optimally," and agents that approximate optimality with training may not represent the suboptimal learning and difficulty to learn that humans demonstrate in most dynamic tasks.

Dynamic environments are often studied with complex interactive computer systems called *microworlds* (Brehmer & Dörner, 1993; Gonzalez et al., 2005). Examples of microworlds include tasks such as commanding a group of firefighters in an unknown environment (e.g., Brehmer & Allard, 1991), determining the procedures to follow in emergency situations (e.g., Joslyn & Hunt, 1998), and managing scarce resources under time constraints and workload (e.g., Gonzalez, 2004, 2005), among others. But years of research with microworlds has made it clear that dynamic complexity exists even in structurally simple tasks (i.e., tasks that consist of a few alternatives, no time constraints, and even less uncertainty) and that more insights regarding dynamic complexity may be obtained from studying tasks that are structurally simple (Gonzalez, 2022; Gonzalez et al., 2017).

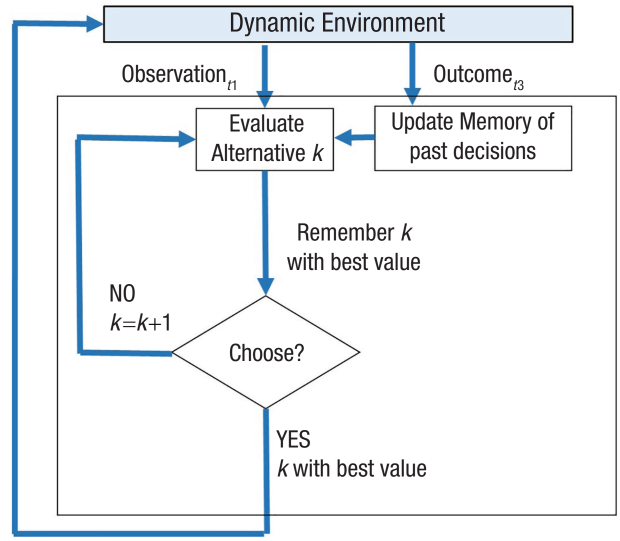

Fig. 2. Closed-loop learning process of choice. This is the general conceptualization of a dynamic decision-making process. This is a control process in which alternatives are evaluated sequentially, and an agent interacts with the environment in a cycle.

A wave of experience-based choice research in simple-choice tasks emerged in the past decades from the observation that human choice depends on how information about a problem is acquired (from description or experience; Hertwig et al., 2004). These efforts focused mostly on binary-choice paradigms (Hertwig & Erev, 2009) rather than microworlds, and the simplicity of these tasks has allowed the discovery of interesting phenomena. A key finding is that, when making sequential decisions from experience, people behave as if low-probability (i.e., rare) events were rarer than they really are rather than behaving as if they were more common than they really are, as predicted by prospect theory. The conclusion is that prospect theory describes human decisions when alternatives are presented simultaneously in simple descriptive gambles (Kahneman & Tversky, 1979) but has less to contribute to the description of how humans make sequential decisions from experience in dynamic environments.

Efforts toward developing comprehensive algorithms that can explain human decisions from experience have resulted in a proliferation of highly task-specific models. These models often can predict or explain choice behavior only in the particular task that they were created for, and they fail to generalize and explain decisions in other, even closely related tasks (see discussions in Bugbee & Gonzalez, 2022; Erev et al., 2017; Gonzalez & Dutt, 2011; Hertwig, 2015; Lejarraga, Dutt, &

Gonzalez, 2012). A general algorithm that has aimed to represent a comprehensive human cognitive process of experiential choice in dynamic tasks is found in IBLT (Gonzalez et al., 2003). This algorithm claims to exhibit human-like decisions in some tasks, and by building on cognitive architectures (Anderson & Lebiere, 1998), this algorithm has reached a level of cognitive precision and systematic calculation to be carried out computationally (Gonzalez, 2022; Hertwig, 2015). However, not all the steps in the IBLT algorithm are precise and concrete enough to implement computationally, and significant research is required to advance human-like algorithms of decision-making.

In what follows, I first address the general question of why it is important to build human-like algorithms: algorithms that emulate human decisions with cognitive precision and that are able to replicate cognitive biases and cognitive constraints when confronted with dynamic environments. Then I introduce the general IBLT algorithm and consider the accumulated evidence that supports human-likeness of the cognitive processes involved in the various steps of IBLT. I also highlight the current gaps that demand more research to advance general algorithmic representations of human-like DDM.

Beyond Getting Inspiration From Humans: Building Human-Like Algorithms of DDM

I build on the premise that it is desirable to build artificial agents that mimic the cognitive process by which humans make decisions in dynamic environments. But given that humans can be poor and slow at learning and improving their choices in dynamic tasks, and given human cognitive constraints, one might ask: Why would we want to build such human-like algorithms? And would building human-like algorithms imply replicating human errors and cognitive biases that make them slow learners? What purpose does it serve to create artificial agents that make mistakes and learn like humans do?

Historically, computational science has attempted to gain inspiration from human behavior and build human-inspired algorithms, with the ultimate goal of finding "optimal computational solutions" to complex tasks (Kochenderfer et al., 2022; Rai et al., 2022; Zhang et al., 2009; Zhou & Chen, 2018). These approaches often look to human behavior to provide a way to improve the algorithms; they attempt to make algorithms more efficient by "imitating" human processes. However, these approaches often use human behavior only as an inspiration, and rarely do they demonstrate any capability to replicate the cognitive process by which humans make decisions.

Creating optimal solutions to DDM problems is essential for providing decision support to humans. Given that humans are often suboptimal at making decisions in dynamic complex tasks, they need optimal solutions that can help them improve their decisions. In other words, AI needs optimal and cognitive algorithms to be able to improve human learning in dynamic environments. However, I suggest that computational algorithms that look for the optimization of solutions in complex tasks would be more efficient if they were informed by process-level cognitive algorithms that emulate and are able to predict human actions. For example, Bayesian principles (Griffiths et al., 2008) and game-theoretic approaches (Roughgarden, 2010) dictate how rational agents should act and update their beliefs on the basis of prior knowledge. But designing solutions under the assumption of human rationality will result in suboptimal outcomes compared or confronted with boundedly rational agents (i.e., humans) acting in dynamic tasks. Generally, the assumption that cognition is approximately optimal to the uncertainty and structure of the environment might be incorrect when dealing with DDM environments, and computationally effective models might not be sufficient to support boundedly rational agents in DDM.

