MODELING THE "CAPITALIST'S DILEMMA" USING MULTIPLE REINFORCEMENT LEARNING SYSTEMS

CHRISTOPHER J. DIAK
Harvard University

Dual-process theories of cognition abound in the psychological literature, mapping distinctions between conscious and unconscious reasoning; analytical and intuitive decision making; and planned versus habitual action. Framing effects are widely thought to emerge downstream of intuitive, habitual, or unconscious systems based on their automatic responses to stimuli. But another possibility is that framing can occur upstream of dual-process systems and shift the balance of analytical and intuitive modes of thought for a given task. Empirical research has demonstrated that investors and general managers use a combination of financial analysis and intuition when appraising investment opportunities or allocating capital, but it is unknown whether framing effects can influence arbitration between analytical and intuitive investment strategies in these contexts. Here I use model-based and model-free reinforcement learning algorithms to formalize the distinction between these two strategies and propose a mixed methods research program investigating whether framing capital as a scarce resource causes investors to favor analytical or intuitive investment strategies.

INTRODUCTION

A fundamental question for organizational behavior is the extent to which cognitive processes shape and are shaped by organizational structures, processes and priorities. Advances in social, cognitive and computational neuroscience have generated a rich and diverse literature drawing out the implications of brain science for strategy and management (Schwenk, 1984; Stubbart, 1987; Tyler and Steensma, 1995; Walsh, 1995) with increased attention in recent years devoted to the role of conscious and nonconscious cognitive processes in the workplace (Weaver et al. 2014; Pratt and Crosina, 2016), often in the form of dual-process cognitive theories (Hodgkinson and Sadler-Smith, 2017). Dual-process theories posit a distinction between intuitive, habitual, and nonconscious cognitive processes on the one hand, and deliberate, conscious, and analytical processes on the other; these processes are often called System 1 and System 2 respectively (Kahneman and Frederick, 2002). Furthermore, dual-process theories can be classified based on the dynamics of the interaction between the subsystems into two classes: a default-interventionist model and a parallel-competitive model (Evans, 2008). Default-interventionist models assume that a set of heuristics, biases, or other cognitive schema (System 1) rapidly and pre-consciously process stimuli upstream of System 2, resulting in “default” behaviors that are subject to change through the downstream “intervention” of System 2 (Hodgkinson and Sadler-Smith, 2017). The archetypal research program in this form is Kahneman and Tversky’s (1981) heuristics and biases approach. This stands in contrast to parallel-competitive models which conceptualize the two subsystems as alternative modes of thought that must enter into a zero-sum competition for cognitive or computational resources (Epstein, 1994; Kool et al., 2017).

Following Kahneman and Tversky (1981) framing effects have often been conceived according to a default-interventionist model (Whitney et al., 2008; Guo et al., 2017) whereby they emerge as a result of default behavior patterns encoded in the heuristics and biases of System 1. And yet, parallel-competitive models of intuition and analysis may provide a more robust representation of the cognitive processes underlying the work of organizational decision makers due to their embodiment of a cost-benefit arbitration process (Hodgkinson and Sadler-Smith, 2017; Kool et al., 2017). In other words, because managers often have opportunities to choose between using slow, analytical processes or quick and intuitive approaches to problem solving, it would be useful to be able to model how people decide whether to use analytical or intuitive methods to solve problems.

MODELING DUAL-PROCESS COGNITION WITH REINFORCEMENT LEARNING

As the subfield of machine learning concerned with the manipulation of dynamical systems, reinforcement learning (RL) provides a powerful tool for researchers interested in modeling intelligent behavior under conditions of uncertainty (Recht, 2018). Recent RL research has formalized the parallel-competitive class of dual-process models in terms of a cost-benefit arbitration process between *model-based* and *model-free* RL systems (Otto et al., 2015; Kool et al., 2017) and emerging neurological evidence indicates this arbitration process is in fact realized in the neural computations underlying decision making (Lee et al., 2014; Weissengruber et al., 2019). Assuming the existence of such a cost-benefit arbitration process, framing effects would be expected to play one of two roles in a parallel-competitive model. Framing effects could either emerge downstream from System 1 due to cognitive load, as the competition for cognitive resources between Systems 1 and 2 would favor computationally cheap intuitive responses over computationally complex behaviors (cf. Whitney et al., 2008) or framing effects could influence the parameters of the cost-benefit arbitration function itself, such that System 1 *or* System 2 would be preferentially activated in response to particular framing effects.

