TESTING BEHAVIORAL THEORIES OF MOTIVATION IN ATARI AGENTS

Maria Gabriela Valeriano & Eshter Luna Colombini
Instituto de Computação
Universidade Estadual de Campinas
Campinas, Brasil
`{mgv, estherlc}@unicamp.br`

ABSTRACT

Motivation plays a central role in biological cognition by shaping attention, valuation, and action selection under internal state constraints. However, in most reinforcement learning systems, motivation is reduced to a fixed external reward signal. In this work, we investigate how different motivational theories from behavioral neuroscience alter agent behavior when embedded into the decision process of a learning agent. We implement motivational modulation mechanisms on top of a model-free reinforcement learner and evaluate their behavioral consequences in interactive environments. Our goal is not to improve benchmark performance, but to study how distinct motivational formulations lead to different patterns of cognitive behavior, exploration, and cue responsiveness. This provides a computational bridge between motivational theory and cognitive decision-making in artificial agents. This work is currently in progress and full experimental results are not yet available. Nevertheless, we argue that the proposed computational formulation, experimental protocol, and theoretical integration between motivational neuroscience and reinforcement learning provide a useful basis for discussion and early feedback within the cognition and motivation research community.

1 INTRODUCTION

Motivated behavior can be understood as goal-directed action, involving changes in the environment to achieve a desired outcome. With mobility and the ability to choose among options, animals rely on motivation as a mechanism for decision-making (Baumeister, 2016). There is no consensus in the literature on a unified theory of motivation (Gudwin, 2019). Various approaches have been proposed to explain and model motivational phenomena. One traditional framework is Hull's drive reduction theory, which conceptualizes motivation as arising from internal physiological needs. These needs disrupt homeostasis and give rise to psychological drives—internal states of tension that motivate behavior aimed at restoring balance (Hull, 1959).

In Hull's framework, the strength of a drive energizes behavior, while learned habits determine its direction. The likelihood of a specific response occurring is governed by the interaction of drive strength and habit strength. One of the most representative experimental applications of this theory comes from maze learning experiments. In these studies, rats were subjected to varying degrees of food deprivation to manipulate the strength of their Hunger drive. When placed in mazes with food rewards at the end, rats that were more deprived, and therefore experiencing stronger drives, learned to navigate the maze more rapidly and accurately demonstrating how drive reduction could serve as a key mechanism for motivation (Hull, 1959).

In humans, basic drives would have evolved into more complex motivational layers. Social life introduces additional drives, such as the need for acceptance, status, and competition. These socially conditioned drives shape behavior in nuanced ways, going beyond mere survival needs (Gudwin, 2019; Sun, 2009). In his theory of human motivation, Maslow (1943) postulated that needs are organized in layers in a hierarchical structure, from the most basic, such as hunger and sleep, to the advanced layers involving esteem needs and self-actualization (Maslow, 1943). Computational models have also drawn upon Hull's drive-based motivation framework to simulate behavior in artificial agents (Berto et al., 2024).

From the perspective of behavioral neuroscience, theories based on drives lack evidence. A simple example is individuals fed intravenously that continues feeling hungry (Berridge, 2004). This indicates that reward is not the same as drive reduction. Additionally, no dedicated brain neurons related to specific motivations were successfully identified. On the other hand, stimulation of a specific brain area could trigger distinct motivated behaviors depending on external situations and individual predispositions and experiences (Berridge, 2004). Motivated behavior is also strongly influenced by hedonic stimuli, which can reinforce actions even when they do not reduce physiological drives (Winkielman et al., 2005). The limitations of drive theories in explaining motivation at the neural level contributed to the emergence of incentive-based approaches. The early contributions of Bindra (1978) and Toates (1994) laid the foundation for understanding motivation as an association between hedonic rewards and predictive cues, where hedonic reward is interpreted as sensory pleasure (Berridge, 2004). As learning evolves, conditioned stimuli can acquire motivational properties, becoming both *liked* and *wanted*. In this direction, Berridge and Robinson proposed that the components of *want* and *like* may be dissociated under some circumstances, the basis of Incentive Salience theory (Berridge & Robinson, 1998; Robinson & Berridge, 2025).

