A Metacognitive Triggering Mechanism for Anticipatory Thinking

Alexander R. Hough[1,2], Othalia Larue[1], and Ion Juvina[1]

[1]Wright State University, Psychology Department, ASTECCA laboratory, Dayton, OH 45435, USA
[2]ORISE at Air Force Research Laboratory, Wright-Patterson Air Force Base, OH 45433, USA
hough.15@wright.edu, othalia.larue@wright.edu, and ion.juvina@wright.edu

Abstract

Current autonomous systems have the ability to adapt to environmental changes in real-time, but limited ability to engage in anticipatory thinking (AT) with the flexibility to generalize and consider hypothetical future situations. We argue that metacognitive processes are important for an and provide supporting literature primarily from psychology. As an example, we present a metacognitive monitoring mechanism implemented in a cognitive model and discuss ways to extend the mechanism to allow for dynamic behavior and anticipatory thinking capabilities.

Anticipatory Thinking

Anticipatory thinking (AT) is the deliberate exploration and consideration of hypothetical future outcomes in order to identify an appropriate action or plan (Amos-Binks and Dannenhauer 2019; Geden et al. 2018). AT involves an array of cognitive processes (Klein et al., 2003; Koziol, Budding, and Chidekel 2012), such as mental simulation, recognition, preparation, and development of expectancies, which are not completely understood prior to an event occurring (Klein, Snowden, and Pin 2011; Warwick and Hutton 2007). It is considered distinct from prediction (Klein et al. 2011) and is described as gambling with attention in hopes of directing it towards the most relevant event (Klein et al. 2007).

Geden et al. (2019) identified three forms of AT: how past states led to current states (retrospective branching), anticipating future states and their indicators (prospective branching), and focusing on a potential future and working backwards (backcasting). Klein et al. (2011) takes a more naturalistic approach to AT emphasizing the detection of discrepancies through recognition and degree of match between past, current, and future situations (pattern matching), using "trajectory" to prepare for the future and extrapolate trends (trajectory tracking), and being mindful of connections, implications, and interdependencies betw een events (Conditional). Geden et al. (2019) and Klein et al. (2011) have slightly different approaches, however, they both identified similar AT processes: recognizing feature and cue relationships between situations, extrapolation or generalization to other states, and construction of mental models based on available evidence.

Anticipating future events is crucial. The real world is dynamic, often unpredictable, may have ill-defined goals, and may involve high stakes. The ability to generate, use, and reason about plans or goals to direct behavior and adapt to changes is important for intelligent behavior (Newell and Simon 1972; Schank and Abelson 1977) and autonomy (Johnson et al. 2016; Vattam et al. 2013). Goal-driven behavior leverages discrepancies between expectations and the environment in real-time, and when detected, they are addressed by modifying goals, reasoning about goals, and learning (Aha 2018; Cox et al. 2016; Muñoz, et al. 2019; Pozanco, Fernández, and Borrajo 2018; Roberts et al. 2018). Amos-Binks and Dannenhauer (2019) suggest most current systems lack AT capabilities, such as the ability to address unknown hypothetical future events by identifying and avoiding errors or discrepancies before they occur, and how to effectively trade off costs of computation and benefits of considering a large number of possible futures. AT systems need to strike a balance between flexibility and stability in order to adapt to dynamic real-world environments before conditions change (Bratman, Israel, and Pollack 1988).

Klein et al. (2011) suggest good AT requires one to be sensitive to the constraints and affordances based on their own beliefs, capabilities, and the current situation. Metacognitive monitoring could help overcome the computation problem and some of the barriers to AT identified by Klein et al. (2011), such as taking a passive stance, becoming fixated on patterns, explaining away evidence or interpretations, and being overconfident. Similar to AT as a metacognitive capability (Amos-Binks and Dannenhauer 2019), we emphasize how the calibration between metacognitive monitoring and reality could help indicate when AT is needed, how to accomplish it efficiently, and reduce the number of futures to consider. We explore psychological metacognitive measurements regarding conflict detection and resolving processes in simpler tasks and discuss how these capabilities could be extended to AT.

A Critical Role for Metacognition

Humans use heuristics to make efficient and accurate decisions (Cosmidies and Tooby 1996; Gigerenzer and Gaissmaier 2011), but this can lead to systematic error in inappropriate, novel, or misleading environments (Evans and Stanovich 2013; Kahneman 2011; Kahneman and Klein 2009). A critical ability is recognizing when an approach is inadequate and suppressing it to come up with an alternative (Stanovich 2018). This ability is metacognition, which serves to detect conflict or mismatch regarding an environment and a strategy, type of processing, or expectation. This detection depends on predictability and cues in the environment, ability to recognize relevant cues, and whether goals are reachable (Dannenhauer et al. 2018; Evans and Stanovich 2013; Johnson et al. 2016; Klein 1998; Klein et al. 2007; Pennycook, Fugelsang, and Koehler 2015a; Stanovich 2018; Vattam, et al. 2013). This process is referred to as metacognitive monitoring or experience, which provides feedback, leads to control decisions, activates knowledge, and can be calibrated through experience leading to better regulation behavior (Efklides 2006; Efklides, Samara, and Petropoulou 1999; Flavell 1979). Conflict often triggers the need for a different approach towards solving a problem or completing a task (Butcher and Sumner 2011; Dannenahuer et al. 2018; Pennycook 2017; Stanovich 2018). However, it does not always lead to efficient processing (Pennycook et al. 2015a, 2015b; Swan, Calvillo, and Revlin 2018) or solutions to resolve the conflict (Novick and Holyoak 1991).

