Abductive Computational Systems: Creative Abduction and Future Directions

Abhinav Sood
University of Sydney
Sydney, Australia
abhinav.sood@sydney.edu.au

Kazjon Grace
University of Sydney
Sydney, Australia
kazjon.grace@sydney.edu.au

Stephen Wan
Data61, CSIRO
Sydney, Australia
stephen.wan@data61.csiro.au

Cecile Paris
Data61, CSIRO
Sydney, Australia
cecile.paris@data61.csiro.au

Abstract

Abductive reasoning, reasoning for inferring explanations for observations, is often mentioned in scientific, design-related and artistic contexts, but its understanding varies across these domains. This paper reviews how abductive reasoning is discussed in epistemology, science and design, and then analyses how various computational systems use abductive reasoning. Our analysis shows that neither theoretical accounts nor computational implementations of abductive reasoning adequately address generating creative hypotheses. Theoretical frameworks do not provide a straightforward model for generating creative abductive hypotheses, and computational systems largely implement syllogistic forms of abductive reasoning. We break down abductive computational systems into components and conclude by identifying specific directions for future research that could advance the state of creative abductive reasoning in computational systems.

Introduction

Abductive reasoning is a form of logical reasoning that deals with explanations for hypotheses, conjectures, anomalies, etc. For example, consider a person who returns to their house to find their belongings scattered around, along with a broken window. If they had left the house clean, they could abduce that a burglar might have broken into their home while they were out. Charles S. Peirce first formulated this form of reasoning as (Peirce 1958):

1. A surprising fact O is observed.
2. But if [hypothesis] H were true, O would follow.
3. Thus, there is a reason to suspect that H is True.

Unlike deductive reasoning, abductive reasoning does not aim to preserve the truth of the derived propositions. Instead, pragmatically, abductive reasoning seeks to explain surprising observations where other forms of reasoning might not be possible. In a computational creativity context, this contrasts with how surprise typically provides an evaluation of creativity (Grace and Maher 2014). Instead, abduction posits that surprises are an initiating event: they are the unexplainable observations for which abduction provides an explanation. This ability to generate plausible explanations in the face of surprise makes abductive reasoning particularly valuable in both scientific, design-related, and artistic contexts: scientists rely on it to formulate hypotheses that might explain unexpected phenomena, while designers use it to envision solutions that address observed needs without complete information about the problem space. Similarly, artists employ abductive reasoning when interpreting ambiguous visual elements and unexpected results, using these surprises as catalysts for developing artistic expressions.

Thus, understanding how computational systems use abductive reasoning can help us identify properties that may increase their utility. To the best of our knowledge, no existing review examines how abductive reasoning is implemented across computational systems in these diverse fields. The main contributions of this paper are:

1. We review and connect literature on abductive reasoning that exists across different, often unlinked fields.
2. We analyse how computational systems implement abductive reasoning and propose a breakdown that identifies simple components for advancing creative abduction.

Abductive Reasoning in Theory

Epistemological Background

Since Pierce's formulation of abduction, several philosophers have further developed the concept of abductive reasoning. Gabbay and Woods (2005) discuss the ignorance-preserving nature of abduction. Ignorance-preserving means that abductive reasoning does not generate a hypothesis that completely resolves the surprising observation. Instead, uncertainty remains about whether the hypothesis is true even after a hypothesis is produced. So, what use does the hypothesis provide? It is a reasoned basis for future action (Gabbay and Woods 2006).

By abductive reasoning, we are not referring to inference to the best explanation (IBE). While abduction is just the reasoning process to create and select a possible hypothesis, IBE is an iterative process that might involve deductive, abductive and inductive reasoning to select and arrive at the best hypothesis available. A model for IBE in diagnostic reasoning is presented in (Magnani 2001, 23). This model uses abduction to propose diagnostic hypotheses based on clinical evidence. Observed clinical data corroborate hypotheses through induction, and the deductive implications

of the hypothesis can provide us with expectations of what our data should look like. These expectations form the basis for future surprises that trigger further abduction.

