On the Semantics of Primary Cause in Hybrid Dynamic Domains

Extended Abstract

Shakil M. Khan
University of Regina
Regina, Saskatchewan, Canada
shakil.khan@uregina.ca

Asim Mehmood
University of Regina
Regina, Saskatchewan, Canada

Sandra Zilles
University of Regina
Regina, Saskatchewan, Canada
sandra.zilles@uregina.ca

ABSTRACT

Reasoning about actual causes of observed effects is fundamental to the study of rationality. This important problem has been studied since the time of Aristotle, with formal mathematical accounts emerging recently. We live in a world where change due to actions can be both discrete and continuous, that is, hybrid. Yet, despite extensive research on actual causation, only few studies looked into causation with continuous change. Building on recent progress, in this paper we outline a definition of primary cause in a hybrid action-theoretic framework, namely the hybrid temporal situation calculus. Our definition formalizes causation through contributions, which can then be verified from a counterfactual perspective using a modified "but-for" test. We show that our definition of causation have some intuitively justifiable properties.

KEYWORDS

Actual cause, Hybrid dynamic systems, Situation calculus

ACM Reference Format:
Shakil M. Khan, Asim Mehmood, and Sandra Zilles. 2026. On the Semantics of Primary Cause in Hybrid Dynamic Domains: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/UETG8421

1 MOTIVATION

A fundamental challenge in dynamic domains is to identify the causes of change, in particular, that of an observed effect becoming true given a history of actions (the "scenario"), a problem known as *actual causation.* Humean counterfactual definition of causation, that "*a cause to be an object, followed by another ⋯ where, if the first object had not been, the second never had existed*",[1] suffers from the problem of preemption: the second might still occur without the first, due to a subsequent third object. Halpern and Pearl's (HP) causal model [13–15] avoids preemption by performing selective counterfactual analysis under carefully chosen contingencies. Since then, others have proposed various extensions to deal with expressive limitations of causal models [1, 4–8, 11, 12, 17, 18, 23].

A distinguishing feature of the real world is that change can be both discrete and continuous. However, despite the enormous body of work on actual causality, almost all have focused on defining causes in discrete domains. Recently in [25], we proposed a first definition of actual cause in a hybrid action theoretic framework. While that definition appeals to intuition, many argue that causation should be rather defined in counterfactual terms. Note that there is strong experimental evidence that humans understand causes using counterfactual reasoning [9, 10].

To deal with this, in this paper we propose a new definition of actual cause in the hybrid temporal situation calculus [2, 3]. We focus on actual primary cause and study causation relative to primitive fluents exclusively. Differently from our previous proposal, our definition formalizes causation through contributions, which can then be verified from a counterfactual perspective using a modified but-for test. Namely, we show that if one were to remove the identified cause along with some contingencies –in particular, actions that are currently preempted by the actual cause– from the given scenario, the outcome would no longer follow. We prove that the proposed definitions are indeed equivalent and that our definitions of causation have some intuitively justifiable properties.

2 ACTUAL ACHIEVEMENT CAUSE

Khan and Lespérance's (KL) recent account of achievement causation [21] in the situation calculus (SC) [26] builds on the original proposal by Batusov and Soutchanski [4]. Both of these assume that the scenario is a linear sequence of actions (modeled using a situation). Since all changes in the SC result from actions, causes are action instances, which are computed relative to a *causal setting* $\langle \mathcal{D}, \varphi, s \rangle$; here $\mathcal{D}$ is a basic action theory (BAT) specifying action dynamics, φ is the observed effect, and s is the scenario (history of actions). Also, it is assumed that s is executable, and φ was false before the execution of the actions in s, but became true afterwards.

The idea behind how causes are computed is as follows. Given an effect φ and scenario s, if some action of the action sequence in s triggers the formula φ to change its truth value from false to true relative to $\mathcal{D}$, and no subsequent actions in s change the value of φ back to false, then this action is the actual *primary* or *direct* cause of achieving φ in s. Now, note that a (primary) cause a might have been non-executable initially. Also, a might have only brought about the effect conditionally and this context condition might have been false initially. Thus earlier actions on the trace that contributed to the preconditions and the context conditions of a cause must be considered as causes as well. KL showed how one can inductively capture both primary and indirect causes; see [21].

