Cognitive Systems Research 71 (2022) 24–40

Contents lists available at ScienceDirect

Cognitive Systems Research

journal homepage: www.elsevier.com/locate/cogsys

A unified framework for interpreting a range of motivation-performance phenomena

Ron Sun[a,*], Sergei Bugrov[a], David Dai[b]

[a] Cognitive Science Department, Rensselaer Polytechnic Institute, 110 Eighth Street, Troy, NY 12180, USA
[b] School of Education, SUNY University at Albany, 1400 Washington Ave, Albany, NY 12222, USA

ARTICLE INFO

Keywords:
Motivation
Cognitive architecture
Simulation
Cognition
Modeling
Clarion

ABSTRACT

Effects of motivation on cognition and performance have been found empirically in different fields. However, the relationship between motivation and performance seems complex and multi-faceted. While seemingly inconsistent or incompatible studies and theories of different disciplinary origins exist in this regard, we show that many of them can potentially be synthesized. Simulations within a unifying computational cognitive architecture account for empirical phenomena from different studies, which demonstrate that a mechanistic (computational) account can potentially unify the interpretations of these studies, largely based on utility calculation from intrinsic motives, and allow for further, more detailed explorations.

1. Introduction and background

Motivation and cognition interact with each other (Dai & Sternberg, 2004; Simon, 1967) and their interaction affects performance. Effects of motivation on cognition and performance have been observed empirically in a variety of different fields. Examining the relevant literature on the relationship between motivation and performance from different disciplines, one would notice the seeming complexity, or sometimes the mutual incompatibility, of empirical findings and phenomena. Looking at various proposed theories in this regard, one would also see the same complexity and apparent lack of coherence.

For instance, in the motivation-performance literature, there have been a number of well-known theories, such as those by Bandura (1977), Locke and Latham (1990; 2013), Ryan and Deci (2000), Steel and König (2006), Vancouver et al. (2010; 2014), Vancouver and Purl (2017), and so on. They were developed within the contexts of social psychology, industrial-organizational psychology, educational psychology, cognitive neuroscience, and other disciplines, respectively. They sometimes lead to different, or even contradictory, interpretations, explanations, or predictions. So how do we reconcile them—making them more compatible and generating a more general theory or theoretical framework?

Empirical findings themselves are complex, multi-faceted, and diverse. Given space limitations and our integrative goal, we can only selectively examine some historically important work. For one instance, "intrinsic" motivation (whereby one is "intrinsically" motivated for an activity if he/she receives no rewards except the activity itself) has been strongly emphasized by some, while others disagree. Early work by Deci (1971) showed that monetary incentives harmed "intrinsic" motivation and effort, while some others showed otherwise (Locke & Latham, 2013; Locke & Schattke, 2019). For another instance, while some claimed that the ability of individuals to make choices enhanced motivation and performance, some others showed that this might not always be the case: Iyengar and Lepper (1999) showed that, while Anglo-American children demonstrated best performance when they had a choice, Asian-American children performed the best when the choice was made by certain others. For yet another instance, some made the general claim that higher assigned goals (i.e., performance targets) led to better performance, while some others showed opposite cases and advocated "do your best" goals instead. Seijts and Latham (2001) showed that the learning goals (which centered on learning, rather than performance, measures) led to better performance than the outcome goals (focusing on performance measures), while the "do your best" goals had mixed results (see also Locke & Latham, 2013 for moderators/mediators). There are also competing accounts that linked motivation to effort (e.g., through changing cost calculation; Meyniel et al., 2013). There are many other issues in the empirical behavioral literature, for example, effects of different types of feedback, effectiveness of priming, role of expectancy (self-efficacy), and so forth. Recent neuroscientific work delves into even more issues such as how individuals learn when/how to exert cognitive

* Corresponding author at: Cognitive Science Department, Rensselaer Polytechnic Institute, 110 Eighth Street, Carnegie 302A, Troy, NY 12180, USA.
E-mail address: dr.ron.sun@gmail.com (R. Sun).

https://doi.org/10.1016/j.cogsys.2021.09.003
Received 8 April 2021; Received in revised form 17 August 2021; Accepted 30 September 2021
Available online 12 October 2021
1389-0417/

control, opportunity cost, and so on (e.g., Holroyd & Yeung, 2012; Inzlicht, Schmeichel, & Macrae, 2014; Shenhav, Musslick, Lieder, Kool, Griffiths, Cohen, & Botvinick, 2017; for more details, see General Discussion).

To resolve these apparent complexities, while we need to analyze circumstances within which these different findings were obtained, we also need to analyze what mechanisms were involved (Craver & Bechtel, 2006) and what processes were going on that led to these findings, in a detailed and precise way. In so doing, one can obtain a more in-depth understanding of underlying mechanisms. Instead of general assertions about behavior, one can take into consideration factors at a more mechanistic level, thereby generating detailed and precise accounts. A mechanistic model (Craver & Bechtel, 2006) precisely specifies a set of (lower-level) entities and their relations, which lead to exact descriptions of how these entities interact to give rise to (higher-level) outcomes. Computational psychology (Sun, 2008) produces precise mechanistic models, with which exact theories may be constructed and tested, complementing empirical work.

A shortcoming of studies of motivation and performance is often the lack of such process-based, mechanistic theories (aside from mediating variables) and corresponding accounts of data and phenomena (with exceptions as discussed later). The present work takes a diverse range of experimental work and provides a coherent interpretation at a mechanistic level.

In order to achieve a coherent interpretation of a variety of empirical phenomena, a fundamental question is *why* one puts more (or less) effort into something. We may attribute decisions to utilities to be maximized (or disparities to be reduced in some cases). But, when invoking utility, it is often unclear how the basic elements of utility calculation (e.g., payoff, reward, cost, value, expectancy, and so on) should be determined (cf. Shenhav et al., 2017). Without utility, we would be at a loss as to what criterion to use. In order to base utility calculation on a more solid footing, we believe that it is necessary to trace it back to intrinsic human motivation (i.e., the evolutionarily acquired set of basic human motives or needs; see, e.g., Murray, 1938; Reiss, 2004), as well as individual differences in this regard (Bretz & Sun, 2018; Sun & Wilson, 2014), which can provide a *mechanistic* (exact and specific, not just conceptual) basis for utility calculation through providing the basis for determining values of outcomes. Otherwise values may become arbitrary parameters in a model. The present work describes such an approach, providing a framework in which a *diverse* range of data can be interpreted.

This work applies a generic computational model (a cognitive architecture) to this task and addresses questions such as:

- Mechanistically, how do individuals decide how much effort they put into a task? What are the criteria that they use?
- What serves as the basis of such criteria? What are roles of basic motives?
- How do we account for, in one unified framework (with such criteria), motivation-performance phenomena that have been emphasized by different theories respectively? How do we simulate these phenomena in the unified framework?

We will show that many seemingly incompatible studies can potentially be synthesized within a computational cognitive architecture and thus be both clarified and unified (i.e., accounted for by the same mechanisms). This synthesis relies on utility calculation from intrinsic needs/motives.

In the remainder of this paper, first, a cognitive architecture is introduced as a unifying framework. How it handles motivation and performance is detailed. Then a number of example simulations show how empirical data from a variety of studies can be described by this framework (with new or refined interpretations), emphasizing coherent integration and what is minimally necessary. Finally, a general discussion completes this article.

Some methodological notes are in order: (1) since the present work aims to provide a formal, structured description of many empirical findings in terms of generic mechanisms, it is not extremely fine-grained (e.g., concerning minute experimental details). Our model is broader and thus coarser by necessity than some other models. (2) The goal of this work is a minimum, integrative model, without unnecessary (or even less relevant) details; such details would not add much to the understanding of the motivation-performance relationship but complicate the model and add parameters. (3) Given space limitations and our integrative goal, we can only selectively examine/simulate experiments important to the goal. Each experiment selected highlights one aspect of the motivation-performance relationship and is often also of historical importance (instead of latest work that might be less well established). The set of selected experiments is also deliberately made as diverse as possible in accordance with our integrative goal.

2. A unifying cognitive architecture

2.1. The Clarion cognitive architecture

A cognitive architecture is a broad, domain-general, individual-invariant, psychological, theoretical framework with working computational instantiations, for interpreting as well as for actually simulating psychological phenomena (e.g., Sun, 2016).

One such cognitive architecture, Clarion, can serve as the basis for unified, mechanistic, process-based interpretations of motivation-performance phenomena. Clarion has been extensively validated empirically (accounting for many different types of tasks; see, e.g., Bretz & Sun, 2018; Helie & Sun, 2010; Sun et al., 2001; 2005). In particular, Clarion accounts for basic human motives that are the basis of cognition and behavior (Sun, 2009) and thus integrates purely cognitive aspects with motivational aspects (as well as personality, emotion, metacognition, sociality, and culture; Bretz & Sun, 2018; Sun, 2020; Sun & Wilson, 2014; Sun et al., 2016).

Below first a general sketch of Clarion (not specific to this work) is presented; then details of Clarion relevant to motivation-performance phenomena are described.

2.2. Overview of Clarion

Clarion consists of four major subsystems: the action-centered subsystem (ACS) for dealing with action selection involving procedural knowledge (cf. Anderson & Lebiere, 1998), the non-action-centered subsystem (NACS) for reasoning and memory involving general (i.e., declarative) knowledge (Helie & Sun, 2010), the motivational

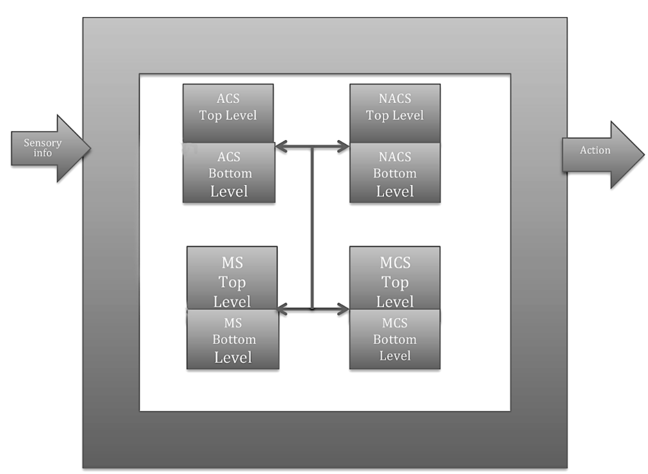

Fig. 1. The Clarion cognitive architecture. The major information flows are shown with arrows. See the text for details.

subsystem (MS) for dealing with motivational processes (Sun, 2009), and the metacognitive subsystem (MCS) for regulating other subsystems (Sun, 2016). See Fig. 1.

Each subsystem consists of two “levels” (two sets of modules). The “top level” carries out explicit (roughly, conscious) processes and encodes explicit knowledge; the “bottom level” carries out implicit (roughly, unconscious) processes and encodes implicit knowledge (for the implicit-explicit distinction and associated theoretical issues, see Reber, 1989; Evans & Frankish, 2009; Sun, 2016). The two levels interact to generate integrated outcomes (Sun et al., 2005).

Below, several subsystems most relevant to this work are sketched. Their mechanisms and processes have been previously justified on the basis of psychological findings; therefore, justifications will not be repeated here (see Sun, 2016 for a full specification as well as justifications; see also Bretz & Sun, 2018; Helie & Sun, 2010; Sun et al., 2001; Sun et al., 2005).

2.2.1. *Action-centered subsystem*

The ACS captures procedural processes (i.e., action selection): Its bottom level carries out implicit procedural processes, and its top level carries out explicit procedural processes (Sun et al., 2001; Sun et al., 2005).

The processes may be described informally as follows (Sun, 2016): One observes the current state (situation) of the world. At the top level, explicit rules recommend actions based on the current state. All *applicable* rules compete based on their utility values. The utility value of an action rule is calculated (implicitly) based on its cost (corresponding to its effort level) and benefit (corresponding to its estimated likelihood and degree of satisfying outstanding needs as represented by activated drives as described later). These utility values are turned into a probability distribution (a Boltzmann distribution, as detailed later), from which a rule is chosen and applied (Sun, 2016). The bottom level of the subsystem automatically (implicitly) computes, on the basis of the current state, values for all possible actions; actions then compete based on these values (also through a Boltzmann distribution, as detailed later). An action is then chosen through selecting the outcome of either the top or the bottom level (with probabilities determined by the relative performance of the two levels; Sun, 2016). Then this cycle of perception and action begins anew.

In this subsystem, the top level is implemented with symbolic action rules , while the bottom level is implemented with Backpropagation neural networks (Rumelhart et al., 1986). Learning can take place at both levels (e.g., Sun et al., 2001). For further details, see Sun (2016).

2.2.2. *Motivational subsystem*

The MS is concerned with why one does what one does. It helps to focus action in ways relevant to one’s functioning in the world (Sun, 2009; Wilson & Sun, 2021). Its resulting goals direct action selection in the ACS. It also influences regulatory functions in the MCS (Sun, 2016).