There are several formalisations of abductive reasoning. The Gabbay Woods Schema (GW-Schema) (Gabbay and Woods 2006, pg. 198) aims to capture several properties of Peircean abduction. Although the GW-Schema illustrates the underlying syllogistic logic of abductive reasoning, it suffers from two major problems. Firstly, it does not explain how the hypotheses themselves are generated. Second, it does not concretely define the conditions that a hypothesis needs to meet to be considered worthy of investigation. Some of the conditions often discussed in the literature (Douven 2021) are simplicity, coherence and explanatory power. However, no consensus exists on whether these specific conditions are necessary. For example, in (Woods 2017, pg. 139), Woods rejects simplicity and "sees no reason why truth favours the uncomplicated". The issues of coming up with a hypothesis and of selecting a hypothesis among candidates based on specific conditions are referred to as the fill-up and cut-down problems (Magnani 2017, pg. 6) respectively. Magnani discusses how the solutions to these problems are "stunningly contextual" (Magnani 2017, pg. 10). While his proposed eco-cognitive model provides us with an understanding of abductive reasoning in more creative, scientific contexts, it does not provide a computable formal structure.

When looking for computable formal structures, some might contest that Abductive Logic Programming (ALP) (Kakas, Kowalski, and Toni 1992) can fill the gap. ALP extends standard logic programming by introducing abducible predicates. These frameworks allow systems to derive explanations by backward-chaining from observations to possible causes, constrained by integrity constraints that filter out implausible explanations. ALP has been shown to be effective for diagnostic reasoning (Koitz-Hristov and Wotawa 2018), but typically operates in closed, well-defined domains and struggles with more open-ended abduction.

The differing theoretical approaches to abductive reasoning have led Woods to argue that "The foundational work for a comprehensive account of abductive reasoning still awaits completion" (Woods 2017, pg. 138). Unfortunately, this means that the epistemological literature on abductive reasoning does not sufficiently address how you can functionally perform the more complex forms of abductive reasoning. As such, it might be worthwhile to look at how creative abductive reasoning emerges in two domains where creativity and hypotheses are central: science and design.

Creative Abductive Reasoning in Science and Design

In the philosophy of science, abductive reasoning is traditionally associated with creative processes in the context of discovery (Schickore 2022). Here, the "context of discovery" refers to the initial generation of theories in science through the conception of hypotheses and research ideas. Logically abductive reasoning deals with producing explanations for surprising observations. We propose that this process of producing explanations is relevant throughout the scientific processes at two distinct levels.

In the "context of discovery", it is relevant macroscopically; abductive reasoning serves as a guiding framework for the entire scientific process, shaping what hypotheses and ideas are investigated. Microscopically, abductive reasoning is used within individual experimental stages of the scientific process. To illustrate this, consider the documentation of the discovery of the urea cycle in (Kulkarni and Simon 1988). Macroscopically, after observing the unexpected effect of ornithine in enhancing urea production, Krebs abductively reasoned that ornithine might play a central role in a cyclical process of urea formation. This explanation accounted for the surprising observation and led him to restructure his entire research program around understanding ornithine's role in urea synthesis. Note that he "did not have a clear hypothesis of a mechanism to account for it [ornithine effect]" that he was testing (Kulkarni and Simon 1988). Instead, a rougher hypothesis was sufficient to guide his investigation. Microscopically, Krebs employed abductive reasoning at specific experimental junctures. When he discovered that the effect was unique to ornithine (as chemical derivatives failed to produce similar results), he generated the explanatory hypothesis that ornithine possessed specific structural properties essential to urea formation.

In design, abduction takes on different characteristics when compared to abductive reasoning in science. Design theorists have identified abduction as the fundamental reasoning pattern for moving from function to form, from "what is needed" to "how to do it" (Kroll and Koskela 2015). Roozenburg (1993) made a distinction between explanatory abduction (explaining existing phenomena) and innovative abduction (creating new solutions), and argued that design is primarily the latter. This innovative form of abduction in design is different from scientific abduction as it aims to create entirely new artefacts to fulfil desired functions. Koskela, Paavola, and Kroll (2018) further elaborated on this view, saying that abduction in design is not just an inference type but a property of many inferences throughout the design process. Using a SAPPhIRE-based model of abductive reasoning in design, Bhatt, Majumder, and Chakrabarti (2021) provide a framework to explain how designers use abductive reasoning to bridge the gap between desired functions and proposed solutions. These models are useful to understand the nature of abduction, but can they be automated? To further understand how abductive reasoning takes place practically, we look at how computational systems implement abductive reasoning. Just as (Boden 2004, pg. 17) argues that computational ideas can inform our understanding of human creativity; computational ideas can inform our understanding of abductive reasoning.