Converging lines of evidence in artificial intelligence and computational neuroscience suggest that cost-benefit arbitration between model-based (System 2) and model-free (System 1) control is a function of the reliability of the respective predictions of each subsystem (Lee et al., 2014; Gershman et al., 2015). In a series of behavioral experiments modeled with multiple RL systems, Kool et al. (2017) demonstrated that people increase their use of model-based control when it affords greater reliability than model-free control and they are especially likely to favor model-based control on “high stakes” trials where the rewards for reliability are enhanced.

However, when model-based and model-free control are equally accurate, the stake-size effect is attenuated and model-based control is not especially favored. To investigate the possibility that framing effects could modulate the parameters of the cost-benefit arbitration function of a parallel-competitive dual-process decision-making model, researchers ought to attempt to modulate the perceived reliability of allocating control to model-based systems, regardless of whether it is the optimal strategy. To distinguish such a model from a default-interventionist model, one would need to show how framing effects shift the balance of model-based and model-free control in the direction of *model-based* control, because it would be implausible to suggest that the structured process of analysis employed in model-based control is a "default" behavior triggered by heuristics or biases. Rather, increased model-based control would be the result of a shift in the perceived reliability of a model-based approach.

CAPITAL ALLOCATION AS A DUAL-PROCESS OF ANALYSIS AND INTUITION

Reviewing the strategic and management literature, Sengul et al. (2019) conceptualize capital allocation as the process of determining, comparing and selecting among multiple investment alternatives at multiple levels within an organization, subject to contextual constraints external to the organization itself. Within the domain of investment comparison and selection, which has received the lion's share of scholarly attention in capital allocation research, "the dominant prescriptive logic of textbook finance for the comparison and selection of investment alternatives is a straightforward one, which uses net present value (NPV)—i.e., the expected stream of cash flows from an investment discounted for their timing and risk—and NPV-based valuation techniques (such as discounted cash flow, discounted earnings, and economic value-added) as the commanding criteria. In the absence of budget constraints, funding should be awarded to an investment whenever its NPV is positive," (Sengul et al., 2019). The authors note that "79% of U.S. CEOs reported that NPV rankings were important or very important when deciding how to allocate capital," despite the fact that the "robustness of NPV-based techniques is severely impaired under environmental uncertainty and complexity," (Sengul et al., 2019). Under such conditions, intuition is recognized as playing a greater role — for better or worse — in the life of the firm and the capital allocation process, with the selection of investments driven by threat perceptions (Christensen and Bower, 1996), framing effects, or other interpretive processes by which "boundedly rational" organizational actors make sense of complexity (Sengul et al., 2019). A dual-process model suggests itself, representing NPV-based capital allocation as model-based and intuitive capital allocation processes as model-free reinforcement learning systems.

Empirical research has demonstrated that when making investment decisions, investors rely on a combination of intuition and formal analysis to achieve desired returns (Huang and Pearce, 2015), but it is an open question how framing effects influence investment strategies and the proclivity of general managers to favor intuitive or formal approaches to the allocation of capital. In circumstances of "instability, unknowable risks, and non-decomposable tasks, when speed is critical and decision makers have complex, domain-relevant experience," Huang and Pearce (2015) suggest that "intuition may be more effective than analytic decision making." Indeed, in a study of 90 experienced angel investors tasked to assess entrepreneurs making pitch presentations, when business viability data did not predict venture survival, growth, subsequent financing, or "homerun" status, investors' intuitive assessments of the entrepreneur *were* predictive of the "homerun status" of investments which were extraordinarily profitable within four years (Huang and Pearce, 2015). These kinds of intuitive judgments can be modeled using model-free RL systems because they represent complex, domain-relevant experience in terms of “cached values” generated by a temporal-difference learning algorithm (Daw et al., 2005). In other words, model-free systems learn by directly retrieving the rewards of past state-action pairs and comparing them with the present, thereby formalizing the notion that past experience informs present action.

