

# Towards a Formal Theory of the Need for Competence via Computational Intrinsic Motivation

Erik M. Lintunen<sup>1</sup> and Nadia M. Ady<sup>1</sup> and Sebastian Deterding<sup>2</sup> and Christian Guckelsberger<sup>1</sup>

<sup>1</sup>Department of Computer Science, Aalto University, Finland, {firstname.lastname}@aalto.fi

<sup>2</sup>Dyson School of Design Engineering, Imperial College London, United Kingdom, s.deterding@imperial.ac.uk

## Abstract

Computational modelling offers a powerful tool for formalising psychological theories, making them more transparent, testable, and applicable in digital contexts. Yet, the question often remains: how should one computationally model a theory? We provide a demonstration of how formalisms taken from artificial intelligence can offer a fertile starting point. Specifically, we focus on the “need for competence”, postulated as a key basic psychological need within Self-Determination Theory (SDT)—arguably the most influential framework for intrinsic motivation (IM) in psychology. Recent research has identified multiple distinct facets of competence in key SDT texts: effectance, skill use, task performance, and capacity growth. We draw on the computational IM literature in reinforcement learning to suggest that different existing formalisms may be appropriate for modelling these different facets. Using these formalisms, we reveal underlying preconditions that SDT fails to make explicit, demonstrating how computational models can improve our understanding of IM. More generally, our work can support a cycle of theory development by inspiring new computational models, which can then be tested empirically to refine the theory. Thus, we provide a foundation for advancing competence-related theory in SDT and motivational psychology more broadly.

**Keywords:** need for competence; self-determination theory; intrinsic motivation; theory development; intrinsically motivated reinforcement learning; computational modelling

## Introduction

Self-Determination Theory (SDT) (Deci and Ryan, 1985) is one of the most influential psychological theories of motivation, specifically intrinsic motivation (IM). Notably, SDT, like many psychological theories, is expressed in “soft” verbal propositions. In light of the current theory crisis (Oberauer and Lewandowsky, 2019), only-verbal theory has drawn increasing critique: it can leave many auxiliary assumptions and boundary conditions unspecified, making it hard to derive hypotheses that can be rigorously tested (Scheel et al., 2021). Many psychology researchers have advocated for computational modelling as a paradigm (e.g., Marsella et al., 2010; Oberauer and Lewandowsky, 2019; Robinaugh et al., 2021) or even as a necessity (Van Dongen et al., 2024) for theory development. Yet formal specification as computational modelling remains largely absent in SDT, judging by, e.g., the most recent integrative volume of the theory (Ryan, 2023).

A major challenge of computational modelling is inferring plausible formalisms for a given theory. We make the case that computational IM, i.e., formalisms developed to drive the behaviour of artificial agents, can inform the development of SDT theory in human motivation research. Computational

IM has been informed by SDT (e.g., Oudeyer and Kaplan, 2007), but has undergone major advancements independent from human psychology, addressing problems such as open-ended development, task-agnostic learning, and dealing with the potential sparsity of rewards (Colas et al., 2022, p. 1161).

Our case study focuses on the *need for competence*, one of the basic psychological needs underpinning IM in SDT. Through conceptual analysis (Machado and Silva, 2007), Deterding et al. (2024) identify four distinct facets of competence in canonic SDT texts, reproduced below in Table 1.

Table 1: **The four distinct facets of competence** identified by Deterding et al. (2024, pp. 7–10).

<b>Effectance (C1)</b>	Observing that one’s action causes consequences, which can be unintended.
<b>Skill use (C2)</b>	Observing <b>(a)</b> an opportunity (an appropriate situation) to use a skill and/or <b>(b)</b> that one uses a particular capacity in the course of one’s action.
<b>Task performance (C3)</b>	Observing that one performs <b>(a)</b> well at an intended task and/or <b>(b)</b> to an extent that requires a certain skill or skill level.
<b>Capacity growth (C4)</b>	Observing gain in the <b>(a)</b> strength and/or <b>(b)</b> range of one’s skills.

In the section *Matching Computational Models With SDT*, we match each distinct facet with examples of formalisms from the computational IM literature and discuss how each example formalises aspects of the theory. These formalisms not only lay a foundation for computationally specifying the need for competence, but, by linking to the corresponding papers, we also unveil a wealth of existing implementations, simulators, and evaluation metrics that could be used in psychological motivation research. The value of such resources is detailed in the section *Future Work and Conclusion*, but, for one, they empower future testing of whether formalisms fitting the facets produce empirical data aligned with propositions from the theory.

## Background: Computational IM & RL

We begin by justifying why reinforcement learning (RL) is a suitable domain from which to draw computational models of competence. In addition, to support understanding of the models we discuss, we introduce key concepts from intrinsically motivated RL—notably, *goals* and *skills* and their relationship to the reward function—and point to further reading.

The idea that IM might be driven by biological reward systems (Barto et al., 2004, p. 113, in reference to Kakade and Dayan, 2002; Dayan and Balleine, 2002) has driven the choice of methods centrally used in computational IM: RL (Baldassarre, 2022, p. 252). Due to their usefulness, intrinsically motivated RL methods have been studied extensively in the computational RL literature at large (see reviews by Colas et al., 2022; Aubret et al., 2023; Lidayan et al., 2024). Central to RL is the construct of *reward function*, and an assumption underlying much work in RL is the *reward hypothesis*: “all of what we mean by goals and purposes can well be thought of as the maximization of the expected value of the cumulative sum of a received scalar signal (called reward)” (Sutton and Barto, 2018, p. 53; cf., Silver et al., 2021, p. 4). This hypothesis informs the formalisms used for competence-based intrinsic motivation (CB-IM), which tend to rely on the assumption that every goal can be sufficiently defined by a reward function. *Intrinsically motivated* RL is characterised by the use of task-agnostic intrinsic reward functions that collate or compare agent-internal variables independently of their semantics (Oudeyer and Kaplan, 2008, p. 3; cf., Berlyne, 1965, p. 246 and Oudeyer and Kaplan, 2007).

