Specific Curiosity is a Holistic Pursuit

Nadia M. Ady
Department of Computing Science, University of Alberta
Alberta Machine Intelligence Institute (Amii)
Edmonton, AB, Canada
nmady@ualberta.ca

Abstract

The development of machine analogs of *specific curiosity*, intrinsic motivation to learn something specific, has the potential to strongly benefit autonomous agents. Earlier work by Ady et al. (2022) demonstrated how three of these properties might be implemented together in a reinforcement learning agent. In this work, we highlight how the behaviour of that agent deteriorates when any one of the included properties is ablated, providing initial evidence for the interconnected nature of the properties of specific curiosity—effective learning behaviour isn't achieved via one or two properties; the properties work together.

Curiosity has long been intertwined with our understanding of intrinsic motivation, since even intrinsic motivation's earliest formal conceptualizations (Deci, 1975, p. 25). Curiosity shapes the learning of humans and animals, motivating behaviour that allows individuals to autonomously develop coherent knowledge about their world (Ady et al., 2022, pp. 15–17). There is an opportunity for machine analogs of curiosity to similarly shape open-ended lifelong learning in autonomous agents and robots.

When a human experiences curiosity, they are usually curious to know *something*, and are satisfied when they learn that thing. Humans take advantage of their learned knowledge of how the world works to learn something specific, answering a question that a curiosity-inducing situation led them to ask. This motivation is sometimes called *specific curiosity* (McNary, 2023, p. 7).

In prior work, Ady et al. (2022) have argued for the existence and importance of five key properties of specific curiosity: directedness towards inostensible referents, cessation when satisfied, voluntary exposure, transience, and coherent long-term learning (pp. 9–17). We further demonstrated the feasibility of three of these properties in a reinforcement learning prototype agent (*ibid.*, p. 26–38). In this abstract, we will focus on presenting evidence that the first three properties in that list work together to generate behaviour characteristic of specific curiosity. We will employ results from an ablation study with the prototype agent.

The original experiments, described in detail by Ady et al. (2022), used the five key properties as guidelines to generate machine behaviour that partially approximates the specific curiosity of animal learners. We make use of the same experiment setup here and summarize it below.

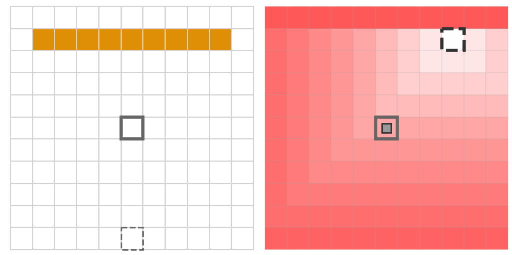

Figure 1: **Left:** Domain used for the experiments. The orange boxes are the potential curiosity-satisfying locations. The heavily-outlined box is the curiosity-inducing location. If the agent leaves the top row, it teleports to the dashed box in the bottom row. **Right:** An example temporary value function, V_{curious}. The curiosity-satisfying location has a heavy dashed outline and a value of zero. Magnitudes of negative values are shown using shades of red.

Experiment domain: The experiments use a grid world shown in Fig. 1. At a given time, the learner is located in one square. The learner has a choice of actions to move to adjacent squares: stay-put, sideward, or upward actions (including diagonals). Any upward action from the top row of the grid teleports the agent to the middle of the bottom row. There is no extrinsic reward.

The properties as implemented are briefly described below. Ady et al. (2022) offer more detail.

Directedness towards inostensible referents In humans, when curiosity is induced, the learner recognizes there is something they do not know, sometimes called an *information gap* (Loewenstein, 1994) or an *inostensible concept* (Inan, 2012). Moreover, humans can roughly specify properties of observations that would rectify such a gap, and even suggest a sequence of actions intended to lead to such observations. Once a learner's curiosity has been induced, the appropriate behaviour is to directly follow such a sequence.

The experiments explored here do not tackle what determines if a situation is curiosity-inducing or how an agent predicts what will satisfy their resulting curiosity. In this case study, the centre of the grid is hard-coded as a curiosity-inducing location: when the agent visits it, a curiosity-satisfying target is generated as a randomly selected locations from a predetermined subset (see Fig. 1). This location