In this work we want to observe how an Atari agent performs when receiving not only extrinsic game rewards. We introduce additional motivational rewards, implemented according to different theoretical perspectives from the motivation literature. Our goal is to observe which behavioral patterns emerge under each approach, and how these motivational mechanisms shape learning and action selection. Our proposal differs from the intrinsic motivation frameworks commonly explored in RL (curiosity-driven exploration, novelty bonuses, prediction error, or information gain) once we investigate the motivational rewards explicitly grounded in behavioral and neuroscientific theories of motivation, aiming to operationalize constructs such as hedonic value and incentive salience. Therefore, rather than treating intrinsic reward as a generic exploration bonus, we model motivation as a theoretically informed reward signal that may generate distinct goal-directed and cue-driven behaviors. We do not claim to provide a direct empirical comparison between psychological theories. Instead, our contribution addresses a gap in the literature regarding computational simulations of motivational theories using reinforcement learning agents. The experiments are currently being conducted and will be submitted to the International Conference on Development and Learning. We first introduce the theories that based our agents, in Section 2. Section 3 describes our proposal.

2 *Want* AND *Like*

A recurrent topic in motivation research is the distinction between *liking* and *wanting* (Baumeister, 2016; Gudwin, 2019; Anselme & Robinson, 2016; Berridge, 2009). Both are components of the motivational system, but serve different functions. *Like* refers to the immediate pleasure experienced in the present, often without any forward-looking component. In contrast, *want* arises in the present but is oriented toward the future, prompting action in pursuit of a desired reward (Gudwin, 2019). Baumeister (2016) argues that motivation is fundamentally rooted in *wanting*. A comprehensive theory of motivation must, according to the same author, explain not only what is *wanted*, but also how this *wanting* emerges and operates. The author also raises questions about how *want* and *like* adapt to the environment. For example, repeated exposure to the same reward (e.g., the same food at the same time and quantity) leads to reduced satisfaction.

This suggests that perception of scarcity and novelty shapes motivation. When a desired object or experience is frequently unavailable, the intensity of the desire may decrease (Baumeister, 2016). Conversely, when gratification is easy and frequent, drive values may increase. This feedback loop suggests that motivation is both a response to the environment and a form of behavioral reinforcement (Baumeister, 2016). To *want* something and then obtain it reinforces the *wanting* mechanism itself. What, then, is the role of *liking*? Gudwin (2019) proposed that *like* is a hedonic measure that tracks the degree to which a drive is being satisfied instantaneously. So while *want* is a previous and direct behavior to meet needs, *like* would have evolved as a form of internal reward, helping organisms learn which behaviors reduce internal imbalances. In this way, *like* supports learning through reinforcement.

Through the lens of the Incentive Motivation of Brindra-Toates *want* and *like* were basically synonyms for the same incentive value. In the Incentive Salience theory, proposed by Berridge, they can be dissociated. We detail this process next.

2.1 THE CORTICAL AND SUBCORTICAL COMPONENTS OF *want* AND *like*

One of the key difficulties in understanding the roles of *want* and *like* in decision-making lies in their inherently subjective nature. Clear operational definitions are crucial, yet often lacking, which can result in contradictory outcomes (Pool et al., 2016). Research involving human participants reveals considerable variability in how psychological constructs are defined and measured, which can lead to potential confounding factors. How can we define and measure these components more effectively?

Some researchers have proposed a distinction between cortical and subcortical components of *want* and *like*. The subcortical aspects, evolutionarily ancient, are labeled within quotes: *'want'* and *'like'*. These refer to the basic, unconscious motivational systems. In contrast, the terms *want* and *like* without quotes refer to the conscious, cognitive interpretations of desire and pleasure (Berridge, 2007). A crucial advantage of this distinction is that subcortical components can be objectively measured in laboratory settings. *'Liking'* reactions, for instance, are observable through specific facial expressions in response to hedonic stimuli. These expressions, such as smiles or grimaces, appear universally in human infants and the young of other species when exposed to sweet or bitter tastes Berridge (2007).

While *want* refers to explicit, conscious desires, *'want'* reflects a more visceral, unconscious drive and can be measured behaviorally by assessing the effort an organism is willing to exert to obtain a reward. In rodent studies, for example, *'wanting'* can be measured by the number of times an animal presses a lever to receive a reward Robinson et al. (2016). To better understand the dynamic between *'like'* and *'want'*, the Incentive Salience theory offers a compelling explanation Robinson & Berridge (1993). According to this theory, the experience of reward consists of three components: *'liking'*, learning as an association of cues with rewards, and *'wanting'* or Incentive Salience. Different neural circuits govern these components and can become dissociated under specific conditions (Berridge, 2009).

