frontiers | Frontiers in Artificial Intelligence

TYPE Perspective
PUBLISHED 13 January 2023
DOI 10.3389/frai.2022.1092053

Check for updates

OPEN ACCESS

EDITED BY
Diego Zapata-Rivera,
Educational Testing Service,
United States

REVIEWED BY
Caitlin Tenison,
Educational Testing Service,
United States
Isabella Poggi,
Roma Tre University, Italy

*CORRESPONDENCE
Baptist Liefooghe
✉ b.liefooghe@uu.nl

SPECIALTY SECTION
This article was submitted to
AI for Human Learning and Behavior Change,
a section of the journal
Frontiers in Artificial Intelligence

RECEIVED 07 November 2022
ACCEPTED 22 December 2022
PUBLISHED 13 January 2023

CITATION
Liefooghe B and van Maanen L (2023) Three levels at which the user's cognition can be represented in artificial intelligence.
Front. Artif. Intell. 5:1092053.
doi: 10.3389/frai.2022.1092053

Three levels at which the user's cognition can be represented in artificial intelligence

Baptist Liefooghe* and Leendert van Maanen

Department of Psychology, Utrecht University, Utrecht, Netherlands

Artificial intelligence (AI) plays an important role in modern society. AI applications are omnipresent and assist many decisions we make in daily life. A common and important feature of such AI applications are user models. These models allow an AI application to adapt to a specific user. Here, we argue that user models in AI can be optimized by modeling these user models more closely to models of human cognition. We identify three levels at which insights from human cognition can be—and have been—integrated in user models. Such integration can be very loose with user models only being inspired by general knowledge of human cognition or very tight with user models implementing specific cognitive processes. Using AI-based applications in the context of education as a case study, we demonstrate that user models that are more deeply rooted in models of cognition offer more valid and more fine-grained adaptations to an individual user. We propose that such user models can also advance the development of explainable AI.

KEYWORDS

human cognition, user model, explainable AI, cognitive modeling, human behavior

Artificial intelligence (AI) has taken an important place in society and offers support in a variety of domains. Many of these domains require interaction between humans and AI systems, ranging from simple recommender systems to more sophisticated diagnostic tools that are driven by machine learning. Whether this interaction is fruitfully adopted in some cases varies with the degree to which AI adapts to what the user wants, thinks, believes and likes (Baker et al., 2010; Bosse and Hoogendoorn, 2015; Rabinowitz et al., 2018; Bonnefon and Rahwan, 2020; Langley et al., 2022; Nguyen and Gonzalez, 2022). Such adaptation requires the AI system to represent the mental states of the user that are not directly observable and use these states to predict the behavior of the user (Premack and Woodruff, 1978).

A key element endowing AI applications with the ability to adapt to a user is the user model. A user model often consists of a decision-making algorithm that is optimized to provide suitable interventions at the right time given the observable behavior of one or more users (Wahlster and Kobsa, 1989). Following this definition, a user model can be conceptualized as a set of input-output mappings that are learned and can be conceptualized as a subset of a broader context model, which includes all possible situational features that may be relevant, such as time of day, previous interactions, or even seasonal fluctuations (Sporrel et al., 2021; Wang et al., 2021a,b). Hence, human behavior observed by the AI system is not necessarily related to a representation that reflects the user's cognitive state that caused that behavior.

Whereas, a user model that is implemented as a set of input-output mappings may be suitable for some applications, such as recommender systems, the question arises whether this suffices for all domains, or whether there are instances where it is necessary to also represent the cognitive state that underlies user behavior. For instance, it has been argued that the detection of deception by means of AI in the context of airport security is intrinsically flawed because the user model does not appropriately incorporate knowledge of human cognition (Jupe and Keatley, 2020). That is, although AI-algorithms can be trained to detect deception in humans on the basis of facial micro-expressions (Rothwell et al., 2006), psychological research has demonstrated that facial micro-expressions have in fact very low validity in predicting deception (DePaulo et al., 2003). Hence, the relation between behavioral proxy and cognitive construct may not be valid, making the application of micro-expressions to detect deception rather tedious. A similar concern may arise in AI-based recruitment applications that analyze candidates' face expressions and speech demeanor to infer traits such as emotional intelligence and personality (Sethumadhavan and Phisuthikul, 2019; Hmoud, 2021). Even for less circumstantial behavior, such as test performance in an e-learning environment, the question arises whether the response of a participant (e.g., the number of correctly recalled items) reflects some relevant cognitive aptitude (e.g., working-memory capacity) or relates to a spurious factor (e.g., fatigue, distraction, stress,…).