Abductive Computational Systems

Natural Language Processing (NLP)

For modern computational systems, recent investigations of abductive reasoning have largely focused on introducing it in Large Language Models (LLMs). Large Language Models (LLMs) are neural networks that generally utilise the trans-

former architecture (Vaswani et al. 2017) to model natural language probabilistically. One of the first papers to investigate abductive reasoning with LLMs was (Bhagavatula et al. 2020). They considered how well LLMs perform abductive reasoning in everyday situations derived from datasets of short stories. The authors created two tasks to test how well LLMs performed. Alpha NLI tested if LLMs could select the correct abductive hypothesis (given a choice of 2) that fits observations that occur before and after the hypothesis. The language model they trained had an accuracy of 68.9%, with humans identifying the correct hypothesis in 91.4% of cases. In our experiments, we found that LLaMa 3.1 70B-Instruct [1] was able to achieve an accuracy of 86.2% with a zero-shot prompt on the test set of the Alpha NLI task, indicating that modern LLMs might perform well in identifying correct abductive hypotheses in everyday situations. These results are merely indicative, as there might be data leakage from the test set, but they point towards the possibility of effective abductive reasoning through LLMs. Beyond everyday situations, Zhao et al. (2024) train LLMs to generate uncommon explanations in everyday situations. Tian et al. (2024) constructs MacGyver-like creative problems to train LLMs. These studies focus on extending the abductive capabilities of LLMs by creating datasets for specific tasks. Such datasets that exclusively focus on abductive reasoning are not yet prevalent in science and design.

Many abductive systems in NLP also focus on knowledge bases. AbductionRules (Young et al. 2022) tests how well transformers reason over logical knowledge bases. Bai et al. (2024) use knowledge graphs to generate hypothesis-observation pairs. There is also research on hypothesis generation in NLP that does not explicitly mention abductive reasoning. These systems have their own host of issues.

In the case of (Si, Yang, and Hashimoto 2025), several limitations of the AI-generated research ideas (inappropriate baselines, unrealistic assumptions, etc.) were pointed out, which brings into question whether the novel ideas generated are usable in practice. With the AI Scientist (Lu et al. 2024), LLMs themselves evaluated the quality of AI-scientist-generated research papers, and such modes of evaluation have been called into question (Koo et al. 2024).

Beyond commonsense reasoning, abductive reasoning has also been used in several computationally creative systems.

Abductive Reasoning in CC Systems

To find computational creativity (CC) systems where abductive reasoning was used, we conducted a review of CC conferences. The start (2010) and end (2024) years for our review mark the first and the most recent ICCC conference. To identify systems that used abductive reasoning, we used the search terms "abductive", "abduce", and "abduction". Among the ICCC proceedings, 14 papers contained the search terms. Of these 8 mentioned abductive for purposes of literature review, argumentation or used the term "abduction" in the physical rather than logical sense (mostly kidnappings in narrative generation papers). We were left with 6 CC systems that either used abductive reasoning computationally or were used to elicit abductive reasoning in the users of the system. These systems are discussed in Table 1.

