The Creative Act: Effective Exploration by Seeking Surprise

Anonymous Author(s)
Affiliation
Address
email

Abstract

That c creates p is a generative relation, resulting from exploration, between a creator c and a possible generative action p where c generates by exploring then recognizing and revealing the possible action p that is surprising, and that we can learn something about the space of possible actions from. This focus on creativity as action allows me to emphasize it's relational nature, rather than trying to separately define what a creative product is or what the creative mental process is independently.

1 Exploration

Exploration is something an agent does in order to expand their knowledge of possible actions and their values [Aronowitz, 2021, 340]. Exploration trades off with other actions that are exploitative: those that involve actualizing the possibility that you believe will have highest value based on what you already know. Within this framework, I argue that creative people are the most effective explorers – they spread their attentional resources into the unknown or otherwise non-salient [Gross et al., 2024, Benedek and Fink, 2019], and discover highly surprising but possible actions that allow us to better understand and manipulate our environment.

If creativity is a kind of effective exploration, and exploration involves seeking the unknown, we are left with the following question about creative search optimization:

> How do creative people optimize their search for creative products?

We cannot use the typical means of search such as Bayesian Hierarchical Search Theory (BHST) because it is "...assumed that there is a prior distribution for the target's location which is known to the searcher..." [Stone, 1975, 1]. So creative people explore unknowns and frequently find fruit; how?

2 Novelty

Novelty is often taken to be the defining or at least 'primary' feature of the creative product [Kant, 2000, 180, 5:308][Boden, 2009, Stokes, 2011, Brainard, 2023, Nanay, 2014]. Given this, novelty is presumably supposed to help us identify creative products. However: "...there can also be original nonsense..." [Kant, 2000, 180, 5:308]. Worse, in some sense, every perception (even an exploitative one) is a little bit novel - it occurs at a different time. So, novelty can't help us narrow our exploratory search.

Mu et al. have suggested (following others) that endowing machines with semantic abstractions (human semantic concepts) can aid in meaningful exploration [Mu et al., 2022, Tam et al., 2022]. If we understand novelty like these RL mechanisms do, then maybe novel objects are just objects associated

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

with novel concepts (observations with novel labels), rather than novel perceptions. However, we cannot treat novelty like the novelty at work in these RL mechanisms. The RL machine benefits from our concepts because, it gets provided with a a fixed set of everyday linguistic concepts from a language oracle. We do not have a fixed set of concepts; I could theoretically label every perception with a concept: <perception p1 at t1>, <perception p2 at t2>... etc.. Worse, every perception could receive multiple conceptual labels: we could overspecify, underspecify, or idiosyncratically specify [Wilson, 1982, Wittgenstein, 1953]. How do we know whether something is really novel, rather than just mislabelled or differently labelled? Although everyday is new, it does not feel new to me – I do not always recognize its newness because I am not always parsing the world with concepts that would highlight its newness (even if I theoretically could).

3 Surprise

The solution is to focus on surprise instead of novelty [Barto et al., 2013, Gonzalez and Haselager, 2005]; prediction instead of direct perception. According to Schmidhuber RL agents should be encouraged to seek only novel stimulations that are surprising and compressible [Schmidhuber, 2010]. Schmidhuber explicitly rejects traditional information-theoretic notions of surprise (like statistical unlikelihood) that would count a static filled t.v. as constantly surprising because the information provided by the t.v. is uncompressible [Schmidhuber, 2010, 9]. Instead, surprising data is data with patterns. A pattern for Schmidhuber exists over some data if there "... exists an algorithm that is significantly shorter than the raw data but is able to encode it without loss of information." [Schmidhuber, 2010, 4]

Of course, creativity can't just be about learning patterns - we do that all the time. It has to be learning sparked by something on the part of the agent, by a physical or mental action; an exploration. This is why we should focus on creativity from an action theoretic perspective, as an *act*. Furthermore, Boden distinguishes between psychological and historical creativity [Boden, 2004, 2, 43-44]; A product p is historically creative iff it would be surprising compressible information relative to human history (taken as an aggregation of human psychological states). So people can be creative by seeking surprise.