Many of our examples operate within the framework of goal-conditioned reinforcement learning (GCRL) (see review by Liu et al., 2022), which extends the standard definition of a reward function to be conditioned on goals. A *goal*, or a *goal-defining variable*, is a parameter to the reward function (cf., Colas et al., 2022, p. 1165; Aubret et al., 2023, p. 6). Then, a *skill* is a policy given a goal, optimising for the return according to the reward conditioned on that goal. Examples of goal-defining variables include indices (e.g., Eysenbach et al., 2019) or elements drawn from a learned distribution (e.g., Nair et al., 2018). Their role is simply to indicate which reward function the agent is aiming to maximise.

Goals can be viewed as “a set of *constraints* ... that the agent seeks to respect” (Colas et al., 2022, p. 1165, emphasis in original). While the most immediate intuition of a goal is often as a desired state for the agent to reach (e.g., Kaelbling, 1993, p. 1; Schaul et al., 2015, p. 2), the formalism allows for a more general set of constraints on behaviour. In effect, any behaviour that can be defined by attempting to maximise some reward function on the environment can be formulated as a goal–skill pairing.

## Matching Computational Models With SDT

We next turn our focus to matching computational formalisms to each of the competence facets in Table 1 by providing examples from RL. The formalisms have been identified

through review of both classic (Oudeyer and Kaplan, 2007) and state-of-the-art (Colas et al., 2022; Aubret et al., 2023) computational IM. An exhaustive matching is beyond the scope of this paper. Rather, we highlight formalisms that we believe to fit the verbal definitions of each facet and reveal preconditions that SDT implicitly assumes. We focus on intuition about how each model fits the theory. For more intuition and detail on how each model is designed, implemented, and evaluated, we direct the reader to the corresponding papers.

### Effectance (C1)

Let us consider the case where competence in SDT refers to *effectance*, i.e., motivation by observing that one’s own actions cause change in the environment. To recognise effectance, an agent needs a mechanism to distinguish whether a change in the environment is caused by the agent itself or some external factor. An example of a computational model fitting effectance is Rewarding Impact-Driven Exploration (RIDE) (Raileanu and Rocktäschel, 2020), as part of which the agent is rewarded for the amount of change it causes via a single action: the “impact” of the agent’s action, measured between two consecutive observations. Specifically, the *impact-driven reward* is defined as:

$$R_t(\mathbf{s}_t, \mathbf{s}_{t+1}) := \frac{\|\phi(\mathbf{s}_{t+1}) - \phi(\mathbf{s}_t)\|_2}{\sqrt{N(\mathbf{s}_{t+1})}}, \quad (1)$$

where  $\phi(\mathbf{s}_t)$  and  $\phi(\mathbf{s}_{t+1})$  are the learned representations of two consecutive states and  $N(\mathbf{s}_{t+1})$  represents or approximates the number (count) of times the latter state has been visited (Raileanu and Rocktäschel, 2020, pp. 4–5).

In RIDE, the agent computes its reward using the learned representation of the observation ( $\phi(\mathbf{s}_t)$ ) as opposed to the raw sensory space (e.g., pixels). The parameters of the representation model  $\phi$  are optimised so that an agent only learns the features it can control—that is, it learns to ignore aspects of the environment it cannot affect (inspired by Pathak et al., 2017). The agent is rewarded for a combination of changes in only those state features that its own actions are able to consistently have some effect on (Eq. 1, numerator). This reward is scaled according to the novelty of the state into which its action leads (Eq. 1, denominator). The effect of the denominator is that the reward obtainable from particular states diminishes the more the agent observes them, even if the agent’s action has the same effective outcome as previously observed. In recent SDT texts, effectance motivation does not require novelty, but the original usage of the term required observations to be novel (White, 1959, p. 322). By having the reward wane the more a particular situation is perceived, RIDE offers a simple formalisation of this original novelty aspect of effectance motivation, while ensuring the agent is only rewarded for changes its own actions have caused.

### Skill use (C2)

We next consider the case in which competence refers to *skill use* or the “exercise ... of one’s capacities” (Ryan and Deci,

2017, p. 86). Following Deterding et al. (2024, p. 6), we cover the specific understanding of “exercise” as implementing or bringing something to bear (Merriam-Webster.com, 2025).<sup>1</sup> As per Table 1, competence as skill-use motivation can be understood in two different ways: observing an opportunity (an appropriate situation) to use a skill (*C2.a*) or that one uses a particular capacity in the course of one’s action (*C2.b*).

**An opportunity (an appropriate situation) to use a skill (C2.a)** Broadly, there are two kinds of computational approaches used to model an agent learning about opportunities to use particular skills: those in which the agent ignores the current state of the world and those in which it does not. In both cases the agent learns a policy over its set of skills to determine which skill to use. In the former case, an agent follows the same policy regardless of its state observation—in this case, an “appropriate” opportunity can be thought of as a learning opportunity, determined not based on the observation of the environment, but on how the agent’s learning is proceeding. In the latter case, identifying an opportunity to use a particular skill can also depend on the observation.

Specifically, the first case of *C2.a*, in which an agent is learning about opportunities to use a skill while ignoring the current state of the world, is well exemplified by computational approaches using Learning Progress (LP), which approximates some derivative of performance over time (initially proposed by Oudeyer et al., 2007, p. 7; see also related earlier work by Schmidhuber, 1991, p. 1461). In systems employing LP, it is typically used as a measure to support selecting intermediate-difficulty goals with respect to the changing skills of an agent over time. In fact, LP, as an intrinsic reward for goal selection, has been well studied across both machine and human reinforcement learners (e.g., Colas et al., 2019; Ten et al., 2021; Molinaro et al., 2024), and could thus offer good starting points for formalising *C2.a*.