Monitoring and Conflict

There are several methods for measuring metacognitive monitoring (e.g., Gascoine, Higgins, and Wall 2017). Two common methods in psychology are the performance-based Cognitive Reflection Test (CRT; Frederick 2005) that requires overriding a primed heuristic response for a more deliberate response, and the subjective-based Feeling of Rightness (FOR; Thompson et al. 2011) that indicates one's accuracy and awareness of their own metacognitive monitoring. Mata, Ferreira, and Sherman (2012) found that those with better metacognitive awareness as measured by the CRT were able to generate heuristic and deliberate answers, more accurately rate performance of others and themselves, and were able to better focus on the most relevant features of a problem. Epstein et al. (1996) found that the ability to shift between heuristic and deliberate thinking was better than exclusively relying on one. Metacognitive monitoring as measured by the CRT, FOR, and related tasks may be more related to actual intelligence than traditional measures, because it includes motivation and ability (Frederick 2005; Toplak, West, and Stanovich 2011). For instance, Barr et al. (2015) found the CRT positively correlates with cognitive ability, need for cognition, analogies, the remote associates test, and negatively with faith in intuition. Furthermore, the CRT correlates with cognitive ability, performance on heuristic and biases tasks, belief bias, rational thinking, set shifting, and working memory, and predicts rational thinking performance independent of intelligence, executive functioning, and thinking dispositions (Toplak et al. 2012).

Conflict associated with metacognitive experience has been measured using response times (De Neys and Glumicic 2008; Pennycook et al. 2015a), the FOR (Thompson et al. 2011), the CRT (Frederick 2005), and activation of specific brain regions including the anterior cingulate cortex (Croxson et al. 2009; Kennerley et al. 2009) and medial prefrontal cortex (Botvinick et al. 1999; Cohen, Botvinick, and Carter 2000) often associated with cognitive control. This conflict is still observed when manipulations are in place to minimize deliberation (Pennycook et al., 2015a; Thompson and Johnson, 2014), and error signals during comprehension (Glenberg, Wilkinson, and Epstein 1982; McNamara et al., 1996) and disfluency appear to prompt similar responses (Alter et al. 2007). Metacognitive monitoring appears to involve both top-down and bottom-up processes, where the willingness or motivation to engage in analytic thinking (e.g., CRT) appears to be top-down, while the detection of conflicts that triggers the engagement (e.g., FOR) appears to be bottom-up (Pennycook et al. 2015b; Stanovich 2018). Here, we focus primarily on bottom-up processes, but plan on further addressing top-down processes in future work.

Dynamic Behavior

Metacognitive monitoring is effective but not perfect. It may fail to detect conflict (Swan et al. 2018) or may result in faulty judgments after conflict is detected. Detection may not direct one to the necessary knowledge to solve the problem or implement a strategy (Novick and Holyoak 1991) and the outcome might be influenced by biases, such as overconfidence with naive individuals (Fischhoff 2012) or confirmation bias with the more experienced (Kahneman 2011). Klein et al. (2006b) acknowledge that detecting such a conflict or recognizing insufficient performance is important, but understanding how to modify thinking processes to address this problem is more valuable. Metacognition could help guide conflict resolution by helping to determine whether the environment calls for more deliberate processing or quick, less elaborate responding. Although humans are often good at sizing up the environment (Klein 1998) and making efficient tradeoffs between speed and accuracy (Payne, Bettman, and Johnson 1988), exerting mental effort is often experienced as aversive (Halpern 2014; Kahneman 2011) and may be avoided based on an individual's subjective cost of effort (Westbrook, Kester, and Braver 2013). Similarly, in AT the environment may favor considering more hypothetical future situations, exploring some more deeply, or by quickly anticipating and preparing for a few. Although typically applied to current events, metacognitive monitoring could be extended to future events to help determine which approach fits better with the environment. For instance, if an individual engages in the three types of AT (i.e., pattern matching, trajectory tracking, and conditionals) identified by Klein et al. (2011) regarding a hypothetical future, the presence of conflict among them could inform whether that hypothetical future is appropriate or if it should be discarded or modified. Metacognitive calibration could help determine the appropriate response based on one's understanding of their own abilities and knowledge, and how that corresponds to a situation. Research addressing the understanding process, sensemaking, critical thinking, forecasting, and counterfactual thinking provide examples of how to identify the source of conflict, how to make sense of and resolve it, and determine which potential outcomes are most likely.

The understanding process was recently defined in a multidisciplinary review as "The acquisition, organization, and appropriate use of knowledge to produce a response directed towards a goal, when that action is taken with awareness of its perceived purpose" (Hough and Gluck forthcoming, p. 11). The review revealed common features of understanding and discussed how computer science, education, psychology, and philosophy all emphasize the importance of metacognition for understanding capabilities. Metacognition was described as a self-evaluative feedback mechanism for identifying faulty knowledge or gaps that triggers additional processing or information search, which helps calibrate mental representations with the environment (Butcher and Sumner 2011; Forbus and Hinrichs 2006; Kirk and Laird 2014; Mayer 1998; Perkins 1998; Perkins and Simmons 1988; Woodward 2003). Better understanding, like expertise, could increase the quality of AT by directing attention toward the most relevant features.

Sensemaking models also involve components of understanding, such as abstraction of knowledge, development of relations, ability to transfer knowledge to distant situations, and often involves leveraging domain and context information to develop frames (Hough and Gluck forthcoming). Pirolli and Card's (2005) sensemaking model includes an information foraging loop (Pirolli and Card 1999) and sensemaking loop (Russell et al. 1993). During foraging, an individual engages in search and filtering of information and then applies effort to give it more structure in an iterative process. During sensemaking, one utilizes schemas to make hypotheses and conclusions, similar to the construction of a mental model (e.g., Johnson-Laird 2013). Although not explicit in the model, there appears to be a metacognitive process. If there is insufficient evidence for a hypothesis, a case cannot be built, or a discrepancy is detected then the agent goes back to the foraging loop to fill in the gaps or gather evidence for a new schema or hypothesis. Similarly, the data/frame model (Klein, Moon, and Hoffman 2006a, 2006b; Klein et al. 2007) does not explicitly mention metacognition, but does involve "questioning the frame" that includes anomaly detection or expectancy violations. If there is a discrepancy, the existing frame can be discarded, elaborated, preserved, reframed, or compared to another.