#	System	Abductive Approach
(1)	The Painting Fool (Colton 2010)	Identifies constraints among partial rectangles, extends patterns to produce abduced versions
(2)	Synthetic Audience (O'Neill and Riedl 2011)	Extracts information from incomplete stories to build a mental model, using abduction to infer character goals
(3)	MILA-S (Goel and Joyner 2015)	Prompts students to engage with abductive explanations through ecological model building.
(4)	ANGELINA (Cook and Colton 2018)	Uses abductive reasoning with answer set programming to generate game rulesets
(5)	Hello (Veiga 2020)	Encourages audience to build narratives through abductive reasoning as the artwork is "incomplete"
(6)	Aris (O'Donoghue et al. 2022)	Creates abduced hypothesised formal specifications using homomorphic matching and topological similarity on knowledge graphs

Table 1: Computational Systems at ICCC that mention Abductive Reasoning

We also reviewed the proceedings of ACM Creativity and Cognition (C&C) and Design Computing and Cognition (DCC) conferences for the same search terms. In our review of C&C and DCC proceedings, we found no computational system implementations: the sole C&C paper used abduction to identify richness (potential to inspire future ideas), inspirational ideas via keywords, while the 11 DCC papers proposed focused on modelling abduction design and were not specifically implemented as CC systems and centered on design theory. We note that our review is not exhaustive. Many computational systems effectively address abductive problems without explicitly framing their approach as abductive reasoning. While these systems may offer valuable insights, their inclusion is beyond the scope of this short paper.

Components of Abductive Computational Systems

Through our review of abductive computational systems, we identify four key components that characterise these systems as visualised in Figure 1.

Abductive Triggers These act as the starting point for performing abduction. Systems (3) and (5) involve using generated artefacts and constructed models, respectively, to encourage abductive reasoning. The other systems do not necessarily possess abductive triggers in a Peircean sense. Instead, they rely on a computational trigger to perform abductive reasoning. For example, (1) conducts abduction when a

[1] https://huggingface.co/meta-llama/Llama-3.1-70B-Instruct

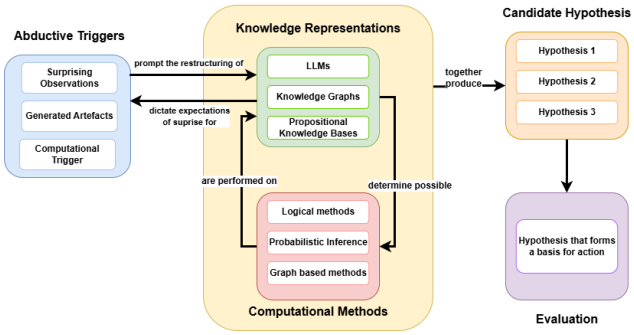

Figure 1: Proposed breakdown for a computational abductive system

user is done inputting a partial configuration of rectangles.

Knowledge Representations In the CC abductive systems we covered, knowledge was encoded as knowledge graphs in (6). In (1), knowledge was construed as constraints that eventually form a constraint satisfaction problem. Abductive reasoning performed in NLP systems we introduced earlier also used knowledge representations. Young et al.(2022) stored knowledge in a propositional form. Bai et al. (2024) used knowledge graphs. In (Bhagavatula et al. 2020), since LLMs were generating the resulting abduced hypothesis, LLMs functioned as black-box stores of knowledge.

Computational Methods :Computational methods perform abductive reasoning in computational systems. These methods heavily depend on the knowledge representation in the CC system. (1) uses extensions of constraints. (4) uses answer set programming to perform abduction. (6) performs abductive reasoning through graph-based methods. The LLM-based systems perform abduction through probabilistic inference.

Evaluating Hypotheses : In the CC systems, the selection process was very specific to the systems. (1) allows users to remove abduced rectangles individually if they are perceived by the user as incorrectly abduced. (6) uses deductive reasoning to verify abduced hypotheses. These steps fall under hypothesis verification, which is a part of IBE but not a part of our definition of abductive reasoning. In (3) and (5), the human performs the evaluation. Few systems actually use abductive properties to evaluate hypotheses. Dalal et al. (2024) used properties like consistency, depth, drift, coherence and linguistic uncertainty to evaluate hypotheses, but the hypotheses are restricted to syllogistic if-then structures.

Future Directions

Our analysis reveals a significant gap between theoretical understandings of abductive reasoning and existing computational implementations. While epistemological accounts emphasise the creative nature of hypothesis generation, computational systems primarily implement restricted, syllogistic forms of abduction. That is, future work should focus on developing systems that can perform creative abduction. One promising direction for this is to create datasets for scientific and design-related abduction, similar to datasets that have already been constructed for commonsense and uncommonsense abductive reasoning. Current systems lack robust methods for evaluating abduced hypotheses, particularly those in natural language form. Developing computationally tractable metrics for properties like simplicity, coherence, and explanatory power could advance our ability to evaluate and compare creative abductive systems.