While a (high-level) reward function prioritises the choice of skills, an agent also needs a mechanism to select a skill from that prioritisation. One commonly used mechanism is Thompson sampling (Thompson, 1933), also known as probability matching, in which an agent draws a skill according to a “belief”—a probability distribution (learned from a reward signal such as LP values) over its set of skills—and then acts greedily with respect to the drawn skill until it has more “evidence” (e.g., a change in LP values) and its “belief” (i.e., skill-selection policy) is updated. Thompson sampling has been shown to explain aspects of human exploration in certain tasks (Gershman, 2018), making it a candidate for a psychologically plausible formalism of human skill selection.

An example of a system formalising the second case of *C2.a*, in which an agent learns about opportunities to use particular skills while accounting for the current state of the world, is Variational Intrinsic Control (VIC). In VIC, the appropriate skill for a given starting state is decided using a separate learned distribution dependant on how empowering the

skill has been in the past. We mean *empowering* in the sense of its definition in the artificial intelligence (AI) literature,<sup>2</sup> as the combined ability of an agent to control its environment and to sense this control (cf., Salge et al., 2014, p. 2; Gregor et al., 2016, p. 2). The decision over skills is conditional on the state of the world, in that the policy for choosing skills is updated based on how empowering the skill turned out to be *for that given situation*. The distribution over skills is updated after skill use, such that more empowering skills become more likely to be chosen (Gregor et al., 2016, p. 4), using the reward (Gregor et al., 2016, p. 2):

$$R_t(\mathbf{s}_0, \mathbf{s}_f, \mathbf{g}) := \underbrace{\log q_{\phi}(\mathbf{g} | \mathbf{s}_0, \mathbf{s}_f)}_{(\alpha)} - \underbrace{\log p_{\theta}(\mathbf{g} | \mathbf{s}_0)}_{(\beta)}. \quad (2)$$

( $\alpha$ ) The discriminator,  $q_{\phi}$ , is an arbitrary variational distribution parametrised by  $\phi$  (Barber and Agakov, 2003, p. 2). Given the first ( $\mathbf{s}_0$ ) and last ( $\mathbf{s}_f$ ) observations induced by the skill corresponding to the goal,  $\mathbf{g}$ , being pursued,  $q_{\phi}$  defines a probability distribution over goals. Discriminating goals in the observation space requires the agent to observe distinct regions of the observation space. If a goal is not discriminable based on observations, two or more skills are producing overlapping behaviours; conversely, if a goal is discriminable, then the corresponding skill is inducing trajectories unique to that skill. Thus, this reward term is maximised when there is no uncertainty in the prediction.

( $\beta$ ) The probability model,  $p_{\theta}(\mathbf{g} | \mathbf{s}_0)$ , parametrised by  $\theta$ , represents the goal-selection policy. As part of the reward, the purpose of this term is to reinforce high entropy: the agent should keep pursuing a wide range of goals, for each starting state,  $\mathbf{s}_0$ , as the reward is higher when the goal had a low probability of being selected.

Given a starting state,  $\mathbf{s}_0$ , a goal,  $\mathbf{g}$ , is selected according to the distribution defined by  $p_{\theta}$ , which the agent pursues until the final state  $\mathbf{s}_f$ . Then,  $p_{\theta}$  is reinforced based on how empowering the corresponding skill was (estimated using Eq. 2). The final state becomes the new starting state,  $\mathbf{s}_0 \leftarrow \mathbf{s}_f$ , and the process is repeated. This can motivate the agent to ( $\alpha$ ) follow skills that lead to distinct final states and ( $\beta$ ) learn a wide range of skills for any given situation. Learning a state-conditional policy over skills fits the understanding of competence as finding suitable opportunities to use particular skills.

**That one uses a particular capacity in the course of one’s action (C2.b)** While *C2.a* is about finding suitable opportunities to use particular skills, *C2.b* is more about an agent being motivated by observing that a skill is in use at all. The first term of the VIC reward function (Eq. 2,  $\alpha$ ) can be adapted to emphasise an agent’s ability to distinguish that a particular skill is in use, by only conditioning on the current state. The reward for one such system, Diversity Is All You Need

<sup>1</sup>An alternative interpretation of “exercise”, as engaging a skill repeatedly to grow it, will be treated under *Capacity growth (C4)*.

<sup>2</sup>“Empowerment” has also been studied in the context of human exploration, e.g., as “exploring options that enable the generation of as many more options as possible” (Brändle et al., 2023, p. 1482).

(DIAYN) (Eysenbach et al., 2019), can be defined<sup>3</sup> as:

$$R_t(\mathbf{s}_{t+1}, \mathbf{g}) := \log q_\phi(\mathbf{g} | \mathbf{s}_{t+1}), \quad (3)$$

where  $q_\phi$  is a discriminator parametrised by  $\phi$  (refer back to the text following Eq. 2 for detail). An agent is rewarded for its ability to distinguish the current goal,  $\mathbf{g}$ ; thus, the overall return is maximised when the agent can perfectly infer which of its skills is in use at any given time.

DIAYN, in further differentiation from VIC, ignores the state of the world in its choice of skill—instead, choosing skills uniformly at random, thus learning skills that are empowering on average over the observation space. Fixing the distribution over skills as uniform, although not an intuitive choice for a psychologically plausible model of goal selection (except maybe in cases such as the absence of prior information) is common in the CB-IM literature. This approach ensures that all skills receive, in expectation, an equal amount of training signal to improve. DIAYN learns a diverse set of skills—running, walking, hopping, flipping, and gliding—across a wide array of situations (Eysenbach et al., 2019, p. 5). It thus shows that recognising which skill is in use, that is, being able to distinguish the use of a particular capacity in the course of one’s action, can drive skill learning.