Dual representation is important: Explicit goals represent specific intentions for action, while drives represent implicit motives or needs (for such a distinction in psychology, see, e.g., Murray, 1938; Reiss, 2004; Sun, 2009; Tolman, 1932; for that within a BDI framework, see, e.g., Dignum, Kinny, & Sonenberg, 2002). Explicit goals derive from implicit drives: For example, a specific goal “*find food*” may be generated based on the “*hunger for food*” drive (Sun, 2009; Tolman, 1932). Note that a generalized notion of drive is adopted here, different from stricter interpretations (e.g., as physiological deficits that require to be reduced; Hull, 1951). It denotes internally felt needs of all kinds that likely lead to corresponding behavior (Sun, 2009).

Primary drives (i.e., basic motives) are essential to any individual (Murray, 1938; Reiss, 2004) and represent basic physiological and psychological needs formed by evolution (i.e., innate needs). Low-level primary drives mostly represent basic physiological needs (e.g., “*food*”, “*water*”, and “*reproduction*”). High-level primary drives mostly represent socially oriented psychological needs (e.g., “*achievement*”, “*affilia-*

Table 1
Primary drives in the motivational subsystem (for details, see Murray, 1938; Reiss, 2004; Sun, 2009).

A. Specifications of primary drives.

food	The drive to consume nourishment.
water	The drive to consume fluid.
sleep	The drive to rest and/or sleep.
reproduction	The drive to mate.
avoiding danger	The drive to avoid situations that have the potential to be or already are harmful.
avoiding unpleasant stimuli	The drive to avoid situations that are physically (or emotionally) uncomfortable or negative in nature.
affiliation & belongingness	The drive to associate with other individuals and to be part of social groups.
dominance & power	The drive to have power over other individuals or groups.
recognition & achievement	The drive to excel and to be viewed as competent at something.
autonomy	The drive to resist control or influence by others.
deference	The drive to willingly follow and serve a person of a higher status.
similance	The drive to identify with other individuals, to imitate others, and to go along with their actions.
fairness	The drive to ensure that one treats others fairly and is treated fairly by others.
honor	The drive to follow social norms and codes of behavior and to avoid blames.
nurturance	The drive to care for, or to attend to the needs of, others who are in need.
conservation	The drive to conserve, to preserve, to organize, or to structure (e.g., one’s environment).
curiosity	The drive to explore, to discover, and to gain new knowledge.

B. Approach- versus avoidance-oriented primary drives.

Approach-oriented Drives	Avoidance-oriented Drives	Both
Food	*Sleep*	*Affiliation & Belongingness*
Water	*Avoiding Danger*	*Similance*
Reproduction	*Avoiding Unpleasant Stimuli*	*Deference*
Nurturance	*Honor*	*Autonomy*
Curiosity	*Conservation*	*Fairness*
Dominance & Power		
Recognition & Achievement		

tion", and "*autonomy*"). These motives are evolutionarily acquired human universals, as validated by empirical and theoretical work in social psychology and other fields (e.g., Maslow, 1943; McClelland, 1951; Murray, 1938; Reiss, 2004; Sun, 2009), although current understanding is still preliminary (not without controversies). These motives may be present to varying extents across individuals (e.g., each with a different prioritization), which are an important source of individual/group psychological differences (Sun & Wilson, 2014). See Table 1(A) for their brief specifications (only the general ideas of these, not exact definitions or exact labels, are relevant here).

Beyond these primary drives, there are also "derived" (non-primary) drives (Sun, 2009; 2016). Not essential to the present work, their details are omitted.

In addition, as indicated by Table 1(B), a drive may be either approach-oriented or avoidance-oriented (or both; Cacioppo, Gardner, & Berntson, 1999; Gray, 1987; Sun, 2016). Approach-oriented drives aim to attain positive consequences, while avoidance-oriented drives aim to avoid negative consequences. The former is sensitive to cues signaling reward, resulting in active approach, while the latter is sensitive to cues of punishment, resulting in avoidance (Cacioppo et al., 1999; Gray, 1987). The *satisfaction* of approach-oriented drives can be measured by a positive number (e.g., between 0 and 1), while the *satisfaction* of avoidance-oriented drives can be measured by a negative number (e.g., between −1 and 0; Sun, 2016).

Processing of drives is such that, roughly, the activation (the strength) of a drive is determined by the product of *stimuluslevel* (a scalar measuring the pertinence of the current situation to the drive) and *deficit* (a scalar measuring the internal inclination towards activating the drive, which captures individual or cultural differences). Relevant technical details are provided in the next subsection.

On the other hand, goals represent explicit, specific intentions for action. Goals may be domain-specific and may be constructed/learned (e.g., on the fly) in a domain-specific way (cf. Anderson & Lebiere, 1998; Gollwitzer, 2012). Goals can have complex internal structures (details are not relevant here and thus omitted; see Sun, 2016).

2.2.3. Metacognitive subsystem

Within the MCS, metacognitive regulations include setting goals (which then direct action selection in the ACS) on the basis of drive activations, setting essential parameters on the basis of drives and goals, and other functions (Sun, 2016; Wilson & Sun, 2021).

Structurally, it is divided into a number of functional modules (Sun, 2016). For instance, the *goal selection* module, in order to select a new goal, first calculates the strength of each possible goal. The strength of a goal is calculated (implicitly) on the basis of the weighted sum of drive strengths (with each weight measuring the relevance of a drive to the selection of a goal); these goal strengths are then turned into a probability distribution (a Boltzmann distribution), from which a new goal is chosen (Sun, 2016; cf. Gollwitzer, 2012). (Although goals can also be chosen explicitly, here we address implicit goal selection only.)

2.3. Details regarding motivation and action in Clarion

Based on the general framework sketched above, to account for the motivation-performance relationship, pertinent details of Clarion (Sun, 2016) are described below. Since our goal is a minimum model accounting for a diverse range of data, aspects not directly relevant are omitted. See Fig. 2 for a diagram of the relevant part of Clarion, which is explained below.

2.3.1. Drive activation within the MS

In the MS, upon receiving inputs concerning the current situation, strengths (activations) of drives are calculated (implicitly; Sun, 2016), basically by the product of *stimuluslevel* and *deficit*:

$$ds_d = g_d \times stimuluslevel_d \times deficit_d + b_d$$

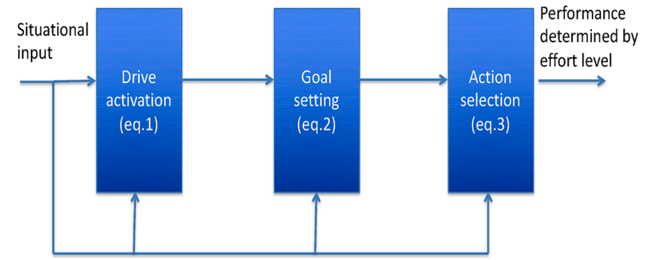

Fig. 2. Diagram of the Clarion motivation-performance model (see the text for explanations; see Table 2 for equations).

where ds_d is the strength of drive d (which is a function of the current situation and personal characteristics), g_d is a scaling parameter, $stimuluslevel_d$ is a scalar value representing the pertinence of the current situation (state) to drive d (McClelland, 1951), $deficit_d$ is a scalar indicating an individual's internal inclination toward activating drive d,[1] and b_d is the situation-invariant baseline strength of drive d.

Note that $deficit_d$ is more individual-specific, while $stimuluslevel_d$ is more situation-specific; they may be estimated separately (e.g., through different instruments). Based on prior work (Sun & Wilson, 2014), b_d is set to 0, and g_d to 1 (reducing degrees of freedom). For the sake of focus, how $stimuluslevel_d$ results from learning/evolution is omitted (Bugrov et al., forthcoming); thus, $stimuluslevel_d$ is treated here as a parameter whose value is determined from interpretations of situations (more later). For $deficit_d$, different individuals (with different cultures, personalities, and other characteristics) may have different values[2] (and therefore may have different drive activations when facing the same situation; Sun & Wilson, 2014). For simplicity, in this work, $deficit_d$ is assumed to follow a normal distribution within a population.

2.3.2. Goal setting by the MCS

Based on drive strengths obtained above, goals strengths (activations) are calculated (implicitly) by the MCS. The strength of a goal is determined by a weighted sum of drive strengths, where each weight is the relevance of a particular drive to the selection of the goal in question (Sun, 2016):

$$gs_g = \sum_d relevance_{s,d\rightarrow g} \times ds_d$$

where gs_g is the strength of goal g, $relevance_{s,d\rightarrow g}$ represents the support that drive d provides to the selection of goal g (given situation s), which measures how relevant goal g is to drive d, ds_d is the strength of drive d (as determined by the MS), and the summation is over all drives. Although $relevance_{s,d\rightarrow g}$ may be learned (Bugrov et al., forthcoming), this is not needed in the present work.

Once calculated, the set of goal strengths is turned into a Boltzmann distribution and a new goal is chosen stochastically from that distribution:

[1] Note that $deficit_d$ represents an internal inclination (a generalized notion), not necessarily a deviation from a set point of a homeostatic mechanism (cf. Hull, 1951). For a physiological drive where there exists an ideal level, $deficit_d$ may represent one's *perception* of the deviation (Sun, 2016). For a high-level drive, however, there is often no set point. $deficit_d$ may change (e.g., due to habituation). $deficit_d$ differs from b_d, as its effect on ds_d varies with situations.

[2] Mostly high-level socially oriented drives (where *deficit* parameters are relatively stable) are responsible for important psychological differences (e.g., personality dimensions such as the Big Five; Sun & Wilson, 2014). Although changes of these *deficits* do occur, their differences across individuals matter more (Sun & Wilson, 2014; Wilson & Sun, 2021). For example, one individual's average value may be higher than another's, although these values fluctuate.

$$p(g) = \frac{e^{gs_g/\tau}}{\sum_i e^{gs_i/\tau}}$$

where *p(g)* is the probability of selecting goal *g*, gs_i is the strength of goal *i*, τ is the "temperature", and *i* ranges over all goals. The "temperature" τ controls the degree of stochasticity (e.g., when the temperature is sufficiently low, the process amounts to always choosing the goal with the highest strength). This corresponds to the choice axiom known for capturing nondeterministic choice behavior (Luce, 1959).

2.3.3. Action selection by the ACS

Within the ACS, each rule is in the form of "*state, goal → action*". In this work, we are concerned with coarse-grained actions that only specify a certain level of effort (a discrete, scalar value, chosen from a set of possible values; e.g., number of minutes devoted to a task). Multi-grained action selection is assumed where each action concerns a decision at a certain level of abstraction only (Sun, 2016). Performance of an individual on a task (by some performance measure) may be assumed to be a function of the chosen effort level. A linear function (the simplest functional form, in the absence of evidence to the contrary; cf. Vancouver et al., 2010) is assumed:

$$performance = p \times effortlevel_j + q$$

where $effortlevel_j$ is a discrete, scalar value and *p* and *q* are parameters.

All applicable action rules (all those matching both the current situation and the current goal) compete to be chosen based on their *utility* values (Sun, 2016). The utility value of a rule is (implicitly) calculated by combining measurements of benefit and cost of applying the rule (e.g., Kahneman & Tversky, 1992). Benefit is determined based on assessment of possible outcomes: how likely and how much they satisfy one's needs (in terms of activated drives); cost is determined by effort level (i.e., the more effort, the higher the cost, although possible complications exist). Utility (and consequently effort allocation) being closely related to one's internal needs/motives (in addition to external situations) is the main underlying assumption here (beyond most existing computational models; cf. Braver et al., 2014).

With some algebraic derivation (as detailed in the appendix), utility of a rule can be calculated as follows:

$$U_j = effortlevel_j \times (a \times value(g) - c)$$

where U_j is the utility of rule *j*, $effortlevel_j$ is a scalar specified by rule *j*, *a* is the effort coefficient, and *c* is the cost coefficient (see the appendix).

The subjective value of achieving goal *g*, denoted as *value(g)*, can be assessed on the basis of satisfying various currently activated drives (i.e., various current needs) when goal *g* is achieved ((Steel and König, 2006); Sun, 2016). It is calculated by the sum of the products each of which involves the strength of a drive and the anticipated satisfaction of the drive (by achieving the said goal):

$$value(g) = \sum_d ds_d \times satisfaction_d(g)$$

where $satisfaction_d(g)$ represents the sense of how well the achievement of goal *g* satisfies drive *d* (which can be learned from experience; Bugrov et al., forthcoming), and the summation is over all activated drives. Note that, while ds_d varies across situations, $satisfaction_d(g)$ is more stable. Thus *value(g)* may be viewed as the overall satisfaction of outstanding needs as a result of achieving goal *g*.

If $a \times value(g) - c > 0$, then a positive relationship exists between effort and utility and thus more effort leads to higher utility. If it is<0, then a negative relationship exists between effort and utility and thus less effort leads to higher utility. This is the main mechanism for deciding on effort levels (e.g., selecting an effort action that maximizes utility).