To improve and accelerate learning in DDM, research needs to increase attention to improving current learning algorithms that imitate human cognitive processes. Furthermore, these models should be able to replicate human constraints and errors and be able to predict when humans are falling victim to biases, when humans are about to commit choice errors, and when humans need support given their cognitive constraints (Fuchs et al., 2022). As an illustration, in the context of making defense decisions in a cybersecurity environment, machine-learning algorithms have recently been shown to result in better solutions when they are informed by models that accurately predict human choices than when those algorithms rely on the assumption of human rationality (Aggarwal et al., 2022). Research frameworks in the context of cognitive cybersecurity propose that new systems can improve human actions when these systems take into account the human cognitive state and the prediction of their biases and choice errors (Gonzalez et al., 2023; Lebiere et al., 2023). The main idea is that we cannot effectively overcome human constraints when building technology that aims at optimizing decisions or at beating the human while considering the human mind as a black box. Human-like cognitive algorithms would allow explaining and predicting how human errors and biases emerge in dynamic environments and how to adapt the environment to the human cognitive constraints.

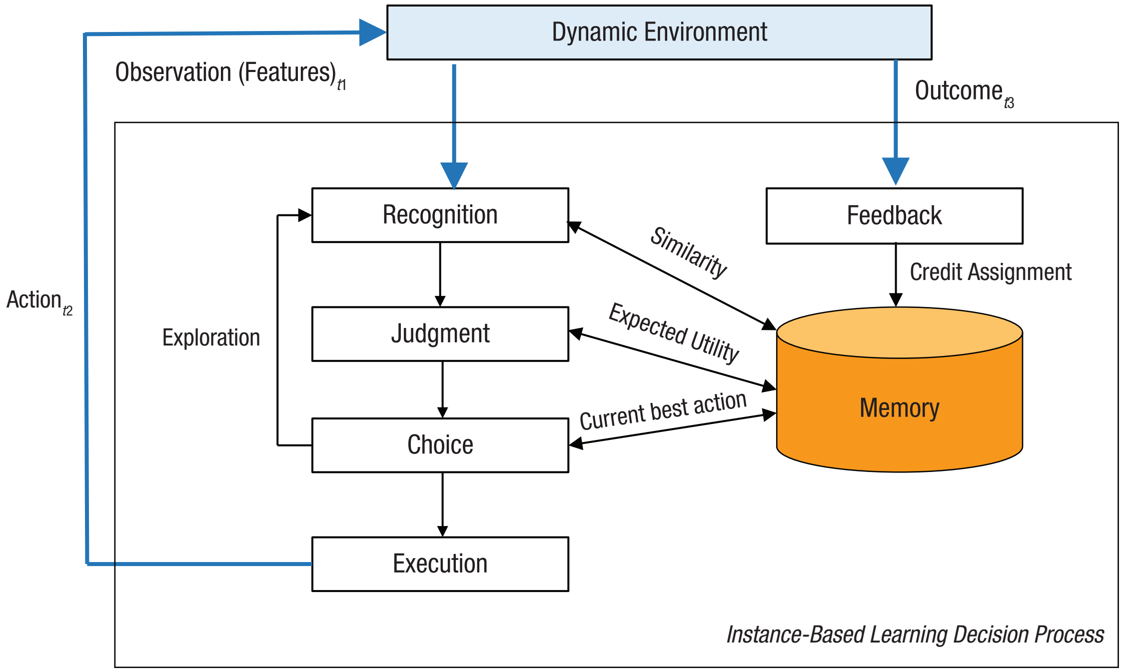

Fig. 3. General cognitive process of decision-making proposed by instance-based learning theory.

To achieve this type of learning algorithm, research also needs to advance the metrics of human-likeness. Most of the research in traditional cognitive computational science relies on outcome metrics: the comparison of choices from model simulations to actual human behavior, using metrics such as the mean squared error or mean squared deviation and correlation. Furthermore, often these comparisons are performed at the aggregate, average level. As discussed in previous research, models of human decision-making must be evaluated at the individual level by both, process and outcome metrics, using multiple criteria rather than relying on a single-outcome comparison (Dutt & Gonzalez, 2015; Farrell & Lewandowsky, 2015; Gluck et al., 2008; Harman et al., 2021).

Many reinforcement-learning (RL) systems demonstrate the ability to replicate the outcome of human decisions (Gershman & Daw, 2017), but these agents are often inadequate for explaining and predicting human adaptation and the learning process in complex environments (Lake et al., 2017; Pouncy et al., 2021). Therefore, a concern has been raised that the advance in RL algorithms is mostly centered on solving computational problems efficiently and optimally rather than on replicating the way humans actually learn (Botvinick et al., 2019; Lake et al., 2017).

Although there might be several ways to replicate human-level decision-making in dynamic tasks, one method relates to the origins of AI: cognitive architectures (Anderson & Lebiere, 1998; Newell, 1992). A common goal of cognitive architectures is to generate computational systems that are capable of demonstrating the same kind of abilities and shortfalls observed in human cognition (Gonzalez et al., 2003; Thomson et al., 2015). Thus, in what follows I present IBLT, a general cognitive algorithm originating from the Adaptive Control of Thought–Rational (ACT-R) cognitive architecture (Anderson & Lebiere, 1998), designed to replicate the cognitive process of human decision-making in dynamic environments. IBLT was designed to mimic how humans use knowledge and memories to make decisions. But as I explain below, there are a number of challenges that need to be addressed to construct human-like agents that adapt and learn in dynamic decision-making situations.

A General Cognitive Algorithm for Human Decision-Making in Dynamic Environments

IBLT is a general postulation of the cognitive processes and the mathematical mechanisms that are globally applicable to DDM tasks (Gonzalez et al., 2003). Specifically, the IBLT algorithm is a process assumed to represent the cognitive steps that humans follow to make decisions in dynamic situations. IBLT is also a computational algorithm that uses mathematical formulations of memory retrieval from ACT-R (Anderson & Lebiere, 1998).

The general cognitive process of decision-making proposed in IBLT is shown in Figure 3. IBLT proposes that decisions are represented in the form of instances involving three parts: state, action, and utility. In general, state is a representation of the features of a decision situation, action is a decision an agent makes in such a state, and utility is an expectation the agent generates from experience or an outcome the agent observes because of such action. The theory generally assumes that instances accumulate over time, and past instances are recalled on the basis of their similarity to a current decision situation. The expected utility of each decision alternative is generated as a function of the utilities in past similar instances and the probability of retrieving those instances from memory. A choice is made for the option that has the highest expected utility. The corresponding mathematically concrete algorithm replicated from Nguyen et al. (2022) and the mathematical formulations are shown in the Appendix.