Beyond DIAYN and VIC, one can turn to other variational empowerment methods (see Choi et al., 2021) for further examples of CB-IMs fitting competence as skill use.

### Task performance (C3)

Next, we discuss computational models for two possible ways of understanding *task performance*: observing that one performs well at an intended task (C3.a) and performing to an extent that requires a certain skill or skill level (C3.b).

**Observing that one performs well at an intended task (C3.a)** Given an agent has set itself a goal (task), if an agent has a mechanism to estimate its progress towards a chosen goal, it can use that estimate as a proxy to track its own performance on the task. An example system of this kind is Reinforcement Learning With Imagined Goals (RIG) (Nair et al., 2018). RIG encodes the agent’s observations into a lower-dimensional latent space from which the agent samples goals to pursue. Defining the reward as the negative distance between the goal and the current state in latent space encourages the agent to move its observations closer to the goal:

$$R_t(\mathbf{s}_{t+1}, \mathbf{g}) := -\|\phi(\mathbf{s}_{t+1}) - \phi(\mathbf{g})\|_{\mathbf{A}}, \quad (4)$$

where  $\phi(\mathbf{s}_{t+1})$  is the learned state representation of  $\mathbf{s}_{t+1}$ ,  $\phi(\mathbf{g})$  denotes the goal drawn from the latent space, and  $\mathbf{A}$  is a matrix that can give more importance to some dimensions of the latent space over others (Nair et al., 2018, p. 5).

In contrast to DIAYN, in which the agent learns to master a fixed number of skills, in RIG, the agent selects from an infinite number of goals (the space is continuous). This

results in a sort of smoothness of the space that the agent can benefit from: if one goal is sampled near another that the agent has already learned to achieve, it may be able to generalise and use some parts of an existing policy to more efficiently learn to achieve the new goal. In the RIG system, *goals* take on the intuitive meaning as desired states for the agent to reach (discussed in *Background: Computational IM & RL*), and “performing well” at an intended task is operationalised as minimising the semantic distance between the agent’s observations and its self-generated goals. Over time, RIG agents learn to generate goals that approximate actual observations they make of their environment, while simultaneously learning the skills necessary to consistently perform well (reach the goals).

**Specifically to an extent that requires a certain skill or skill level (C3.b)** While RIG captures the idea of motivation by performance on a self-selected task (C3.a), it does not take into account the level of skill needed to perform well on the task (C3.b). For this, we can look to other system designs inspired by the idea of preferring to focus on goals or skills at an appropriate level of difficulty, given what the agent already knows. To identify skills of appropriate difficulty, such systems can make use of Automatic Curriculum Learning (ACL) (for a review see Portelas et al., 2021), enabling an agent to self-organise their goal selection. This can significantly reduce the number of training examples required for an agent to learn to reach its goals (Florensa et al., 2018, p. 1).

For example, in the Continual Universal Reinforcement learning with Intrinsically motivated sUBstitutionS (CURI-IOUS) system by Colas et al. (2019), the system considers its capability to achieve its goals. Specifically, the agent is motivated to pursue goals in certain parts of the sensory space in which it believes itself to be increasingly or decreasingly successful at achieving its goals, estimated by measuring how successful it has been in the recent past—that is, its LP. The underpinning logic of the system is that such goals will be of optimal difficulty, neither too easy nor too difficult. In the CURI-IOUS experiments, an agent manipulates cubes by learning to control a robotic arm (Colas et al., 2019, Supplementary p. 3). Each observation includes information about the gripper and the objects of interest (e.g., their position and rotation), and each goal is a desired observation. The high-dimensional observation space makes learning challenging. Furthermore, some parts of the sensory space are irrelevant for achieving certain goals—for example, if the goal is to manipulate only one of several cubes, the agent might not need information about the other cubes. To ignore such irrelevant information in its formulation of goals, CURI-IOUS modularises the observation space, which is also the goal space, with each module being a specific, hand-defined subspace of this space. A CURI-IOUS agent estimates its own subjective LP for each module, and uses the absolute values of its LP estimates to compute a probability distribution for selecting the module to generate a goal within (Colas et al., 2019, p. 4). This mechanism helps the agent avoid modules containing

<sup>3</sup>See Eysenbach et al., 2019, Appendix A, p. 12 for discussion on a constant term omitted here.

goals that are too difficult or for which it already has sufficient skill, as well as to return to skills it is forgetting (Colas et al., 2019, pp. 2–3). In this way, the selection of different goals is driven by self-assessment of task performance.

The *subjective LP* for a given module,  $LP(M_i)$ , is defined as the derivative of *subjective competence*,  $C(M_i)$ , with respect to time, where  $C(M_i)$  is defined as the mean over results from self-evaluated trials performed by the agent (see Colas et al., 2019, p. 4). The results are stored in a queue of preset length, and each result is binary: either 1 for a successful trial or 0 for a failure. This can be viewed as the frequentist estimate of the probability of success. Finally, the probability of choosing a module is computed as:

$$p(M_i) := \varepsilon \times \frac{1}{N} + (1 - \varepsilon) \times \frac{|LP(M_i)|}{\sum_{j=1}^N |LP(M_j)|}, \quad (5)$$

in which  $\varepsilon \in [0, 1]$  is a hyperparameter that can be used to ensure that even poorly performing modules are occasionally selected, and  $N$  corresponds to the total number of modules. As  $\varepsilon$  approaches 1, the probability distribution over modules resembles a uniform distribution. When  $\varepsilon$  is smaller than 1, the modules with the highest absolute LP will have the highest probability of being selected.