In [25], we adapted this definition for the hybrid temporal situation calculus (HTSC) [3], which supports both discrete and continuous change. The HTSC takes inspiration from hybrid systems in control theory, which are based on discrete transitions between states that continuously evolve over time. In HTSC, (atemporal) SC

[1]The second-half of the quoted definition is also known as the "but-for" test.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/UETG8421

fluents are preserved to provide a context within which the values of temporal fluents can change. Contexts are mutually exclusive to make sure continuous fluents do not assume two different values at the same time. HTSC modifies SC's BAT by including the axioms for time of actions and start time of situations. Moreover, in addition to successor-state axioms [26], which specify how fluents change as a result of actions, HTSC introduces *state evolution axioms* (SEA) [2], which defines how a temporal fluent's value changes over time.

In HTSC, we first defined a notion of causal setting. In a hybrid domain, in general, one can query the causes of an observed effect at any time-point within a situation s, i.e. at any time-point in between the start-time and the end-time of s, inclusive. To simplify, we assumed that the query is posed relative to the starting time of s. Also, we dealt with effects φ that are conditions on the values of primitive temporal fluents (e.g. $temperature(P) > 1000$) exclusively.

For temporal effects, unlike discrete ones where the primary cause acts immediately, the effect may be realized only after some delay and intervening irrelevant actions. This requires the identification of the situation in which the effect is actually achieved (the "achievement situation"). Recall that in the HTSC, the values of temporal fluents can change only when certain relevant contexts are enabled, which in turn are affected as a result of action execution. Thus, when determining the primary cause of some temporal fluent having a certain value, we first need to identify the last context γ that was enabled before the fluent acquired this value, i.e., the context γ which was true in the achievement situation s_φ of the effect φ, and then figure out the action a that caused/enabled this context in s_φ. Since contexts are mutually exclusive (no two contexts can be true at the same time) and discrete, γ must have been the only enabled context in the achievement situation s_φ, which ensures that the action a is unique. Additionally, a must have been the last action that enabled γ, and whose contribution brought about the temporal effect under consideration.

Following this logic, we defined an action instance a to be the direct cause of an effect φ in scenario s iff the achievement situation of φ in s is s_φ, and a executed in some earlier situation directly (discretely) caused the active context γ_i^f for the temporal fluent f mentioned in φ in scenario s_φ. See [22] for a formalization.

3 A MODIFIED DEFINITION & BUT-FOR TEST

The above definition of actual cause of temporal effects appeals to intuition, but it neither formalizes an actual cause in counterfactual terms [19, 20], nor does it relate to the regularity accounts (e.g. [24]), the two most popular formalizations of causation. As a consequence, it is not clear how this definition compares with these common but different approaches to actual causation. To deal with this, we now give a second definition using the notion of contributions, which can then be related to counterfactuals.

As detailed in a longer version [22], to this end we can first define *direct possible contributors* as actions that directly initiate the required change in the values of temporal fluents. Using this, we can define *direct actual contributors*, which are direct possible contributors that are contained within a given setting/scenario. Finally, we define primary cause as a direct actual contributor such that, after its contribution, the effect is achieved and persists until the end of a scenario—i.e. a direct actual contributor to the active context in the achievement situation. Formally, an action instance α is the primary cause of an effect φ in scenario σ iff s_φ is the achievement situation of φ in σ, and α executed in some situation s_α is a direct actual contributor of φ in σ. We can show that this new definition is equivalent to the previous one.

In our attempt to intuitively justify the proposed definitions, we now formalize a modified but-for test for validating actual cause. Recall that the but-for test fails due to the presence of preempted actions. Namely, even if one were to remove the actual cause from the scenario, the effect might still follow due to a subsequent action in the scenario whose effect in the original scenario was preempted by the actual cause. Our modification handles this issue by removing all such preempted actions before testing for the effect. A particular complication that arises in the context of hybrid domains is that such preempted actions can even occur *before* the direct cause. To address this problem, we first define *counterfactual scenarios*, which are worlds that would have been realized had actions/events been different from what actually occurred. Using this, we identify the *preempted contributors*, i.e. actions/contributors that would have still caused the effect if the actual cause was absent. Technically, the primary preempted contributor is the primary cause in the counterfactual scenario obtained by removing the actual cause from the given scenario; this idea can be used to inductively find all preempted contributors. We then introduce the concept of a "defused" situation, where the actual cause along with all preempted contributors are replaced with *noOp* actions, i.e. actions that are always possible and that have no effect, thereby removing their effects. This substitution allows us to isolate the impact of the primary cause by removing the influence of the cause and its preempted contributors. See [22] for a formalization.