Our breakdown of computational abductive systems also presents opportunities for the exploration of collaborative and co-creative systems in science. Consider that a literature review of co-creative systems produced 92 systems (Rezwana and Maher 2023); not one was in the domain of science. For abductive triggers, human-machine collaboration could involve the system identifying potentially surprising patterns in data that a human might overlook, while humans could guide the system toward observations that they believe are relevant to scientific questions. For knowledge representations, humans could supply contextual domain knowledge that might be missing from the system's formal representations, while the system could visualise its knowledge structures in ways that reveal unexpected connections.

Conclusion

This paper covered how abductive reasoning is viewed and discussed across epistemology, science, design and computational systems. Our analysis shows that while abductive reasoning is widely held to be key to creative processes, computational implementations of abduction are restricted to syllogistic forms that leave little room for creativity. By breaking down abductive computational systems into abductive triggers, knowledge representations, computational methods and evaluation, we provide a basis to understand and implement more advanced computational creative abduction systems that close the gap between the theoretical understanding and computational implementation of abduction.

Author Contributions

Author 1 (AS) conducted the review and prepared the manuscript. Author 2 (KG), Author 3 (SW) and Author 4 (CP) supervised the research and provided feedback and direction for the development of this work.

Acknowledgement

We gratefully acknowledge the Collaborative Intelligence (CINTEL) Future Science Platform for their support towards this work.

References

Bai, J.; Wang, Y.; Zheng, T.; Guo, Y.; Liu, X.; and Song, Y. 2024. Advancing abductive reasoning in knowledge graphs through complex logical hypothesis generation. In Ku, L.-W.; Martins, A.; and Srikumar, V., eds., *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1312–1329. Bangkok, Thailand: Association for Computational Linguistics.

Bhagavatula, C.; Bras, R. L.; Malaviya, C.; Sakaguchi, K.; Holtzman, A.; Rashkin, H.; Downey, D.; tau Yih, W.; and Choi, Y. 2020. Abductive commonsense reasoning. In *International Conference on Learning Representations*.

Bhatt, A. N.; Majumder, A.; and Chakrabarti, A. 2021. Analyzing the modes of reasoning in design using the sapphire model of causality and the extended integrated model of designing. *Ai Edam* 35(4):384–403.

Boden, M. A. 2004. *The creative mind: Myths and mechanisms: second edition*. Routledge.

Colton, S. 2010. The painting fool teaching interface. In *International Conference on Computational Creativity 2010*, 289.

Cook, M., and Colton, S. 2018. Redesigning computationally creative systems for continuous creation. In *International Conference on Computational Creativity 2018*, 32–39. Association for Computational Creativity (ACC).

Dalal, D.; Valentino, M.; Freitas, A.; and Buitelaar, P. 2024. Inference to the best explanation in large language models. In Ku, L.-W.; Martins, A.; and Srikumar, V., eds., *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 217–235. Bangkok, Thailand: Association for Computational Linguistics.

Douven, I. 2021. Abduction. In Zalta, E. N., ed., *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2021 edition.

Gabbay, D., and Woods, J. 2005. *The Reach of Abduction: Insight and Trial (A Practical Logic of Cognitive Systems, Vol. 2)*. Elsevier.

Gabbay, D., and Woods, J. 2006. Advice on abductive logic. *Logic Journal of IGPL* 14(2):189–219.

Goel, A., and Joyner, D. 2015. Impact of a creativity support tool on student learning about scientific discovery processes. In *Proceedings of the Sixth International Conference on Computational Creativity*.

Grace, K., and Maher, M. L. 2014. What to expect when you're expecting: The role of unexpectedness in computationally evaluating creativity. In *ICCC*, 120–128. Ljubljana.

Kakas, A. C.; Kowalski, R. A.; and Toni, F. 1992. Abductive logic programming. *Journal of logic and computation* 2(6):719–770.