In the same vein, an emerging ACL paradigm, Unsuper-vised Environment Design (UED) (for a review see Rutherford et al., 2024), formalises the problem of automatically generating environments for agents based on some criterion, for instance, the difference between the current agent and (in practice, an approximation of) an optimal agent (Rutherford et al., 2024, p. 1). In UED, environments with their own reward functions can be thought of as goals for an agent to pursue. As part of one such system, Open-endedness via Models of human Notions of Interestingness with Environments Programmed in Code (OMNI-EPIC), agents self-generate novel, interesting, and solvable goals, appropriate for an agent’s current capabilities (Faldor et al., 2024, p. 2).

While a significant proportion of ACL methods optimise task difficulty or competence progress with respect to current capabilities, ACL can also motivate agents to optimise goal selection based on other criteria that can be interpreted as proxies for “skill level” (from *C3.b*). Pong et al. (2020) extended RIG (introduced in *C3.a*) by weighting the selection of previously visited states used in learning the goal-generation model, encouraging the agent to generate goals resembling less frequently visited states. This weighting can be seen as a heuristic for attending to tasks that are expected to be non-trivial for the agent. As a second example, Lintunen et al. (2024) introduce a method motivating agents to select goals based on how much the agent believes pursuing them will diversify its entire set of skills. As ACL systems often consider the level of skill needed to perform well on the task, they offer a good starting point for formalising *C3.b*.

## Capacity growth (C4)

Competence, understood as *capacity growth*, can be interpreted in two different ways: observing gain in the strength

(*C4.a*) and/or range (*C4.b*) of one’s skills. While we saw that a reward reflecting a derivative of some measure of performance (e.g., LP) can direct the agent towards goals requiring a certain level of skill (*C3.b*), the same derivative directly encourages capacity growth or improvement (*C4.a*). While these two facets appear to have different meanings, from a computational perspective, they can be closely aligned.

An example of a system that explores both *C4.a* and *C4.b*—both repeatedly engaging in skills that need strengthening and extending the agent’s skill repertoire—is Intrinsically Motivated Reinforcement Learning (IMRL) described by Barto et al. (2004) and Singh et al. (2004). IMRL extends the agent’s repertoire of skills (*C4.b*): each time a “salient” event is observed, the agent adds a new skill associated with the salient event.<sup>4</sup> After this first observation, the agent learns how to return to observe the salient event again from any other situation via experience over time—this is the new skill associated with the salient event. Strictly speaking, the skill repertoire expands as a matter of course, not because the agent is intrinsically rewarded for expanding it.

The intrinsic reward used in the IMRL system (Singh et al., 2004, p. 4) can be defined as:

$$R_t(\mathbf{s}_t, \mathbf{s}_{t+1}, \mathbf{g}_{\mathbf{s}_{t+1}}) := \begin{cases} 1 - P^{\mathbf{g}_{\mathbf{s}_{t+1}}}(\mathbf{s}_{t+1} | \mathbf{s}_t) & \text{if } \mathbf{s}_{t+1} \text{ is salient} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where  $P^{\mathbf{g}_{\mathbf{s}_{t+1}}}$  is what Precup and Sutton (1997) call a *multi-time model* for skill  $\mathbf{g}_{\mathbf{s}_{t+1}}$ .  $P^{\mathbf{g}_{\mathbf{s}_{t+1}}}$  is not *quite* a probability estimate for the transition from  $\mathbf{s}_t$  to  $\mathbf{s}_{t+1}$  given the agent is following skill  $\mathbf{g}_{\mathbf{s}_{t+1}}$ , but that is close to the correct intuition.

The agent only receives a non-zero intrinsic reward when it reaches a salient state. The intrinsic reward here can be thought of as modelling how “surprising” (Singh et al., 2004, p. 6) it is that the agent reached  $\mathbf{s}_{t+1}$  immediately from state  $\mathbf{s}_t$ —which it has just done—by following the skill  $\mathbf{g}_{\mathbf{s}_{t+1}}$ . If the agent expected, without uncertainty, to transition directly from  $\mathbf{s}_t$  to  $\mathbf{s}_{t+1}$ , then  $P^{\mathbf{g}_{\mathbf{s}_{t+1}}}$  is 1 and the reward is 0 (the transition is very “unsurprising”). If the agent did not expect it could transition from  $\mathbf{s}_t$  to  $\mathbf{s}_{t+1}$  by following skill  $\mathbf{g}_{\mathbf{s}_{t+1}}$ , then  $P^{\mathbf{g}_{\mathbf{s}_{t+1}}}$  is 0 and the reward is 1 (maximally “surprising”). If the agent is uncertain,  $P^{\mathbf{g}_{\mathbf{s}_{t+1}}}$  will be somewhere in between. However, if the agent expected that it could reach  $\mathbf{s}_{t+1}$  from  $\mathbf{s}_t$  by following skill  $\mathbf{g}_{\mathbf{s}_{t+1}}$ , just that it would take more than a single step,  $P^{\mathbf{g}_{\mathbf{s}_{t+1}}}$  will be discounted based on the length of the path. This reward encourages the agent to decrease the number of steps it takes to get from one state to another, strengthening its skill of reaching a particular state.

Another example of *C4.b*, that is, being motivated by growing the number of skills in the agent’s repertoire, is the aforementioned VIC (see *C2.a*). The VIC reward function motivates the agent to grow its number of effective skills: in expectation, the second term of the reward function (Eq. 2,

<sup>4</sup>Barto et al. (2004) leave to future work what “salient” means and how to determine whether an event is salient or not. They simply hard-code events like turning on a light or ringing a bell as salient.

$\beta$ ) increases as the distribution over skills becomes more uniform. In theory, this encourages the distribution over skills not to collapse to a small number of skills,<sup>5</sup> maintaining or increasing the size of the repertoire.