To select an action rule (specifying an effort level), a Boltzmann distribution is constructed from the set of rule utility values:

$$p(j) = \frac{e^{U_j/\tau}}{\sum_i e^{U_i/\tau}}$$

where *p(j)* is the probability of selecting rule *j*, U_i is the utility of rule *i*, τ is the "temperature" (as explained before), and *i* ranges over all applicable rules. One rule is selected from this distribution and applied.

The process above can be applied repeatedly, one step at a time, capturing a "feedback loop" (situation → action → situation, etc.). However, for the sake of brevity and focus, simulations in this work will be courser-grained; step-wise details are not explicitly modeled (although they can be incorporated).

Note that there has been prior work on effort allocation based on utility. For example, Kool and Botvinick (2014) argued that mental effort, similar to physical labor, was based on cost-benefit analysis. Griffiths et al. (2015) and Shenhav et al. (2017) outlined related general arguments.

Note also that, in the context of the experiments addressed in the present work, individuals did not have sufficient experience with these settings, so the bottom level of the ACS was not sufficiently trained and thus did not contribute significantly to performance. In other circumstances, the bottom level can be effective and can be governed by the same cost-benefit consideration.

2.4. Summary of the motivation-performance model

Thus, our motivation-performance model can be summarized as four equations, as shown in Table 2 (see also Fig. 2).

Note that different drive activations (i.e., ds_d, resulting from $deficit_d$ and $stimuluslevel_d$) lead to different goal outcome values (i.e., *value(g)*), which in turn affect utility calculation, consequent rule selection, and thus resulting effort level and performance. For example, a higher "*achievement*" drive may lead to a higher goal outcome value and thus more effort being put into a task (e.g., Phillips & Gully, 1997).

Within this model, roles of key parameters are hypothesized as follows:

- The parameter $stimuluslevel_d$ (in the drive strength equation) represents external conditions that are relevant to a drive (including, e.g., a performance target; (Locke & Latham, 1990, 2013)). A higher *stimuluslevel* value likely leads to a higher activation (strength) of the corresponding drive and consequently a higher/lower goal outcome value (i.e., *value(g)*) and a higher/lower utility (depending on other parameters).
- Effects of externally assigned goals (targets) may be such that, for some drives, a high *stimuluslevel* value may result from a high assigned target and a low *stimuluslevel* value from a low (or no) assigned target (Locke & Latham, 1990). These drives can be either approach- or avoidance-oriented, and thus different effects ensue.
- An assigned learning goal (which focuses on some learning, rather than performance, measures; e.g., Seijts & Latham, 2001) likely leads (in part) to a high *stimuluslevel* for some approach-oriented drives (e.g., "*curiosity*"), due to its learning focus; an assigned performance goal (focusing on some performance measure) likely leads (in part)

Table 2
The motivation-performance model. See the text for definitions and other details.

E1: $ds_d = stimuluslevel_d \times deficit_d$
E2: $gs_g = \sum_d relevance_{s,d \to g} \times ds_d$
(A Boltzmann distribution of the gs_g's is then used to select a goal)
E3: $U_j = effortlevel_j \times (a \times \sum_d (ds_d \times satisfaction_d(g)) - c)$
(A Boltzmann distribution of the U_j's is then used to select a rule and consequently an effort level)
E4: $performance = p \times effortlevel_j + q$

to a high *stimuluslevel* for some avoidance-oriented drives (e.g., concerning avoiding performance failure), because of its focus on outcome (Brooks et al., 2012; Chen & Latham, 2014; Locke & Latham, 1990; 2013; Seijts & Latham, 2001; Wilson & Sun, 2021).

- The *deficit* parameters capture a major source of individual (or cultural) differences (Sun & Wilson, 2014). The *deficit* parameters are relatively stable, determined internally (regardless of *stimuluslevel*). For example, some achievement-oriented individuals may have higher *deficit* values for some approach-oriented drives (e.g., "*achievement*"; Sun & Wilson, 2014). Higher *deficit* values for approach-oriented drives likely lead to corresponding drives being more highly activated and consequently higher goal outcome values and higher utilities for effort.
- When "intrinsically" motivated (as defined by Ryan & Deci, 2000), multiple needs (drives) of an individual may be highly activated (e.g., "*achievement", "autonomy*", etc.) and their simultaneous satisfactions under a goal lead to higher goal outcome values and higher utilities. When not "intrinsically" motivated, some of the drives may be highly activated and satisfied but some others may not, thus resulting in lower values (Krapp, 2005; Ryan & Deci, 2000). Note that in Clarion, all drives are considered intrinsic; goals and actions resulting from them are (at least in part) intrinsically motivated (a view broader than Ryan & Deci's; cf. Locke & Schattke, 2019).

Note that several parameters are associated with each drive and there are multiple relevant drives in each situation. This number of parameters is inevitable, given the broad framework that invokes essential human motives resulting from the history of human evolution. While simplicity is attractive, models with more parameters may also be needed, because (1) one may need a broad framework addressing many different tasks and situations, and (2) one may need to do so at a level sufficiently detailed—with mechanistic process details, which lead to more parameters.

One possible doubt is that computational models may not provide more insights or explanatory depth than existing verbal theories. To address that, note that verbal theories are often vague, in the sense that some (or most) process details of a theory are left out of consideration, and the theory may thus be somewhat vacuous, inconsistent, or otherwise problematic. These problems are often not discovered until a computational model is developed (Hintzman, 1990). Given the complexity of the mind, it has proven difficult to explore its fine-grained details (Sun, 2008); computational models remedy such problems. See also earlier discussion of mechanistic models in general (Craver & Bechtel, 2006).

Another doubt may be that, since equations here provide a mathematical description , there is no role for Clarion and its mechanisms. Note that equations can specify (and may even constitute) mechanisms. In this work, neuroscience is not the focus, so no biological mechanisms are specified, only abstract mechanisms expressed by equations and/or algorithms from Clarion. Moreover, to properly situate these equations or algorithms, an overarching framework is needed. Such a framework may be explicitly stated (as in Clarion), or implicitly assumed (if not explicitly stated). It is better to make it explicit so that there is no unnecessary ambiguity.

One may also argue that effort allocation could (or should) be done implicitly. However, "effort" in this work refers mainly to time spent on a task; therefore no distinction is made between explicit versus implicit effort (cf. Shenhav et al., 2017). Likewise, in terms of effort level selection, explicit and implicit processes can be guided by the same utility function (e.g., embodied by a neural network), thus making no difference in the present context.

A related point is that implicit-explicit interactions should not be limited to implicit motives giving rise to explicit goals. In fact, there are dynamic, multi-faceted interactions between implicit and explicit processes (e.g., Helie & Sun, 2010; Reber, 1989; Sun et al., 2001; Sun et al., 2005). Such interactions, although not addressed here, have been explored extensively in Clarion (see Sun, 2016 for details).

Using this model, a set of experiments and their simulations are examined below, to demonstrate how the model provides unified interpretations of different findings.

3. Simulations

For each simulation below, first a human experiment is summarized. Then the general idea for capturing it is sketched. Simulation setups are specified based on the idea. The result of the simulation is compared with the data from the human experiment.

Each experiment simulated below was selected to demonstrate a particular aspect of the motivation-performance relationship (mostly from industrial-organizational and social psychology); we cannot possibly cover all relevant experiments. The fact that our model is capable of simulating an experiment is used to suggest that the corresponding aspect of the motivation-performance relationship can be accounted for by the model.

Our model, as stated before, is meant to be broadly scoped and thus coarse-grained by necessity (for space and model complexity considerations). Thus we did not attempt to capture all the details in each experiment, emphasizing instead model generality.

Since this work is meant to provide a computational account encompassing a variety of phenomena, it is necessary to make some general assumptions about underlying entities, relations, and mechanisms (based on prior work cited earlier). Theoretical (computational) and experimental work are different in nature.

Since each simulation was meant to convey ideas concerning general mechanisms, parameter tweaking was kept to a minimum. Simulation results were not sensitive to exact parameter values used (Veksler et al., 2015). Settings of parameter values were justified by the existing literature (see below).

Note that, while each simulation below provides a plausible interpretation of data, it may not be the only one possible, nor necessarily the most accurate, but together these simulations lead to a *unified* framework, which has merits of its own.

3.1. *Effects of external rewards on "intrinsic" motivation*

3.1.1. *Human experiment*

In Deci's (1971) classical experiment, the task was to solve a puzzle called Soma (believed to be interesting to college students). The participants were 24 introductory psychology students, who were randomly distributed to the experimental and the control group (12 each). The experiment was divided into three sessions. For each session, the participants were asked to solve four puzzles as fast as possible. The first session was the same for both groups. For the control group, the second and third sessions were the same as the first. During the second session, the experimental group was paid $1 per solved puzzle. During the third session, the experimental group was told that there was no money to pay them. That is, the experimental group was given an extrinsic motivation (a monetary reward) during the second session that was rescinded during the third session.

In the middle of each session, participants were given eight minutes of free time: They could do whatever they wanted (e.g. read magazines, relax, or work on the puzzles). However, the time they spent on solving the puzzles was recorded as the primary dependent measure. The puzzles given for these periods were impossible to solve, so that the time

Table 3
Mean number of seconds spent working on the puzzles during the eight-minute free choice periods.

	Session 1	Session 2	Session 3	S_3 - S_1
Experimental	248.2	313.9	198.5	−49.7
Control	213.9	205.7	241.8	27.9

that one spent on the puzzles represented motivation as opposed to ability to solve puzzles.

The experimental results (Table 3) showed that the withdrawal of monetary reward significantly decreased performance (effort): For the experimental group, performance of session 1 (where there was no monetary reward) was higher than performance of session 3 (where the monetary reward was rescinded). Furthermore, the difference in performance between session 3 and session 1 (S_3 - S_1) was statistically different across the two groups.[3] No other statistically significant effect was reported.

3.1.2. Simulation

Conceptual Description. As hypothesized earlier, individuals choose a level of effort through rules each of which specifies a different effort level, which compete based on their utilities (i.e., evaluations of cost and benefit, in a mostly implicit way; section 2.2.1). Benefit of a rule is assessed based on an individual's outstanding needs (i.e., currently activated drives) and their possible satisfactions by outcomes obtained through the level of effort recommended by the rule (sections 2.3.1 and 2.3.3). Cost of a rule is assessed on the basis of the effort level recommended by the rule (section 2.3.3). Thus, the utilities of rules can be calculated and a rule is chosen on that basis, which recommends a certain effort level, which in turn determines a performance level (section 2.2.1 and 2.3.3).

In this experiment, relevant drives presumably include "*achievement*" and "*curiosity*" (sections 2.2.2). For the experimental group, the monetary reward given amounts to adding an additional motive (activation of a drive for money or other resources[4]). Rescinding the monetary reward leads to a significant reduction of motives for the group (i.e., loss aversion—an inherent human tendency; Kahneman & Tversky, 1992; Baumeister et al., 2001). Hence the performance difference between the experimental and the control group during sessions 2 and 3 (i.e., better performance by the experimental group in session 2, and worse performance by the experimental group in session 3). Clarion thus provides a plausible interpretation of the phenomenon (see mechanistic details below).

The reader, if not interested in technical details, may skip "Simulation Setup" below and go directly to "Simulation Results".

Simulation Setup. 48 simulated participants were involved, divided into two groups of 24 each. In the model of each participant, the ACS, the MS, and the MCS were involved. For each drive within the MS, the product of *deficit* and *stimuluslevel* determined the drive strength. A goal was then chosen by the MCS based on the goal strengths, which were determined from the drive strengths. Anticipated *satisfactions* of these drives, along with their *drive strengths*, determined the *utility* of each rule for accomplishing the goal within the ACS. The *utility* values of the rules in turn determined the rule chosen within the ACS and thus the corresponding *effortlevel* and consequently the performance.

Specifically, the situational input to a simulated participant was simply the session setting (initial, monetary rewards introduced, or monetary rewards rescinded). The following was done: First, the *deficit* of each drive involved was drawn from a normal distribution (defined by its mean and standard deviation; Table 4) for capturing individual differences. The *stimuluslevel* was determined based on the current situation (see Table 5).[5] Then, *drive strengths* were obtained according to equation E1. The goal of performing the task was selected based on goal strengths obtained according to equation E2 (there was only one relevant goal in this case; the *relevance* parameters were all 1′s; Boltzmann temperature = 0.7). The value of accomplishing the task goal was then assessed (within equation E3), based on *drive strengths* and their *satisfactions* (Table 6); then the *utility* of each rule was obtained according to equation E3 ($a = 1$, $c = 0.185$). Each rule specifies an *effortlevel*: *no effort, low effort, median effort, high effort,* or *maximum effort* (mapped to a scalar value between 1 and 10). A rule (and thus an *effortlevel*) was selected based on these *utility* values (through a Boltzmann distribution, with temperature = 0.7). Finally, performance (i.e., time spent on the puzzle) was determined as a function of the chosen *effortlevel* (equation E4; $p = 50$, $q = 0$).