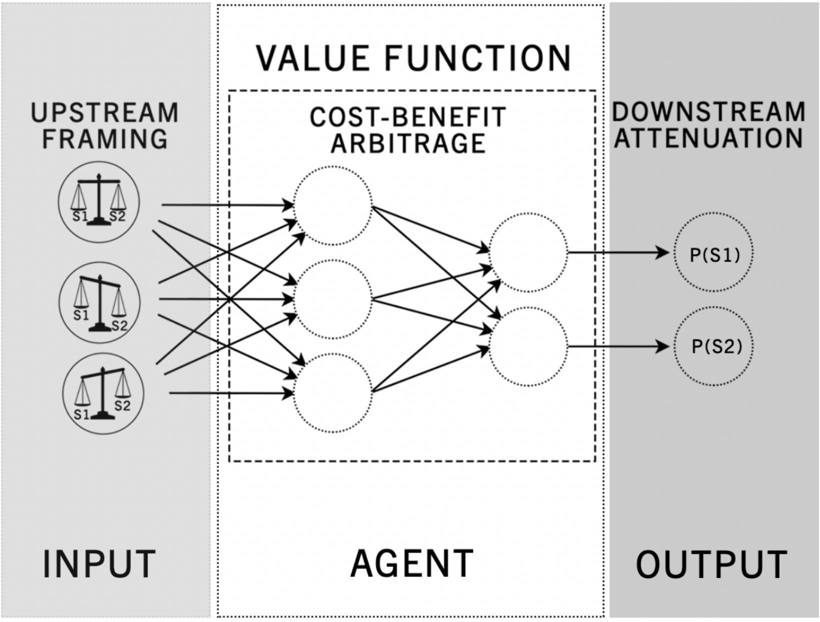
In the research tradition of heuristics and biases, intuition is generally conceived negatively as a detriment to optimal decision making (Kahneman and Tversky, 1971; Hirshleifer, 2001). But one prominent articulation of the positive roles that intuition can play in managerial decision making comes from Epstein and colleagues’ *cognitive-experiential self-theory*, a dual-process model of human information processing consisting of an “experiential” or intuitive mode and a “rational” or analytic mode (Epstein, 1985, 1994; Epstein et al., 1996; Hodgkinson and Sadler-Smith, 2017). Hodgkinson and Sadler-Smith (2017) point out that the cognitive-experiential self-theory, unlike many dual-process theories, posits a variety of interactions between the intuitive and analytical subsystems, beyond conflict and competition, including cooperation and compromise. This insight was prefigured in the strategic-decision making literature of the 1970s and 1980s which viewed “cognitive simplification processes” as a means by which managers could channel cognitive biases or heuristics into productively useful forms, thereby managing complexity (Schwenk, 1984).

Depending on the manner in which the problem is formulated, cognitive simplification processes (CSPs) could be modeled using model-free or model-based RL systems; on the one hand, CSPs transform environmental rewards into action based on a low-resolution map of the structure of the environment (akin to model-based RL systems) but on the other hand, agents which are assumed to learn through experience can be modeled successfully using model-free approaches such as a direct policy search or dynamic programming (Recht, 2018). This is just to say that the choice to represent intuitions as a model-free system, as is done in our present study, leaves out features of the reasoning process — such as CSPs — which other machine learning models may better capture. Our interest is in formalizing a parallel-competitive dual-process model of decision making to test the degree to which framing effects modulate the arbitration of control between two well-defined reasoning systems in the neuro-computational literature (Daw et al., 2005; Lee et al., 2014; Kool et al., 2017; Weissengruber et al., 2019), transposing this work into the field of organizational behavior by studying the ways these effects manifest in a core managerial function: the capital allocation process.

THE CAPITALIST'S DILEMMA

Counterintuitively, framing effects may play a role in the predominance of NPV-based capital allocation. According to a proposal by Christensen and van Bever (2014), termed the *Capitalist's Dilemma*, when capital is framed as a "scarce resource" managers are hypothesized to favor the use of financial metrics such as RONA, ROIC, NPV and IRR when making investment decisions to ensure capital is allocated efficiently from a "balance sheet perspective," (Christensen and van Bever, 2014). Incentivized to maximize performance on these metrics, managers tend to make investments that take capital and assets off the balance sheet by outsourcing, making their operations more efficient, or investing in short-term projects, patterns of behavior which Christensen and van Bever believed were contributing factors to stagnating growth in the U.S. economy, (Christensen and van Bever, 2014). The result of this process, they argued, was a net loss of jobs coupled with increased productivity, increased capital availability, a disincentive to pursue growth opportunities that require capital or long-term assets on the balance sheet, and an incentive to reproduce these outcomes in the next investment cycle. A parallel-competitive dual-process model of the capital allocation process could explain how upstream framing effects from System 1 (e.g. that capital is "scarce") or System 2 (e.g. a bias toward being "data-driven" and "fact-based") can modulate the propensity to utilize System 1 or 2 reasoning strategies by modifying the parameters of the value function realizing the cost-benefit arbitration *between* Systems 1 and 2 (See Figure 1).