### Advancing Competence Modelling in SDT

What is gained for human motivation researchers by identifying computational formalisms that match aspects of verbal theory? For one, translating a concept like competence into an algorithmic description induces a different way of thinking about theory: the process forces us to specify abstract and informal verbal theories to the point that they can be turned into a system that can actually produce observable behaviour. Observing how a computational model of a theory behaves can “reveal implicit assumptions and hidden complexities” (Marsella et al., 2010, p. 4) and resolve jingle-jangle fallacies (Murayama and Jach, 2025, pp. 3–4, 6; Pekrun, 2024, p. 26).

For instance, LP-based methods can simultaneously exemplify skill use as (*C2.a*) an opportunity to use a particular skill, (*C3.b*) task performance as requiring a certain skill or skill level, and (*C4.a*) capacity growth as observing gains in the strength of one’s skills. Given that SDT does not acknowledge distinct facets, finding such a formalism that fits multiple facets supports the theory. On the other hand, in our matching, some facets are formally distinct. For example, it is unclear how (*C1*) effectance, (*C4.b*) being motivated by extending one’s repertoire of skills, or (*C3.a*) being motivated by observing that one performs well at an intended task, equate to each other and other facets. By identifying formalisms for each of the facets, and testing whether they result in demonstrably different observable behaviours (e.g., using simulation), we can prompt review of SDT—is one term, “competence”, being used for multiple different constructs?

Another benefit of this matching is that it reveals underlying preconditions that SDT fails to specify, such as that a competence-motivated agent must be able to recognise when its actions cause effects (*C1*) and differentiate between distinct skills (*C2*, *C3*, *C4*). Similarly, formalisms of concepts such as novelty, diversity, and progress towards goals, underlie competence-driven learning in computational IM, yet are often overlooked in SDT. By drawing on such formalisms from computational agents that demonstrably and autonomously develop capacities in diverse environments, this paper shines light on overlooked preconditions and what potential mechanisms might computationally fulfil them.

We do not argue that identifying AI algorithms that demonstrably perform some capacities will mean that we have identified which algorithms best approximate human cognition (Van Rooij et al., 2024, p. 626). Rather, our work can support a cycle of theory development by inspiring new computational models, which can then be tested to refine the theory.

---

<sup>5</sup>Eysenbach et al. (2019, Appendix E, pp. 17–18) showed empirically that, in practice, since there are numerous other factors that affect how agents learn, this collapse may still happen.

### Future Work and Conclusion

Showing that RL formalisms can be aligned with SDT uncovers a wealth of resources for motivation research, with existing implementations of models, simulators, and evaluation metrics in the papers associated with each formalism. These resources can, by enabling large-scale simulation studies, allow us to explore computational IMs as candidates for model fitting to human data. We have used the different facets of competence outlined by Deterding et al. (2024) as starting points for selecting promising computational formalisms. However, it remains an open question whether the facets capture different cognitive mechanisms or even require separate formalisms to generate associated explananda, such as exploratory and challenge-seeking behaviour. Towards investigating part of this question, the computational RL literature offers different methods for composing reward functions (e.g., Bagot et al., 2020), but how best to combine rewards and what methods would be psychologically plausible still require work. Such methods, in combination with the aforementioned resources, allow us to study how distinct facets of competence, either individually or in conjunction, contribute to competence-driven behaviour and motivation more widely.

These avenues of research are contingent on developing effective means to compare artificial agents with biological agents (a project partially undertaken by researchers in *representational alignment*; e.g., Sucholutsky et al., 2024). While work exists comparing behaviour induced by different intrinsic reward functions (e.g., Santucci et al., 2012; Biehl et al., 2018; Linke et al., 2020), as does work studying different choices of state representation (e.g., Burda et al., 2019), more research is needed on how different models behave. It is also an open question how to best evaluate intrinsically motivated open-ended learning (Colas et al., 2022, pp. 1167–1169).

Most formalisms we identified relate to pursuing goals, so we need to understand how humans represent and generate goals (Colas et al., 2024, p. 1). The topic has not been well studied for various reasons, e.g., computational intractability (Byers et al., 2024, p. 1). Excitingly, goals have recently emerged as a multidisciplinary area of research spanning AI, cognitive science, and developmental psychology (e.g., Colas et al., 2022; Chu et al., 2024; Davidson et al., 2024). Molinaro and Collins (2023, p. 1150) suggest that goals affect many agent dynamics, such as representing the environment, choosing relevant actions, and evaluating rewards.

In summary, we demonstrate that computational models of IM can plausibly match verbal definitions of competence in SDT. Specifically, we show that four distinct facets of competence—effectance, skill use, task performance, and capacity growth—can be aligned with existing formalisms from intrinsically motivated RL, and in the process, highlight the potential of computational models to provide deeper insights into motivated behaviour. While this is a promising first step, further research is required to empirically study these computational models and consider other potential models, inviting collaboration from motivation researchers across disciplines.

## Acknowledgments

We thank the members of the Intrinsically Motivation Open-ended Learning (IMOL) community for fun and illuminating discussions during workshops in Paris (2023) and Vancouver (2024). We also thank the members of the Autotelic Interaction Research (AIR) group, in particular Niki Pennanen, for feedback. EML and CG received financial support from the Research Council of Finland (NEXT-IM, grant 349036) and NMA from the Helsinki Institute for Information Technology.