The process considers one decision alternative at a time and starts with the observation of its environmental state at time t_1 and the determination of whether there are past experiences in memory (i.e., instances) that are similar to the current environmental state (i.e., *recognition*). If similar past instances are found in memory, the expected utility of the decision alternative (i.e., *judgment*) is calculated via a process of *blending* past instances from memory, but if there are no similar past instances, then a heuristic is used to generate the expected utility instead. The *current best action* (i.e., the alternative with the highest expected utility) is maintained in short-term memory, and a decision is made as to whether to continue exploring new alternatives (i.e., *exploration loop*) or to stop exploring to execute the action of the current best alternative (i.e., *choice*). When the exploration loop ends, the choice associated with the alternative that has the highest expected utility is executed at time t_2, affecting the environment accordingly (i.e., *execution*). As the *execution loop* continues and new decision alternatives emerge over time, a result from previous decisions may be observed from the environment, either immediately or with delay from the execution of a choice (i.e., *feedback*) at time t_3. Such a *decision result* is used to update the utility of past instances in memory through a *credit-assignment* process.

Recognition: When is it possible to reuse experience?

The process of recognition is the ability to discriminate among familiar classes of objects (Langley & Simon, 1981). In the field of naturalistic decision-making, recognition is the primary process that experts use to make decisions under uncertainty, reusing decisions that have worked previously according to their own experience (Klein et al., 1993; Zsambok & Klein, 2014). IBLT made this process more precise by using a "partial-matching" mechanism in the ACT-R (Anderson & Lebiere, 1998), defining *similarity* metrics and a threshold of similarity that applies to a current decision situation and the state in past instances stored in memory (see Appendix).

Significantly more empirical and computational work is required to determine the human-likeness of this recognition process of IBLT. For example, although in cognitive science there is an abundance of work on judgments of similarity, there is much less emphasis on the connection between similarity to choice. Tversky (1977) provided foundational ideas regarding feature-based models of similarity, in which features represented objects and a similarity metric was based on matching and mismatching those features relative to past decision-making situations. Such ideas have been essential to advancing knowledge of the recognition process and context effects in decision-making, particularly about how improved recognition emerges from experience. Classic studies suggest that experts are very selective in using features to guide their decisions, whereas novices engage in a more thorough search to determine the similarity of experience applicable to a problem situation (Chase & Simon, 1973; Chi et al., 1981; de Groot, 2014; Gobet & Simon, 1996). Similar patterns of novice and expert behaviors are observed in naturalistic settings (Chase & Simon, 1973; Klein, 1999; Klein et al., 1993).

Research is emerging on computational models of how features in a decision situation influence choice from experience (e.g., Trueblood, 2022; Trueblood et al., 2014; Yearsley et al., 2022). Yet these models focus on the importance of the features (i.e., attribute weights) and do not shed much light on the similarity metrics that are required to evaluate past experience of such features. Research is required to integrate models that rely on the storage of exemplars in memory and that use the similarity of a stimulus to stored exemplars to make decisions (Medin & Schaffer, 1978; Nosofsky, 1986). Such models measure similarity as the distance between objects in memory and those under consideration in a multidimensional psychological space (e.g., Nosofsky, 1986; Shepard, 1987). In addition, beyond feature-based models and geometric multidimensional models, there is a rich literature on psychological models of similarity that would need to be considered and integrated into current work on decision-making (see Goldstone & Son, 2012). New metrics should consider that attribute weights may differ depending on the similarity judgments beyond full matching and mismatching of features.

IBLT's current algorithm often uses a linear-distance metric of similarity for each feature in an instance. In addition, each feature can be weighted so that the sum of the weights of the various features adds up to 1. Then the sum of the weighted similarities across all features are multiplied by a "mismatch-penalty" parameter, as defined in the partial-matching mechanism of the ACT-R (see Appendix). Although some evidence exists regarding the human-likeness of this similarity-functional form (Gonzalez et al., 2003; Gonzalez & Quesada, 2003), the formalization of the recognition process and the generalization of such a process across multiple tasks needs to be investigated further. In particular, future research needs to validate the similarity metrics, attribute weights and how these representations may vary over time, study how these representations are context-dependent, and determine the psychological accuracy of the mathematical representations of similarity (Nachshon et al., 2022).

Judgment: how to determine the value of an alternative?

The value of choice options has historically been formulated as some form of expected value. For example, decision theory from the economics perspective generally describes how rational agents should act through a calculation of a utility value. However, as discussed above, such models do not describe how humans actually evaluate choice options. Humans are boundedly irrational (or "predictably irrational"; Ariely, 2009), and their behavior can be described by a large set of cognitive biases and by relaxed traditional SEU. However, a significant disadvantage of the SEU algorithms is that they do not explain details of the process by which beliefs are formed or from which values are determined (Hastie, 2001).

Computational approaches commonly address the sequential decision problem under uncertainty by representing the problem as a Markov decision process and resolving the decisions through RL or stochastic optimization (Powell, 2022). From this perspective, a problem is represented by state variables that include information that is known to make a decision, the decision itself (i.e., policy), information learned after a decision is made (i.e., exogenous information), a transition function (i.e., equations needed to update each element of the state), and an objective function (i.e., reward) to optimize. The goal of these algorithms is to determine the best method for making a decision (i.e., an optimal policy) given some approximation of the impact that such a decision will have in the future, such that some objective function is optimized.

Generally, it is important to determine what the best decision is in a particular situation. It is expected that humans would want to make the best decision possible, and thus, knowing how to calculate the best value of each alternative is very important. However, because humans are only boundedly rational, they can make the best decision only at a particular point in time given their constraints on knowledge, memory factors, and exogenous variables such as complexity and remaining time (Gonzalez et al., 2003). Thus, the general focus of the computational optimization of policies is valuable, but it is an approach unlikely to represent human-like decisions. The metrics these approaches use to evaluate their models relate to optimal decisions, not to psychological accuracy.

Work with RL models has aimed at endowing RL algorithms with cognitive and psychological characteristics, such as human processing of episodic memory, that would allow these type of models to approximate the challenges that humans face in more naturalistic tasks, such as learning from sparse data and connecting actions and rewards over time (Gershman & Daw, 2017). The primary focus of such studies has been on the similarities between neural processes and computational mechanisms (Gershman & Daw, 2017; Niv, 2009). Although the results have been encouraging, they leave significant room for replicating or explaining the cognitive plausibility and psychological accuracy of such algorithms through behavioral metrics (Botvinick et al., 2019; Lake et al., 2017).