Critical thinking is described as the ability to explain, justify, extrapolate, relate, and apply in ways that go beyond knowledge and skill, and training in critical thinking and metacognitive monitoring can enhance understanding and generalizability (Halpern 1998, 2014; Willingham 2007). Similar to metacognitive monitoring, after controlling for cognitive ability, critical thinking correlates with the ability to avoid cognitive biases by thinking logically even when it conflicts with prior beliefs and thinking dispositions (West, Toplak, and Stanovich 2008). To better understand and teach these skills, Halpern (2010) developed a comprehensive measure, called the Halpern Critical Thinking Assessment (HCTA), which includes decision making, problem solving, hypothesis testing, argument analysis, likelihoods and uncertainties, and verbal reasoning. The HCTA has generalizability across various populations, positively correlates with years of education, and negatively correlates with the frequency of negative life events in a real-world outcome inventory (Butler 2012). Critical thinking has some similarities to sensemaking, but may be more generalizable and appropriate for AT with little available knowledge.

Forecasting involves predicting the probability that specific events will occur. Its associated processes could be used during AT to help reduce the unnecessary consideration of unlikely future outcomes or to determine which are more likely and should be better prepared for. Accurate forecasters typically have higher cognitive ability, motivation, CRT scores, and open-minded thinking (Juvina et al. under review; Mellers et al. 2015; Tetlock and Gardner 2015). In addition, they often respond faster, have better discriminability and calibration, and learn faster.

Counterfactual thinking occurs after an event is experienced and involves considering forgone outcomes (Byrne 2016; Kahneman and Miller 1986). It is more typical after failures or shortcomings (Hur 2001; Roese and Olsen 1997; Sanna and Turley 1996; Sanna and Turley-Ames 2000) and often involves ways to correct or improve upon previous behaviors (Markman et al. 1993; Roese 1997; Roese, Hur, and Pennington 1999). Epstude and Roese (2008) suggest that this may depend on the realization that there is a problem or goals are not sufficiently met, which is a form of metacognitive monitoring. Improving future outcomes may be achieved through goal-oriented reasoning (Epstude and Roese 2008; Roese and Epstude 2017) or by increases in motivation, persistence, and performance (Dyczewski and Markman 2012; Markman, McMullen, and Elizaga 2008). Although after the fact, this type of thinking could provide experience to help calibrate metacognitive processes, provide more constructive ways to think about future events, and help identify relevant alternative possibilities.

A Cognitive Model with Metacognitive Monitoring

We previously developed a model of the Wason card selection task (Wason 1966, 1968) with initial aspects of metacognitive monitoring (see Larue, Hough, and Juvina 2018 for a full description). Our approach was informed by mental models (Johnson-Laird, 2013) and dual process theories, specifically Stanovich's (2009) tripartite framework. Stanovich's (2009) framework provides an explanation of how reflective and adaptive (characterized by reactivity) human behavior emerges from the interaction of three distinct cognitive levels or "minds". The autonomous mind, responsible for fast behaviors, includes instinctive and over-learned processes, domain-specific knowledge, and emotional regulation. The algorithmic mind, responsible for cognitive control, can affect decoupling (i.e., simulation) and serial associative processes. The reflective mind, responsible for deliberative processing, can trigger or suppress the algorithmic minds' decoupling and serial associative processes. In this framework, the reflective mind would be the center for metacognitive monitoring.

Here we briefly describe our model and in the next paragraph, discuss how this model could be augmented and applied to AT. In the Wason card selection task, two cards (A and 7) out of four (A, D, 3, and 7) must be flipped over to verify a rule: If "A" is on one side, then there is a "3" on the other. "A" and "3" are intuitively compelling to flip over because they are both present in the rule. Flipping over "A" is necessary because it can falsify the rule, however, flipping over "3" is unnecessary because it may confirm the rule but it cannot falsify it. Two types of logical errors are common: the selection of the unnecessary card (3), and the non-selection of the necessary card (7). We believe selecting the unnecessary card results from a metacognitive failure to detect its inadequacy and not selecting the necessary card, involves incomplete decoupling after the detection and override of selecting the unnecessary card. The incomplete decoupling may result from participants applying *modus ponens* (if P then Q), but failing in the application of *modus tollens* (if not Q then not P) in a "partial insight" (Evans 1977; Wason 1969). Only flipping the correct intuitively compelling card (A) occurs because modus tollens needs to simulate more intermediary mental models compared to the modus ponens.

Our model was implemented in the ACT-R cognitive architecture (Anderson, 2007) with a core affect mechanism (see Juvina, et al. 2018) and a FOR component to drive decoupling behavior. Rethinking times, answer changes, and fluency are functions of the FOR. The FOR is computed based on the time required to achieve the initial retrieval of the answer for the two intuitively compelling cards through the initial priming rule (autonomous mind), and serves as a gateway for further processing. We use the temporal module in ACT-R (Taatgen, van Rijn, and Anderson 2007) to measure time in ticks, which are noisy and increase in time in a fashion similar to human time estimation. In the Wason task, the FOR is computed as a function of the time required to achieve the initial retrieval of "A" and "3" through the initial priming rule. When the FOR is high, the model goes with the initial answer (i.e., heuristic processing), but when low cognitive decoupling is launched by the reflective mind and carried out by the algorithmic mind. In the model, the time required to achieve the initial retrievals is assigned to FOR-inverse (e.g., higher time means lower FOR). When FOR-inverse is below threshold (see Figure 1), the model goes with the initial answer (i.e., type 1 processing). When FOR-inverse is above threshold, cognitive decoupling is launched and the model engages in further processing (i.e., type 2 processing) with representations that are copied from its working memory based on activation during open retrieval (those with highest activation are retrieved). The representations are used in an inner cognitive simulation to indicate which rules from the reflective mind can be applied. The process by which representations are copied and used in a separate inner simulation is "cognitive decoupling". The importance of further processing is a function of the FOR, which determines the extent a decoupling result (i.e., wrong, partial, and complete) is taken into account in the final answer. The model produces an answer when the valuation of a representation is above a certain threshold. The valuation and arousal values, which are sub-symbolic quantities added to the current sub-symbolic equations of ACT-R, help to define the core affect. When a reward is triggered, valuations are updated. Rewards are a function of the initial FOR (negative factor in the case of negative reward), which affects answer selection ("yes" or "no" answers are produced according to how the model "feels" about the answer) (see Figure 1).