## References

- Aubret, A., Matignon, L., and Hassas, S. (2023). An information-theoretic perspective on intrinsic motivation in reinforcement learning: A survey. *Entropy*, 25(2).
- Bagot, L., Mets, K., and Latré, S. (2020). Learning intrinsically motivated options to stimulate policy exploration. In *Lifelong Machine Learning Workshop at ICML 2020*.
- Baldassarre, G. (2022). Intrinsic motivations for open-ended learning. In *Cognitive Robotics*, page 251–270. The MIT Press.
- Barber, D. and Agakov, F. (2003). The IM algorithm: a variational approach to information maximization. In *Advances in Neural Information Processing Systems*, pages 201–208.
- Barto, A. G., Singh, S., and Chentanez, N. (2004). Intrinsically motivated learning of hierarchical collections of skills. In *International Conference on Development and Learning*, pages 112–120.
- Berlyne, D. (1965). *Structure and Direction in Thinking*. John Wiley & Sons.
- Biehl, M., Guckelsberger, C., Salge, C., Smith, S. C., and Polani, D. (2018). Expanding the active inference landscape: More intrinsic motivations in the perception-action loop. *Frontiers in Neurorobotics*, 12.
- Brändle, F., Stocks, L. J., Tenenbaum, J. B., Gershman, S. J., and Schulz, E. (2023). Empowerment contributes to exploration behaviour in a creative video game. *Nature Human Behaviour*, 7(9):1481–1489.
- Burda, Y., Edwards, H., Pathak, D., Storkey, A., Darrell, T., and Efros, A. A. (2019). Large-scale study of curiosity-driven learning. In *International Conference on Learning Representations*.
- Byers, J. B., Zhao, B., and Niv, Y. (2024). Modeling goal selection with program synthesis. In *Intrinsically Motivated Open-ended Learning Workshop at NeurIPS 2024*.
- Choi, J., Sharma, A., Lee, H., Levine, S., and Gu, S. S. (2021). Variational empowerment as representation learning for goal-conditioned reinforcement learning. In *International Conference on Machine Learning*, pages 1953–1963.
- Chu, J., Tenenbaum, J. B., and Schulz, L. E. (2024). In praise of folly: flexible goals and human cognition. *Trends in Cognitive Sciences*, 28(7):628–642.
- Colas, C., Chu, J., Molinaro, G., and Hawkins, R. (2024). What should i do now? goal-centric outlooks on learning, exploration, and communication. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.
- Colas, C., Fournier, P., Chetouani, M., Sigaud, O., and Oudeyer, P.-Y. (2019). CURIOUS: Intrinsically motivated modular multi-goal reinforcement learning. In *International Conference on Machine Learning*, pages 1331–1340.
- Colas, C., Karch, T., Sigaud, O., and Oudeyer, P. (2022). Autotelic agents with intrinsically motivated goal-conditioned reinforcement learning: a short survey. *Journal of Artificial Intelligence Research*, 74:1159–1199.
- Davidson, G., Todd, G., Togelius, J., Gureckis, T. M., and Lake, B. M. (2024). Goals as reward-producing programs. *arXiv preprint arXiv:2405.13242*.
- Dayan, P. and Balleine, B. W. (2002). Reward, motivation, and reinforcement learning. *Neuron*, 36(2):285–298.
- Deci, E. L. and Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. Springer US.
- Deterding, S., Lintunen, E. M., Guckelsberger, C., and Ady, N. M. (2024). Why self-determination theory needs formal modelling: The case of competence and balanced challenge. *PsyArXiv preprint 10.31234/osf.io/n6x8s\_v1*.
- Eysenbach, B., Gupta, A., Ibarz, J., and Levine, S. (2019). Diversity is all you need: Learning skills without a reward function. In *International Conference on Learning Representations*.
- Faldor, M., Zhang, J., Cully, A., and Clune, J. (2024). Omnipic: Open-endedness via models of human notions of interestingness with environments programmed in code. *arXiv preprint arXiv:2405.15568*.
- Florensa, C., Held, D., Geng, X., and Abbeel, P. (2018). Automatic goal generation for reinforcement learning agents. In *International Conference on Machine Learning*, pages 1515–1528.
- Gershman, S. J. (2018). Deconstructing the human algorithms for exploration. *Cognition*, 173:34–42.
- Gregor, K., Rezende, D. J., and Wierstra, D. (2016). Variational intrinsic control. *arXiv preprint arXiv:1611.07507*.
- Kaelbling, L. P. (1993). Learning to achieve goals. In *International Joint Conference on Artificial Intelligence*, pages 1094–8.
- Kakade, S. and Dayan, P. (2002). Dopamine: generalization and bonuses. *Neural Networks*, 15(4):549–559.
- Lidayan, A., Dennis, M., and Russell, S. (2024). BAMDP shaping: a unified theoretical framework for intrinsic motivation and reward shaping. *arXiv preprint arXiv:2409.05358*.
- Linke, C., Ady, N. M., White, M., Degris, T., and White, A. (2020). Adapting behavior via intrinsic reward: A survey and empirical study. *Journal of Artificial Intelligence Research*, 69:1287–1332.