Koitz-Hristov, R., and Wotawa, F. 2018. Applying algorithm selection to abductive diagnostic reasoning. *Applied Intelligence* 48(11):3976–3994.

Koo, R.; Lee, M.; Raheja, V.; Park, J. I.; Kim, Z. M.; and Kang, D. 2024. Benchmarking cognitive biases in large language models as evaluators. In *Findings of the Association for Computational Linguistics ACL 2024*, 517–545.

Koskela, L.; Paavola, S.; and Kroll, E. 2018. The role of abduction in production of new ideas in design. *Advancements in the Philosophy of Design* 153–183.

Kroll, E., and Koskela, L. 2015. On abduction in design. In *Design Computing and Cognition'14*, 327–344. Springer.

Kulkarni, D., and Simon, H. A. 1988. The processes of scientific discovery: The strategy of experimentation. *Cognitive science* 12(2):139–175.

Lu, C.; Lu, C.; Lange, R. T.; Foerster, J.; Clune, J.; and Ha, D. 2024. The ai scientist: Towards fully automated open-ended scientific discovery.

Magnani, L. 2001. *Abduction, Reason and Science*. Springer US.

Magnani, L. 2017. *The Abductive Structure of Scientific Creativity: An Essay on the Ecology of Cognition*, volume 37 of *Studies in Applied Philosophy, Epistemology and Rational Ethics*. Springer International Publishing.

O'Donoghue, D. P.; Brady, C.; Smullen, S.; and Crowe, G. 2022. Novelty assurance by a cognitively inspired analogy approach to uncover hidden similarity. In *ICCC*, 277–281.

O'Neill, B., and Riedl, M. O. 2011. Simulating the everyday creativity of readers. In *International Conference on Computational Creativity 2011*, 153–158.

Peirce, C. S. 1958. *Collected papers*, volume 5. Belknap Press of Harvard University Press.

Rezwana, J., and Maher, M. L. 2023. Designing creative ai partners with cofi: A framework for modeling interaction in human-ai co-creative systems. *ACM Trans. Comput.-Hum. Interact.* 30(5).

Roozenburg, N. F. 1993. On the pattern of reasoning in innovative design. *Design studies* 14(1):4–18.

Schickore, J. 2022. Scientific Discovery. In Zalta, E. N., and Nodelman, U., eds., *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2022 edition.

Si, C.; Yang, D.; and Hashimoto, T. 2025. Can LLMs generate novel research ideas? a large-scale human study with 100+ NLP researchers. In *The Thirteenth International Conference on Learning Representations*.

Tian, Y.; Ravichander, A.; Qin, L.; Le Bras, R.; Marjieh, R.; Peng, N.; Choi, Y.; Griffiths, T. L.; and Brahman, F. 2024. Macgyver: Are large language models creative problem solvers? In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, 5303–5324.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, 6000–6010. Red Hook, NY, USA: Curran Associates Inc.

Veiga, P. A. d. 2020. Hello, an interactive cinematic generative artwork. In *Proceedings of the Eleventh International Conference on Computational Creativity*, 469–475. Association for Computational Creativity (ACC).

Woods, J. 2017. Reorienting the logic of abduction. In Magnani, L., and Bertolotti, T., eds., *Springer Handbook of Model-Based Science*. Cham: Springer International Publishing. 137–150.

Young, N.; Bao, Q.; Bensemann, J.; and Witbrock, M. 2022. AbductionRules: Training transformers to explain unexpected inputs. In Muresan, S.; Nakov, P.; and Villavicencio, A., eds., *Findings of the Association for Computational Linguistics: ACL 2022*, 218–227. Dublin, Ireland: Association for Computational Linguistics.

Zhao, W.; Chiu, J.; Hwang, J.; Brahman, F.; Hessel, J.; Choudhury, S.; Choi, Y.; Li, X.; and Suhr, A. 2024. UNcommonsense reasoning: Abductive reasoning about uncommon situations. In Duh, K.; Gomez, H.; and Bethard, S., eds., *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, 8487–8505. Mexico City, Mexico: Association for Computational Linguistics.