A training procedure was used to simulate individual differences in heuristic and analytical behavior, where the different degrees of reinforcement allowed the model to learn logical skills and vary in FOR. Metacognitive monitoring implemented by the FOR determined if and how much additional processing occurred. Model simulations produced

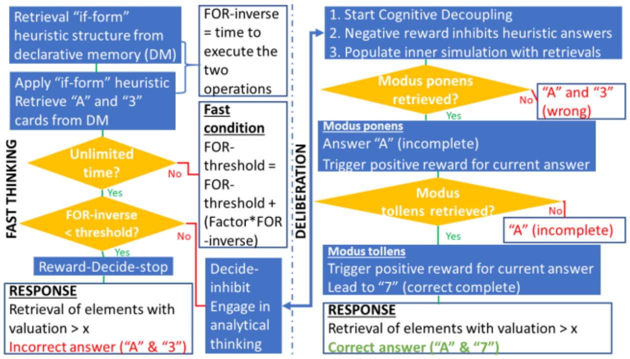

Figure 1. Answer Selection Processes in Our Cognitive Model

three types of outcomes that are typically observed in humans: (1) Complete reliance on the autonomous mind (no decoupling) leading to the observed common error (correctly selecting "A" and incorrectly selecting "3" guided by confirmation bias), (2) partial decoupling (partial insight) leading to correctly selecting "A" and not falling for the confirmation bias (already simulated in a possible world), and (3) complete decoupling allowing for the activation of the counter-information rule which is less often activated.

Discussion

We presented some perspectives in metacognitive monitoring research and some dynamic higher-level cognitive processes. This research informed the development of our model and we believe the interaction between the FOR (metacognitive monitoring) and decoupling (mental simulation) in our model could be applied to AT.

In the model discussed here, the FOR determines whether there is a need for more deliberation or a different approach to complete the task. The FOR could be extended to simulate future situations through decoupling and include how much the agent "knows" about the environment. For instance, the FOR could determine how long decoupling should continue (e.g., generating counterfactuals) and when it should stop. Sensemaking and critical thinking emphasize generating hypotheses and gathering evidence to indicate the degree each is supported. The FOR could be informative when there is a lack of knowledge, failure to find matching strategies or procedures, or a lack of evidence. A partial matching procedure could be used based on the degree of the FOR, a lower FOR could increase the breadth of future situations to consider, and a strategy or approach could be chosen out of a hierarchy based on the FOR. Learning can occur over time and counterfactuals could be generated and utilized for learning what could have happened based on the hypothetical actions corresponding to a given situation or the environment. Although these processes typically apply to situations in real time, they could be extended and applied to hypothetical future states through mental simulation. Furthermore, when hypothetical futures are considered they could be weighted based on predicting the probability of their occurrence through forecasting. This process in combination with the FOR could also help inform when enough hypothetical futures are considered. These types of processes could help highlight the most relevant and likely future outcomes, so that less preparation and planning is required. As the architecture learns more and is placed in context, it reacts to events that it might have previously encountered. If there is a match (i.e., full or partial) with a previously developed strategy from cognitive decoupling, this strategy is recorded and reinforced in the architecture. The reinforcement of this strategy will lead to its prioritization in procedural memory over some other strategies and its declarative components will be faster to retrieve in declarative memory. This means that the next time this strategy is used, the FOR will better calibrated, resulting in more accurate and adaptive behavior with the potential for AT capabilities.

Acknowledgements

This research was supported in part by Alex Hough's appointment to the Student Research Participation Program at the U.S. Air Force Research Laboratory, 711th Human Performance Wing, Cognitive Science, Models, and Agents Branch, administered by the Oak Ridge Institute for Science and Education. Alex would like to thank the Chief Scientist's office of the 711th Human Performance Wing for the opportunity to participate in the Repperager internship program. We also thank Kevin Gluck and two anonymous reviewers for constructive feedback on an earlier draft.

References

Alter, A. L.; Oppenheimer, D. M.; Epley, N.; and Eyre, R. N. 2007. Overcoming Intuition: Metacognitive Difficulty Activates Analytic Reasoning. *Journal of Experimental Psychology: General* 136(4): 569-576. psycnet.apa.org/doiLanding?doi=10.1037%2F0096-3445.136.4.569

Anderson, J. R. 2007. *How Can the Human Mind Occur in the Physical Universe?* New York: Oxford University Press.

Amos-Binks, A.; and Dannenhauer, D. 2019. Anticipatory Thinking: A Metacognitive Capability. arxiv.org/pdf/1906.12249.pdf

Barr, N.; Pennycook, G.; Stolz, J. A.; & Fugelsang, J. A. 2015. Reasoned Connections: A Dual-process Perspective on Creative Thought. *Thinking & Reasoning* 21(1): 61-75. doi.org/10.1080/13546783.2014.895915

Botvinick, M.; Nystrom, L. E.; Fissell, K.; Carter, C. S.; and Cohen, J. D. 1999. Conflict Monitoring Versus Selection-For-Action in Anterior Cingulate Cortex. *Nature* 402(6758): 179. doi.org/1710/13fb3f03a2bb0d707d2c73597f436e280728

Bratman, M. E.; Israel, D. J.; and Pollack, M. E. 1988. Plans and Resource-Bounded Practical Reasoning. *Computational Intelligence* 4: 349-355.
doi.org/10.1111/j.1467-8640.1988.tb00284.x

Butcher, K.; and Sumner, T. 2011. Self-Directed Learning and The Sensemaking Paradox. *Human-Computer Interaction* 26: 123-159. www.tandfonline.com/doi/abs/10.1080/07370024.2011.556552.