When dopamine transmission is blocked in rodents, either by receptor antagonists or by inhibiting dopamine production, *'wanting'* behavior drastically reduces or disappears. The animals no longer approach food (Berridge, 2007). However, when a sweet solution is directly administered into their mouths, they still exhibit normal 'liking' facial responses. Conversely, electrical stimulation of specific brain regions can trigger intense *'wanting'* (compulsive eating) without any corresponding signs of 'liking,' and even with signs of aversion (Berridge, 2007). The theory posits that it is the assignment of Incentive Salience that transforms a merely pleasurable stimulus into something that is actively *wanted* (Robinson et al., 2016). Moreover, this attribution of *'want'* is re-triggered every time the stimulus is encountered, with its intensity modulated by both past experiences and the organism's current physiological state. This helps explain why *'want'* can fluctuate independently of *'like'*.

3 PROPOSAL

The key starting point we propose here is to explore how integrating motivational theories into reinforcement learning (RL) frameworks can enhance autonomous decision-making. The integration of a motivational theory within an RL framework sounds promising, also due to the historical relationship between RL and animal behavior (Sutton & Barto, 2018). For this purpose, we adopt the Atari game Ms. Pac-Man implemented using the Arcade Learning Environment (Bellemare et al., 2013). In this game, the agent moves around a maze collecting food and avoiding ghosts. In the frightened mode, after eating a Power Pellet, ghosts can be eaten as well. The objective of Pac-Man is to maximize cumulative score while avoiding termination, which occurs when Pac-Man loses all lives.

Formally, at each episode, the agent must learn a policy that balances:

- Reward collection: eating pellets, power pellets, and ghosts;

- Risk avoidance: avoiding non-frightened ghosts.

The episode ends when all pellets are consumed, at level completion, or Pac-Man loses all available lives. The reward is proportional to the score increase, and it is increasingly positive when eating a pellet, a power pellet or a frightened ghost also increasing per ghost. The reward is negative when getting caught by a ghost, losing a life or ending the episode. Survival is not directly rewarded.

We adopt a model-free reinforcement learning approach, which means the agent does not learn or use explicit model of the environment dynamics. A Deep Q-learning (DQN) is adopted for training following the approach adopted by (Mnih et al., 2013). Unlike the original pixel-based training, we use RAM state information as input, that reduces state dimensionality and allow the computation of behavioral metrics to compare agents.

In the work of (Mnih et al., 2013) there is a distinction among Atari games regarding DQN performance. Some games are more challenging due to delayed rewards and stochastic environments. In Pong and BreakOut, for example, DQN beats the human performance. However it remains far below in games that require building a long-run strategy. Therefore in Ms. Pac-Man achieving human-level performance is not expected. This motivates our choice: relying only on extrinsic reward would not reach optimal learning. Notably, agent57, which incorporates intrinsic motivation and directed exploration mechanisms, achieved human or superhuman performance across the Atari benchmark (Badia et al., 2020), reinforcing the potential benefits of intrinsic motivational signals. We therefore believe there is opportunity to improve behavior by adding intrinsic motivation signals.

We are also modeling environments employing Minecraft, where the same concepts used here can be adapted to diverse RL tasks.

3.1 INTRODUCING INTRINSIC REWARDS

The game rewards described above are received from the environment. What we propose is to add motivation to the agents, adding intrinsic components to the reward function. Intrinsic rewards are received from the agent itself, observing its internal state. These components will be designed based on the motivation theories derived from the literature. Each theory tested here is subjected to deeper exploration in itself, however our intention is not to exhaustively examine each one but to offer a collection of experiments that allow a first observation on how they influence agent behavior. Four agents will be tested, with the following reward modifications:

1. Vanilla agent: will receive only extrinsic rewards, from the environment. Then the reward function is $R = R_e$, where R is the total reward and R_e is the extrinsic reward that follows game score.

2. Hull's drive agent: this agent will be modeled with a hunger drive that follows the level of energy. It will receive extrinsic rewards, but also a feedback from its internal state that means: how hungry it is! The drive level D ranges from 0 to 50, with the homeostasis level set at $D^* = 30$. The energy consumption per time step is -0.5. The reward is now given by $R = R_e + R_i$, where R_i is the intrinsic component. The reward accounts for the motivation by drive reduction. As we want the intrinsic rewards to cooperate in the game task we do not penalize eating over the homeostasis. Accordingly, the Hunger drive is:

$$R_i = \begin{cases} -\left(\frac{D^* - D}{D^*}\right)^2, & D < D^* \\ \left(\frac{D - D^*}{D^*}\right), & D \geq D^* \end{cases}$$

In this formulation, the reward is inside the interval [-1, 1], that follows the same dimension as the clipped extrinsic rewards. The same holds for the following agents.