Interestingly, creative people receive less intrinsic reward from surprising content [Gross et al., 2024]. Perhaps they seek it more often [Schelling, 1988, Nagel]. Furthermore, brilliant or creative moves in chess are those that elude weaker players, and seem to be bad moves earlier on in the game [Zaidi and Guerzhoy]. In other words, they're surprisingly good (unexpected).

4 Further Work and Open Questions

There is obviously more to be done. What is it about the *creative* surprising moves that makes them creative rather than normally surprising?

1. Kant claims that creative products are exemplars. They propogate themselves and allow for substantial future work. How might we cash this out in terms of the information value of creative products and acts?

2. Marta Halina suggests that AI creativity results in domain specific knowledge, while human creativity results in domain general knowledge [Halina, 2021]. How do we cash out domain generality informationally?

3. What does framing creativity as a kind of exploration mean for the relationship between creativity and stochasticity? One can randomly explore and get surprising results? Does this mean machines can be creative by taking stochastic actions?

4. What does linking creativity to exploration mean for the relationship between creativity and attention? Do creatives engage in more exploratory forms of attention?

5. What's the relationship between creativity and expertise [Gaut, 2012] that emerges from this view?

References

Sara Aronowitz. Exploring by Believing. *The Philosophical Review*, 130(3):339–383, July 2021. ISSN 0031-8108. doi: 10.1215/00318108-8998825. URL https://doi.org/10.1215/00318108-8998825.

Andrew Barto, Marco Mirolli, and Gianluca Baldassarre. Novelty or Surprise? *Frontiers in Psychology*, 4, 2013. ISSN 1664-1078. URL https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00907.

Mathias Benedek and Andreas Fink. Toward a neurocognitive framework of creative cognition: the role of memory, attention, and cognitive control. *Current opinion in behavioral sciences*, 27: 116–122, 2019. ISSN 2352-1546. doi: 10.1016/j.cobeha.2018.11.002.

M. A. Boden. Chapter Thirteen. Creativity: How Does It Work? In *The Idea of Creativity*, pages 235–250. Brill, January 2009. ISBN 978-90-474-2790-2. URL https://brill.com/display/book/edcoll/9789047427902/Bej.9789004174443.i-348_014.xml.

Margaret A. Boden. *The Creative Mind: Myths and Mechanisms*. Psychology Press, 2004. ISBN 978-0-415-31452-7.

Lindsay Brainard. The curious case of uncurious creation. *Inquiry*, pages 1–31, September 2023. ISSN 0020-174X, 1502-3923. doi: 10.1080/0020174X.2023.2261503. URL https://www.tandfonline.com/doi/full/10.1080/0020174X.2023.2261503.

B. Gaut. Creativity and Art: Three Roads to Surprise. *The British Journal of Aesthetics*, 52(1): 116–119, January 2012. ISSN 0007-0904, 1468-2842. doi: 10.1093/aesthj/ayr041. URL https://academic.oup.com/bjaesthetics/article-lookup/doi/10.1093/aesthj/ayr041.

Maria Eunice Quilici Gonzalez and Willem (Pim) Ferdinand Gerardus Haselager. Creativity: Surprise and Abductive Reasoning. *Semiotica*, 2005(153):325–342, February 2005. ISSN 1613-3692. doi: 10.1515/semi.2005.2005.153-1-4.325. URL https://www.degruyter.com/document/doi/10.1515/semi.2005.2005.153-1-4.325/html.

Madeleine E. Gross, James C. Elliott, and Jonathan W. Schooler. Why creatives don't find the oddball odd: Neural and psychological evidence for atypical salience processing. *Brain and Cognition*, 178:106178, August 2024. ISSN 02782626. doi: 10.1016/j.bandc.2024.106178. URL https://linkinghub.elsevier.com/retrieve/pii/S0278262624000551.

Marta Halina. Insightful artificial intelligence. *Mind & Language*, 36(2):315–329, 2021. ISSN 1468-0017. doi: 10.1111/mila.12321. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/mila.12321.

Immanuel Kant. *Critique of the Power of Judgment*. Cambridge University Press, 2000.