Butler, H. A. 2012. Halpern Critical Thinking Assessment Predicts Real-World Outcomes of Critical Thinking. *Applied Cognitive Psychology* 26: 721-729. doi.org/10.1002/acp.2851

Byrne, R. M. 2016. Counterfactual Thought. *Annual Review of Psychology* 67: 135-157. doi.org/10.1146/annurev-psych-122414-033249

Cohen, J. D.; Botvinick, M.; and Carter, C. S. 2000. Anterior Cingulate and Prefrontal Cortex: Who's in Control?. *Nature Neuroscience* 3(5): 421-423. doi.org/10.1038/74783

Cosmides, L.; and Tooby, J. 1996. Are Humans Good Intuitive Statisticians After All? Rethinking Some Conclusions from the Literature on Judgment Under Uncertainty. *Cognition* 58(1): 1-73. doi.org/10.1.1.131.8290.

Croxson, P. L.; Walton, M. E.; O'Reilly, J. X.; Behrens, T. E.; and Rushworth, M. F. 2009. Effort-Based Cost–Benefit Valuation and the Human Brain. *Journal of Neuroscience 29*(14): 4531-4541. https://doi.org/10.1523/JNEUROSCI.4515-08.2009

Cox, M. T.; Alavi, Z.; Dannenhauer, D.; Eyorokon, V.; Munoz-Avila, H.; and Perlis, D. 2016. MIDCA: A Metacognitive, Integrated Dual-Cycle Architecture for Self-Regulated Autonomy. In Proceedings of the Thirteenth AAAI Conference on Artificial Intelligence. Phoenix, AZ: AAAI Press. www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/download/12292/12151

Dannenhauer, D.; Cox, M. T.; and Munoz-Avila, H. 2018. Declarative Metacognitive Expectations for High-Level Cognition. *Advances in Cognitive Systems* 6: 231-250. www.cogsys.org/papers/ACSvol6/papers/paper-6-15.pdf

De Neys, W.; and Glumicic, T. 2008. Conflict Monitoring in Dual Process Theories of Thinking. *Cognition*, 106: 1248–1299. doi.org/10.1016/j.cognition.2007.06.002

Dyczewski, E. A.; and Markman, K. D. 2012. General Attainability Beliefs Moderate the Motivational Effects of Counterfactual Thinking. *Journal of Experimental Social Psychology* 48(5): 1217-1220. doi.org/10.1016/j.jesp.2012.04.016

Efklides, A. 2006. Metacognition and Affect: What Can Metacognitive Experiences Tell Us About the Learning Process? *Educational Research Review* 1(1): 3-14.
doi.org/10.1016/j.edurev.2005.11.001

Efklides, A.; Samara, A.; and Petropoulou, M. 1999. Feeling of Difficulty: An Aspect of Monitoring That Influences Control. *European Journal of Psychology of Education* 14(4): 461-476. http://www.jstor.org/stable/23420265

Epstein, S.; Pacini, R.; Denes-Raj, V.; and Heier, H. 1996. Individual Differences in Intuitive-Experiential and Analytical-Rational Thinking Styles. *Journal of Personality and Social Psychology* 71(2): 390-405.
psycnet.apa.org/doiLanding?doi=10.1037%2F0022-3514.71.2.390

Epstude, K.; and Roese, N. J. 2008. The Functional Theory of Counterfactual Thinking. *Personality and Social Psychology Review* 12(2): 168-192.

Evans, J. S. B. 1977. Linguistic Factors in Reasoning. *Quarterly Journal of Experimental Psychology* 29(2): 297-306. www.tandfonline.com/doi/abs/10.1080/14640747708400605

Evans, J. S. B.; and Stanovich, K. E. 2013. Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives On Psychological Science* 8(3): 223-241.
doi.org/10.1080/14640747708400605

Fischoff, B. 2012. Judgment and Decision Making. London and New York: Earthscan

Flavell, J. H. 1979. Metacognition and Cognitive Monitoring: A New Area of Cognitive–Developmental Inquiry. *American Psychologist* 34(10): 906-911.
dx.doi.org/10.1037/0003-066X.34.10.906

Forbus, K. D., and Hinrichs, T. R. 2006. Companion Cognitive Systems: A Step Toward Human-level AI. *AI Magazine 27*: 83-95. doi.org/10.1609/aimag.v27i2.1882

Frederick, S. 2005. Cognitive Reflection and Decision Making. *Journal of Economic Perspectives* 19(4): 25-42. pubs.aeaweb.org/doi/pdfplus/10.1257/089533005775196732

Gascoine, L.; Higgins, S.; and Wall, K. 2017. The Assessment of Metacognition in Children Aged 4–16 Years: A Systematic Review. *Review of Education* 5(1): 3-57. doi.org/10.1002/rev3.3077

Geden, M.; Smith, A.; Campbell, J.; Amos-Binks, A.; Mott, B.; Feng, J.; and Lester, J. 2018. Towards Adaptive Support for Anticipatory Thinking. *In Proceedings of the Technology, Mind, and Society*. Washington, DC: ACM. doi.org/10.1145/3183654.3183665

Geden, M.; Smith, A.; Campbell, J.; Spain, R.; Amos-Binks, A.; Mott, B.; …Lester, J. 2019. Construction and Validation of an Anticipatory Thinking Assessment. PsyAriXiv.
doi.org/10.31234/osf.io/9reby