The previous examples thus indicate that in some applications user models need to represent user behavior as well as the mental states underlying that behavior. However, inferring a particular mental state on the basis of a particular observable behavior is often invalid. This is a general problem in cognitive sciences (Borsboom et al., 2004; Poldrack, 2006; De Houwer, 2011; IJzerman et al., 2020) that may contaminate the validity of AI applications as well. The solution to this problem we propose is based on the work of Oberauer and Lewandowsky (2019). These authors distinguish between two types of research in cognitive sciences: discovery-oriented research and theory-testing research. In discovery-oriented research, cognitive models define a search space for the discovery of (behavioral) proxies, but do not entail strong hypotheses by which they can be tested and falsified through the use of these proxies. Theory-testing research relies on cognitive models that do strongly imply such hypotheses and the relation between cognition and behavior is often explicated by formalizing cognitive processes. In the current perspective paper, we argue that implementing user models on the basis of theory-testing research with formal models of human cognition offers a greater insurance that the correct inferences are made by AI applications about a user.

We identify three levels of integration between cognitive theory and user models, ranging from very loose (based on only anecdotal knowledge of cognitive processes) to very tight (implementing the hypothesized cognitive processes in the user model). An overview of these levels is presented in Figure 1. In the next section we elaborate on these levels of integration by considering user models in AI-based applications that assist instruction in educational contexts. We demonstrate that user models that are more strongly integrated with formal models of human cognition, offer greater insurance that AI makes more valid and more fine-grained adaptations to an individual user.

Three levels of Integration

At the loosest level of integration, the user model is mainly inspired by discovery-oriented research and does not involve any formal specification of human cognition and how it leads to a specific behavior. For instance, it is a generally accepted fact that human short-term memory is limited in capacity (Cowan, 2001). This knowledge could inspire the elaboration of a user model without any explicit specification of the cognitive machinery underlying this capacity limitation. In other words, only general or anecdotal knowledge is used. Hence, we refer to this level as the *anecdotal level* of integration. The second level of integration, the *computational level*, involves user models that consider the computations that are hypothesized to be executed by the users, but ignore the specification of the mental processes giving rise to those computations (Marr, 1982). Building upon our previous example, the user model thus now represents specific short-term memory capacity limits of a user, but does not identify the cognitive processes giving rise to these limitations. The tightest integration can be found at the *algorithmic level*, where a formal specification of mental processes that lead to mental computations is taken into account (Marr, 1982). Within our example, the mental processes leading to a particular short-term memory capacity, such as rehearsal processes, are thus now also formalized in the user model.

We further illustrate these levels in the domain of education in which AI-based applications are used to instruct and learn factual knowledge such as word pairs when learning a foreign language, or a list of all European capitals (Pavlik, 2007; Van Rijn et al., 2009; Sense et al., 2015, 2016, 2018). These applications are inspired by the classic flashcard method [e.g., Pimsleur Language Programs; Leitner (1972) system] in which to-be-learned facts are written out on a deck of cards that are presented one by one. Typically, a question is written on one side of the card, with the answer on the other side. If the learner answers correctly, then the card is put aside, but if the learner answers incorrectly, the card is placed at the bottom of the deck for repetition. When all cards are put aside, the deck is picked up again and the procedure is repeated (Leitner, 1972).

User models of computerized versions of the flashcard method [e.g., SuperMemo, (Wozniak and Gorzelanczyk, 1994)] keep track of users' performance (e.g., errors) so that retention can be improved in three ways. First, by increasing the frequency by which facts are repeated. Consequently, learners display