3. *Like* and *want* agent: this agent will also be modeled with a Hunger drive, serving as *want* system, but adding a *liking* component. This is an attempt at modeling the *want* and *like* dichotomy as presented by Baumeister (2016) and Gudwin (2019), that proposes modeling the *like* system as the intensity of drive reduction. It will receive the extrinsic rewards, and the internal state now will be composed by the *want* and *like* components $R_i = R_{iw} + R_{il}$. The *want* component is defined as the reward designed above, while the *like* component is given by $R_{il} = \frac{D_t - D_{t-1}}{D^*}$. Where D_t is the drive value at the instant t.

4. Incentive Salience agent: The intrinsic component of this agent does not receive a drive, but instead it is modeled in terms of incentive aspects. The internal modulation is divided into a hedonic reward term and a decision-time value modulation term. The *'like'* component is a fixed value, indicating the sensory hedonic pleasure, then $R_{il} = 1$ whenever the agent eats.

Incentive Salience requires dynamic motivational revaluation, which is not naturally supported in standard model-free RL. We therefore introduce a salience modulation layer that re-weights cached action values based on internal motivational state. The *"want"* component is composed of the Incentive Salience value of the Cue intensified by the κ variable that represents the physiological state of the agent (Zhang et al., 2009). Then $Q^{\text{sal}}(s, a) = Q(s, a) + \kappa\, C(s, a)$ being $Q(s, a)$ learned cached value (model-free RL), $C(s, a)$ cue salience signal, κ physiological-dependent amplification factor and $Q^{\text{sal}}(s, a)$ the decision-time motivated value. The Cue value given in a state s, in theory, is learned by Pavlovian association. As a future work we want to add Pavlovian conditioning to the model. In this first stage we implement them deterministically. The Cue value include in a first implementation is given by the number of non-eaten pellets the agent can reach in the next movements in each direction.

The κ value represents the physiological state and results of diverse components, we can imagine for example that stress and sleep deprivation also affects the decision for food. Hunger itself is an element of the physiological state, so after eating food cues are less salient. Then we modulate κ varying accordingly with the level of energy, in a way that the highest value would be when $D = 0$ and the lowest value when $D \geq 30$. This can be formulated as $\kappa = \max\Big(1, 1 + \frac{D^* - D}{D^*}\Big)$.

Each agent will be trained through a million steps starting with five distinct random seeds. After training each agent will be evaluated in 100 trials with learning disabled and varying seeds. Each episode ends when the agent dies.

3.2 EVALUATING EXPERIMENTS

Our objective with the modified agents is not to improve game score, but to observe different behaviors when playing. The agents will be evaluated according to the following metrics, that aim to capture behavioral nuances.

Average return: differences in the average curves of reward against episodes evolution indicates that the motivational components influenced agent behavior. To a better understanding of each component of the reward, extrinsic and intrinsic, they will be recorded separately also in their sub components of *want* and *like*.

Average lifetime: very short lifetimes indicate death, and very long lifetime may point to inefficient strategy. The comparison of average lifetime will indicate the progress of learning and may help to capture risk sensitivity decisions.

Pellet efficiency: the rate of pellets eaten by the number of time steps in the episode indicate how the agent strategically collect pellets.

Ghost-eating efficiency: after eating a power pellet the agent is able to eat ghosts for a short time. So eating a power pellet without ghosts around is not a good strategy. This metric evaluates the level of strategies of the agent.

Danger-zone time: the metric aim to capture how frequently the agent stays near ghosts when not in the frightened mode. High values may indicate a risk behavior, mainly if in short lifetime. Very low values indicate higher risk-sensitivity resulting in not frequently eating ghosts after power pellets.

Backtracking rate: the percentage of steps where agent returns to a recently visited cell. High values may indicate policy instability.

Beyond performance outcomes, observing how each motivation theory impacts behavior may help design motivational models that perform more effectively. We hope to contribute with insights to behavioral neuroscience and also with the design of more human-like and efficient agents. Early feedback from the community will help refine both the modeling assumptions and behavioral evaluation protocol.