Gigerenzer, G.; and Gaissmaier, W. 2011. Heuristic Decision Making. *Annual Review of Psychology* 62: 451-482.
doi.org/10.1146/annurev-psych-120709-145346

Gilovich, T. 1983. Biased Evaluation and Persistence in Gambling. *Journal of Personality and Social Psychology* 44(6): 1110. https://pdfs.semanticscholar.org/f644/196083e55f0f824f849625f6e5592aa1d686.pdf

Glenberg, A. M.; Wilkinson, A. C.; and Epstein, W. 1982. The illusion Of Knowing: Failure in The Self-Assessment of Comprehension. *Memory and Cognition* 10: 597-602. https://link.springer.com/content/pdf/10.3758%2FBF03202442.pdf

Halpern, D. F. 1998. Teaching Critical Thinking for Transfer Across Domains: Disposition, Skills, Structure Training, and Metacognitive Monitoring. *American Psychologist* 53(4): 449.
doi.org/10.1037/0003-066x.53.4.449

Halpern, D. F. 2010. Halpern Critical Thinking Assessment. SCHUHFRIED (Vienna Test System): Moedling, Austria. www.schuhfried.com/viennatestsystem10/tests-test-sets/all-tests-from-a-z/test/hcta-halpern-critical-thinking-assessment-1/

Halpern, D. F. 2014. *Thought and Knowledge: An Introduction to Critical Thinking*. New York: Psychology Press.

Hough, A. R.; and Gluck, K. Forthcoming. The understanding problem in cognitive science. *Advances in Cognitive Systems* 7.

Hur, T. (2001). The Role of Regulatory Focus in Activation of Counterfactual Thinking. *Korean Journal of Social and Personality Psychology* 15: 159-171.

Johnson, B.; Roberts, M.; Apker, T.; and Aha, D. W. 2016. Goal Reasoning with Information Measures. In *Proceedings of the Fourth Conference on Advances in Cognitive Systems* 4: 1-14.

Johnson-Laird, P. N. 2013. The Mental Models Perspective. In *The Oxford Handbook Of Cognitive Psychology*, edited by D. Reisberg, 650-667. New York: Oxford University Press.

Juvina, I.; Larue, O.; and Hough, A. 2017. Modeling Valuation and Core Affect in A Cognitive Architecture: The Impact of Valence and Arousal on Memory and Decision-Making. *Cognitive Systems Research* 48: 4-24. doi.org/10.1016/j.cogsys.2017.06.002

Juvina, I.; Larue, O.; Widmer, C.; Ganapathy, S.; Nadella, S.; and Minnery, B. under review. Computer-supported Collaborative Information Search for Geopolitical Forecasting**.**

Kahneman, D. 2011. Thinking, Fast and Slow. New York: Farrar, Straus, and Giroux.

Kahneman, D.; and Klein, G. 2009. Conditions for Intuitive Expertise: A Failure to Disagree. *American Psychologist* 64(6): 515 526. doi.org/10.1037/a0016755

Kahneman, D.; and Miller, D. T. 1986. Norm Theory: Comparing Reality to Its Alternatives. *Psychological Review* 93(2): 136. pdfs.semanticscholar.org/9809/8ee48700173e2f09aeff48c4

06ef943918b5.pdf

Kennerley, S. W.; Dahmubed, A. F.; Lara, A. H.; and Wallis, J. D. 2009. Neurons in the Frontal Lobe Encode the Value of Multiple Decision Variables. *Journal of Cognitive Neuroscience* 21(6): 1162-1178.www.mitpressjournals.org/doi/abs/10.1162/jocn.2009.21100

Kirk, J. R.; and Laird, J. 2014. Interactive Task Learning for Simple Games. *Advances in Cognitive Systems 3*: 13-30. www.cogsys.org/pdf/paper-9-3-33.pdf

Klein, G. 1998. *Sources of Power: How People Make Decisions*. Cambridge, MA: MIT Press.

Klein, G.; Moon., B.; and Hoffman, R. R. 2006a. Making Sense of Sensemaking 1: Alternative Perspectives. *IEEE Intelligent Systems* 21(4): 70-73. doi.org/10.1109/MIS.2006.75

Klein, G.; Moon., B.; and Hoffman, R. R. 2006b. Making Sense of Sensemaking 2: A Macro Cognitive Model. *IEEE Intelligent Systems* 21(5): 88-92. doi.org/10.1109/MIS.2006.100

Klein, G.; Phillips, J. K.; Rall, E.; and Peluso, D. A. 2007. A Data/Frame Theory of Sensemaking. In *Expertise out of context: Proceedings of the 6th international conference on naturalistic decision making*, edited by R. R. Hoffman, 118-160. New York: Psychology Press.

Klein, G.; Ross, K. G.; Moon, B. M.; Klein, D. E.; Hoffman, R. R.; and Hollnagel, E. (2003). Macrocognition. *IEEE Intelligent Systems, 18*(3), 81-85. doi.org/10.1109/MIS.2003.1200735

Klein, G.; Snowden, D.; and Pin, C. L. 2007. Anticipatory Thinking. In *Proceedings of the Eighth International NDM Conference*, edited by K. Mosier and U. Fischer, 1-8. Pacific Grove, CA, June 2007.

Kurtz, C.F.; and Snowden, D.J. 2003. The New Dynamics of Strategy: Sensemaking In A Complex and Complicated World. *e-Business Management,* 42. doi.org/10.1147/sj.423.0462

Larue, O.; Hough, A.; and Juvina, I. (2018). A Cognitive Model of Switching Between Reflective and Reactive Decision Making in The Wason Task. In *Proceedings of the Sixteenth International Conference on Cognitive Modeling*, edited by I. Juvina, J. Houpt, and C. Myers, 55-60. Madison, WI: University of Wisconsin.