FIGURE 1
Framing Effects Modulate Perceived Costs and Benefits of System 1 and System 2 Strategies



Notes: Upstream cognitive schemata in the form of framing effects alter the perceived cost or benefit of System 1 (S1) or System 2 (S2) strategies. Here, the scales represent antecedent states of the brain favoring S1 or S2 systems based on framing effects. These states serve as inputs to a set of neural computations (cf. Lee et al. 2014) that realize a process of cost-benefit arbitrage which can be modeled with a value function. The parameters of this function are modified by inputs embodying differing framing effects. Ultimately, the value function outputs probabilities of selecting S1 or S2 reasoning strategies.

Arbitration between model-based and model-free control is a function of the perceived reliability of each strategy relative to the task at hand (Lee et al., 2014; Gershman et al., 2015). Consider, for example, a student trying to complete a mathematics exam. Model-free control would amount to making educated guesses about the answers to problems; model-based control would involve precise, time-intensive computations. If time is scarce, the student may choose to engage in model-free reasoning and attempt to intuit her way through the exam. That is because the cost of engaging in model-based reasoning, in terms of time, is judged to be too great to warrant the investment in a more reliable strategy. Christensen and van Bever (2014) proposed that when *capital* is framed as a scarce resource, managers will prefer what amounts to model-based strategies for valuing investments: NPV, IRR, RONA, and ROIC. If this is true, it may be because the perceived reliability of model-based strategies far outweighs the costs of the mental computation and analysis involved, or it may be due to a lack of perceived reliability for model-free strategies. Based on this line of reasoning, I propose the following hypothesis:

> *Hypothesis 1:* When capital is framed as a scarce resource, investors will systematically increase their use of model-based control to make decisions relative to model-free control.

Additionally, research has shown that people are sensitive to reward amplification when there is a reliability benefit to model-based control (Kool et al., 2017), therefore I propose:

> *Hypothesis 2:* When capital is framed as a scarce resource and there is a reliability benefit of a model-based strategy there will be an interaction effect increasing the use of model-based control.

Arbitration between the two systems could also be influenced by the relative costs (perceived or actual) of engaging in either model-based or model-free reasoning in organizational contexts. Given the ubiquity of the spreadsheet, increased model-based reasoning in the form of NPV calculations *in situ* could be a result of the relatively low *cognitive* cost of engaging in these behaviors. Therefore, I propose the following hypotheses:

> *Hypothesis 3a:* Increasing the cognitive cost of model-based decision-making (i.e. by requiring handwritten NPV calculations) will not attenuate the interaction effect proposed in Hypothesis 2.
>
> *Hypothesis 3b:* Controlling for the costs of model-based strategies, when there are no reliability benefits for model-based reasoning and capital is not framed as a scarce resource, investors will not favor a model-based strategy over a model-free strategy.
>
> *Hypothesis 3c:* Controlling for cost, when there are no reliability benefits for a model-based strategy and capital is framed as a scarce resource, investors will favor a model-based strategy over a model-free strategy.

This set of predictions aligns with a parallel-competitive dual-process model of analytical and intuitive decision-making embodying a cost-benefit tradeoff that could be modulated due to upstream framing effects. By formalizing the psychological mechanism proposed by Christensen and van Bever (2014) in terms of the modulation of an arbitration function between model-based and model-free reinforcement learning systems, the groundwork is laid for a research program synthesizing cognitive and computational neuroscience with organizational behavior.

MODELING THE CAPITALIST'S DILEMMA WITH MULTIPLE RL SYSTEMS

RL algorithms use a mathematical framework known as a *Markov decision process* (MDP) to define the interaction between an agent and its environment in terms of states, actions, and rewards (Sutton and Barto, 2018). As the subfield of machine learning concerned with the manipulation of dynamical systems, RL can be used to model the cost-benefit arbitration of parallel-competitive dual-process models over time in terms of the state-space evolution of a dynamical system (Recht, 2018). In general, the goal of constructing RL systems is to find a sequence of inputs in the form of state-action pairs that will optimize the behavior of a dynamical system to maximize the notion of reward over *N* amount of states, assuming minimal prior knowledge of the dynamics of the system (Recht, 2018). Borrowing Recht's (2018) exposition, we can mathematically represent a dynamical system with a difference equation of the form:

$$s_{t+1} = f_t(s_t, a_t, e_t)$$

Where, at time *t*, s_t is the *state* of the system;
a_t is the *action* taken by the agent;
e_t is a *random disturbance*;
and f_t is a rule mapping the current state, action, and disturbance at time *t* to a new state s_{t+1}.