In IBLT the value of each sequential alternative is determined through a process called blending (Gonzalez et al., 2003). Blending is a mechanism for combining past experiences, which has its origin in general ideas for "blended retrievals," through which past memories generate a continuous value of probability judgment (Lebiere, 1999). In IBLT, blending is defined as the sum of utilities of past similar instances weighted by each instance's probability of memory recall (see Appendix; Gonzalez & Dutt, 2011; Lejarraga et al., 2012). Importantly, the use of blending depends on the result of the recognition process. Blending is used to determine the expected utility of the current alternative when there are past instances in memory that are similar to the attributes of such an alternative. If there are no past instance in memory, a general heuristic or "prepopulated instances" that represent previous expectations can be used to determine the expected utility of the current alternative (Gonzalez & Dutt, 2011; Gonzalez et al., 2003; Lejarraga et al., 2012; Nguyen et al., 2022).

Since its inception, the decisions resulting from IBLT have been compared to the decisions made by humans. This is in fact a traditional way to demonstrate the psychological accuracy of cognitive models (Busemeyer & Diederich, 2010; Farrell & Lewandowsky, 2015). The IBLT process has shown a close relationship with the decisions resulting from humans, and this has been demonstrated in a large diversity of tasks, domains, time scales, and data-aggregation levels (for summaries, see Gonzalez, 2013, 2022; Gonzalez & Dutt, 2011; Gonzalez et al., 2017; Lejarraga et al., 2012). These models are commonly evaluated by the output of this process (i.e., decisions predicted by IBLT and decisions actually made by the human), and given the accuracy of the IBLT predictions, the blending process itself displays human-like characteristics. However, the psychological accuracy of the judgment process in IBLT requires fine-grained psychological evaluation methods and metrics, such as evaluating the attributes or the activation and retrieval probability of the instances used to make each decision. This line of research would be worth pursuing in the future.

Choice: when to stop exploration?

In sequential decision-making the decision maker faces an important decision: when to stop evaluating potential alternatives to make a choice. Research has shown that people do not accurately decide when to stop evaluating alternatives so that they optimize their decisions. Instead, people generally show a bias from the optimal stopping point (Bhatia et al., 2021; Seale & Rapoport, 2000), and frequently humans explore less than is optimal (Wulff et al., 2018). Current psychological models suggest that humans adjust their aspirations throughout a sequence of choices according to a threshold, but the functional form of such a threshold is under debate (Baumann et al., 2020; Guan & Lee, 2018; M. D. Lee, 2006). Furthermore, these models do not describe the cognitive process behind stopping decisions; they are usually applicable to specific tasks and cannot generalize to other tasks without significant modifications (Bugbee & Gonzalez, 2022).

Generally, substantial research is required to determine how people make stopping decisions by learning in dynamic environments. New research is emerging to investigate whether humans are able to approach an optimal stopping point after repeated experience (Bugbee & Gonzalez, 2022; D. G. Goldstein et al., 2020). There is also some evidence for adaptation of the stopping decisions to changing conditions of the environment (M. D. Lee & Courey, 2021), consistent with research findings that people are able to learn with repeated experience (D. G. Goldstein et al., 2020).

Other lines of research have investigated how revealing features of various choice alternatives in changing environments influences the stopping point (Lee et al., 2014). Generally, the authors found that stopping decisions change according to environmental dynamics, and their modeling work suggests that decision confidence instead of error may be a regulatory mechanism for the stopping decisions. However, significant efforts are required to explain the inductive generation of the decision to stop the search process in dynamic environments.

The decision of whether to explore new alternatives in search for higher rewards or to exploit already known options (select the option that is best so far) is a fundamental and pervasive problem in our understanding of adaptive behavior (Cohen et al., 2007; Mehlhorn et al., 2015). This is also a problem that has been studied extensively in RL computational approaches (Kochenderfer et al., 2022). But once again, this dilemma is often addressed through algorithms such as those using optimal stopping time (e.g., Gittins index: optimal strategy for exploration), searching for the optimal policy in bandit problems (Kochenderfer et al., 2022; Russell & Norvig, 2010), or assuming a threshold adjustment according to some functional form (Baumann et al., 2020; Guan & Lee, 2018; M. D. Lee, 2006). These algorithms are unlikely to represent human strategies and the inductive learning process for adjustment of the stopping point in changing environments.

IBLT provides an integrated cognitive account of the learning process of stopping decisions in sequential tasks. IBLT does not rely on explicit thresholds but rather proposes that the decision of when to stop exploring is learned from experience. The current alternative-value, external factors such as time and the number of alternatives remaining in the sequence and the change in the available options trigger the decision to stop exploring without relying on the concept of a threshold (Bugbee & Gonzalez, 2022). In IBLT, the transition from exploration to exploitation emerges naturally from the detected consistencies in the environment and as a consequence of the information experienced without explicitly defining choice rules between exploring or exploiting (Gonzalez & Dutt, 2016; Gonzalez et al., 2003). The human-likeness of such a process has been demonstrated by comparing simulated individual choices produced by IBLT to empirical stopping decisions from human data (Bugbee & Gonzalez, 2022). Because the same process can be used across different decision environments, it is expected that IBLT will provide an integrated cognitively plausible process through which stopping decisions are made in sequential decision tasks (Bugbee & Gonzalez, 2022; Bugbee & Gonzalez, 2022; Gonzalez & Dutt, 2016; Mehlhorn et al., 2015). Future research should investigate the generalizability of this process and the robustness of the proposed algorithms (Gonzalez & Aggarwal, in press).

Feedback: How do we learn from past actions?

Learning the relationship between actions and outcomes is essential to behavioral adaptation and decision-making in dynamic environments (Gonzalez, 2005; Gonzalez et al., 2003). However, in dynamic-decision environments, accounting for feedback can be complicated because the knowledge of outcomes is often delayed and the associations of the outcomes to the decisions that produced such outcomes is not unique (Brehmer, 1992; Sterman, 1989). It is well known that longer feedback delays negatively affect long-term performance in decision tasks (Einhorn & Hogarth, 1978; Gonzalez, 2005). Thus, an important fundamental question is how agents learn from their own experience in dynamic decisions under uncertainty when feedback delays are pervasive.