Markman, K. D.; Gavanski, I.; Sherman, S. J.; and McMullen, M. N. 1993. The Mental Simulation of Better and Worse Possible Worlds. *Journal of Experimental Social Psychology* 29: 87-109. doi.org/10.1006/jesp.1993.1005

Markman, K. D.; McMullen, M. N.; and Elizaga, R. A. 2008. Counterfactual thinking, Persistence, and Performance: A Test of the Reflection and Evaluation Model. *Journal of Experimental Social Psychology* 44(2): 421-428. doi.org/10.1016/j.jesp.2007.01.001

Mata, A.; Ferreira, M. B.; and Sherman, S. J. 2013. The Metacognitive Advantage of Deliberative Thinkers: A Dual-Process Perspective on Overconfidence. *Journal of Personality and Social Psychology* 105(3): 353. dx.doi.org/10.1037/a0033640

Mayer, R. E. 1998. Cognitive, Metacognitive, and Motivational Aspects of Problem Solving. *Instructional Science 26*: 49-63. link.springer.com/content/pdf/10.1023%2FA%3A1003088013286.pdf

McNamara, D. S.; Kintsch, E.; Songer, N. B.; and Kintsch, W. 1996. Are Good Texts Always Better? Interactions of Text Coherence, Background Knowledge, and Levels of Understanding in Learning from Text. *Cognition and Instruction* 14: 1-43. doi.org/10.1207/s1532690xci1401_1

Mellers, B.; Stone, E., Murray, T.; Minster, A.; Rohrbaugh, N.; Bishop, M.; ... and Ungar, L. 2015. Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions. *Perspectives on Psychological Science 10*(3): 267-281. doi.org/10.1177/1745691615577794

Muñoz-Avila, H.; Dannenhauer, D.; and Reifsnyder, N. 2019. Is Everything Going According to Plan? Expectations in Goal Reasoning Agents. In *Proceedings of the AAAI Conference on Artificial Intelligence* 33: 9823-9829. doi.org/10.1609/aaai.v33i01.33019823

Muñoz -Avila, H.; and Cox, M. T. 2008. Case-based Plan Adaptation: An Analysis and Review. *IEEE Intelligent Systems* 23: 75–81. doi.org/10.1109/MIS.2008.59

Newell, A.; and Simon, H. 1972. Human Problem Solving. Englewood Cliffs, NJ: Prentice-Hall.

Novick, L. R.; and Holyoak, K. J. 1991. Mathematical Problem Solving by Analogy. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 17(3): 398-415. https://pdfs.semanticscholar.org/6bc0/a20b01f0ebfc243d00de3e55f6387f8a77b8.pdf

Payne, J. W.; Bettman, J. R.; and Johnson, E. J. 1988. Adaptive Strategy Selection in Decision Making. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14(3): 534-552. apps.dtic.mil/dtic/tr/fulltext/u2/a170858.pdf

Pennycook, G. 2017. A Perspective on the Theoretical Foundation of Dual Process Models. In *Dual Process Theory 2.0*, edited by W. De Neys, 6-26. New York: Routledge.

Pennycook, G.; Fugelsang, J. A.; and Koehler, D. J. 2015a. What Makes Us Think? A Three-Stage Dual-Process Model of Analytic Engagement. *Cognitive Psychology* 80: 34-72. doi.org/10.1016/j.cogpsych.2015.05.001

Pennycook, G.; Fugelsang, J. A.; and Koehler, D. J. 2015b. Everyday Consequences of Analytic Thinking. *Current Directions in Psychological Science* 24(6): 425-432. doi.org/10.1177/0963721415604610

Perkins, D. N. 1998. What is Understanding? In *Teaching for Understanding: Linking Research with Practice*, edited by M. S. Wiske, 39-57. San Francisco, CA: Jossey-Bass.

Perkins, D. N.; and Simmons, R. 1988. Patterns of Misunderstanding: An Integrative Model for Science, Math, and Programming. *Review of Educational Research* 58: 303-326. doi.org/10.3102/00346543058003303

Pirolli, P.; and Russell, D. M. 2011. Introduction to This Special Issue on Sensemaking. *Human-Computer Interaction* 26: 1-8. www.tandfonline.com/doi/full/10.1080/07370024.2011.556557

Pirolli, P.; and Card, S. K. 1999. Information Foraging. *Psychological Review* 106: 643-675.

Pirolli, P.; and Card, S. K. 2005. The Sensemaking Process and Leverage Points for Analyst Technology. Paper presented at the *International Conference on Intelligence Analysis*, McLean, VA. pdfs.semanticscholar.org/f6b5/4379043d0ee28bea45555f481af1a 693c16c.pdf

Pozanco, A.; Fernández, S.; and Borrajo, D. 2018. Learning-driven Goal Generation. *AI Communications* 31: 137–150.

Roberts, M.; Borrajo, D.; Cox, M. T.; and Yorke-Smith, N. 2018. Special Issue on Goal Reasoning. *AI Communications* 31: 115–116. http://doi.org/10.3233/AIC-180754

Roese, N. J.; and Epstude, K. 2017. The Functional Theory of Counterfactual Thinking: New Evidence, New Challenges, New Insights. In *Advances in Experimental Social Psychology: Vol. 56*, edited by J. M. Olson, 1-79. Cambridge, MA: Academic Press. doi.org/10.1016/bs.aesp.2017.02.001

Roese, N. J.; and Olson, J. M. 1997. Counterfactual Thinking: The Intersection of Affect and Function. In *Advances in experimental social psychology: Vol.* 29, edited by M. P. Zanna, 1-59. San Diego, CA: Academic Press. doi.org/10.1016/S0065-2601(08)60015-5