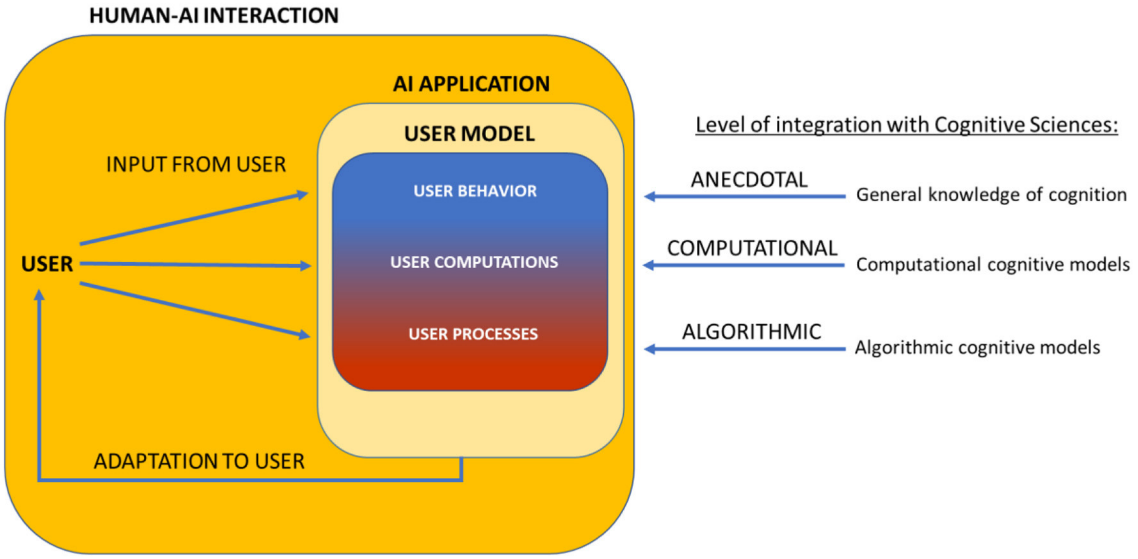

FIGURE 1
User models in Human-AI interaction can represent the user to different degrees depending on how strongly they are integrated with models from cognitive sciences. At the anecdotal level the user model is inspired from general knowledge of human cognition and does not represent the cognitive state of the user. At the computational level, computations of the user are represented. At the algorithmic level, the user model also incorporates the cognitive processes of the user.

a learning curve or repetition effect, which indicates that performance improves with practice and learning (Ebbinghaus, 1885; Newell and Rosenbloom, 1981; Anderson et al., 1999). Second, by spacing out the presentation of facts evenly in time, which also improves retention [i.e., spacing effect, (Ebbinghaus, 1885; Pavlik and Anderson, 2005)]. Third, by repeatedly testing subjects, which improves recall [i.e., testing effect; (Roediger and Karpicke, 2006a,b)]. These user models have been developed with different levels of integration and vary in the degree to which they can adapt to the user.

At the anecdotal level, the user model registers learning performance and adapts the frequency of repetitions, spacing and number of tests. However, the user model in itself does not include a representation of the mental state of the user or the cognitive processes that are mediating the user's performance. The application thus only adapts to the user's performance on the basis of general principles (repetition effect, spacing effect, and testing effect). Because user performance is directly used as input for this adaptation it is furthermore difficult to control whether this performance reflects the cognitive construct of interest or some uncontrolled mediator (e.g., fatigue) and if the adaption is thus valid.

At the computational level of integration, a representation is made of the user's ability, which underlies the performance of that user, while they execute the task [e.g., (Zhang et al., 2016)]. User models at the computational level can be situated in the Fechnerian tradition of mathematical modeling, which aims to discover functional relationships between observable and metaphysical quantities. To this end, behavioral measures are mapped onto psychological concepts *via* mathematical principles. From this perspective, the user model defines psychological concepts mathematically and behavior is interpreted in light of these concepts. Such an approach is related to cognitive psychometrics [e.g., (Riefer et al., 2002)].

Recent applications in fact learning, which employ user models that are situated at the computational level of integration [e.g., LanguageGarden, (Klinkenberg et al., 2011)] are based on item-response theory [IRT; (Rasch, 1960)]. The idea behind IRT is that the probability that a person has retained a fact (and will answer correctly when tested) is a combination of the difficulty of that fact and the learning ability of the individual. A user model based on IRT simultaneously updates the estimated difficulty of the facts and the estimated ability of the learners by comparing the probability of retaining a fact with the actual outcome on a given test (Klinkenberg et al., 2011; Pelánek, 2016; Pelánek et al., 2017). These estimates are then used to select the next to-be-presented fact. Such a specification ensures that individual learners are presented with facts that are within their reach, but also that facts are repeated in a spaced schedule. This follows because when an individual correctly recalls a fact, both the individual's ability and the item's difficulty are re-estimated, such that the difference between ability and difficulty increases. The adaptations made by the application are thus now based on an explicit representation of the user's cognitive computations that are hypothesized to underlie the observed psychological effects of repetition, spacing and testing.

However, a parameter in a computational model in itself does not guarantee that it reflects a particular feature of cognition and always needs to be validated empirically (Heathcote et al., 2015). Greater validity can be obtained with user models that integrate cognition at the algorithmic level aim to represent the cognitive processes of the user, rather than only the outcome or computations of these processes. To this end, these user models are based on formal models that describe specific cognitive processes [e.g., (van Maanen and Marewski, 2009; Van Maanen et al., 2010)]. An example of an influential formal model that has been used in the domain of fact learning is the ACT-R theory of declarative memory (Anderson and Schooler, 1991a; Pavlik and Anderson, 2005). This theory proposes that memory traces of declarative facts reflect the probability of requiring to recall these facts in the immediate future[1]. This probability or activation is computed on the basis of previous encounters with declarative facts. In particular, the activation is considered to be the highest immediately after a successful recall moment or immediately after a study moment. Following these recall and/or test moments, activation decays with a particular forgetting rate that is specific to the difficulty of the item that is learned and the learning ability of the user. The sum of activation to all encounters of a specific fact determines the probability of needing that fact in the immediate future, as well as a probability of recall of that fact (Anderson and Milson, 1989; Anderson and Schooler, 1991b). This cognitive model of declarative memory has been shown to predict response times and accuracy scores in numerous experiments, including standard memory paradigms (Anderson et al., 1999; Pavlik and Anderson, 2005; Schneider and Anderson, 2012), but also extending to more complex cognitive behavior that involves retrieval of information from memory (Van Rij et al., 2010; Schneider and Anderson, 2011; Banks, 2013).