In computational science, this problem is known as temporal credit assignment, the challenge to assign credit to intermediate actions within a sequence (Minsky, 1961). Computational science has proposed a number of approaches to handle delayed feedback. One of the most prominent mechanisms to address the credit-assignment problem is the temporal-difference (TD) mechanism for RL models (Sutton & Barto, 2018). According to the TD approach, an agent predicts the value of intermediate states in the absence of final feedback and uses prediction errors over small intervals to update their future predictions. Psychological research has addressed the question of how well a model similar to the TD mechanism is aligned with humans in tasks involving feedback delays (Walsh & Anderson, 2011, 2014). In addition, the TD credit-assignment methods have been incorporated into cognitive architectures to emulate how humans process feedback delays in sequential decision-making tasks (Fu & Anderson, 2006), suggesting that the human evaluation of intermediate states in terms of future rewards, as predicted by TD learning, may be psychologically accurate (Walsh & Anderson, 2011). However, the primary focus of such studies has been on the similarities between neural processes and computational mechanisms rather than the comparison to observed human behavior. In recent work, the TD mechanism was used in instance-based learning models across a variety of tasks and resulted in similar learning results to that of humans, suggesting that the TD mechanism may indeed be a human-like process (Nguyen et al., 2022). However, the robustness of the TD and potentially competing credit-assignment mechanisms that are psychologically accurate needs to be addressed further. The development and comparison of human-like mechanisms for credit assignment in IBLT is still in the early stages of exploration (Nguyen et al., 2023).

Conclusions

An initial goal of AI was to build algorithms that emulate human behavior (Lake et al., 2017; Simon, 1983; Turing, 1950). However, the focus of AI in more recent decades has been dominated by computational algorithms that aim at making optimal decisions in complex tasks and beating the human rather than replicating human decisions and replicating their cognitive decision-making process. In this article, I claim that we need to focus on building learning algorithms that simulate human cognitive processes so that optimization machine-learning AI can be devised to improve and accelerate human learning.

Generally, DDM involves a complex process of learning and making decisions by retrieving solutions from past similar memories (Gonzalez et al., 2003). DDM is very challenging for humans, and most research suggests difficulties in learning and improving decision-making in dynamic tasks. Building human-like artificial agents in dynamic tasks is important for overcoming such limitations of human behavior. However, current computational algorithms are far from achieving the level of cognitive specificity and precision that is required to create human-like algorithms for DDM. Although many research challenges are computational, many other challenges are behavioral. Basic research should focus on understanding the learning process in humans and why human learning is slow and inefficient in DDM tasks.

Historically, computational algorithms have been centered on the efficiency and optimality of decisions rather than replicating or explaining human behavior (Botvinick et al., 2019; Lake et al., 2017). Many of these algorithms make strong assumptions about the environment and about human capabilities. They ignore human cognitive constraints, assuming that humans do not forget, are able to remember, learn, store and recall information optimally, and ultimately choose rationally. However, the main issue relies on the psychological and cognitive decision-making research that has lagged behind in the study of complex, dynamic environments. To achieve human-like decision-making algorithms, research on behavioral decision-making needs to advance to the study of dynamic environments. The body of accumulated evidence on the dynamics of decision processes must then be mathematically formulated to be carried out computationally.

Although there are many efforts intended to create computational algorithms that replicate the cognitive process of human DDM, one method that is historically linked to AI is cognitive architectures. Cognitive architectures aim at enabling the construction of artificial agents that are capable of exhibiting intelligent behavior from the foundational mechanisms that underlie human cognition. IBLT originated in the ACT-R (Anderson &

Lebiere, 1998) and relies on the ACT-R's proposed mechanism for the activations of memories (Gonzalez et al., 2003). IBLT provides a general cognitive algorithm for decisions under uncertainty in dynamic tasks, addressing many of the essential elements of naturalistic decisions, including uncertainty, recognition of decision alternatives through similarity, bounded rationality, exploration-exploitation trade-offs, and feedback delays, among other elements. The IBLT algorithm has been implemented computationally in a large number of tasks from simple bandit-type binary-choice problems to significantly more complex tasks such as real-time dynamic allocation of limited resources and navigation and coordination in teams (for examples, see Nguyen et al., 2022).

In some ways, models that derive from IBLT have achieved human-like decision-making, demonstrated by the comparison of decisions that the algorithm predicts to decisions that humans actually make. However, as discussed above, there are a number of challenges that cognitive and computational decision science needs to address to demonstrate the human-likeness of IBLT or other algorithms. First, several of the steps in IBLT need to be investigated, formalized, and validated. To mention two important issues: The recognition and feedback processes are the least developed in this theory. Second, advanced methods for verifying human-likeness are required. Evaluation should not be limited to the comparison of outcomes (i.e., the actions predicted by a model and the actions made by human actors). New metrics should involve the comparison of the cognitive decision-making process, including the prediction of the errors and biases that a model makes compared with the human psychological processes that led to such predictions.

Appendix

For an agent, an option $k = (s, a)$ is defined by taking action a after observing state s. At time t, assume that there are n_{kt} different instances (k_i, x_{ikit}) for $i = 1, \ldots, n_{kt}$, associated with k. Each instance i in memory has an *activation* value that represents how readily available that information is in memory and expressed as follows (Anderson & Lebiere, 1998):

$$\Lambda_{ik_it} = \ln\left(\sum_{t'\in T_{ik_it}} (t-t')^{-d}\right) + \alpha \sum_j Sim_j\left(f_j^k, f_j^{k_i}\right) + \sigma \ln\frac{1-\xi_{ik_it}}{\xi_{ik_it}}, \qquad (1)$$

where d, α, and σ are the decay, mismatch penalty, and noise parameters, respectively, and $T_{ikit} \subset \{0, \ldots, t-1\}$ is the set of the previous timestamps in which the instance i was observed, f_j^k is the j-th attribute of the state s, and Sim_j is a similarity function associated with the j-th attribute. The second term is a partial matching process reflecting the similarity between the current state s and the state of the option k_i. The rightmost term represents a noise for capturing individual variation in activation, and ξ_{ikit} is a

Algorithm 1: Pseudo Code of Instance-based Learning process

Input: default utility x_0, a memory dictionary $\mathcal{M} = \{\}$, global counter $t = 1$, step limit L, a flag *delayed* to indicate whether feedback is delayed.

repeat
    Initialize a counter (i.e., step) $l = 0$ and observe state s_l
    while s_l *is not terminal and* $l < L$ **do**
        Execution Loop
            Exploration Loop $a \in A$ **do**
                Compute activation values $\Lambda_{i(s_l^i,a)t}$ of instances $((s_l^i, a), x_{i(s_l^i,a)t}, T_{i(s_l^i,a)t})$ by (1)
                Compute retrieval probabilities $P_{i(s_l^i,a)t}$ by (2)
                Compute blended values $V_{(s_l,a)t}$ corresponding to (s_l, a) by (3)
            end
            Choose an action $a_l \in \arg\max_{a\in A} V_{(s_l,a)t}$
        end
        Take action a_l, move to state s_{l+1}, observe s_{l+1}, and receive outcome x_{l+1}
        Store t into instance corresponding to selecting (s_l, a_l) and achieving outcome x_{l+1} in $\mathcal{M}$
        If *delayed* is true, update outcomes using a *credit assignment* mechanism
        $l \leftarrow l + 1$ and $t \leftarrow t + 1$
    end
until *task stopping condition*

Figure 4. Pseudo code of the instance-based learning process. This process assumes the use of prepopulated instances in a memory dictionary (i.e., "default utility") rather than the use of heuristics (see the *Judgment* step in the discussion of IBLT, above). The formalization of Equations 1 to 3 are shown below. This figure is replicated from Nguyen et al. (2022).

random number drawn from a uniform distribution $U(0, 1)$ at each time step and for each instance and option.