The model was robust and thus the exact parameter values shown in the tables were not very important; what mattered more was the relative ordering of parameter values across conditions (Veksler et al, 2015). The ordering of parameter values was not arbitrary and can be readily justified as follows. For instance, for the experimental group, in accordance with the hypothesis in "Conceptual Description", the *stimuluslevel* for "*resource*" across the three sessions should be: medium, high, low; this was because the introduction of monetary reward during session 2 increased significantly its *stimuluslevel,* but then during session 3 the removal of monetary reward led to lower *stimuluslevel* than before (due to the perception of loss; Kahneman & Tversky, 1992; Baumeister et al., 2001).[6] The exact values were not as important as the above ordering.

Table 4
The mean and the standard deviation of the *deficit* for each relevant drive. (Note that the model was robust; exact parameter values were not important.)

	Achievement	*Curiosity*	*Resource*
Mean	0.7	0.7	0.7
Standard Deviation		0.1	

Table 5
The *stimuluslevel* for each relevant drive of the experimental group. The columns represent drives. The rows represent situational inputs. For the control group, situations remained the same as session 1. (The model was robust; exact parameter values were not as important as the ordering of these values across conditions.)

	Achievement	*Curiosity*	*Resource*
Session 1 (initial)	0.7	0.7	0.2
Session 2 (monetary reward introduced)	0.6	0.6	0.8
Session 3 (monetary reward rescinded)	0.5	0.5	0.1

Table 6
The anticipated *satisfaction* of each relevant drive of the experimental group. The columns represent drives. The rows represent goals. For the control group they remained the same as session 1. (The model was robust; exact parameter values were not as important as the ordering of these values across conditions.)

	Achievement	*Curiosity*	*Resource*
Task Goal for Session 1	0.7	0.7	0.3
Task Goal for Session 2	0.7	0.7	0.9
Task Goal for Session 3	0.5	0.5	0.2

[3] This effect ($p < 0.10$) was marginally significant but considered sufficient in the 1970s (although not at present). However, since then, many similar or related experiments have showed the effect in a statistically significant way (see, e.g., Murayama et al., 2010; Ryan & Deci, 2000).

[4] With regard to money, often both approach- and avoidance-oriented drives are involved (Sun, 2016). However, in this simple simulation, we did not address such subtleties and only included a derived drive ("*resources*") for it (as a summary of multiple primary drives involved; Sun, 2016). Our (unifying) interpretation is also somewhat different from Deci's (which cannot be detailed here, given the scope). Alternatively, parameters can be learned, which makes moot the issue of what parameters to use to capture a phenomenon.

[5] Note that the mean and the standard deviation of the *deficit* of each drive were determined based on the authors' consensus. So was the *stimuluslevel* of each drive.

[6] Changing *stimuluslevel* leads to changing the drive strength and, along with *satisfaction*, changing the utility values, which determine action. As noted earlier, parameters can be learned, which makes moot the issue of what parameters to use to capture a phenomenon. The details of loss perception (e.g., violation of an internal expectation) are beyond the scope of the present paper.

Likewise, the anticipated *satisfaction* of "*resource*" should also follow this pattern for the same reason. For the other two drives of the experimental group, the *stimuluslevel* during sessions 2 and 3 should be lower than during session 1, due to the higher competing *stimuluslevel* for "*resource*" (during session 2) or due to the general perception of loss (during session 3, as explained before). Their anticipated *satisfaction* should essentially stay the same, except during session 3 due to the perception of loss. Note that technically only one of the two drives, "*achievement*" or "*curiosity*", is minimally needed to account for the data.

Note that, as the model is meant to provide an integrative description and interpretation of a diverse range of data, it is not pertinent to evaluate parameters with regard to simulation accuracy for an individual experiment (but see Bretz & Sun, 2018). Also, although parameters (e.g., *relevance* or *satisfaction*) can be learned (Bugrov et al., forthcoming), to focus on the main issue and to avoid complicating the model, learning is not included here. Likewise, as the model is coarse, the role of the NACS is not detailed either.

Simulation Results. The simulation results (Table 7) demonstrated the negative effect when the reward was rescinded after it was introduced. For the experimental group, performance of session 1 (where there were no monetary reward) was higher than that of session 3 (where the monetary reward was rescinded; S_3 - S_1 = -37.5), the same as in the human data. The difference between the two groups in terms of performance difference between session 3 and session 1 (i.e., in terms of S_3 - S_1) was statistically significant ($F(1, 46) = 20.90$, $p < 0.0001$), similar to the human data. See Table 7 for the simulation results.

3.1.3. Discussion

Deci (e.g., 1971) has been emphasizing the importance of "intrinsic" motivation (i.e., without separate reward, based on internal needs for *achievement*, *affiliation*, *autonomy,* and so on) as opposed to "extrinsic" factors (including, but not limited to, monetary reward). He speculated on the inner working of these aspects and their effects. But one often does not know whether or how such speculations work, and that is why modeling and simulation are useful.

The insight from the simulation derives from the fact that the model provides a unifying account of a diverse range of empirical phenomena, which is achieved through *mechanistically* linking utility calculation with basic motives/needs. As the very definition of "intrinsic" motivation is questionable (as alluded to earlier), modeling helps to substantiate and clarify it mechanistically.

One might argue that with enough parameters, any pattern of data, including opposite results, can be simulated by a model. This is not the case here, unless one uses parameter values that are unreasonable and unjustified by the existing literature. Furthermore, while alternative interpretations of this phenomenon are possible, the present simulation should be evaluated in the context of unified interpretations of many other phenomena. In addition, the Clarion framework can also facilitate formal comparisons of different accounts (see, e.g., Bretz & Sun, 2018).

One might also argue that ambiguities abound in simulation in terms of which drives are actually involved in a situation. But this would boil down essentially to an argument about definitions or labels of drives involved in a simulation, which is tangential to our focus here.

Another possible objection is that simulation results may derive from values of input parameters and thus the equations provide no value. To counter this point, note that, regardless of the importance of input parameters, a precise mechanism mapping these parameters to the outcome is needed. The present work provides such a mechanism in a set of simple equations. Can these equations be made even simpler and still account for all the empirical results described in the present paper? The answer is no, based on all the arguments, justifications, and data discussed in the paper. This work integrates a diverse range of phenomena; no simpler mechanism would be sufficient for that.

Note that even though one equation (i.e., E3) in one subsystem is important to this simulation, it works within the context of other equations and other subsystems. One equation or one subsystem alone cannot account for the data (e.g., E3 relied on E1 and E2, etc.; see Table 2).

In addition, some researchers have found more complex relationships among "intrinsic" motivation, extrinsic factors, and performance (e.g., Deci, Koestner, & Ryan, 1999). Locke and Latham (2002) argued that externally assigned goals quickly led to "intrinsic" motivation. Some others (e.g., Amabile & Pratt, 2016) distinguished "controlling" versus "informational" extrinsic motivation. The literature on this issue (e.g., Deci, Koestner, & Ryan, 1999; Murayama et al., 2010; Ryan & Deci, 2000 etc.) cannot be detailed here due to lengths. These findings, however, should be accounted for in the future.

Table 7
Mean number of seconds spent on the puzzles during the eight-minute free choice periods (in simulation).

	S_1	S_2	S_3	S_3 - S_1
Experimental	237.5	247.91	200.00	−37.5
Control	243.75	229.17	247.91	4.17

3.2. Effects of self choices on motivation

3.2.1. Human experiment

In Iyengar and Lepper (1999), the task was to solve as many anagrams as possible. Participants were 52 Asian-American and 52 Anglo-American children (from 7 to 9 y.o.). Each group was randomly distributed into 3 conditions: personal choice, experimenter's choice, and mom's choice. In the personal choice group, children were given the opportunity to choose a category of anagrams. In the experimenter's choice group, an experimenter chose a category for them. In the mom's choice group, children were told that their mothers picked the category. Then, the children were given six minutes to solve anagrams in the chosen category. Performance was measured by the total number of correctly solved anagrams.

The results (see Table 8) showed that Anglo-American children performed the best when they had a choice. On the contrary, Asian-American children performed the best when the choice was made by their mothers. An Ethnicity × Condition ANOVA showed statistical significance of experimental condition ($F(2, 99) = 21.77$, $p < 0.0001$), ethnicity ($F(1, 99) = 24.33$, $p < 0.0001$), and interaction of these two factors ($F(2, 99) = 22.68$, $p < 0.0001$).

3.2.2. Simulation

Conceptual Description. As before, an individual decides on an effort level based on utility (section 2.2.1): Utilities of rules are calculated; a rule (specifying an effort level) is chosen on that basis, which in turn determines performance (section 2.3.3). The motivational difference between Asian-American and Anglo-American children is captured through activations of drives (see section 2.4): While Anglo-American children value autonomy more (Henrich et al., 2010; Hofstede, 2001; Nisbett et al., 2001) and have higher inclinations towards activating the "*autonomy*" drive (sections 2.2.2 and 2.3.1), Asian-American children have higher inclinations towards activating the "*deference*" drive (Henrich et al., 2010; Hofstede, 2001). (There are also other drives, such as "*achievement*", that do not differ systematically.) This difference in drive

Table 8
Mean number of anagrams correctly solved by experimental condition and ethnicity.

	Personal Choice	Experimenter's Choice	Mom's Choice
Anglo American	M = 7.39, SD = 1.88	M = 3.06, SD = 1.89	M = 2.94, SD = 1.84
Asian American	M = 6.47, SD = 2.10	M = 4.28, SD = 2.65	M = 8.78, SD = 2.24

Table 9
The mean and the standard deviation of the *deficit* of each relevant drive.

	Achievement	*Autonomy*	*Deference*
Mean for Anglo-American	0.6	0.8	0.1
Mean for Asian-American	0.6	0.3	0.5
Standard Deviation		0.1	

Table 10
The *stimuluslevel* for each relevant drive. The columns represent drives. The rows represent situational inputs.

	Achievement	*Autonomy*	*Deference*
Personal choice	0.7	0.7	0.1
Experimenter's choice	0.7	0.6	0.1
Mom's choice	0.7	0.6	0.6

Table 11
The anticipated *satisfaction* of each relevant drive. The columns represent drives. The rows represent goals.

	Achievement	*Autonomy*	*Deference*
Task goal with personal choice	0.7	0.0	0.0
Task goal with experimenter's choice	0.7	-0.7	0.2
Task goal with mom's choice	0.7	-0.7	1.0

Table 12
Mean number of anagrams correctly completed in simulation by experimental condition and ethnicity.

	Personal Choice	Experimenter's Choice	Mom's Choice
Anglo American	M = 6.52, SD = 2.30	M = 2.82, SD = 1.74	M = 4.14, SD = 1.77
Asian American	M = 5.38, SD = 1.99	M = 3.17, SD = 1.63	M = 6.26, SD = 1.52

activations translates into different valuations of outcomes and thus different utility values; different utility values lead to different selections of rules (section 2.3.3). Therefore, individuals can show different effort levels (as a result of different rules) in these circumstances (e.g., choice made by oneself versus choice made by mothers, each circumstance suited for a different cultural orientation), based on their different drive activations (i.e., their different cultural orientations: emphasizing "*autonomy*" versus emphasizing "*deference*"); effort levels are the highest when circumstances match cultural orientations (whereby outcome valuations are the highest; e.g., choice made by oneself matches emphasizing "*autonomy*"). Clarion thereby provides a process-based, mechanistic interpretation of this cultural difference (see details below).

Simulation Setup. 102 simulated participants (roughly the same as in the human experiment) were divided into six groups of 17 each. The process within each simulated participant was the same as in the previous simulation (involving the ACS, the MS, and the MCS).

Specifically, the input to each simulated participant was the task situation. For each simulated participant, as before, the *deficit* of each drive was drawn from a normal distribution (Table 9) to capture individual/group differences. The *stimuluslevel* of each drive was determined by the task situation (Table 10). Then, *drive strengths* were obtained according to equation E1. The task goal was then selected based on goal strengths (equation E2; only one relevant goal; *relevance* = 1). The value of the task goal was assessed based on the *strength* and the anticipated *satisfaction* of each drive (Table 11); *utility* values of rules (each specifying a different *effortlevel*) were obtained according to equation E3 (a = 1, c = 0.185). A rule was selected based on *utility* values (through a Boltzmann distribution, with temperature = 0.7). Performance was determined by the chosen *effortlevel* (equation E4; p = 1.5, q = 0).