In the domain of fact learning, the ACT-R theory of declarative memory has been applied in RuggedLearning (Van Rijn et al., 2009; Sense et al., 2015, 2016, 2018). This system uses the activation values of all facts for a particular user to determine which fact has a probability of recall that will drop below a particular threshold in the immediate future. This fact is then selected for the subsequent test. The activation value of that fact is also used to predict a response time of the test. The deviation from the response time is used to calibrate the parameters of the model to best predict the observed recall and response times (Van Rijn et al., 2009). The user model thus now represents assumptions of cognitive processes and adapts the task parameters accordingly. As a result, the user model predicts the effects of repetition, spacing and testing based on the presumed cognitive processes of an individual.

Discussion

We have illustrated that user models and cognitive models can be integrated at three different levels. Whereas, each level has specific characteristics that can help in the design of a user model, we emphasize that the boundaries between each level are not strict and intermediate levels of integration can be conceived. When considering the different levels of integration in the domain of fact learning, user models designed at the algorithmic and computational level offer greater insurance that the behavior of a user is related to the correct mental states underlying that behavior. The reason for this is that these user models are grounded in formal models of human cognition in which the pathway by which a mental construct leads to a particular behavior is explicated. Accordingly, these models are safer to use in the context of AI applications in which the reverse inference is made, namely from observable behavior to mental construct.

An additional advantage of user models at the computational and algorithmic level is that they offer new avenues for developing explainable AI. Explainable AI refers to AI systems that attempt to provide insight in their decision-making steps to human operators (Gunning et al., 2019; Babic et al., 2021). Not only does explainable AI serve to improve human-AI interaction, but it also helps to make decisions made by AI-algorithms more transparent in society (Ritter et al., 2017), which has recently been identified as one of the main challenges in the future of AI (Schwartz et al., 2022). In recent years it has been advocated that social sciences plays an important role in improving explainable AI. On the one hand, social sciences have offered insights about what constitutes a good explanation (Miller et al., 2017). On the other hand, social sciences and more specifically experimental cognitive psychology has developed research methods that can help unravel the decision processes that are fulfilled by deep neural networks (Taylor and Taylor, 2021a). Here, we argue that explanations offered by an AI system are incorrect when the relation between observable behavior and underlying associated construct is wrong. For instance, if an AI application targets an invalid behavioral proxy, e.g., facial expression as in indicator of threat, then this also invalidates also the explanation provided by that system for the decision it made.

[1] The ACT-R theory of declarative memory is based on a rational analysis of memory (Anderson and Milson, 1989; Anderson and Schooler, 1991b; Oaksford and Chater, 1994). Rational analyses of cognition are typically considered computational models, as they relate the computations of a cognitive system to environmental demands, such as the probability that facts are required. The ACT-R theory of declarative retrieval specifies potential algorithms that provide such computations. Thus, a rational analysis of memory not only provides estimates of cognitive computations or outcomes such as IRT, but also determines explicit assumptions about the cognitive processes that give rise to these parameters.

We propose that by using computational or algorithmic user models, AI decision-making can be understood by examining the cognitive models that explicate the reasoning steps taken by the AI system. For example, the aforementioned RuggedLearning application estimates a *rate of forgetting parameter* for each individual user. This parameter informs the decision to adapt the sequence of factual information that needs to be learned. Because the forgetting parameter reflects a relevant cognitive process (that is, memory persistence), it helps to formulate a transparent explanation for the changes made by the AI system. That is, the instruction system provides more learning opportunities, not because general principles of cognition were implemented in the user model (anecdotal level), but because we can identify a parameter value that represents a cognitive process (algorithmic level). Specifically, RuggedLearning provides more learning opportunities when the rate of forgetting parameter is low, *because* the individual has more difficulty retaining the facts.

Of course, we acknowledge that user models not only require the formal representation of cognitive traits such as the user's learning ability. Also more social traits of the user, such as attitudes, likes and dislikes are important. Hence, user models will need to find inspiration in a broad range of (social) cognitive models that formalize personality traits and attitudes [e.g., (Broekens et al., 2013; Moutoussis et al., 2014; Bosse, 2017; Dalege et al., 2018)] in addition to cognitive processes. For instance, a formal theory specifying how particular facial expressions relate to their corresponding mental state would advance the intelligent detection of deception on the basis of facial expressions (Rothwell et al., 2006), validating the decisions made by the AI application (DePaulo et al., 2003) and increasing trust in the system (Ishowo-Oloko et al., 2019). In addition, even if the development of cognitive models leads to a greater scrutiny in the interpretation of behavioral proxies, it remains possible that, over time, a computational or algorithmic model of human cognition proves wrong or incomplete. Consequently, a tight integration with a particular theory of cognitive processing might eventually yield incorrect or suboptimal decisions by the AI system [cf. alternative models in the fact learning domain (Khajah et al., 2014; Lindsey et al., 2014)]. Similarly, algorithmic and even computational models may not be readily available, forcing developers to adopt the anecdotal level until tighter levels of integration become available.