At each moment in time, the RL system will be presented with a reward $R(s_t, a_t)$ that maps the current state and action to a value function, and it is the goal of an RL system to maximize this notion of cumulative reward over *N* moments in time (Recht, 2018). To understand how it does so, it is helpful to have a notion of *trajectory,* a sequence of states and actions generated by the dynamical system and here represented as $\tau = (s_1, \dots s_{t-1}, a_0, \dots a_t)$ and the notion of a *policy* (π)which represents a function that takes a trajectory as an input and outputs an action *a* (Recht, 2018). This can be represented with the following mathematical formalism:

$$\begin{array}{ll} \text{maximize} & \mathbb{E}_{e_t}\left[\sum_{t=o}^{N} R_t(s_t, a_t)\right] \\ \text{subject to} & s_{t+1} = f_t(s_t, a_t, e_t),\ s_t = \pi_t(\tau_t) \\ & (s_0 \text{ given}). \end{array}$$

The goal of an RL system is thus to tune π_t until an optimal policy is achieved for transitioning between state-action pairs (Recht, 2018; Sutton and Barto, 2018).

Model-based and model-free reinforcement learning systems go about this optimization process in fundamentally different ways, either by fitting a model to previously observed data or by mapping observations of the environment to actions, respectively (Recht, 2018). To model the Capitalist's Dilemma, I would use a validated dual-system RL model (Daw et al., 2011) and build an investing task modeled on a four-state, two-stage game design (Kool et al., 2017) with three conditions 1) a condition wherein model-based control is the optimal strategy, and 2) a condition wherein neither model-based nor model-free control is favored, and 3) a condition wherein model-free control is favored. The model-free RL system will use a simple temporal-difference learning algorithm based on Rummery and Niranjan (1994) and the model-based system will learn a transition function that maps first-stage state-action pairs to a probability distribution over the following states, thereby learning which first-stage action leads to which second-stage state, before then computing the state-action values (Kool et al., 2017, Supplemental Information). Thousands of simulations will be performed on the game, verifying that the reward-maximizing strategies for each of the three conditions are 1) 100% model-based control, 2) 50% model-based control, 50% model-free control and 3) 100% model-free control, using linear regression to estimate the strength of the relationship between the idealized strategy and the game conditions (Kool et al., 2017). Behavioral tests of ~150 human beings playing the same game on Amazon MTurk would then be analyzed and factorial ANOVAs run that measure increase in model-based or model-free control relative to optimal RL policies for each condition.

2 x 3 Factorial ANOVA Design		IV 1 ***Capital Framing***	
		Scarce	Not Scarce
IV 2 ***Reliability Benefit to Model-Based Control***	+ Reliability	MB >> MF	MB > MF
	= Reliability	MB > MF	MB = MF
	- Reliability	MB > MF	MB < MF

Table 1. Proposed 2 x 3 factorial ANOVA. Categorical IVs = Reliability Benefit to Model-Based Control, classed as +, = or -. DV = % increase in model-based (MB) or model-free (MF) control relative to dual RL system. Expected direction of effects denoted with >, =, < with >> indicating an interaction effect.

References

Bain and Company. 2011. A world awash in money: capital trends through 2020. ***Bain and Company.*** https://media.bain.com/Images/BAIN_REPORT_A_world_awash_in_money.pdf

Christensen, C.M., van Bever, D. C. M. 2014. The Capitalist's Dilemma. ***Harvard Business Review***.

Christensen, C., Bower, J. 1996. Customer Power, Strategic Investment, and the Failure of Leading Firms. ***Strategic Management Journal**,* 17(3), 197-218.

Daw N.D., Niv Y., Dayan P. 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. ***Nature Neuroscience***. Dec;8 (12): 1704-11.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. ***Neuron***, 69, 1204-1215.

Epstein, S. 1985. The Implications of Cognitive-experiential Self-theory for Research in Social Psychology and Personality. ***Journal for the Theory of Social Behaviour***, 15(3): 283–310.

Epstein S.1994. Integration of the cognitive and the psychodynamic unconscious. ***American Psychologist***. Aug 49 (8): 709-24.

Epstein, S., Pacini, R., Denes-Raj, V., Heier, H. 1996. Individual differences in intuitive experiential and analytical-rational thinking styles. ***Journal of Personality and Social Psychology***, 71, 390-405.