The activation of an instance i is used to determine the probability of retrieval of an instance from memory. The probability of an instance i is defined by a soft-max function as follows:

$$p_{ik_it} = \frac{e^{\Lambda_{ik_it/T}}}{\sum_{j=1}^{n_{kt}} e^{\Lambda_{jk\,jt/T}}}, \tag{2}$$

where τ is the Boltzmann constant (i.e., the "temperature") in the Boltzmann distribution. For simplicity, τ is often defined as a function of the same σ used in the activation equation $T = \sigma\sqrt{2}$.

The expected utility of option k is calculated on the basis of *blending* as specified in choice tasks (Gonzalez & Dutt, 2011; Lejarraga et al., 2012):

$$V_{kt} = \sum_{i=1}^{nkt} P_{ik_it} x_{ik_it}. \tag{3}$$

The choice rule is to select the option that corresponds to the maximum blended value. In particular, at the l-th step of an episode, the agent selects the option (s_l, a_l) with

$$a_l = \arg\max_{\alpha \in A} V_{(s_l,\alpha),t}. \tag{4}$$

The flag *delayed* on line 14 of Figure 4 is true when the agent knows the real outcome after making a sequence of decisions without feedback. In such a case, the agent updates outcomes by using one of the credit-assignment mechanisms (Nguyen et al., 2023). It is worth noting that when the flag *delayed* is true depends on a specific task. For instance, *delayed* can be set to true when the agent reaches the terminal state or when the agent receives a positive reward.

Transparency

Action Editor: Melanie Mitchell
Editor: Interim Editorial Panel
Declaration of Conflicting Interests
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.
Funding
This work was supported by Australia-US MURI Grant W911NF-20-S-000, Army Research Laboratory Grant W911NF-13-2-0045, and National Science Foundation AI Institute for Societal Decision Making Grant IIS 2229881.

ORCID iD

Cleotilde Gonzalez https://orcid.org/0000-0002-6244-2918

Notes

1. An algorithm is a sequence of steps used to accomplish a task, and a computational algorithm requires a level of precision and systematic calculation to be carried out automatically (Chabert, 1999).
2. This perspective disregards work on "embodied cognition" (Foglia & Wilson, 2013), the neural basis of decision-making (D. Lee et al., 2012), and most of the work under the perspective of complex problem-solving related to motivations and emotions (Dörner & Güss, 2022), but I acknowledge that "body" functions (i.e., vision, modalities for execution), neuroscience research, and emotions are relevant to the dynamic decision-making process discussed here.

References

Aggarwal, P., Thakoor, O., Jabbari, S., Cranford, E. A., Lebiere, C., Tambe, M., & Gonzalez, C. (2022). Designing effective masking strategies for cyberdefense through human experimentation and cognitive models. *Computers & Security*, *117*, Article 102671. https://doi.org/10.1016/j.cose.2022.102671

Anderson, J. R., & Lebiere, C. J. (1998). *The atomic components of thought* (1st ed.). Psychology Press. https://doi.org/10.4324/9781315805696

Ariely, D. (2009). The end of rational economics. *Harvard Business Review*, *87*(7–8), 78–84.

Baumann, F., Lorenz-Spreen, P., Sokolov, I. M., & Starnini, M. (2020). Modeling echo chambers and polarization dynamics in social networks. *Physical Review Letters*, *124*(4), Article 048301. https://doi.org/10.1103/PhysRevLett.124.048301

Bhatia, A., Svegliato, J., & Zilberstein, S. (2021, August 2–13). *Tuning the hyperparameters of anytime planning: A deep reinforcement learning approach* [Paper presentation]. ICAPS 2021 Workshop on Heuristics and Search for Domain-Independent Planning, Guangzhou, China. https://openreview.net/forum?id=c7hpFp_eRCo

Botvinick, M., Ritter, S., Wang, J. X., Kurth-Nelson, Z., Blundell, C., & Hassabis, D. (2019). Reinforcement learning, fast and slow. *Trends in Cognitive Sciences*, *23*(5), 408–422.

Brehmer, B. (1980). In one word: Not from experience. *Acta Psychologica*, *45*(1–3), 223–241.

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica*, *81*(3), 211–241.

Brehmer, B., & Allard, R. (1991). Dynamic decision making: The effects of task complexity and feedback delay. In J. Rasmussen, B. Brehmer, & J. Leplat (Eds.), *Distributed decision making: Cognitive models for cooperative work* (pp. 319–334). John Wiley & Sons.

Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior*, *9*(2–3), 171–184.

Bugbee, E. H., & Gonzalez, C. (2022). Making predictions without data: How an instance-based learning model predicts sequential decisions in the balloon analog risk task. In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the*

Cognitive Science Society. Cognitive Science Society. https://escholarship.org/uc/item/5rj08314

Busemeyer, J. R., & Diederich, A. (2010). *Cognitive modeling*. Sage.

Chabert, J.-L. (Ed.). (1999). Euclid's algorithm. In *A history of algorithms* (pp. 113–138). Springer.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81.

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152.

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 933–942.

De Groot, A. D. (2014). *Thought and choice in chess*. De Gruyter Mouton.

Diehl, E., & Sterman, J. D. (1995). Effects of feedback complexity on dynamic decision making. *Organizational Behavior and Human Decision Processes*, *62*(2), 198–215.

Dörner, D., & Güss, C. D. (2022). Human error in complex problem solving and dynamic decision making: A taxonomy of 24 errors and a theory. *Computers in Human Behavior Reports*, *7*, Article 100222. https://doi.org/10.1016/j.chbr.2022.100222

Dutt, V. & Gonzalez, C. (2015). Accounting for outcome and process measures and the effects of model calibration. *Journal of Dynamic Decision Making*, *1*(2), 1–10. https://doi.org/10.1016/j.chb.2015.01.039

Edwards, W. (1962). Dynamic decision theory and probabilistic information processings. *Human Factors*, *4*(2), 59–74.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgement and choice. *Annual Review of Psychology*, *32*(1), 53–88.

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review*, *85*(5), 395.