In [22], we show that: **(1)** as expected, the but-for test fails; that is, if an action instance a is a cause of φ in scenario s, it is not guaranteed that removing a's effect yields either a non-executable counterfactual situation or one in which φ no longer holds. **(2)** In hybrid domains, there may be no primary causes (e.g. when the context of the achievement situation was enabled initially). **(3)** If there is a primary cause, there is a unique defused scenario. **(4)** If additionally it is known that all the contexts of φ are inactive initially, then it must be the case that φ is false in the defused situation s', unless s' has become non-executable. Theorem (4) can be viewed as a modified but-for test stating that if one removes causes and preempted actions from the scenario, under certain reasonable and intuitive conditions (i.e., that no context of the temporal fluent holds initially and that the actions in s' are still executable), the effect will disappear.

4 CONCLUSION

We showed that causation in hybrid domains, even in relatively simple settings, requires careful modeling and reasoning. Also, a useful variant of the but-for test can be formulated. To our knowledge, the only other work connecting actual causation with continuous change is Halpern and Peters' generalized structural equation models (GSEM) [16], which allow variables indexed by time or other continuous parameters but do not provide a definition of actual cause. In future work, we plan to study direct and indirect causes of arbitrary compound effects, both discrete and temporal.

ACKNOWLEDGMENTS

This work is partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and by the University of Regina.

REFERENCES

[1] Carlos Aguilera-Ventura, Xinghan Liu, Emiliano Lorini, and Dmitry Rozplokhas. 2025. A Non-Interventionist Approach to Causal Reasoning Based on Lewisian Counterfactuals. In *Proceedings of the 34th International Joint Conference on Artificial Intelligence, IJCAI 2025, Montreal, Quebec, Canada, August 16-22, 2025*. 4301–4310.

[2] Vitaliy Batusov, Giuseppe De Giacomo, and Mikhail Soutchanski. 2019. Hybrid Temporal Situation Calculus. In *Advances in Artificial Intelligence - 32nd Canadian Conference on Artificial Intelligence, Canadian AI 2019, Kingston, ON, Canada, May 28-31, 2019, Proceedings (Lecture Notes in Computer Science, Vol. 11489)*, Marie-Jean Meurs and Frank Rudzicz (Eds.). Springer, 173–185.

[3] Vitaliy Batusov, Giuseppe De Giacomo, and Mikhail Soutchanski. 2019. Hybrid temporal situation calculus. In *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, SAC 2019, Limassol, Cyprus, April 8-12, 2019*, Chih-Cheng Hung and George A. Papadopoulos (Eds.). ACM, 1162–1164.

[4] Vitaliy Batusov and Mikhail Soutchanski. 2018. Situation Calculus Semantics for Actual Causality. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, Sheila A. McIlraith and Kilian Q. Weinberger (Eds.). AAAI Press, 1744–1752.

[5] Sander Beckers and Joost Vennekens. 2018. A Principled Approach to Defining Actual Causation. *Synthese* 195, 2 (2018), 835–862.

[6] Alexander Bochman. 2003. A Logic For Causal Reasoning. In *IJCAI-03, Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, Acapulco, Mexico, August 9-15, 2003*, Georg Gottlob and Toby Walsh (Eds.). Morgan Kaufmann, 141–146. http://ijcai.org/Proceedings/03/Papers/020.pdf

[7] Alexander Bochman. 2018. Actual Causality in a Logical Setting. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, Jérôme Lang (Ed.). ijcai.org, 1730–1736. https://doi.org/10.24963/IJCAI.2018/239

[8] Alexander Bochman. 2018. On Laws and Counterfactuals in Causal Reasoning. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018, Tempe, Arizona, 30 October - 2 November 2018*, Michael Thielscher, Francesca Toni, and Frank Wolter (Eds.). AAAI Press, 494–503. https://aaai.org/ocs/index.php/KR/KR18/paper/view/17990

[9] Tobias Gerstenberg, Noah D. Goodman, David A. Lagnado, and Joshua B. Tenenbaum. 2014. From counterfactual simulation to causal judgment. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society, CogSci 2014, Quebec City, Canada, July 23-26, 2014*, Paul Bello, Marcello Guarini, Marjorie McShane, and Brian Scassellati (Eds.). cognitivesciencesociety.org. https://escholarship.org/uc/item/7kk1g3t8

[10] Tobias Gerstenberg, Noah D. Goodman, David A. Lagnado, and Joshua B. Tenenbaum. 2015. How, whether, why: Causal judgments as counterfactual contrasts. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society, CogSci 2015, Pasadena, California, USA, July 22-25, 2015*, David C. Noelle, Rick Dale, Anne S. Warlaumont, Jeff Yoshimi, Teenie Matlock, Carolyn D. Jennings, and Paul P. Maglio (Eds.). cognitivesciencesociety.org.