Cognitive theories and models may become invalid over time and are often restricted to a specific domain (e.g., attention, language,...) (Newell, 1973). Accordingly, it could be argued that user models will benefit more by only representing the dynamics of user behavior by means of functional models that formalize the relation between context and behavior without calling upon mediating cognitive processes (Chiesa, 1992). Skinners' conceptualization of teaching machines (Skinner, 1961) offers an early example of how user models can be devised on the basis of the experimental analysis of behavior and formal models are also available in that domain (Mazur, 2006). As pointed out in the Introduction, many user models are functional in nature and useful in a variety of AI applications. However, we believe that such models have difficulties to guarantee that an AI application makes valid inferences about user behavior as they do no consider the cognitive processes underlying that behavior. Furthermore, it has been questioned whether it is computationally possible for AI to make complex inferences about a user, when only observable behavior is available (Armstrong and Mindermann, 2018) and using (formal) models of human cognition may be helpful to mitigate this problem (Hélie and Pizlo, 2022; Langley et al., 2022).

The current perspective can be considered in view of recent developments to endow AI with a Theory of Mind (Premack and Woodruff, 1978) [ToM, e.g., Baker et al., 2010; Cuzzolin et al., 2020; Nguyen and Gonzalez, 2022; for a review see, Langley et al. (2022)]. Whereas a review of these developments is beyond the scope of the present endeavor, we believe that the levels of integration we propose can be helpful when researchers in AI seek inspiration in cognitive (neuro)sciences to develop Machine ToM. Advances in research on ToM are sometimes based on discovery-oriented research (e.g., Wang et al., 2017), which is may be tedious when making inferences about cognition on the basis of behavior (Oberauer and Lewandowsky, 2019). Developing AI models on the basis of such research will result in an anecdotal level of integration between AI and cognitive sciences. As such, researchers in AI may rather use formal of models of human cognition as a basis for their developments [see Nguyen and Gonzalez (2022) for an example]. In general, our framework can thus help in searching for and critically interpreting research in cognitive (neuro)sciences, which has been under heavy debate (Nosek et al., 2015; IJzerman et al., 2020).

To conclude, previous work already highlighted the importance of cognitive psychology (Taylor and Taylor, 2021b) and cognitive neurosciences (Hassabis et al., 2017) in further advancing insights in AI. Typically, these disciplines are considered to be useful in disentangling the so-called black box of artificial cognition by providing research approaches that were developed for the study of human cognition. In the present endeavor we join this position by demonstrating that cognitive models not only help to understand what was engineered, but can also contribute to the engineering itself.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

Both authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Anderson, J. R., Fincham, J. M., and Douglass, S. (1999). Practice and retention: a unifying analysis. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 1120–1136.

Anderson, J. R., and Milson, R. (1989). Human memory: an adaptive perspective. *Psychol. Rev.* 96, 703.

Anderson, J. R., and Schooler, L. J. (1991a). Reflections of the environment in memory. *Psychol. Sci.* 2, 396–408.

Anderson, J. R., and Schooler, L. J. (1991b). Reflections of the environment in memory. *Psychol. Sci.* 2, 396–408.

Armstrong, S., and Mindermann, S. (2018). "Occam' s razor is insufficient to infer the preferences of irrational agents," in *Advances in Neural Information Processing Systems,* vol. 31 (Red Hook, NY: Curran Associates, Inc.).

Babic, B., Gerke, S., Evgeniou, T., and Cohen, I. G. (2021). Beware explanations from AI in health care. *Science* 373, 284–286. doi: 10.1126/science.abg1834

Baker, C. L., Saxe, R., and Tenenbaum, J. (2010). "Bayesian theory of mind," in *Proceedings of the Annual Meeting of the Cognitive Science Society* (Boston, MA), 33.