Erev, I., Ert, E., Plonsky, O., Cohen, D., & Cohen, O. (2017). From anomalies to forecasts: Toward a descriptive model of decisions under risk, under ambiguity, and from experience. *Psychological Review*, *124*(4), 369–409. https://doi.org/10.1037/rev0000062

Estes, W. K. (1964). Probability learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 89–128). Academic Press.

Estes, W. K. (1976). The cognitive side of probability learning. *Psychological Review*, *83*(1), 37–64. https://doi.org/10.1037/0033-295X.83.1.37

Farrell, S., & Lewandowsky, S. (2015). An introduction to cognitive modeling. In B. U. Forstmann & E.-J. Wagenmakers (Eds.), *An introduction to model-based cognitive neuroscience* (pp. 3–24). Springer.

Foglia, L., & Wilson, R. A. (2013). Embodied cognition. *WIREs Cognitive Science*, *4*(3), 319–325.

Fu, W. T., & Anderson, J. R. (2006). From recurrent choice to skill learning: A reinforcement-learning model. *Journal of Experimental Psychology: General*, *135*(2), 184–206. https://doi.org/10.1037/0096-3445.135.2.184

Fuchs, A., Passarella, A., & Conti, M. (2022). *Modeling human behavior part II: Cognitive approaches and uncertainty*. *arXiv*. https://doi.org/10.48550/arXiv.2205.06483

Gershman, S. J., & Daw, N. D. (2017). Reinforcement learning and episodic memory in humans and animals: An integrative framework. *Annual Review of Psychology*, *68*, 101–128.

Gigerenzer, G. E., Hertwig, R. E., & Pachur, T. E. (2011). *Heuristics: The foundations of adaptive behavior*. Oxford University Press.

Gigerenzer, G. E., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. In G. Gigerenzer, P. M. Todd, & The ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 3–34). Oxford University Press.

Gluck, K., Bello, P., & Busemeyer, J. (2008). Introduction to the special issue. *Cognitive Science*, *32*(8), 1245–1247. https://doi.org/10.1080/03640210802473582

Gobet, F., & Simon, H. A. (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology*, *31*(1), 1–40.

Goldstein, D. G., McAfee, R. P., Suri, S., & Wright, J. R. (2020). Learning when to stop searching. *Management Science*, *66*(3), 1375–1394.

Goldstein, W. M., & Hogarth, R. (1997). *Research on judgment and decision making*. Cambridge University Press.

Goldstone, R. L., & Son, J. Y. (2012). *Similarity*. Oxford University Press.

Gonzalez, C. (2004). Learning to make decisions in dynamic environments: Effects of time constraints and cognitive abilities. *Human Factors*, *46*(3), 449–460.

Gonzalez, C. (2005). Decision support for real-time, dynamic decision-making tasks. *Organizational Behavior and Human Decision Processes*, *96*(2), 142–154.

Gonzalez, C. (2013). The boundaries of instance-based learning theory for explaining decisions from experience. *Progress in Brain Research*, *202*, 73–98.

Gonzalez, C. (2017). Decision making: A cognitive science perspective. In S. E. F. Chipman (Ed.), *The Oxford handbook of cognitive science* (pp. 249–263). Oxford University Press.

Gonzalez, C. (2022). Learning and dynamic decision making. *Topics in Cognitive Science*, *14*(1), 14–30.

Gonzalez, C., & Aggarwal, P. (in press). Sequential decisions from sampling: Inductive generation of stopping decisions using instance-based learning theory. In K. Fiedler, P. Juslin, & J. Denrell (Eds.), *Sampling theories continue to inspire novel judgment and decision research*.

Gonzalez, C., Aggarwal, P., Cranford, E. A., & Lebiere, C. (2023). Adaptive cyberdefense with deception: A human-AI cognitive approach. In T. Bao, M. Tambe, & C. Wang (Eds.), *Cyber deception, techniques, strategies, and human aspects* (*Vol. 89*, pp. 41–57). Springer. https://doi.org/10.1007/978-3-031-16613-6_3

Gonzalez, C., & Dutt, V. (2011). A generic dynamic control task for behavioral research and education. *Computers in Human Behavior*, *27*(5), 1904–1914.

Gonzalez, C., & Dutt, V. (2016). Exploration and exploitation during information search and experimential choice.

Journal of Dynamic Decision Making, *2*, Article 2. https://doi.org/10.11588/jddm.2016.1.29308

Gonzalez, C., Fakhari, P., & Busemeyer, J. (2017). Dynamic decision making: Learning processes and new research directions. *Human Factors*, *59*(5), 713–721.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, *27*(4), 591–635.

Gonzalez, C., & Quesada, J. (2003). Learning in dynamic decision making: The recognition process. *Computational & Mathematical Organization Theory*, *9*(4), 287–304.

Gonzalez, C., Vanyukov, P., & Martin, M. K. (2005). The use of microworlds to study dynamic decision making. *Computers in Human Behavior*, *21*(2), 273–286.

Griffiths, T., Kemp, C. B., & Tenenbaum, J. (2008). *Bayesian models of cognition*. Carnegie Mellon University. https://doi.org/10.1184/R1/6613682.v1

Guan, M., & Lee, M. D. (2018). The effect of goals and environments on human performance in optimal stopping problems. *Decision*, *5*(4), 339–361. https://doi.org/10.1037/dec0000081

Harman, J. L., Yu, M., Konstantinidis, E., & Gonzalez, C. (2021). How to use a multi-criteria comparison procedure to improve modeling competitions. *Psychological Review*, *128*(5), 995–1005. https://doi.org/10.1037/rev0000274

Hastie, R. (2001). Problems for judgment and decision making. *Annual Review of Psychology*, *52*(1), 653–683.

Hertwig, R. (2015). Decisions from experience. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making* (Vol. 1, pp. 240–267). John Wiley & Sons.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*(8), 534–539.

Hertwig, R., & Erev, I. (2009). The description-experience gap in risky choice. *Trends in Cognitive Sciences*, *13*(12), 517–523.

Joslyn, S., & Hunt, E. (1998). Evaluating individual differences in response to time-pressure situations. *Journal of Experimental Psychology: Applied*, *4*(1), 16–43. https://doi.org/10.1037/1076-898X.4.1.16.

Kahneman, D., Slovic, S. P., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.

Kahneman, D., & Tversky, A. (1979). On the interpretation of intuitive probability: A reply to Jonathan Cohen. *Cognition*, *7*(4), 409–411.

Kerstholt, J. H., & Raaijmakers, J. G. (1997). *Decision making: Cognitive models and explanations*. Routledge.

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 138–147). Ablex Publishing.

Klein, G. A. (1999). *Sources of power: How people make decisions*. MIT Press.