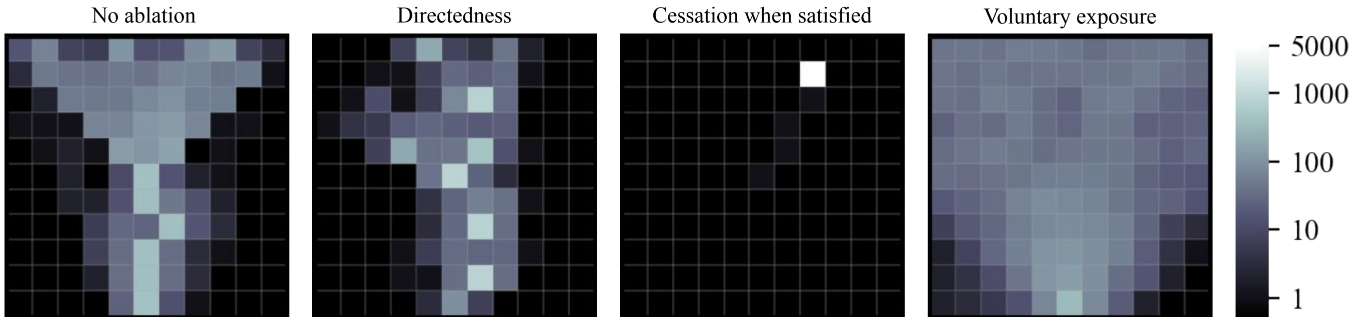

Figure 2: Visit counts for a single 5000-step trial, for the properties in combination (a) and the ablations (b-d). (b) When directedness towards inostensible referents is ablated by having the agent follow only its persistent value function even when curious, states visited by the agent accrue persistent value. The agent initially behaves mostly randomly, so the increasing value causes the agent to revisit these arbitrary states. (c) When cessation when satisfied is ablated by leaving the temporary value function in place, the agent never leaves its first satisfying location. (d) When voluntary exposure is ablated by removing the TD modification (Eq. 1), the learner only finds the curiosity-inducing location by chance and rarely benefits from curiosity, spending most time in a random walk.

is provided to the agent. The agent is also provided with a model of the environment so it can generate a non-positive temporary value function (Fig. 1) that guides it directly to that target.

Cessation when satisfied Curiosity is experienced by the agent as binary: either in effect or not. Directedness towards inostensible referents is temporary and should not persist when curiosity is not in effect. Once a learner has found a situation that satisfies their curiosity, they do not need to experience that same situation again, so curiosity's effect ends. Examples documenting the satisfiability of curiosity are prevalent in the literature, for example by Wiggin, Reimann, and Jain (2019, p. 1194). In our design, the learner follows a temporary value function while curious and uses a separate persistent value function to capture long-term preferences. The temporary value function is different each time curiosity is induced and it always leads the learner directly to the new curiosity-satisfying location. Once the learner visits the target, the temporary value function is zeroed out and the agent acts ϵ-greedily (Sutton and Barto, 2018, pp. 27–28) with respect to its persistent value function until curiosity is next induced.

Voluntary exposure Choosing to partake in activities likely to induce curiosity—for example, picking up puzzles or mysteries—has been called voluntary exposure (Loewenstein, 1994, p. 84). Though curiosity should cease when satisfied, learners should develop enduring preferences for curiosity-inducing situations, following the property of voluntary exposure. To create an enduring preference in the persistent value function, V, we make a small modification to the standard temporal-difference (TD) learning algorithm (Sutton and Barto, 2018, p. 120), modifying the TD error update:

$$\delta \leftarrow R + \gamma V(x') - [\boldsymbol{V(x)} + \boldsymbol{V_{\text{curious}}(x)}] \tag{1}$$

where V_{curious} refers to the temporary value function, which should be non-positive everywhere (as in Fig. 1) so V accrues positive value.

Ablations With these three properties implemented in combination, the learner's behaviour has important characteristics of specific curiosity: when curious, the learner rapidly satisfies its curiosity, spending minimal time in curiosity-satisfying locations, and instead learning a direct path to the curiosity-inducing location (voluntary exposure). Further detail on the learner's developing preference (value function) is provided by Ady et al. (2022, pp. 34–38).

In the ablation studies, we removed each design element associated with a key property in turn and observed the resulting behaviour. The resulting visit behaviour in the grid is shown in Fig. 2 and briefly described in the caption.

In the context of this prototype agent, the properties of directedness, cessation when satisfied, and voluntary exposure appear to work together; curious behaviour is noticeably impaired when any one property is missing. Our ablation study provides initial evidence for the interconnected nature of the properties of specific curiosity—effective learning behaviour isn't achieved via one or two properties; the properties work together. Indeed, the benefits of each property are so interwoven that they are best understood via their combined influence on the whole of specific curiosity.

References

Ady, N. M.; Shariff, R.; Günther, J.; and Pilarski, P. M. 2022. Five properties of specific curiosity you didn't know curious machines should have. *arXiv:1703.01732*.

Deci, E. L. 1975. *Intrinsic Motivation*. Springer.

Inan, I. 2012. *The Philosophy of Curiosity*. Routledge.

Loewenstein, G. 1994. The psychology of curiosity: A review and reinterpretation. *Psychol. Bull.* 116(1):75–98.

McNary, L. 2023. Curiosity: A conceptual re-analysis for improved measurement. *Current Psychology*.

Sutton, R. S., and Barto, A. G. 2018. *Reinforcement Learning: An Introduction*. MIT Press, second edition.

Wiggin, K. L.; Reimann, M.; and Jain, S. P. 2019. Curiosity Tempts Indulgence. *J. Consum. Res.* 45(6):1194–1212.