Banks, A. P. (2013). The influence of activation level on belief bias in relational reasoning. *Cogn. Sci.* 37, 544–577. doi: 10.1111/cogs.12017

Bonnefon, J.-F., and Rahwan, I. (2020). Machine thinking, fast and slow. *Trends Cogn. Sci.* 24, 1019–1027. doi: 10.1016/j.tics.2020.09.007

Borsboom, D., Mellenbergh, G. J., and van Heerden, J. (2004). The concept of validity. *Psychol. Rev.* 111, 1061–1071. doi: 10.1037/0033-295X.111.4.1061

Bosse, T. (2017). "Chapter 13: on computational models of emotion regulation and their applications within HCI," in *Emotions and Affect in Human Factors and Human-Computer Interaction*, ed M. Jeon (New York, NY: Academic Press), 311–337. doi: 10.1016/B978-0-12-801851-4.00013-6

Bosse, T., and Hoogendoorn, M. (2015). Special issue on advances in applied artificial intelligence. *Appl. Intell.* 42, 1–2. doi: 10.1007/s10489-014-0588-z

Broekens, J., Bosse, T., and Marsella, S. C. (2013). Challenges in computational modeling of affective processes. *IEEE Trans. Affect. Comput.* 4, 242–245. doi: 10.1109/T-AFFC.2013.23

Chiesa, M. (1992). Radical behaviorism and scientific frameworks: from mechanistic to relational accounts. *Am. Psychol.* 47, 1287–1299.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci.* 24, 87–114. doi: 10.1017/S0140525X01003922

Cuzzolin, F., Morelli, A., Cîrstea, B., and Sahakian, B. J. (2020). Knowing me, knowing you: theory of mind in AI. *Psychol. Med.* 50, 1057–1061. doi: 10.1017/S0033291720000835

Dalege, J., Borsboom, D., van Harreveld, F., and van der Maas, H. L. (2018). The attitudinal entropy (AE) framework as a general theory of individual attitudes. *Psychol. Inq.* 29, 175–193. doi: 10.1080/1047840X.2018.1537246

De Houwer, J. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspect. Psychol. Sci.* 6, 202–209. doi: 10.1177/1745691611400238

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., and Cooper, H. (2003). Cues to deception. *Psychol. Bull.* 129, 74–118. doi: 10.1037/0033-2909.129.1.74

Ebbinghaus, H. (1885). *Über das gedächtnis: untersuchungen zur experimentellen psychologie*. Berlin: Duncker and Humblot.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., and Yang, G. Z. (2019). XAI—explainable artificial intelligence. *Sci. Robot.* 4, eaay7120. doi: 10.1126/scirobotics.aay7120

Hassabis, D., Kumaran, D., Summerfield, C., and Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron* 95, 245–258. doi: 10.1016/j.neuron.2017.06.011

Heathcote, A., Brown, S. D., and Wagenmakers, E.-J. (2015). "An introduction to good practices in cognitive modeling," in *An Introduction to Model-Based Cognitive Neuroscience*, eds B. U. Forstmann, and E.-J. Wagenmakers (New York, NY: Springer), 25–48. doi: 10.1007/978-1-4939-2236-9_2

Hélie, S., and Pizlo, Z. (2022). When is psychology research useful in artificial intelligence? A case for reducing computational complexity in problem solving. *Top. Cogn. Sci.* 14, 687–701. doi: 10.1111/tops.12572

Hmoud, B. (2021). "The adoption of artificial intelligence in human resource management and the role of human resources," in *Forum Scientiae Oeconomia. Vol. 9.* Wydawnictwo Naukowe Akademii WSB, 105–118.

IJzerman, H., Lewis, N. A., Przybylski, A. K., Weinstein, N., DeBruine, L., Ritchie, S. J., et al. (2020). Use caution when applying behavioural science to policy. *Nat. Hum. Behav.* 4, 1092–1094. doi: 10.1038/s41562-020-00990-w

Ishowo-Oloko, F., Bonnefon, J. F., Soroye, Z., Crandall, J., Rahwan, I., and Rahwan, T. (2019). Behavioural evidence for a transparency–efficiency tradeoff in human–machine cooperation. *Nat. Mach. Intell.* 1, 517–521. doi: 10.1038/s42256-019-0113-5

Jupe, L. M., and Keatley, D. A. (2020). Airport artificial intelligence can detect deception: or am i lying? *Secur. J.* 33, 622–635. doi: 10.1057/s41284-019-00204-7

Khajah, M. M., Lindsey, R. V., and Mozer, M. C. (2014). Maximizing students' retention *via* spaced review: practical guidance from computational models of memory. *Top. Cogn. Sci.* 6, 157–169. doi: 10.1111/tops.12077

Klinkenberg, S., Straatemeier, M., and van der Maas, H. L. (2011). Computer adaptive practice of maths ability using a new item response model for on the fly ability and difficulty estimation. *Comput. Educ.* 57, 1813–1824. doi: 10.1016/j.compedu.2011.02.003

Langley, C., Cirstea, B. I., Cuzzolin, F., and Sahakian, B. J. (2022). Theory of mind and preference learning at the interface of cognitive science, neuroscience, and AI: a review. *Front. Artif. Intell.* 5, 62. doi: 10.3389/frai.2022.778852

Leitner, S. (1972). *So Lernt Man Lernen: Der Weg zum Erfolg—Google Scholar*. Alberta: Herder.