The ordering of parameter values across conditions/groups can be justified as follows. Anglo-Americans would have a higher *deficit* (i.e., internal inclination) for the "*autonomy*" drive than for "*deference*", while the reverse would be true of Asian-Americans, in accordance with the hypothesis in "Conceptual Description" (Henrich et al., 2010; Hofstede, 2001; Nisbett et al., 2001). Furthermore, the *stimuluslevel* for "*deference*" was the highest in the mom's choice condition (as a result of invoking mothers)[7], while the *stimuluslevel* for "*autonomy*" was the highest in the personal choice condition (because personal choice implied autonomy). The anticipated *satisfaction* was such that the satisfaction of the "*deference*" drive was somewhat higher in the experimenter's choice condition compared with the personal choice condition, and much higher in the mom's choice condition (as naturally implied by these conditions), while the *satisfaction* of the "*autonomy*" drive was low in both the experimenter's choice and the mom's choice condition (because both meant lack of autonomy). No other parameters were significant in this simulation (based on definitions and assumptions in section 2).

Simulation Results. This simulation (Table 12) successfully reproduced the major findings from the human experiment: Simulated Anglo-American children performed the best when they themselves chose the category (M = 6.52), while their performance was lower under the experimenter's choice and the mom's choice condition (M = 2.82 and M = 4.14, respectively). Simulated Asian-American children performed the best when the choice was made by their mothers (M = 6.26), better than under the personal choice and the experimenter's choice condition (M = 5.38 and M = 3.17, respectively). Statistical analysis of the simulation results showed a significant difference of condition ($F(2,99) = 11.97$, $p < 0.0001$) as in the human data, as well as an interaction of condition and ethnicity ($F(2,96) = 12.51$, $p < 0.0001$) as in the human data. (The difference across ethnicity was not pertinent to our focus; the model was broad and thus coarse-grained.)

3.2.3. *Discussion*

This phenomenon concerning cultural differences has not been addressed in depth by any theories mentioned before. However, this phenomenon is important, for both theoretical and practical reasons (see, e.g., Henrich et al., 2010; Hofstede, 2001; Nisbett et al., 2001); for instance, practical prescriptions from theories that ignore such differences may not work well, and sometimes may even have opposite effects than intended.

While emphasizing "intrinsic" motivation in general, Ryan and Deci (2000) did not explore extensively how it might vary across individuals or cultures exactly. They attributed their notion of intrinsic motivation to the need for achievement, autonomy, affiliation, and so on, largely invariant across cultures. Experiments such as the one addressed here (as well as others, e.g., Thompson et al., 1993) show that this might not be the case and differences and variability need to be taken into consideration.

The present simulation provides one possible interpretation of this phenomenon, in the context of similarly accounting for a variety of other phenomena. Although one may argue that with enough parameters any pattern of data can be accounted for, this is not the case here, unless one uses parameter values unjustified by the existing literature.

[7] In many Asian cultures, family is the most important and the most basic unit, which calls for relationships of deference and so on. Family relationships also map onto a number of other significant social relationships, but their range is limited. "Experimenter" (and other minor or ad hoc roles) has no significant status in these cultures. Thus one's mother is far more important in this context.

3.3. Effects of different types of goal setting

3.3.1. Human experiment

In Seijts and Latham (2001), the task was to complete class schedules, over three trials. 96 participants were randomly distributed into six groups of approximately equal size. Each group was assigned a different goal: (1) "do your best" outcome goal, (2) "do your best" learning goal, (3) distal outcome goal, (4) distal learning goal, (5) distal outcome goal with proximal goals, or (6) distal learning goal with proximal goals. An outcome goal is commonly known as a performance goal (or target), specifying a target number of schedules to be completed, while a learning goal specifies the number of strategies to be discovered. Distal goals specify the total number to be achieved over three trials, while proximal goals specify the number for each trial.

The experimental results (Table 13A) showed that, excluding the "do your best" goals, on average, the learning goals led to better performance than the outcome goals ($t(62) = 3.71$, $p < 0.001$). Adding proximal goals did not change performance significantly (regardless of whether an outcome or a learning goal was involved). The "do your best" goals had middling results, mitigating effects of learning and outcome goals: The "do your best" outcome goal led to better performance than a specific outcome goal ($t(46) = 3.71$, $p < 0.001$), while the "do your best" learning goal led to worse performance than a specific learning goal ($t(46) = 2.27$, $p < 0.05$).

To highlight the main point of interest, these conditions were grouped into three categories: the learning goal condition (both distal and distal-proximal), the outcome goal condition (both distal and distal-proximal), and the "do your best" goal condition (including both variants). Among the three, the learning goal condition performed the best, and the outcome goal condition performed the worst; see Table 13B. (We avoided delving into minute details of the experiment for reasons explained in Introduction.)

3.3.2. Simulation

Conceptual Description. To capture the main point of this experiment, the three aggregate conditions are considered. As before, individuals choose effort that they put into a task through utility calculation. The three conditions likely lead to different drive activations (see section 2.4): The learning goal condition triggers more approach-oriented drives (e.g., "*curiosity*"; section 2.2.2) because of its learning focus; the outcome goal condition triggers more avoidance-oriented drives (e.g., "*honor*", concerning avoiding performance failure; see section 2.2.2) because of its emphasis of a high outcome (Locke & Latham, 1990; 2002; 2013). The expected satisfactions of these drives are such that the avoidance-oriented drives (e.g., "*honor*") are least likely satisfied in the outcome goal condition (because of the demand of a high outcome), and the approach-oriented drives (e.g., "*curiosity*") are most likely satisfied in the learning goal condition (because of its focus on learning). Thus, in the learning goal condition, higher utility values are likely obtained for higher effort levels (because of a higher goal outcome value, as determined by the afore-described drive strengths and drive satisfactions; section 2.3.3); in the outcome goal condition, higher utility values are likely obtained for lower effort levels (because of a lower goal outcome value, as determined by the drive strengths and drive satisfactions); the "do your best" goal condition lies somewhere in between. An effort level is chosen based on utility values in these conditions respectively—that is, a higher effort level in the learning goal condition, a lower effort level in the outcome goal condition, and so on, which in turn determines performance in these conditions.

Simulation Setup. 96 simulated participants (the same number as in the human experiment) were divided into 3 groups of equal size. The simulation setup was basically the same as the previous simulations.

For each drive involved, the *drive strength* was obtained from its *deficit* (Table 14) and *stimuluslevel* (Table 15), according to equation E1. The goal of performing the task was selected based on the *goal strength* (obtained from equation E2; *relevance* = 1). The value of accomplishing the task goal was assessed based on the *drive strength* and the anticipated *satisfaction* of each drive (Table 16); the *utility* of each rule (with an *effortlevel*) was obtained according to equation E3 ($a = 1$, $c = 0.185$). Then, an *effortlevel* was chosen based on the *utility* values (through a Boltzmann distribution, with temperature = 0.7); performance was determined by the *effortlevel* (equation E4; $p = 1.825$, $q = 0$).

In accordance with the hypothesis in "Conceptual Description", the *stimuluslevel* parameters (Table 15) were set so that the learning goal condition triggered more an approach-oriented drive ("*curiosity*"), while the outcome goal condition triggered more an avoidance-oriented drive ("*honor*"; Locke & Latham, 2002; section 2.4). The "do your best" goal condition led to the approach-oriented drive to lie somewhere in between, but did not trigger an avoidance-oriented drive (because there was no pressure to avoid failure). The anticipated *satisfactions* of these drives (Table 16) can also be justified in accordance with the hypothesis earlier. The total satisfaction of these drives (the value of the goal outcome) under the learning goal was thus higher than under the outcome goal, while the "do your best" goal lay in between. Note that technically one approach-oriented and one avoidance-oriented drive are minimally needed to account for the data, although the "*achievement*" drive was also included. No actual learning per se was simulated. No other parameters were relevant (based on definitions and assumptions in section 2).

Simulation Results. This simulation (Table 17) reproduced the important findings from the human experiment: Different goals significantly affected performance ($F(2,93) = 11.64$, $p < 0.00001$). The learning goal demonstrated the highest performance, as in the human data. The outcome goal resulted in lower performance than the learning goal, as in the human data ($t(62) = 4.86$, $p < 0.00001$). The performance of the "do your best" goal was in between, as in the human data: It was higher than that of the outcome goal ($t(62) = 2.68$, $p = 0.009$), but lower than that of the learning goal ($t(62) = 2.06$, $p = 0.04$).

3.3.3. Discussion

It was found by some researchers that people performed better when they were simply asked to do their best (e.g., Kanfer & Ackerman, 1989), while others argued against this point (e.g., Locke & Latham, 2002). Locke and Latham (2002) made the distinction between performance (outcome) goals/targets and learning goals/targets and argued that learning goals/targets often led to better performance (especially when a task was new and necessary skills were lacking). The experiment simulated here tested these ideas in one task setting. The simulation

Table 13
Mean of the performance measure (number of class schedules completed).

A. Mean performance in six conditions.

Goal Condition	Performance
do your best outcome goal	8.71
distal outcome goal	6.19
distal outcome goal with proximal goals	6.93
do your best learning goal	7.00
distal learning goal	8.17
distal learning goal with proximal goals	8.95

B. Mean performance in three categories.

Goal condition	Performance
Outcome goals	6.56
Learning goals	8.56
"Do your best" goals	7.85

Table 14
The mean and the SD of the *deficit* of each drive involved.

	Achievement	*Honor*	*Curiosity*
Mean	0.7	0.6	0.7
Standard Deviation		0.1	

Table 15
The *stimuluslevel* for each drive involved. The columns represent drives. The rows represent situational inputs.

	Achievement	*Honor*	*Curiosity*
Outcome goal	0.8	0.3	0.1
Learning goal	0.8	0.1	0.5
'Do your best' goal	0.7	0.0	0.3

Table 16
The anticipated *satisfaction* of each drive involved. The columns represent drives. The rows represent goals.

	Achievement	*Honor*	*Curiosity*
Outcome goal	0.7	−0.4	0.1
Learning goal	0.7	−0.1	0.6
'Do your best' goal	0.7	0.0	0.5

Table 17
Simulated mean performance.

Goal Condition	**Performance**
Outcome goal	6.59
Learning goal	8.60
'Do your best' goal	7.65

shows that such findings can possibly be accounted for by a unified model that also accounts for many other findings.

Furthermore, Elliot and Harackiewicz (1994) claimed that a performance (outcome) goal/target enhanced "intrinsic" motivation for achievement-oriented individuals, but not for others. Those low in achievement orientation showed the highest levels of "intrinsic" motivation when provided with learning goals/targets. See also Phillips and Gully (1997). There have also been explorations of the approach vs. the avoidance orientation of goals (e.g., Higgins, 1997). Our framework can be used to explore these issues; Clarion has addressed some personality differences and their consequences (e.g., Sun & Wilson, 2014).

There are also other patterns of goal effects, with various moderators or mediators (e.g., goal commitment; Locke & Latham, 2013). Given the broad scope and limited space of this paper, these effects cannot be detailed here.

3.4. *Effects of different types of priming*

3.4.1. *Human experiment*

In the experiment of Chen and Latham (2014), instead of external goal assignment, priming was used. Participants performed a class scheduling task. 88 participants were randomly distributed into four groups of approximately equal size: (1) learning priming, (2) performance priming, (3) both learning and performance priming, and (4) control. During the preparation stage, the participants were presented with a prime: (1) a picture of "The Thinker" (supposedly priming learning), (2) a picture of a racer (supposedly priming performance), (3) pictures of both "The Thinker" and the racer (priming both learning and performance), or (4) pictures of trees and rocks (the control condition).

The experiment (Table 18) showed that learning priming improved performance: *t*-test showed a significant difference of the learning priming group as compared to the control group ($t(41) = 2.22$, $p = 0.032$) and as compared to the performance priming group ($t(38) = 2.23$, $p = 0.032$). Performance priming alone did not significantly improve performance compared with the control group ($t(43) = 0.35$, *n.s.*).

Table 18
Mean number of correct schedules generated.

Group	**Number of Schedules**
learning priming	9.28
performance priming	7.09
learning and performance priming	7.71
control group	6.68

3.4.2. *Simulation*

Conceptual Description. Compared with performance priming, learning priming leads to less avoidance-oriented drive activation and more approach-oriented drive activation (e.g., "*curiosity*"), because, similar to the externally assigned learning goal condition in the previous simulation, learning priming entails more focus on exploration and less on performance (see section 2.4; Locke & Latham, 1990; 2002; 2013). Anticipated satisfactions of these drives under learning priming are also higher (for the same reason as explained for the previous simulation). Thus, learning priming leads to higher utilities for higher effort levels (based on these drive strengths and satisfactions) and results in higher levels of effort being put into the task.

Both learning and performance priming lead to drive activations at a level somewhere between performance priming and learning priming (for obvious reasons). The anticipated satisfactions of these drives also lie somewhere between performance priming and learning priming. Thus, it leads eventually to effort levels somewhere in between.

Simulation Setup. 88 simulated participants (the same number as in the human experiment) were divided into four groups of 22 each.