Kochenderfer, M. J., Wheeler, T. A., & Wray, K. H. (2022). *Algorithms for decision making*. MIT Press.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*, Article E253. https://doi.org/10.1017/S0140525X16001837

Langley, P., & Simon, H. A. (1981). The central role of learning in cognition. In J. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 361–380). Lawrence Erlbaum Associates.

Lebiere, C. (1999). The dynamics of cognition: An ACT-R model of cognitive arithmetic. *Kognitionswissenschaft*, *8*(1), 5–19.

Lebiere, C., Cranford, E. A., Aggarwal, P., Cooney, S., Tambe, M., & Gonzalez, C. (2023). Cognitive modeling for personalized, adaptive signaling for cyber deception. In T. Bao, M. Tambe, & C. Wang (Eds.), *Cyber deception, techniques, strategies, and human aspects* (Vol. 89, pp. 59–82). Springer. https://doi.org/10.1007/978-3-031-16613-6_4

Lee, D., Seo, H., & Jung, M. W. (2012). Neural basis of reinforcement learning and decision making. *Annual Review of Neuroscience*, *35*, 287–308.

Lee, M. D. (2006). A hierarchical Bayesian model of human decision-making on an optimal stopping problem. *Cognitive Science*, *30*(3), 1–26.

Lee, M. D., & Courey, K. A. (2021). Modeling optimal stopping in changing environments: A case study in mate selection. *Computational Brain & Behavior*, *4*, 1–17.

Lee, M. D., Newell, B. R., & Vandekerckhove, J. (2014). Modeling the adaptation of search termination in human decision making. *Decision*, *1*(4), 223–251. https://doi.org/10.1037/dec0000019

Lejarraga, T., Dutt, V., & Gonzalez, C. (2012). Instance-based learning: A general model of repeated binary choice. *Journal of Behavioral Decision Making*, *25*(2), 143–153. https://doi.org/10.1002/bdm.722

Lejarraga, T., Hertwig, R., & Gonzalez, C. (2012). How choice ecology influences search in decisions from experience. *Cognition*, *124*(3), 334–342.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238.

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., Hausmann, D., Fiedler, K., & Gonzalez, C. (2015). Unpacking the exploration–exploitation tradeoff: A synthesis of human and animal literatures. *Decision*, *2*(3), 191–215.

Minsky, M. (1961). Steps toward artificial intelligence. *Proceedings of the IRE*, *49*(1), 8–30.

Nachshon, Y., Cohen, H., Ben-Artzi, M., & Maril, A. (2022). *A model of similarity: Metric in a patch*. PsyArXiv. https://doi.org/10.31234/osf.io/fyd4q

Newell, A. (1992). Unified theories of cognition and the role of soar. In J. A. Michon & A. Akyürek (Eds.), *Soar: A cognitive architecture in perspective: A tribute to Allen Newell* (pp. 25–79). Springer.

Nguyen, T. N., McDonald, C., & Gonzalez, C. (2023). *Credit assignment: Challenges and opportunities in developing human-like AI agents*. arXiv:2307.08171. https://doi.org/10.48550/arXiv.2307.08171

Nguyen, T. N., Phan, D. N., & Gonzalez, C. (2022). *SpeedyIBL: A comprehensive, precise, and fast implementation of instance-based learning theory*. arXiv. https://doi.org/10.48550/arXiv.2111.10268

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, *53*(3), 139–154.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge University Press.

Pouncy, T., Tsividis, P., & Gershman, S. J. (2021). What is the model in model-based planning? *Cognitive Science*, *45*(1), Article e12928. https://doi.org/10.1111/cogs.12928

Powell, W. B. (2022). *Reinforcement learning and stochastic optimization: A unified framework for sequential decisions*. John Wiley & Sons.

Rai, R., Das, A., Ray, S., & Dhal, K. G. (2022). Human-inspired optimization algorithms: Theoretical foundations, algorithms, open-research issues and application for multi-level thresholding. *Archives of Computational Methods in Engineering*, *29*(7), 5313–5352.

Roughgarden, T. (2010). Algorithmic game theory. *Communications of the ACM*, *53*(7), 78–86.

Russell, S. J., & Norvig, P. (2010). *Artificial intelligence: A modern approach* (3rd ed.). Pearson Education.

Seale, D. A., & Rapoport, A. (2000). Optimal stopping behavior with relative ranks: The secretary problem with unknown population size. *Journal of Behavioral Decision Making*, *13*(4), 391–411.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99–118.

Simon, H. A. (1957). *Models of man; Social and rational*. John Wiley & Sons.

Simon, H. A. (1983). Why should machines learn? In J. G. Carbonell, R. S. Michalski, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (pp. 25–37). Springer-Verlag.

Sterman, J. D. (1989). Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management Science*, *35*(3), 321–339.

Sterman, J. D. (1994). Learning in and about complex systems. *System Dynamics Review*, *10*(2–3), 291–330.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT Press.

Thomson, R., Lebiere, C., Anderson, J. R., & Staszewski, J. (2015). A general instance-based learning framework for studying intuitive decision-making in a cognitive architecture. *Journal of Applied Research in Memory and Cognition*, *4*(3), 180–190.

Trueblood, J. S. (2022). Theories of context effects in multialternative, multiattribute choice. *Current Directions in Psychological Science*, *31*(5), 428–435. https://doi.org/10.1177/09637214221109587

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychological Review*, *121*(2), 179–205.

Turing, A. M. (1950). I.—Computing machinery and intelligence. *Mind*, *59*(236), 433–460.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, *185*(4157), 1124–1131.

Walsh, M. M., & Anderson, J. R. (2011). Modulation of the feedback-related negativity by instruction and experience. *Proceedings of the National Academy of Sciences, USA*, *108*(47), 19048–19053.

Walsh, M. M., & Anderson, J. R. (2014). Navigating complex decision spaces: Problems and paradigms in sequential choice. *Psychological Bulletin*, *140*(2), 466–486.

Wulff, D. U., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin*, *144*(2), 140–176.

Yearsley, J. M., Pothos, E. M., Barque-Duran, A., Trueblood, J. S., & Hampton, J. A. (2022). Context effects in similarity judgments. *Journal of Experimental Psychology: General*, *151*(3), 711–717.

Zhang, L. M., Dahlmann, C., & Zhang, Y. (2009). Human-inspired algorithms for continuous function optimization. In *2009 IEEE international conference on intelligent computing and intelligent systems* (Vol. 1, pp. 318–321). Institute of Electrical and Electronics Engineers.

Zhou, J., & Chen, F. (Eds.). (2018). *Human and machine learning*. Springer.

Zsambok, C. E., & Klein, G. (Eds.). (2014). *Naturalistic decision making*. Psychology Press.