- Lintunen, E. M., Ady, N. M., and Guckelsberger, C. (2024). Diversity progress for goal selection in discriminability-motivated RL. In *Intrinsically Motivated Open-ended Learning Workshop at NeurIPS 2024*.
- Liu, M., Zhu, M., and Zhang, W. (2022). Goal-conditioned reinforcement learning: Problems and solutions. In *International Joint Conference on Artificial Intelligence*, pages 5502–5511.
- Machado, A. and Silva, F. J. (2007). Toward a richer view of the scientific method: The role of conceptual analysis. *American Psychologist*, 62(7):671.
- Marsella, S., Gratch, J., and Petta, P. (2010). Computational models of emotion. In *A Blueprint for Affective Computing: A sourcebook and manual*. Oxford University Press.
- Merriam-Webster.com (2025). Retrieved from <https://www.merriam-webster.com>.
- Molinaro, G., Colas, C., Oudeyer, P.-Y., and Collins, A. (2024). Latent learning progress drives autonomous goal selection in human reinforcement learning. In *Advances in Neural Information Processing Systems*, pages 32251–32280.
- Molinaro, G. and Collins, A. G. (2023). A goal-centric outlook on learning. *Trends in Cognitive Sciences*, 27(12):1150–1164.
- Murayama, K. and Jach, H. K. (2025). A critique of motivation constructs to explain higher-order behavior: We should unpack the black box. *Behavioral and Brain Sciences*, 48:24.
- Nair, A. V., Pong, V., Dalal, M., Bahl, S., Lin, S., and Levine, S. (2018). Visual reinforcement learning with imagined goals. In *Advances in Neural Information Processing Systems*, pages 9191–9200.
- Oberauer, K. and Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review*, 26(5):1596–1618.
- Oudeyer, P.-Y. and Kaplan, F. (2007). What is intrinsic motivation? A typology of computational approaches. *Frontiers in Neurobotics*, 1:1–14. Article 6.
- Oudeyer, P.-Y. and Kaplan, F. (2008). How can we define intrinsic motivation? In *International Conference on Epigenetic Robotics*.
- Oudeyer, P.-Y., Kaplan, F., and Hafner, V. V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, 11(2):265–286.
- Pathak, D., Agrawal, P., Efros, A. A., and Darrell, T. (2017). Curiosity-driven exploration by self-supervised prediction. In *International Conference on Machine Learning*, pages 2778–2787.
- Pekrun, R. (2024). Overcoming fragmentation in motivation science: Why, when, and how should we integrate theories? *Educational Psychology Review*, 36(1):27.
- Pong, V., Dalal, M., Lin, S., Nair, A., Bahl, S., and Levine, S. (2020). Skew-fit: State-covering self-supervised reinforcement learning. In *International Conference on Machine Learning*, pages 7783–7792.
- Portelas, R., Colas, C., Weng, L., Hofmann, K., and Oudeyer, P.-Y. (2021). Automatic curriculum learning for deep rl: a short survey. In *International Joint Conference on Artificial Intelligence*, pages 4819–4825.
- Precup, D. and Sutton, R. S. (1997). Multi-time models for temporally abstract planning. In *Advances in Neural Information Processing Systems*, pages 1050–1056.
- Raileanu, R. and Rocktäschel, T. (2020). RIDE: Rewarding impact-driven exploration for procedurally-generated environments. In *International Conference on Learning Representations*.
- Robinaugh, D. J., Haslbeck, J. M. B., Ryan, O., Fried, E. I., and Waldorp, L. J. (2021). Invisible Hands and Fine Calipers: A Call to Use Formal Theory as a Toolkit for Theory Construction. *Perspectives on Psychological Science*, 16(4):725–743.
- Rutherford, A., Beukman, M., Willi, T., Lacerda, B., Hawes, N., and Foerster, J. N. (2024). No regrets: Investigating and improving regret approximations for curriculum discovery. In *Advances in Neural Information Processing Systems*, pages 16071–16101.
- Ryan, R. M. (2023). *The Oxford Handbook of Self-Determination Theory*. Oxford University Press.
- Ryan, R. M. and Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford.
- Salge, C., Glackin, C., and Polani, D. (2014). Empowerment—an introduction. In *Guided Self-Organization: Inception*, pages 67–114. Springer Berlin Heidelberg.
- Santucci, V. G., Baldassarre, G., and Mirolli, M. (2012). Intrinsic motivation mechanisms for competence acquisition. In *International Conference on Development and Learning and Epigenetic Robotics*, pages 1–6.
- Schaul, T., Horgan, D., Gregor, K., and Silver, D. (2015). Universal value function approximators. In *International Conference on Machine Learning*, pages 1312–1320.
- Scheel, A. M., Tiokhin, L., Isager, P. M., and Lakens, D. (2021). Why hypothesis testers should spend less time testing hypotheses. *Perspectives on Psychological Science*, 16(4):744–755.
- Schmidhuber, J. (1991). Curious model-building control systems. In *IEEE International Joint Conference on Neural Networks*, pages 1458–1463.
- Silver, D., Singh, S., Precup, D., and Sutton, R. S. (2021). Reward is enough. *Artificial Intelligence*, 299.
- Singh, S., Barto, A., and Chentanez, N. (2004). Intrinsically motivated reinforcement learning. In *Advances in Neural Information Processing Systems*, pages 1281–1288.

- Sucholutsky, I., Muttenthaler, L., Weller, A., Peng, A., Bobu, A., Kim, B., Love, B. C., Cueva, C. J., Grant, E., Groen, I., Achterberg, J., Tenenbaum, J. B., Collins, K. M., Hermann, K. L., Oktar, K., Greff, K., Hebart, M. N., Cloos, N., Kriegeskorte, N., Jacoby, N., Zhang, Q., Marjeh, R., Geirhos, R., Chen, S., Kornblith, S., Rane, S., Konkle, T., O'Connell, T. P., Unterthiner, T., Lampinen, A. K., Müller, K.-R., Toneva, M., and Griffiths, T. L. (2024). Getting aligned on representational alignment. *arXiv preprint arXiv:2310.13018*.
- Sutton, R. S. and Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. The MIT Press, second edition.
- Ten, A., Kaushik, P., Oudeyer, P.-Y., and Gottlieb, J. (2021). Humans monitor learning progress in curiosity-driven exploration. *Nature communications*, 12(1):5972.
- Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3–4):285–294.
- Van Dongen, N., van Bork, R., Finnemann, A., Haslbeck, J., van der Maas, H. L., Robinaugh, D. J., de Ron, J., Sprenger, J., and Borsboom, D. (2024). Productive explanation: A framework for evaluating explanations in psychological science. *Psychological review*.
- Van Rooij, I., Guest, O., Adolfs, F., de Haan, R., Kolokolova, A., and Rich, P. (2024). Reclaiming ai as a theoretical tool for cognitive science. *Computational Brain & Behavior*, 7(4):616–636.
- White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*, 66(5):297–333.