For each drive involved, the *drive strength* was obtained (according to equation E1) based on the *deficit* (Table 19) and the *stimuluslevel* (Table 20). The goal of performing the task was selected based on *goal strengths* (equation E2; *relevance* = 1). The value of accomplishing the task goal was determined based on the *drive strength* and the anticipated *satisfaction* (Table 21) of each drive involved; *utility* values of different rules (each specifying a different *effortlevel*) were obtained according to equation E3 ($a = 1$, $c = 0.185$). One rule was chosen (through a Boltzmann distribution; temperature = 0.7); performance was determined by the chosen *effortlevel* (equation E4; $p = 1.93$, $q = 0$).

The justifications of the parameter values were essentially the same as those for the previous simulation. Basically, as explained in "Conceptual Description" (see also section 2.4), the *stimuluslevel* parameters (Table 20) were such that learning priming triggered more the approach-oriented drive ("*curiosity*"); performance priming triggered more the avoidance-oriented drive ("*honor*"); performance and learning priming together resulted in drive activations somewhere in between. Also as explained before, the anticipated *satisfaction* of the "*honor*" drive was the highest under learning priming, the lowest under performance priming, and somewhere in between under both learning and performance priming (Table 21). (The "*achievement*" drive could also be included here as before, but would not make much difference.)

Simulation Results. This simulation (Table 22) captured the major effects in the human data: The difference across conditions was significant ($F(3, 84) = 7.92$, $p = 0.0001$). Performance was the highest under learning priming as in the human data (Table 22). Pairwise comparisons (Table 23) showed that learning priming led to significantly better performance than performance priming and the control condition, as in the human data. Performance priming alone did not significantly improve performance compared with the control condition, as in the human data.

Table 19
The mean and the SD of the *deficit* of each drive involved.

	Curiosity	*Honor*
Mean	0.7	0.7
Standard Deviation	0.1	

Table 20
The *stimuluslevel* for each drive involved. The columns represent drives. The rows represent situational inputs.

	Curiosity	*Honor*
Learning priming	1.	0.1
Performance priming	0.8	0.7
Learning and performance priming	0.9	0.4
Control	0.4	0.1

Table 21
The anticipated *satisfaction* for each drive involved. The columns represent drives. The rows represent goals.

	Curiosity	*Honor*
Task goal in learning priming	1.	0.0
Task goal in performance priming	1.	−0.55
Task goal in learning and performance priming	1.	−0.35
Task goal in control	1.	0.0

Table 22
Mean number of schedules in simulation.

Group	**Number of Schedules**
Learning priming	9.56
Performance priming	7.19
Learning and performance priming	8.59
Control	6.57

Table 23
Pairwise analysis of simulation results.

	Learning	Performance	Learning & Performance	Control
Learning	–	$t(42) = 3.9$, $p = 0.0003$	$t(42) = 2.4$, $p = 0.02$	$t(42) = 5.0$, $p < 0.0001$
Performance		–	$t(42) = 1.87$, $p = 0.06$	$t(42) = 0.73$, $p = 0.46$
Learning & Performance			–	$t(42) = 2.5$, $p = 0.02$
Control				–

3.4.3. Discussion

Significance of unconscious (implicit) influences on the human mind has been gaining attention in recent decades (e.g., Helie & Sun, 2010; Reber, 1989; Sun et al., 2005). Such influences have been taken into consideration in studying motivation and performance (e.g., Chen & Latham, 2014). This simulation shows that our model accounts for this aspect (along with other aspects). Furthermore, effects of explicit goals and implicit priming seem similar and are captured by the same mechanisms in the model.

Priming can also be linked to other processes in Clarion (e.g., in the ACS or the NACS) so as to better explore implications of unconscious/implicit processes for motivation and performance. So far, Clarion has accounted for a variety of implicit processes, ranging from those in skill learning (e.g., Sun et al., 2001; Sun et al., 2005) to those in problem solving and reasoning (e.g., Helie & Sun, 2010). Therefore it can be useful in further exploring implicit processes.

In the experimental literature, there are many other findings regarding priming (e.g., Papies, 2016, and others). Due to the broad nature and limited space of this paper, they cannot be examined here.

3.5. Effects of goal prioritization

3.5.1. Human experiment

The experiment of Schmidt and DeShon (2007) used a class scheduling task. 252 participants were involved. Each had to build 30 schedules: 15 for ABC College, and 15 for XYZ College. Each participant was given 30 min to complete the task. Although the original experiment was divided into different conditions, we consider only one condition here (for space considerations) whereby the two subtasks were the same and an incentive was given when a subtask was completed.

For each subtask, there was a line of students whose schedules were needed. Each subtask began with a line of five students. A student was removed when his/her schedule had been completed. New students were added periodically. Participants had difficulty meeting the goals for both subtasks, thus creating a trade-off.

The experimental results showed that prioritization, between the subtask with more schedules to be completed and the subtask with fewer schedules to be completed, changed over time: The participants started with spending more time on the subtask that they performed relatively worse (with more schedules to be completed), but towards the end, they actually spent more time on the subtask that they performed relatively better (with fewer schedules to be completed). The relative discrepancy was a statistically significant factor ($F(1, 116) = 85.64$, $p < 0.001$), and a significant interaction was found between time and discrepancy ($F(3, 247) = 84.05$, $p < 0.001$). See Fig. 3 for the X-shaped pattern.

3.5.2. Simulation

Conceptual Description. In this task, at each step, a goal specifying one of the two subtasks is selected based on goal strengths, which are determined by the drive strengths and the relevance of the drives. The goal that reduces the larger discrepancy (i.e., working on the college that has more schedules to be completed) can lead to better need (i.e., drive, such as "*achievement*") fulfilment and thus higher relevance (cf. Vancouver et al., 2010). However, relevance also depends on what one can realistically finish by the deadline (cf. Vancouver et al., 2010): When one perceives that it is less likely to have enough time to complete all the schedules for a college (which usually happens towards the end), the relevance of working on that college is reduced significantly.

Once a goal (of working on a subtask) is selected, an individual chooses how much effort to put into the selected subtask through rules. As before, utility of each rule can be calculated, which in turn leads to rule selection and effort allocation (i.e., time spent on the subtask).

Therefore, a pattern emerges that individuals start by spending more time on the subtask that they performed relatively worse (because a larger discrepancy leads to higher perceived relevance of the subtask, which becomes more likely to be selected), but towards the end, they spend more time on the subtask that they performed relatively better (because it has a better chance of being completed and is thus more relevant).

Simulation Setup. The number of simulated participants was 42 (the same as the average number of human participants in one condition in

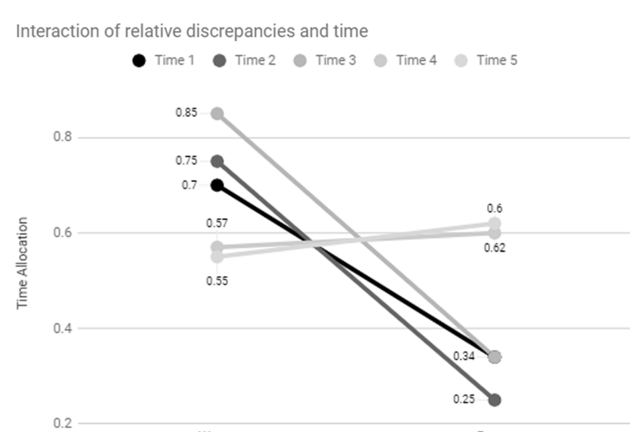

Fig. 3. Effort allocation between the two subtasks in the experiment of Schmidt and DeShon (2007).

Table 24
The mean and the SD of the *deficit* of the drive.

	Achievement
Mean	0.8
Standard Deviation	0.1

Table 25
The *stimuluslevel* for the drive. Each column represents a drive. Each row represents a situational input.

	Achievement
the task setting	0.6

Table 26
The *relevance* parameters. Each column represents a drive. Each row represents a goal. The relevance of working on the subtask with the largest discrepancy is 0.9; the relevance of working on a subtask with a smaller discrepancy is 0.1. When one perceives that a subtask is unlikely to be completed by the deadline, its relevance becomes 0.

	Achievement
Work on ABC	0.9/0.1/0.0
Work on XYZ	0.9/0.1/0.0

the original experiment).

At each step, the input was the information concerning the two subtasks (including the number of schedules to be done for each subtask and the time remaining). For each drive involved, the *deficit* (Table 24) and the *stimuluslevel* (Table 25) determined the *drive strength* (equation E1). Goal strengths were calculated (equation E2) based on the *strength* and *relevance* of each drive (Table 26). A goal was selected based on goal strengths. The value of accomplishing the selected goal (the selected subtask) was assessed based on the *drive strengths* and their *satisfactions* (Table 27), and the *utility* of each rule was thus obtained (equation E3; $a = 1$, $c = 0.185$). A rule (specifying an *effortlevel*) was selected based on these *utility* values. The time spent on the chosen subtask at the current step was determined by the chosen *effortlevel* (E4; $p = 0.4$, $q = 0$). The step described above was repeated until the 30 min given were used up.

The justifications of the parameter values were similar to those for the earlier simulations. The only important parameter value ordering was specified in Table 26: The *relevance* of working on the subtask with the larger discrepancy was greater than the *relevance* of working on the subtask with the smaller discrepancy, as explained in "Conceptual Description". During the later steps (the last few minutes), the fact that the deadline was approaching became apparent: If one perceived that a subtask was less likely to be completed by the deadline, its *relevance* was reduced (because *relevance* was in part a function of probability of accomplishing a goal, which was a function of time remaining and number of schedules to complete).

Simulation Results. The results of the simulation (Fig. 4) matched the X-shaped pattern found in the human data. During the early period of the simulation, more time was allocated to the subtask with the larger discrepancy. Towards the end, however, more time was spent on the subtask with the smaller discrepancy (Fig. 4). The relative discrepancy was a statistically significant factor for effort allocation (F(1, 125) = 87, $p < 0.0001$), as in the human data. A significant interaction was also found between time and discrepancy (F(1,125) = 27.9, $p < 0.0001$), also as in the human data.

Table 27
The anticipated *satisfaction* of the drive. Each column represents a drive. Each row represents a goal.

	Achievement
Work on ABC	0.7
Work on XYZ	0.7

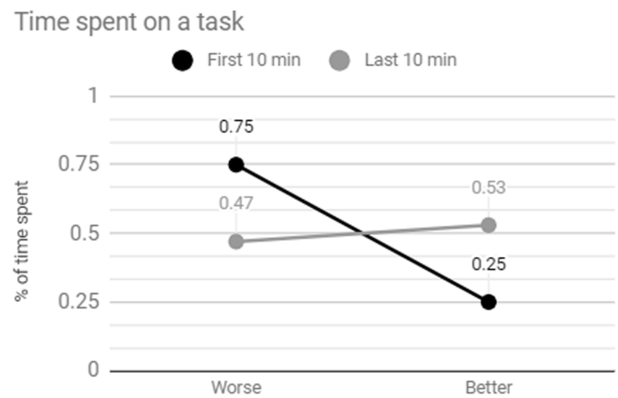

Fig. 4. Effort allocation in the simulation of Schmidt and DeShon (2007), in terms of percent of time working on a subtask when the performance on that subtask was worse or better at that moment.

3.5.3. Discussion

The results from this simulation matched the key finding from Schmidt and DeShon (2007) that relative discrepancy positively correlated with allocation of effort early on, but the opposite was true toward the end. This simulation showed that our model could account for such a phenomenon.

This phenomenon was also captured by Vancouver et al. (2010) using their control-theory-based model. A comparison with Vancouver et al.'s model can be found in section 4.2. There are also other related empirical findings and models (e.g., Ballard et al., 2018; Neal et al., 2017).

4. General discussion

4.1. Contributions and future work

The present work develops a framework for relatively coherent interpretations of a diverse range of experimental work on motivation and performance, in a mechanistic, process-based way (Craver & Bechtel, 2006), by utilizing an existing cognitive architecture. It describes, accounts for, and simulates these data. It thereby reconciles (i.e., accounts for, using the same underlying mechanisms) phenomena separately described by various existing theories (e.g., Ryan & Deci, 2000, Locke & Latham,1990; 2013, Vancouver et al., 2010, and so on; see also Braver et al. 2014). It shows that these (seemingly disparate) phenomena may have common underlying mechanisms (by virtue of our model).

The model accounts for empirical data based on factors at a more mechanistic level, thereby generating deeper and more unified descriptions and interpretations. To do so, it traces the notion of utility back to basic human motivation (the evolutionarily acquired set of essential human motives; Maslow, 1943; McClelland, 1951; Murray, 1938; Reiss, 2004; Sun, 2009), putting utility calculation on a more solid footing in our model. Basic human motivation provides the mechanistic basis for determining values of different outcomes (in relation to satisfying basic human needs) and thereby provides a mechanistic basis for utility calculation. This work demonstrated how a diverse range of data could be interpreted on that basis (see individual simulations earlier for refined, revised, or new interpretations in each case).