ACKNOWLEDGMENTS

MG is funded by Unicamp's vice-rectorate of research. EC is partially funded by CNPq PQ-2 grant (309760/2025-9), AS is partially funded by CNPq PQ-2 grant (312323/2022-0). This project was supported by the Brazilian Ministry of Science, Technology and Innovations, with resources from Law n^o 8,248, of October 23, 1991, within the scope of PPI-SOFTEX, coordinated by Softex and published Arquitetura Cognitiva (Phase 3), DOU 01245.003479/2024-10.

REFERENCES

Patrick Anselme and Mike J. F. Robinson. "Wanting," "liking," and their relation to consciousness. *Journal of Experimental Psychology: Animal Learning and Cognition*, 42(2):123–140, 2016. ISSN 2329-8464, 2329-8456. doi: 10.1037/xan0000090. URL `https://doi.apa.org/doi/10.1037/xan0000090`.

Adrià Puigdomènech Badia, Bilal Piot, Steven Kapturowski, Pablo Sprechmann, Alex Vitvitskyi, Zhaohan Daniel Guo, and Charles Blundell. Agent57: Outperforming the atari human benchmark. In *International conference on machine learning*, pp. 507–517. PMLR, 2020.

Roy F Baumeister. Toward a general theory of motivation: Problems, challenges, opportunities, and the big picture. *Motivation and Emotion*, 40:1–10, 2016.

Marc G Bellemare, Yavar Naddaf, Joel Veness, and Michael Bowling. The arcade learning environment: An evaluation platform for general agents. *Journal of artificial intelligence research*, 47: 253–279, 2013.

Kent C Berridge. Motivation concepts in behavioral neuroscience. *Physiology & behavior*, 81(2): 179–209, 2004.

Kent C Berridge. The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology*, 191:391–431, 2007.

Kent C Berridge. Wanting and liking: Observations from the neuroscience and psychology laboratory. *Inquiry*, 52(4):378–398, 2009.

Kent C Berridge and Terry E Robinson. What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain research reviews*, 28(3):309–369, 1998.

Letícia Berto, Paula Costa, Alexandre Simões, Ricardo Gudwin, and Esther Colombini. A motivational-based learning model for mobile robots. *Cognitive Systems Research*, 88:101278, 2024.

Dalbir Bindra. How adaptive behavior is produced: a perceptual-motivational alternative to response reinforcements. *Behavioral and brain sciences*, 1(1):41–52, 1978.

Ricardo R Gudwin. A review of motivational systems and emotions in cognitive architectures and systems. *Artificial Intelligence: 5th RAAI Summer School, Dolgoprudny, Russia, July 4–7, 2019, Tutorial Lectures*, pp. 65–84, 2019.

Clark Leonard Hull. Principles of behavior: An introduction to behavior theory. 1959.

Abraham Harold Maslow. A theory of human motivation. *Psychological review*, 50(4):370, 1943.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.

Eva Pool, Vanessa Sennwald, Sylvain Delplanque, Tobias Brosch, and David Sander. Measuring wanting and liking from animals to humans: A systematic review. *Neurosci Biobehav Rev*, 63: 124–142, 2016.

Mike James F Robinson, Agneta M Fischer, Aarit Ahuja, Ellen N Lesser, and Hannah Maniates. Roles of "wanting" and "liking" in motivating behavior: gambling, food, and drug addictions. *Behavioral neuroscience of motivation*, pp. 105–136, 2016.

Terry E Robinson and Kent C Berridge. The neural basis of drug craving: an incentive-sensitization theory of addiction. *Brain research reviews*, 18(3):247–291, 1993.

Terry E Robinson and Kent C Berridge. The incentive-sensitization theory of addiction 30 years on. *Annual review of psychology*, 76, 2025.

Ron Sun. Motivational representations within a computational cognitive architecture. *Cognitive Computation*, 1:91–103, 2009.

Richard S. Sutton and Andrew G. Barto. Psychology. In *Reinforcement Learning: An Introduction*, chapter 14. MIT Press, Cambridge, MA, 2nd edition, 2018.

Frederick Toates. Comparing motivational systems—an incentive motivation perspective. 1994.

Piotr Winkielman, Kent C Berridge, and Julia L Wilbarger. Unconscious affective reactions to masked happy versus angry faces influence consumption behavior and judgments of value. *Personality and social psychology bulletin*, 31(1):121–135, 2005.

Jun Zhang, Kent C Berridge, Amy J Tindell, Kyle S Smith, and J Wayne Aldridge. A neural computational model of incentive salience. *PLoS computational biology*, 5(7):e1000437, 2009.