Roese, N. J.; Hur, T.; and Pennington, G. L. 1999. Counterfactual Thinking and Regulatory Focus: Implications for Action Versus Inaction and Sufficiency Versus Necessity. *Journal of Personality and Social Psychology* 77(6): 1109. doi.org/10.1037//0022-3514.77.6.1109

Russell, D. M.; Stefik, M. J.; Pirolli, P.; and Card, S. K. 1993. The Cost of Structure of Sensemaking. Paper presented at the *INTERCHI 1993 conference on Human Factors in Computing Systems*, Amsterdam, the Netherlands. doi.org/10.1145/169059.169209

Sanna, L. J.; and Turley, K. J. 1996. Antecedents to Spontaneous Counterfactual Thinking: Effects of Expectancy Violation and Outcome Valence. *Personality and Social Psychology Bulletin* 22: 906-919. doi.org/10.1177%2F0146167296229005

Sanna, L. J.; and Turley-Ames, K. J. 2000. Counterfactual Intensity. *European Journal of Social Psychology* 30: 273-296. doi.org/10.1002/(SICI)1099-0992(200003/04)30:2%3C273::AID-EJSP993%3E3.0.CO;2-Y

Schank, R. C.; and Abelson, R. P. 1977. Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures. Hillsdale, NJ: Lawrence Erlbaum Associates.

Stanovich, K. E. 2009. Distinguishing the Reflective, Algorithmic, And Autonomous Minds: Is It Time for A Tri-Process Theory. In *Two minds: Dual processes and beyond*, edited by J. S. B. T. Evans and K. Frankish, 55-88. New York: Oxford University Press.

Stanovich, K. E. 2018. Miserliness in Human Cognition: The Interaction of Detection, Override and Mindware. *Thinking and Reasoning* 24(4): 423-444. doi.org/10.1080/13546783.2018.1459314

Swan, A. B.; Calvillo, D. P.; and Revlin, R. 2018. To Detect or Not to Detect: A Replication and Extension of the Three-stage Model. *Acta Psychologica* 187: 54-65. doi.org/10.1016/j.actpsy.2018.05.003

Taatgen, N. A.; Van Rijn, H.; and Anderson, J. 2007. An Integrated Theory of Prospective Time Interval Estimation: The Role of Cognition, Attention, and Learning. *Psychological Review* 114(3): 577-598. https://psycnet.apa.org/doiLanding?doi=10.1037%2F0033-295X.114.3.577.

Tetlock, P. E.; and Gardner, D. 2015. *Superforcasting: The Art and Science of Prediction*. New York, NY: Crown.

Thompson, V. A.; and Johnson, S. C. 2014. Conflict, Metacognition, and Analytic Thinking. *Thinking and Reasoning* 20(2): 215-244. doi.org/10.1080/13546783.2013.869763

Thompson, V. A.; Evans, J. S. B.; and Campbell, J. I. 2013. Matching Bias on the Selection Task: It's Fast and Feels Good. *Thinking and Reasoning* 19: 431-452. doi.org/10.1080/13546783.2013.820220

Thompson, V. A.; Prowse Turner, J. A.; and Pennycook, G. 2011. Intuition, Reason, and Metacognition. *Cognitive Psychology* 63: 107-140. doi.org/10.1016/j.cogpsych.2011.06.001

Toplak, M. E.; West, R. F.; and Stanovich, K. E. 2011. The Cognitive Reflection Test as A Predictor of Performance on Heuristics-And-Biases Tasks. *Memory and Cognition* 39(7): 1275-1289. link.springer.com/article/10.3758%2Fs13421-011-0104-1

Vattam, S.; Klenk, M.; Molineaux, M.; and Aha, D. W. 2013. Breadth of Approaches to Goal Reasoning: A Research Survey. In *Goal Reasoning: Papers from the ACS Workshop*, edited by D. W. Aha, M. T. Cox, and H. Muñoz-Avila. Technical Report CS-TR-5029. College Park, MD: University of Maryland, Department of Computer Science. apps.dtic.mil/dtic/tr/fulltext/u2/a603400.pdf

Warwick, W.; and Hutton, R. J. B. 2007. Computational and Theoretical Perspectives on Recognition-Primed Decision Making. In *Expertise out of context: Proceedings of the sixth international conference on naturalistic decision making*, edited by R. R. Hoffman, 429-452. New York: Psychology Press.

Wason, P. C. 1966. Reasoning. In *New horizons in psychology*, edited by B. Foss, 135-151. London, UK: Penguin.

Wason, P. C. 1968. Reasoning About a Rule. *Quarterly Journal of Experimental Psychology* 20: 273-281. doi.org/10.1080%2F14640746808400161

Wason, P. C. 1969. Regression in Reasoning? *British Journal of Psychology* 60(4): 471-480. doi.org/10.1111/j.2044-8295.1969.tb01221.x

Weick, K.E.; and Sutcliffe, K.M. 2001. *Managing the Unexpected: Assuring High Performance in An Age of Complexity*. San Francisco: Jossey-Bass.

West, R. F.; Toplak, M. E.; and Stanovich, K. E. 2008. Heuristics and Biases as Measures of Critical Thinking: Associations with Cognitive Ability and Thinking Dispositions. *Journal of Educational Psychology* 100(4): 930-941. https://psycnet.apa.org/doiLanding?doi=10.1037%2Fa0012842

Westbrook, A.; Kester, D.; and Braver, T. S. 2013. What is the Subjective Cost of Cognitive Effort? Load, Trait, and Aging Effects Revealed by Economic Preference. PloS one 8(7): e68210. doi.org/10.1371/journal.pone.0068210

Willingham, D. T. 2007. Critical Thinking: Why is it So Hard to Teach? *American Educator* 31(3): 8-19. doi.org/10.3200/AEPR.109.4.21-32

Woodward, J. 2003. *Making Things Happen: A Theory of Casual Explanation.* New York, NY: Oxford University Press.