A question is: Can the hypotheses underlying the model be falsified? In this regard, note that the goal of the present work is a unifying framework and it is not about a narrowly focused hypothesis. It is more in line with the Kuhnian notion of a "paradigm", which may be falsified (in a sense) when it fails to meaningfully account for many important phenomena. In addition, for each individual simulation, it is conceivable that new human experiments can be developed that test motivational underpinnings of participants to confirm or falsify hypotheses.

Another question concerns "cherry picking": While many experiments were simulated, how many have failed? This is a legitimate concern. We have simulated other experiments and there has been (at least so far) no failure. These other simulations cannot be included in the present paper: The goal of this paper is to present a unifying framework that integrates disparate phenomena and it has to do so using a limited set of representative experiments. Simulation studies are time-consuming to conduct and lengthy to describe, so this paper can only address one example of each selected phenomenon. Nevertheless the present work can lead to more extensive and more specific explorations of these phenomena.

Note that a broadly scoped paper cannot possibly address in depth all details. Therefore simulations were highly simplified. For example, section 3.3 argued that outcome goals might trigger avoidance drives, but the fact that learning goals (especially higher learning goals) might also trigger avoidance drives was not discussed. Nevertheless, the point is that outcomes goals are more likely to trigger avoidance drives than learning goals (as suggested by the literature; Locke & Latham, 1990; 2013). Likewise, in section 3.4, the "*achievement*" drive was omitted in the simulation, because it was essentially the same across groups (as suggested by the literature) and did not contribute to the explanation of performance differences across groups.

This work is only a step towards a reconciliation of the motivation-performance literature, which would require much more theoretical and empirical work. The literature involves many seemingly inconsistent claims, and much more work is needed to address these. Since the cognitive architecture provides formal means to capture or unify different accounts, it may serve as a tool for resolving theoretical issues. The present work will lead to further explorations.

Future work needs to account for other motivation-performance phenomena that have not yet been accounted for, in addition to further exploring these discussed in the present paper. For instance, "self-efficacy" is important (Bandura, 1977). It has been argued that in case a difficult performance goal (target) is present and self-efficacy is high, performance improves (Locke & Latham, 1990). The question remains why and how such a phenomenon happens, in a mechanistic and process sense. More work is needed, although Clarion showed some promise in this regard (e.g., Brooks et al., 2012).

Effects of various types of feedback (e.g., Cianci et al., 2010; Shih & Alexander, 2000) also need to be understood. Our model can capture a number of such effects (which are not detailed here due to space), but more explorations are needed.

Personality also has significant effects on motivation and performance. For instance, Judge and Ilies (2002) showed that *neuroticism* and *conscientiousness* (two of the Big Five personality dimensions) were strong correlates of performance motivation (see also Elliot & Harackiewicz, 1994; Phillips & Gully, 1997). Although Clarion includes a model of personality (Sun & Wilson, 2014), more modeling of personality in relation to motivation and performance is needed.

Exceptions and boundary conditions may also need to be accounted for. For instance, one may take into consideration possible roles of impulses and habits (e.g., Hofmann et al., 2009), which may override cost-benefit calculation but may also be counteracted by rational considerations (cf. Griffiths et al., 2015; Shenhav et al., 2017).

Finally, learning, including learning of parameters of drives, goals, and action rules (as touched upon earlier), should be developed further (Bugrov et al., forthcoming). This is a challenge for computational psychology and can benefit from recent advances in machine learning.

By tackling these issues above, this model can be extended in the future towards a general theory of the motivation-performance relationship. As is, this work is only an attempt at a unified account of a set of phenomena.

4.2. Comparisons

It should be noted that the present work is not aimed at replacing empirical theories. It operates at a different level of abstraction; computational and experimental work are different in nature. This work aims to show that often the same mechanisms are underlying a diverse range of different phenomena and thereby attempts to unify them.

In Carver and Scheier's (1998) control theory, based on ideas from cybernetics, the source of motivation is believed to be a negative feedback loop that reduces goal–performance discrepancies. The theory is broad and can account for many phenomena. However, some researchers questioned some basic assumptions of the theory. Bandura and Cervone (1986) found that both goals and self-efficacy mediated feedback effects. Psychological plausibility of explicit, intentional discrepancy reduction is also an issue; deeper explanations may be needed. Also, according to the theory, in the absence of negative feedback, the natural state of an organism is rest, which may not be realistic. Some of these criticisms have been countered by control theory advocates.

The model of Vancouver et al. (2010) follows that idea, but it provides detailed computational simulation. The central concept in their model is the discrepancy-reducing negative feedback loop whereby discrepancies between desired and perceived states determine outcomes such as effort exerted. Their model often describes system-level behavior in terms of subsystems each of which detects and reduces its own discrepancies. They later incorporated learning to some extent and extended the model to deal with more scenarios (e.g., Ballard et al., 2018). Our model, like their model, tends to reduce discrepancies, but it does so in relation to satisfying basic needs/motives, without an explicit, intentional feedback loop. In our model, discrepancy has an effect through drive activation, goal setting, action selection, and utility calculation. Our model thus avoids those criticisms leveled against the control theory, while it also accounts for a wider range of other motivation-related phenomena (e.g., Bretz & Sun, 2018; Sun, 2020; Sun & Wilson, 2014; Sun et al., 2016; Wilson & Sun, 2021). It is possible that their model, with a different focus, may be incorporated into Clarion, leading to a more complete model.

We can also relate our model to Locke and Latham's (1990; 2013) goal setting theory, a broad summary of many empirical data patterns. For one thing, it generally states that the more difficult the goal (target) is, the better the performance (while it takes into account moderators and mediators, such as goal commitment, goal type, self-efficacy, and so on). Our model accounts for their key insight that more difficult goal setting often leads to better performance, through hypothesizing that higher drive activations (especially approach-oriented ones), as a result of more difficult goals, often lead to higher values for the corresponding outcomes, thus leading to better performance. Our model also accounts for other possible consequences in other circumstances (e.g., triggering higher activations of avoidance-oriented drives and thus leading to worse performance). Kanfer and Ackerman (1989), for example, found that people might perform better when they were simply asked to do their best (which some argued against). Locke and Latham (2002) found that learning goals often led to better performance, while different learning goals might have different effects (Locke & Latham, 2013). Our model incorporated some such subtleties (see section 3.3; see also Brooks et al., 2012), although not yet the full theory .

We can also compare our model with Vroom's (1964) valence–instrumentality–expectancy theory, which argues that the force to act is a multiplicative combination of valence (anticipated satisfaction), instrumentality (the belief that performance will lead to rewards), and expectancy (the belief that effort will lead to the performance needed to attain the rewards). Our model is consistent with the general idea of Vroom's, but it delved deeper, among other things, by linking anticipated satisfaction to basic human motives/needs. It also took cost into consideration.

Comparisons with Ryan and Deci's (2000) theory have been discussed earlier (see also Deci, Koestner, & Ryan, 1999). Ryan and Deci emphasized the role of "intrinsic" motivation as opposed to extrinsic reward or external pressure. Our model accounts for it mechanistically, along with other phenomena, with a different view on intrinsic motivation (as discussed earlier). Furthermore, our model is capable of accounting for more complex relationships among intrinsic motivation, extrinsic motivation, situation, individual difference, and performance (such as Amabile & Pratt, 2016; Iyengar & Lepper, 1999; Thompson, et al., 1993), which was touched upon in section 3.2 (see also Sun, 2020; Sun & Wilson, 2014).

Relatedly, educational psychology deals with interest-based learning. It was shown empirically that interest was highly correlated with basic motives, such as *achievement* or *autonomy* (e.g., Krapp, 2005), consistent with our model. According to Krapp (2005), interest can be explained by a dual system in which rational and emotional control coexist. Thus, such interest can be explained in-depth, mechanistically, based on cognitive architectures, especially those involving dual processes (Sun, 2016).

Braver et al. (2014) discussed the integration of studies of motivation-cognition interaction from different disciplines, which is also our ultimate goal. Yee and Braver (2018) explored effects of extrinsic motivation, while the present work addresses *internal* needs/motives and interprets behavior on that basis.

Cost-benefit analysis and utility maximization have had a long history (Griffiths et al., 2015). While a comprehensive survey is impractical here, it is worth noting that recent work explores, through psychological, neuroscientific, and computational means, issues such as how people learn when to exert cognitive control and how much effort to spend (based on external rewards, but not intrinsic needs as in the present work; e.g., Lieder et al., 2018; Verguts et al., 2015), the role of reinforcement learning in cognitive control (e.g., Holroyd & Yeung, 2012), the notions of opportunity cost and mental effort (Kurzban et al., 2013), neural computational implementation (Verguts et al., 2015), and so on. Such work is complementary to, and provides general support for, our approach. In particular, reinforcement learning of cost and benefit can be incorporated into our model (the mechanism of which is already present in Clarion; Sun, 2016). Our framework adds to the existing (mostly neuroscientific) work linking motivation and effort (Braver et al., 2014; Meyniel et al., 2013).

Finally, another line of comparison is with other existing cognitive architectures (see Kotseruba & Tsotsos, 2020 for a comprehensive review). For instance, ACT-R and Soar are such cognitive architectures, but they do not have sufficient built-in motivational mechanisms compared with Clarion. Even though the notion of goal exists in ACT-R and Soar, their setting of goals has not been based on basic human motives (which are absent in them). Consequently, although ACT-R has utility calculation built-in, it is not linked to basic internal motives (Anderson & Lebiere, 1998). Besides a better developed motivational subsystem, Clarion also has a better developed metacognitive subsystem; both are crucial to accounting for the motivation-performance relationship (Sun, 2016). Although, as a general programming environment, anything can be programmed into ACT-R or Soar (e.g., Belavkin, 2001; Nagashima et al., 2020), they do not provide explanations of the motivation-performance relationship based on the cognitive architectures per se. In contrast, Clarion links purely cognitive aspects with motivation, metacognition, emotion, personality, sociality, and culture (see Sun, 2016). (In that regard, it is closer to MicroPsi.)

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by ARI grant W911NF-17-1-0236.

Appendix A. Derivation of utility

In Clarion, utility is calculated based on *cost* and *benefit,* as follows. The utility of rule *j* is determined by:

$$U_j = benefit_j - \upsilon \times cost_j$$

where υ is a scaling factor balancing measurements of *benefit* and *cost* (Sun, 2016).

For assessing *cost*, for the sake of simplicity, assuming a linear cost function, we have:

$$cost_j = f\left(effortlevel_j\right)$$

$$= c \times effortlevel_j$$

where *c* is a coefficient that may vary with individuals as well as task settings.[8]

On the other hand, for assessing *benefit*, we have (cf. Kahneman & Tversky, 1992; Vroom, 1964),

$$Benefit_j = \sum_y prob(y) \times value(y)$$

$$= prob(g) \times value(g) + (1 - prob(g)) \times value(\text{not } g)$$

where *y* represents all possible outcomes, *prob(y)* denotes the probability of *y*, *value(y)* denotes the subjective value of *y*, and *g* represents achieving the goal that dictates accomplishing the task objectives given. Note that *value(y)* can be positive or negative, so can $benefit_j$.

In particular, we can have:

$$Benefit_j = prob(g) \times value(g)$$

[8] Cost is a function of effort. In the absence of evidence to the contrary, the simplest functional form is used here (as in, e.g., Verguts et al., 2015). Cost or utility are difficult to determine empirically, often confounded with other constructs (e.g., Hofmann et al., 2009; Kurzban et al., 2013; Shenhav et al., 2017).

where g represents achieving the goal that comprises the task objectives given, when achieving such a goal is assumed to be the only outcome that is valuable.

One may subjectively estimate the probability of achieving a goal as follows:

$$prob(g) = f(CL,\ effortlevel_j, \ldots.)$$

$$= a \times CL \times effortlevel_j$$

where CL is a measure of confidence level (e.g., *self-efficacy* or a function of it; Bandura, 1977) and a is a scaling parameter (both may vary with regard to individuals and task settings). For the sake of simplicity, a linear function of the product of CL and $effortlevel_j$ is assumed here (which might not be the case beyond a certain point).

On the other hand, *value(g)* can be internally assessed on the basis of satisfying activated drives (i.e., internal needs) when goal g is achieved (Sun, 2016):

$$value(g) = B(g) + \sum ds_d \times satisfaction_d(g)$$

$$= \sum ds_d \times satisfaction_d(g)$$

where $satisfaction_d(g)$ (which can be simply −1, 0 or 1, or continuous within the range) measures how well the achievement of goal g satisfies the need represented by drive d, and $B(g)$ (which is assumed to be 0 here) is the "bonus" for achieving the chosen goal g (Sun, 2016).