Lindsey, R. V., Shroyer, J. D., Pashler, H., and Mozer, M. C. (2014). Improving students' long-term knowledge retention through personalized review. *Psychol. Sci.* 25, 639–647. doi: 10.1177/0956797613504302

Marr, D. (1982). *Vision*. New York, NY: W. H. Freeman

Mazur, J. E. (2006). Mathematical models and the experimental analysis of behavior. *J. Exp. Anal. Behav.* 85, 275–291. doi: 10.1901/jeab.2006.65-05

Miller, T., Howe, P., and Sonenberg, L. (2017). Explainable AI: Beware of inmates running the asylum or: How I learnt to stop worrying and love the social and behavioural sciences. *arXiv. Preprint.* arXiv:1712.00547.

Moutoussis, M., Trujillo-Barreto, N. J., El-Deredy, W., Dolan, R., and Friston, K. (2014). A formal model of interpersonal inference. *Front. Hum. Neurosci.* 8, 160. doi: 10.3389/fnhum.2014.00160

Newell, A. (1973). "You can't play 20 questions with nature and win: projective comments on the papers of this symposium," in *Visual Information Processing* (Amsterdam: Elsevier), 283–308. doi: 10.1016/B978-0-12-170150-5.50012-3

Newell, A., and Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. *Cogn. Ski. Their Acquis.* 1, 1–55.

Nguyen, T. N., and Gonzalez, C. (2022). Theory of mind from observation in cognitive models and humans. *Top. Cogn. Sci.* 14, 665–686. doi: 10.1111/tops.12553

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., et al. (2015). Promoting an open research culture. *Science* 348, 1422–1425. doi: 10.1126/science.aab2374

Oaksford, M., and Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Pshycol. Rev.* 24, 608.

Oberauer, K., and Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychon. Bull. Rev.* 26, 1596–1618. doi: 10.3758/s13423-019-01645-2

Pavlik, P. I. (2007). Understanding and applying the dynamics of test practice and study practice. *Instr. Sci.* 35, 407–441. doi: 10.1007/s11251-006-9013-2

Pavlik, P. I., and Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: an activation-based model of the spacing effect. *Cognit. Sci.* 29, 559–586. doi: 10.1207/s15516709cog0000_14

Pelánek, R. (2016). Applications of the Elo rating system in adaptive educational systems. *Comput. Educ.* 98, 169–179. doi: 10.1016/j.compedu.2016.03.017

Pelánek, R., Papoušek, J., Rihák, J., Stanislav, V., and NiŽnan, J. (2017). Elo-based learner modeling for the adaptive practice of facts. *User Model. User-Adapt. Interact.* 27, 89–118. doi: 10.1007/s11257-016-9185-7

Poldrack, R. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends Cogn. Sci.* 10, 59–63. doi: 10.1016/j.tics.2005.12.004

Premack, D., and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* 1, 515–526.

Rabinowitz, N., Perbet, F., Song, F., Zhang, C., Eslami, S. A., and Botvinick, M. (2018). "Machine theory of mind," in *Proceedings of the 35th International Conference on Machine Learning* (Valencia: PMLR), 4218–4227.

Rasch, G. (1960). *Studies in Mathematical Psychology: I. Probabilistic Models for Some Intelligence and Attainment Tests.* Nielsen & Lydiche.

Riefer, D. M., Knapp, B. R., Batchelder, W. H., Bamber, D., and Manifold, V. (2002). Cognitive psychometrics: assessing storage and retrieval deficits in special populations with multinomial processing tree models. *Psychol. Assess.* 14, 184–201. doi: 10.1037/1040-3590.14.2.184

Ritter, S., Barrett, D. G., Santoro, A., and Botvinick, M. M. (2017). "Cognitive psychology for deep neural networks: a shape bias case study," in *International Conference on Machine Learning* (Valencia: PMLR), 2940–2949.

Roediger, H. L., and Karpicke, J. D. (2006a). Test-enhanced learning: TAKING memory tests improves long-term retention. *Psychol. Sci.* 17, 249–255. doi: 10.1111/j.1467-9280.2006.01693.x

Roediger, H. L., and Karpicke, J. D. (2006b). The power of testing memory: basic research and implications for educational practice. *Perspect. Psychol. Sci.* 1, 181–210. doi: 10.1111/j.1745-6916.2006.00012.x

Rothwell, J., Bandar, Z., O'Shea, J., and McLean, D. (2006). Silent talker: a new computer-based system for the analysis of facial cues to deception. *Appl. Cogn. Psychol.* 20, 757–777. doi: 10.1002/acp.1204