Jesse Mu, Victor Zhong, Roberta Raileanu, Minqi Jiang, Noah Goodman, Tim Rocktäschel, and Edward Grefenstette. Improving Intrinsic Exploration with Language Abstractions, February 2022. URL http://arxiv.org/abs/2202.08938.

Jennifer Nagel. Natural Curiosity. In Artūrs Logins and Jacques-Henri Vollet, editors, *Putting Knowledge to Work: New Directions for Knowledge-First Epistemology*. Oxford University Press. URL https://philarchive.org/rec/NAGNCZ-2.

Bence Nanay. An Experiential Account of Creativity. In Elliot Samuel Paul and Scott Barry Kaufman, editors, *The Philosophy of Creativity: New Essays*, page 0. Oxford University Press, May 2014. ISBN 978-0-19-983696-3. doi: 10.1093/acprof:oso/9780199836963.003.0002. URL https://doi.org/10.1093/acprof:oso/9780199836963.003.0002.

T. C. Schelling. THE MIND AS A CONSUMING ORGAN. In Amos Tversky, David E. Bell, and Howard Raiffa, editors, *Decision Making: Descriptive, Normative, and Prescriptive Interactions*, pages 343–357. Cambridge University Press, Cambridge, 1988. ISBN 978-0-521-36851-3. doi: 10.1017/CBO9780511598951.017. URL https://www.cambridge.org/core/books/decision-making/mind-as-a-consuming-organ/DBAA2ED740FD03A60E35D6E6A2FF7182.

Jürgen Schmidhuber. Formal Theory of Creativity, Fun, and Intrinsic Motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development*, 2(3):230–247, September 2010. ISSN 1943-0604, 1943-0612. doi: 10.1109/TAMD.2010.2056368. URL http://ieeexplore.ieee.org/document/5508364/.

Dustin Stokes. Minimally Creative Thought. *Metaphilosophy*, 42(5):658–681, 2011. ISSN 1467-9973. doi: 10.1111/j.1467-9973.2011.01716.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9973.2011.01716.x.

Lawrence D. Stone. *Theory of optimal search*. Number v. 118 in Mathematics in science and engineering. Academic Press, New York, 1975. ISBN 978-0-12-672450-9.

Allison C. Tam, Neil C. Rabinowitz, Andrew K. Lampinen, Nicholas A. Roy, Stephanie C. Y. Chan, D. J. Strouse, Jane X. Wang, Andrea Banino, and Felix Hill. Semantic Exploration from Language Abstractions and Pretrained Representations, May 2022. URL http://arxiv.org/abs/2204.05080.

Mark Wilson. Predicate Meets Property. *The Philosophical Review*, 91(4):549–589, 1982. ISSN 0031-8108. doi: 10.2307/2184801. URL https://www.jstor.org/stable/2184801.

Ludwig Wittgenstein. *Philosophical Investigations*. Basil Blackwell, Oxford, 1953.

Kamron Zaidi and Michael Guerzhoy. Predicting User Perception of Move Brilliance in Chess.

NeurIPS Paper Checklist

You should answer [Yes] , [No] , or [NA] .

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The claims made in the abstract are simply a summary of the claims made in the short paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: It's short. I point out some remaining questions.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Justification: Nothing that requires 'proofs' in the usual sense.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [NA]

   Justification: No experiment was performed.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [NA]

   Justification: the paper does not include experiments requiring code

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [NA]

   Justification: the paper does not include experiments.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

   Justification: the paper does not include experiments.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [NA]

   Justification: the paper does not include experiments.

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: the work is theoretical, I perform no experiments that require adherance to the code of ethics, nor do I produce models that are potentially harmful.

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

    Justification: This is a really short paper. It's hard to add everything. I think learning more about creativity can make us more creative people, and can also help us better understand AI. Those are both social impacts this work might have.

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: the paper poses no such risks.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [NA]

    Justification: the paper does not use existing assets.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: the paper does not release new assets.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: the paper does not involve crowdsourcing nor research with human subjects.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: the paper does not involve crowdsourcing nor research with human subjects.