[11] Maksim Gladyshev, Natasha Alechina, Mehdi Dastani, Dragan Doder, and Brian Logan. 2025. Temporal Causal Reasoning with (Non-Recursive) Structural Equation Models. In *AAAI-25, Sponsored by the Association for the Advancement of Artificial Intelligence, February 25 - March 4, 2025, Philadelphia, PA, USA*, Toby Walsh, Julie Shah, and Zico Kolter (Eds.). AAAI Press, 14949–14957. https://doi.org/10.1609/AAAI.V39I14.33639

[12] Clark Glymour, David Danks, Bruce Glymour, Frederick Eberhardt, Joseph D. Ramsey, Richard Scheines, Peter Spirtes, Choh Man Teng, and Jiji Zhang. 2010. Actual Causation: A Stone Soup Essay. *Synthese* 175, 2 (2010), 169–192.

[13] Joseph Y. Halpern. 2000. Axiomatizing Causal Reasoning. *Journal of Artificial Intelligence Research* 12 (2000), 317–337.

[14] Joseph Y. Halpern. 2016. *Actual Causality.* MIT Press.

[15] Joseph Y. Halpern and Judea Pearl. 2005. Causes and Explanations: A Structural-Model Approach. Part I: Causes. *The British Journal for the Philosophy of Science* 56, 4 (2005), 843–887.

[16] Joseph Y. Halpern and Spencer Peters. 2022. Reasoning about Causal Models with Infinitely Many Variables. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022.* AAAI Press, 5668–5675.

[17] Mark Hopkins. 2005. *The Actual Cause: From Intuition to Automation.* Ph.D. Dissertation. University of California Los Angeles.

[18] Mark Hopkins and Judea Pearl. 2007. Causality and Counterfactuals in the Situation Calculus. *Journal of Logic and Computation* 17, 5 (2007), 939–953.

[19] Mark Hopkins and Judea Pearl. 2007. Causality and Counterfactuals in the Situation Calculus. *Journal of Logic and Computation* 17, 5 (2007), 939–953. https://doi.org/10.1093/logcom/exm048

[20] David Hume. 1748. *An Enquiry Concerning Human Understanding.*

[21] Shakil M. Khan and Yves Lespérance. 2021. Knowing Why - On the Dynamics of Knowledge about Actual Causes in the Situation Calculus. In *AAMAS '21: 20th International Conference on Autonomous Agents and Multiagent Systems, Virtual Event, United Kingdom, May 3-7, 2021*, Frank Dignum, Alessio Lomuscio, Ulle Endriss, and Ann Nowé (Eds.). ACM, 701–709.

[22] Shakil M. Khan, Asim Mehmood, and Sandra Zilles. 2026. On the Semantics of Primary Cause in Hybrid Dynamic Domains. *ArXiv* (2026). arXiv:2602.14994 [cs.AI]

[23] Shakil M. Khan and Mikhail Soutchanski. 2020. Necessary and Sufficient Conditions for Actual Root Causes. In *ECAI 2020 - 24th European Conference on Artificial Intelligence, 29 August-8 September 2020, Santiago de Compostela, Spain, August 29 - September 8, 2020 - Including 10th Conference on Prestigious Applications of Artificial Intelligence (PAIS 2020) (Frontiers in Artificial Intelligence and Applications, Vol. 325)*, Giuseppe De Giacomo, Alejandro Catalá, Bistra Dilkina, Michela Milano, Senén Barro, Alberto Bugarín, and Jérôme Lang (Eds.). IOS Press, 800–808.

[24] John Leslie Mackie. 1965. Causes and Conditions. *American Philosophical Quarterly* 2, 4 (1965), 245–264.

[25] Asim Mehmood and Shakil M. Khan. 2024. Towards a Definition of Primary Cause in Hybrid Dynamic Domains. In *Proceedings of the Canadian Conference on Artificial Intelligence.* Canadian Artificial Intelligence Association (CAIAC). https://caiac.pubpub.org/pub/wr72f8ae.

[26] Raymond Reiter. 2001. *Knowledge in Action. Logical Foundations for Specifying and Implementing Dynamical Systems.* MIT Press, Cambridge, MA, USA.