Schneider, D. W., and Anderson, J. R. (2011). A memory-based model of hick's law. *Cognit. Psychol.* 62, 193–222. doi: 10.1016/j.cogpsych.2010.11.001

Schneider, D. W., and Anderson, J. R. (2012). Modeling fan effects on the time course of associative recognition. *Cognit. Psychol.* 64, 127–160. doi: 10.1016/j.cogpsych.2011.11.001

Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., and Hall, P. (2022). *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*. National Institute of Standards and Technology Special Publication 1270, U.S Department of Commerce. doi: 10.6028/NIST.SP.1270

Sense, F., Behrens, F., Meijer, R., and Rijn, H. (2015). *Stability of Individual Parameters in a Model of Optimal Fact Learning*. Groningen: University of Groningen.

Sense, F., Behrens, F., Meijer, R. R., and van Rijn, H. (2016). An individual's rate of forgetting is stable over time but differs across materials. *Top. Cogn. Sci.* 8, 305–321. doi: 10.1111/tops.12183

Sense, F., Meijer, R. R., and van Rijn, H. (2018). Exploration of the rate of forgetting as a domain-specific individual differences measure. *Front. Educ.* 3, 112. doi: 10.3389/feduc.2018.00112

Sethumadhavan, A., and Phisuthikul, A. (2019). Can machines detect emotions? *Ergon. Des.* 27, 30. doi: 10.1177/1064804619847190

Skinner, B. F. (1961). Teaching machines. *Sci. Am.* 205, 90–106.

Sporrel, K., Nibbeling, N., Wang, S., Ettema, D., and Simons, M. (2021). Unraveling mobile health exercise interventions for adults: scoping review on the implementations and designs of persuasive strategies. *JMIR MHealth UHealth* 9, e16282. doi: 10.2196/16282

Taylor, J. E. T., and Taylor, G. W. (2021a). Artificial cognition: how experimental psychology can help generate explainable artificial intelligence. *Psychon. Bull. Rev.* 28, 454–475. doi: 10.3758/s13423-020-01825-5

Taylor, J. T., and Taylor, W. T. (2021b). "Software architecture," in *Patterns in the Machine* (New York, NY: Springer), 63–82. doi: 10.1007/978-1-4842-6440-9_5

van Maanen, L., and Marewski, J. N. (2009). Recommender systems for literature selection: a competition between decision making and memory models, *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (Austin, TX: Cognitive Science Society), 2914–2919.

Van Maanen, L., Van Rijn, H., Van Grootel, M., Kemna, S., Klomp, M., and Scholtens, E. (2010). Personal publication assistant: abstract recommendations by a cognitive model. *Cogn. Syst. Res.* 11, 120–129. doi: 10.1016/j.cogsys.2008.08.002

Van Rij, J., Van Rijn, H., and Hendriks, P. (2010). Cognitive architectures and language acquisition: a case study in pronoun comprehension. *J. Child Lang.* 37, 731–766. doi: 10.1017/S0305000909990560

Van Rijn, H., Van Maanen, L., and Van Woudenberg, M. (2009). "Passing the test: improving learning gains by balancing spacing and testing effects," in *Proceedings of the 9th international conference of cognitive modelling* (Manchester), 7–6.

Wahlster, W., and Kobsa, A. (1989). "User models in dialog systems," in *User Models in Dialog Systems,* eds A. Kobsa, and W. Wahlster (New York, NY: Springer), 4–34. doi: 10.1007/978-3-642-83230-7_1

Wang, S., Scheider, S., Sporrel, K., Deutekom, M., Timmer, J., and Kröse, B. (2021a). What are good situations for running? A machine learning study using mobile and geographical data. *Front. Public Health* 8, 536370. doi: 10.3389/fpubh.2020.536370

Wang, S., Sporrel, K., van Hoof, H., Simons, M., de Boer, R. D., Ettema, D., et al. (2021b). Reinforcement learning to send reminders at right moments in smartphone exercise application: a feasibility study. *Int. J. Environ. Res. Public. Health* 18, 6059. doi: 10.3390/ijerph18116059

Wang, Y., Zhang, Y. B., Liu, L. L., Cui, J. F., Wang, J., Shum, D. H., et al. (2017). A meta-analysis of working memory impairments in autism spectrum disorders. *Neuropsychol. Rev.* 27, 46–61. doi: 10.1007/s11065-016-9336-y

Wozniak, P. A., and Gorzelanczyk, E. J. (1994). Optimization of repetition spacing in the practice of learning. *Acta Neurobiol. Exp.* 4, 59.

Zhang, C., van Wissen, A., Lakens, D., Vanschoren, J., De Ruyter, B., and IJsselsteijn, W. A. (2016). "Anticipating habit formation: a psychological computing approach to behavior change support," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct* (Heidelberg), 1247–1254. doi: 10.1145/2968219.2968439