Evans, J. S. B. T. 2008. Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. ***Annual Review of Psychology***, 59(1), 255–278.

Gershman S.J., Horvitz E.J., Tenenbaum J.B. 2015. Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. ***Science***. Jul 17; 349 (6245): 273-8.

Guo L, Trueblood JS, Diederich A. 2017. Thinking Fast Increases Framing Effects in Risky Decision Making. ***Psychological Science***. 28 (4): 530-543.

Hirshleifer, D. 2001. Investor Psychology and Asset Pricing. ***The Journal of Finance**, 56(4), 1533–1597.*

Hodgkinson, G.P., Sadler-Smith, E. 2018. The dynamics of intuition and analysis in managerial and organizational decision making. ***Academy of Management Perspectives***, 32 (4): 473-492

Huang, L., Pearce, J. L. 2015. Managing the Unknowable. ***Administrative Science Quarterly**,* 60(4), 634–670.

Kahneman, D., Frederick, S. 2005. A Model of Heuristic Judgment. In K. J. Holyoak & R. G. Morrison (Eds.), ***The Cambridge Handbook of Thinking and Reasoning*** (p. 267–293). Cambridge University Press.

Kool W., Gershman S.J., Cushman F.A. 2017. Cost-Benefit Arbitration Between Multiple Reinforcement-Learning Systems. ***Psychological Science***. Sep;28 (9): 1321-1333.

Lee S.W., Shimojo S., O'Doherty J.P. 2014. Neural computations underlying arbitration between model-based and model-free learning. ***Neuron***. Feb 5;81 (3): 687-99.

Levinthal, D. A. 2011. A behavioral approach to strategy-what's the alternative? ***Strategic Management Journal,*** 32 (13): 1517–1523.

Mankins, M., Harris, K., Harding, D. 2017. Strategy in the Age of Superabundant Capital. ***Harvard Business Review.***

Otto A.R., Skatova A., Madlon-Kay S., Daw N.D. 2015. Cognitive control predicts use of model-based reinforcement learning. ***Journal of Cognitive Neuroscience.*** Feb;27 (2): 319-33.

Pratt, M. G., & Crosina, E. (2016). The nonconscious at work. ***Annual Review of Organizational Psychology and Organizational Behavior***, 3, 321-347.

Recht, B. 2018. A Tour of Reinforcement Learning: The View from Continuous Control. ***Annual Review of Control, Robotics, and Autonomous Systems***, Vol. 2:253-279.

Rummery, G., & Niranjan, M. (1994). On-line Q-learning using connectionist systems. ***Cambridge University***.

Schwenk, C.R. 1984. Cognitive Simplification Processes in Strategic Decision Making. ***Strategic Management Journal.*** Vol. 5, 111-128.

Schwenk, C. R. (1988). The Cognitive Perspective on Strategic Decision Making. ***Journal of Management Studies***, 25(1), 41–55.

Sengul, M., Almeida Costa, A., & Gimeno, J. 2018. The allocation of capital within firms: A review and integration toward a research revival. ***Academy of Management Annals***. doi:10.5465/annals.2017.0009

Stubbart, C.I. 1987. Cognitive Science and Strategic Management Theoretical and Methodological Issues. ***Academy of Management Proceedings,*** Vol. 1.

Sutton, R.S., Barto, A.G. 2018. Reinforcement Learning: An Introduction. Second Edition. The MIT Press.

Tyler, B. B., Steensma, H.K. 1995. Evaluating technological collaborative opportunities: A cognitive modeling perspective. ***Strategic Management Journal,*** 16(S1), 43–70.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. ***Psychological Bulletin,*** 76(2), 105–110.

Walsh, J.P. 1995. Managerial and Organizational Cognition: Notes from a Trip Down Memory Lane. ***Organization Science,*** 6(3), 280–321.

Weissengruber S., Lee S.W., O'Doherty J.P., Ruff C.C. 2019. Neurostimulation Reveals Context-Dependent Arbitration Between Model-Based and Model-Free Reinforcement Learning. ***Cerebral Cortex***. Dec 17; 29(11):4850-4862.

Weaver, G. R., Reynolds, S. J., & Brown, M. E. 2014. Moral intuition: Connecting current knowledge to future organizational research and practice. ***Journal of Management***, 40, 100- 129.

Whitney, P., Rinehart, C.A., Hinson, J.M. 2008. Framing effects under cognitive load: The role of working memory in risky decisions. ***Psychonomic Bulletin & Review*** 15, 1179–1184.