Thus, utilities for different rules that specify different effort levels (when the goal is to accomplish the task objectives) are determined as follows:

$$U_j = benefit_j - \upsilon \times cost_j$$
$$= a \times CL \times effortlevel_j \times value(g) - \upsilon \times c \times effortlevel_j$$
$$= effortlevel_j \times (a \times CL \times value(g) - \upsilon \times c)$$
$$= effortlevel_j \times (a \times value(g) - c)$$

where υ is assumed to be 1 (or absorbed by c) and CL is assumed to be 1 (or absorbed by a; note that CL is not addressed in this work).

References

Amabile, T. M., & Pratt, M. G. (2016). The dynamic componential model of creativity and innovation in organizations: Making progress, making meaning. *Research in Organizational Behavior, 36*, 157–183.

Anderson, J. R., & Lebiere, C. (1998). *The Atomic Components of Thought*. Mahwah, NJ: Lawrence Erlbaum Associates.

Ballard, T., Vancouver, J. B., & Neal, A. (2018). On the pursuit of multiple goals with different deadlines. *Journal of Applied Psychology, 103*(11), 1242–1264.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215.

Bandura, A., & Cervone, D. (1986). Differential engagement of self-reactive influences in cognitive motivation. *Organizational Behavior and Human Decision Processes, 38*(1), 92–113.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*(4), 323–370.

Belavkin, R. V. (2001). The role of emotion in problem solving. In *In: Proceedings of the AISB'01 Symposium on emotion, cognition and affective computing, York, England* (pp. 49–57).

Braver, T. S., Krug, M. K., Chiew, K. S., Kool, W., Westbrook, J. A., Clement, N. J., et al. (2014). Mechanisms of motivation–cognition interaction: Challenges and opportunities. *Cognitive, Affective, and Behavioral Neuroscience, 14*(2), 443–472.

Bretz, S., & Sun, R. (2018). Two models of moral judgment. *Cognitive Science, 42*(S1), 4–37.

Brooks, J. D., Wilson, N., & Sun, R. (2012). The effects of performance motivation: A computational exploration of a dynamic decision making task. Proceedings of the First International Conference on Brain-Mind. pp.7-14. East Lansing, MI: BMI Press.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology, 76*(5), 839–855.

Carver, C., & Scheier, M. (1998). On the Self-Regulation of Behavior. Cambridge, UK: Cambridge University Press.

Chen, X., & Latham, G. P. (2014). The effect of priming learning vs. performance goals on a complex task. *Organizational Behavior and Human Decision Processes, 125*(2), 88–97.

Cianci, A. M., Klein, H. J., & Seijts, G. H. (2010). The Effect of Negative Feedback on Tension and Subsequent Performance: The Main and Interactive Effects of Goal Content and Conscientiousness. *Journal of Applied Psychology, 95*(4), 618–630.

Craver, C. F., & Bechtel, W. (2006). Mechanism. In S. Sarkar, & J. Pfeifer (Eds.), *Philosophy of science: an encyclopedia* (pp. 469–478). New York: Routledge.

Dai, D. Y., & Sternberg, R. J. (Eds.). (2004). *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development*. Mahwah, NJ, US: Lawrence Erlbaum Associates.

Deci, E. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology., 18*(1), 105–115.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*(6), 627–668.

Dignum, F., Kinny, D., & Sonenberg, L. (2002). Motivational attitudes of agents: On desires, obligations, and norms. In B. Dunin-Keplicz, & E. Nawarecki (Eds.), *From Theory to Practice in Multi-Agent Systems* (pp. 83–92). Berlin: Springer.

Elliot, A. J., & Harackiewicz, J. M. (1994). Goal Setting, Achievement Orientation, and Intrinsic Motivation: A Mediational Analysis. *Journal of Personality and Social Psychology, 66*(5), 968–980.

Evans, J., & Frankish, K. (Eds.). (2009). *In Two Minds: Dual Processes and Beyond*. Oxford, UK: Oxford University Press.

Gollwitzer, P. M. (2012). Mindset theory of action phases. In P. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 526–545). London, UK: Sage.

Gray, J. A. (1987). *The Psychology of Fear and Stress* (2nd Ed). New York: Cambridge.

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science, 7*(2), 217–229.

Helie, S., & Sun, R. (2010). Incubation, insight, and creative problem solving: A unified theory and a connectionist model. *Psychological Review, 117*(3), 994–1024.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Beyond WEIRD: Towards a broad-based behavioral science. *Behavioral and Brain Sciences., 33*(2-3), 111–135.

Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist, 52*(12), 1280–1300.

Hintzman, D. (1990). Human learning and memory: Connections and dissociations. In: Annual Review of Psychology, pp.109-139. Palo Alto, CA: Annual Reviews Inc.

Hofmann, W., Friese, M., & Strack, F. (2009). Impulse and self-control from a dual-systems perspective. *Perspect. Psychol. Sci., 4*(2), 162–176.

Hofstede, G. (2001). *Culture's Consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

Holroyd, C. B., & Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends in Cognitive Sciences, 16*(2), 122–128.

Hull, C. (1951). *Essentials of Behavior*. New Haven, CT: Yale University Press.

Inzlicht, M., Schmeichel, B. J., & Macrae, C. N. (2014). Why self-control seems (but may not be) limited. *Trends in cognitive sciences, 18*(3), 127–133.

Iyengar, S. S., & Lepper, M. R. (1999). Rethinking the Value of Choice: A Cultural Perspective on Intrinsic Motivation. *Journal of Personality and Social Psychology, 76*(3), 349–366.

Judge, T. A., & Ilies, R. (2002). Relationship of Personality to Performance Motivation: A Meta-Analytic Review. *Journal of Applied Psychology, 87*(4), 797–807.

Kahneman, D., & Tversky, A. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty., 5*(4), 297–323.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology, 74*(4), 657–690.

Kool, W., & Botvinick, M. (2014). A labor/leisure tradeoff in cognitive control. *J Exp Psychol Gen., 143*(1), 131–141.
Kotseruba, I., & Tsotsos, J. K. (2020). 40 years of cognitive architectures: Core cognitive abilities and practical applications. *Artificial Intelligence Review, 53*(1), 17–94. https://doi.org/10.1007/s10462-018-9646-y
Krapp, A. (2005). Basic needs and the development of interest and intrinsic motivational orientation. *Learning and Instruction, 15*, 381–395.
Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences, 36*(6), 661–679.
Lieder, F., Shenhav, A., Musslick, S., & Griffiths, T. L. (2018). Rational metareasoning and the plasticity of cognitive control. PLoS Computational Biology, 14(4), e1006043.
Locke, E. A., & Latham, G. P. (1990). *A Theory of Goal Setting and Task Performance.* Englewood Cliffs, NJ: Prentice Hall.
Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*(9), 705–717.
Locke, E. A., & Latham, G. P. (2013). *New Developments in Goal Setting and Task Performance*. New York: Routledge.
Locke, E. A., & Schattke, K. (2019). Intrinsic and Extrinsic Motivation: Time for Expansion and Clarification. *Motivation Science, 5*(4), 277–290. https://doi.org/10.1037/mot0000116
Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.
Maslow, A. H. (1943). A theory of human motivation. *Psychological Review, 50*(4), 370–396.
McClelland, D. (1951). *Personality*. New York: Dryden Press.
Meyniel, F., Sergent, C., Rigoux, L., Daunizeau, J., & Pessiglione, M. (2013). Neurocomputational account of how the human brain decides when to have a break. *Proceedings of the National Academy of Sciences, 110*(7), 2641–2646.
Murayama, K., Matsumoto, M., Izuma, K., & Matsumoto, K. (2010). Neural basis of the undermining effect of monetary reward on intrinsic motivation. *Proceedings of the National Academy of Sciences, 107*(49), 20911–20916.
Murray, H. (1938). *Explorations in Personality*. New York: Oxford University Press.
Nagashima, K., Morita, J., & Takeuchi, Y. (2020). Modeling Intrinsic Motivation in ACT-R: Focusing on the Relation Between Pattern Matching and Intellectual Curiosity. *Proceedings of the 18th Annual Meeting of the International Conference on Cognitive Modelling*.
Neal, A., Ballard, T., & Vancouver, J. B. (2017). Dynamic self-regulation and multiple-goal pursuit. *Annual Review of Organizational Psychology and Organizational Behavior, 4*(1), 401–423.
Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic vs. analytic cognition. *Psychological Review, 108*, 291–310.
Papies, E. K. (2016). Goal priming as a situated intervention tool. *Current Opinion in Psychology, 12*, 12–16. https://doi.org/10.1016/j.copsyc.2016.04.008
Phillips, J. M., & Gully, S. M. (1997). Role of goal orientation, ability, need for achievement, and locus of control in the self-efficacy and goal-setting process. *Journal of Applied Psychology, 82*(5), 792–802.
Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*(3), 219–235.
Reiss, S. (2004). Multifaceted nature of intrinsic motivation: The theory of 16 basic desires. *Review of General Psychology, 8*(3), 179–193.
Rumelhart, D., McClelland, J., & the PDP Research Group, (1986). Parallel Distributed Processing: Explorations in the Microstructures of Cognition, Cambridge, MA: MIT Press.
Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78.
Schmidt, A. M., & DeShon, R. P. (2007). What to Do? The Effects of Discrepancies, Incentives, and Time on Dynamic Goal Prioritization. *Journal of Applied Psychology, 92*(4), 928–941.
Seijts, G. H., & Latham, G. P. (2001). The effect of distal learning, outcome, and proximal goals on a moderately complex task. *Journal of Organizational Behavior, 22*(3), 291–307.
Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., et al. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience, 40*(1), 99–124.
Shih, S., & Alexander, J. M. (2000). Interacting effects of goal setting and self- or other-referenced feedback on children's development of self-efficacy and cognitive skill within the Taiwanese classroom. *Journal of Educational Psychology, 92*(3), 536–543.
Simon, H. A. (1967). Motivational and emotional controls of cognition. *Psychological Review, 74*(1), 29–39.
Steel, P., & König, C. J. (2006). Integrating theories of motivation. *Academy of Management Review, 31*(4), 889–913.
Sun, R. (Ed.). (2008). *The Cambridge Handbook of Computational Psychology*. New York: Cambridge University Press.
Sun, R. (2009). Motivational representations within a computational cognitive architecture. *Cognitive Computation, 1*(1), 91–103.
Sun, R. (2016). *Anatomy of the Mind: Exploring Psychological Mechanisms and Processes with the Clarion Cognitive Architecture*. New York: Oxford University Press.
Sun, R. (2020). Exploring culture from the standpoint of a cognitive architecture. *Philosophical Psychology, 33*(2), 155–180.
Sun, R., Merrill, E., & Peterson, T. (2001). From implicit skills to explicit knowledge: A bottom- up model of skill learning. *Cognitive Science, 25*(2), 203–244.
Sun, R., Slusarz, P., & Terry, C. (2005). The interaction of the explicit and the implicit in skill learning: A dual-process approach. *Psychological Review, 112*(1), 159–192.
Sun, R., & Wilson, N. (2014). A model of personality should be a cognitive architecture itself. *Cognitive Systems Research, 29-30*, 1–30.
Sun, R., Wilson, N., & Lynch, M. (2016). Emotion: A unified mechanistic interpretation from a cognitive architecture. *Cognitive Computation, 8*(1), 1–14.
Thompson, E. P., Chaiken, S., & Hazlewood, J. D. (1993). Need for cognition and desire for control as moderators of extrinsic reward effects: A person × situation approach to the study of intrinsic motivation. *Journal of Personality and Social Psychology, 64*, 987–999.
Tolman, E. C. (1932). *Purposive Behavior in Animals and Men*. New York: Century.
Vancouver, J. B., & Purl, J. (2017). A computational model of self efficacy's various effects on performance. *Journal of Applied Psychology, 102*(4), 599–616.
Vancouver, J. B., Weinhardt, J. M., & Schmidt, A. M. (2010). A formal, computational theory of multiple-goal pursuit: Integrating goal-choice and goal-striving processes. *Journal of Applied Psychology, 95*(6), 985–1008.
Vancouver, J.B., Weinhardt, J.M., & Vigo, R. (2014). Change one can believe in: Adding learning to computational models of self-regulation. Organizational Behavior and Human Decision Processes, 124 (1), 56-74.
Veksler, V. D., Myers, C. W., & Gluck, K. A. (2015). Model flexibility analysis. *Psychological Review, 122*(4), 755–769.
Verguts, T., Vassena, E., & Silvetti, M. (2015). Adaptive effort investment in cognitive and physical tasks: A neurocomputational model. *Frontiers in Behavioral Neuroscience, 9*, 57.
Vroom, V. (1964). *Work and Motivation*. New York: Wiley.
Wilson, N. R., & Sun, R. (2021). A mechanistic account of stress-induced performance degradation. *Cognitive Computation, 13*(1), 207–227.
Yee, D. M., & Braver, T. S. (2018). Interactions of motivation and cognitive control. *Current Opinion in Behavioral Sciences, 19*, 83–90.