出國報告（出國類別：國際會議）

因果與機率推理國際會議
(Causal and Probabilistic Reasoning Conference)

服務機關： 國立中正大學哲學系

姓名職稱：王一奇 教授

派赴國家：德國 (慕尼黑)

出國期間：**104** 年 **6** 月 **16** 日至 **104** 年 **6** 月 **22** 日

報告日期：**104** 年 **7** 月 **3** 日

摘要：

本次出國目的為參加國際會議發表論文「Paradigmatic Causation and Multitudes of Trumping」(中文譯名：典範因果關係與壓迫性特例的多面性)，本人為論文的第二作者，在會議上演講口頭報告論文。此次的會議名稱為: 「因果與機率推理國際會議」 **(Causal and Probabilistic Reasoning Conference)**，會議地點在德國慕尼黑大學，主辦單位為慕尼黑數學化哲學研究中心 (Munich Center for Mathematical Philosophy)，會議全程使用英文。本會議為因果推理與機率推理的專業會議，舉辦的因緣為伯爾 (Pearl) 及史伯帝等人 (Spirtes *et. al.*) 關於因果及機率推理重要著作的出版及再版 15 年。本會議有超過 30 位國際上研究因果推理及機率推理的學者參與，本人在會議中以英文口頭發表論文，獲得許多重要的建議及指正。

目次：

本文：

一、目的

本次出國目的為參加國際會議發表英文論文「Paradigmatic Causation and Multitudes of Trumping」 (中文譯名：典範因果關係與壓迫性特例的多面性)，本人為論文的第二作者，在會議上演講口頭報告論文。此次的會議名稱為: 因果與機率推理會議 **(Causal and Probabilistic Reasoning Conference)**，會議地點在德國慕尼黑大學，主辦單位為慕尼黑數學化哲學研究中心 (Munich Center for Mathematical Philosophy)，會議全程使用英文。本會議為因果推理與機率推理的專業會議，舉辦的因緣為為伯爾 (Pearl) 及史伯帝等人 (Spirtes et. al.) 關於因果及機率推理重要著作的出版及再版 15 年。本會議有超過 30 位國際上研究因果推理及機率推理的學者參與，本人在會議中以英文口頭發表論文。發表英文論文"典範因果關係與壓迫性特例的多面性"內容主要分為五部分。第一部分介紹與論文內容相關的概念工具，第二部分介紹西區考克 (Hitchcock) 的典範因果關係 (paradigmatic causation) 的理論，第三部分指出西區考克理論的缺陷，並提出一個解決方案，第四部分介紹西區考克對壓迫性特例的理論並指出他理論中的難題，第五部分提出針對西區考克關於壓迫性特例理論難題的解決方案。本論文的內容與因果模型及因果推理有關，因此參加此國際會議以獲得機會與國際學者交流，希望獲得與會專家學者的建議及指正，以期能更加完善本論文。

二、過程

104 年 6 月 16 日：出發前往德國慕尼黑『第二屆亞洲哲學邏輯工作坊』。

104 年 06 月 18 日至 143 年 06 月 20 日：在德國慕尼黑大學參加『因果與機率推理國際會議』，包括發表論文、參與其他學者的演講並進行討論。

104年06月18日:上午於會議發表論文「Paradigmatic Causation and Multitudes of Trumping」(中文譯名：典範因果關係與壓迫性特例的多樣性)，英文口頭報告 45 分鐘；在發表論文後，針對論文中的重要議題，和與會學者進行細緻的討論，取得很多的重要意見。

104 年 06 月 22 日：搭機返臺。

以下將從會議議程及議場主題、與會內容重點及心得、個人報告內容及交流三方面說明與會過程。

三、會議議程及議場主題

「因果與機率推理國際會議」主題是與因果推理及機率推理相關的議題，會議參與者包含各個不同社會科學領域的學者 (心理學領域佔多數)，以及哲學的學者，這是一個跨領域的國際會議。會議中最重要的主題是因果模型的理論與實務，涵蓋機率與與非關機率的層面，與會學者總共發表了24篇論文 (詳細議程請見附錄1)，包含三個主題演講，數量非常豐富。在24篇會議論文中，內容涵蓋了因果與機率推論的多個重要面向，包含(一)因果模型在語言與因果推理的運用、（二）非因果圖形模型的理論與應用、（三）機率推理的理論與運用，尤其是在心理學中的運用。

四、與會內容重點及心得

在24篇會議論文報告中，許多與會學者的報告內容令人印象深刻，以下僅對幾篇最令本人印象深刻的內容及本人的學習收穫提供大概的說明。

（一）許邦 (Wolfgang Spohn) 教授”關於評價因果理論的15個面向” (Fifteen Dimensions of Evaluating Theories of Causation) 的演講，主要針對因果模型

(causal model) 及排序理論 (ranking) 比較研究，提出15個兩種理論基礎的差異，並論證偏好排序理論。在演講後的私下討論爭，為對排序理論有些許的同情，似乎在某些地方言之成理。許邦特別提供，他在1970年代的博士論文中，已清楚證明圖像模型中的馬可夫條件 (Markov Condition) 與機率獨立性的關係，比伯爾及史伯帝等人至少早了十多年，但但他的貢獻在文獻中幾乎沒有提到。 許邦教授雖沒有明說，但他顯然有所報怨美國的學術沙文主義。我下次論文中寫到相關議題時，會引用許邦教授的文章。

（二）單克斯 (David Danks) 關於"圖像模型、認知表徵、及語意的不同質性" (Graphic Models, Cognitive Representation, and Semantic Heterogeneity) 說明除了因果圖像模型，還有其他圖像模型，並說明在非因果圖像模型中，如何進行推論。他特別強調，除了一般在因果模型中常利用的硬干預 (hard intervention)，還有其它形式的干預，如軟干預 (soft intervention)。有趣的是，有其它演講討論的就是非因果圖像模型，動機來自於在某些情況下，例如量子狀態，因果馬可夫條件 (causal Markov condition) 不成立。

（三）漢斯 (Ulrike Hahn) 關於"因果論證" (causal argument) 的演講分析了關於因果論證的各種形式，特別推薦孔恩 (Kuhn, 1991) 論證的技藝 (The Skills of Argument) 一書值得一讀。演講人認為，從各種書面資料，包含語料庫，所收集到的因果論證形式，如何與哲學與心理學中的因果概念形成連結，是一個值得開發，但尚未被開發的新研究領域。漢斯教授及許邦教授對相關議題的研究方法針鋒相對，漢斯教授認為應該用由上而下 (top-down) 的研究方式，許邦堅持由下而上 (bottom-up) 的研究方式，非常有趣。

（四）賀提格 (Ralph Hertwig) 教授演講主題 "在不確定的微光中航行" (Navigating the Twilight Uncertainty)，演講主要探討人們對不確定事件的評估與反應，尤其是對於稀有 (rare) 事件的評估與反應，這是一個非常熱門的議題，屬於新興的決策行為科學的領域。這個演講中包含了幾個不同的心理學實驗，包含以描述為基礎的決策實驗，和以經驗為基礎的決策實驗，以及這些心理學實驗

所帶來的挑戰。

（五）布魯說 (Peter Brossel) 教授演講主題 “神經科學中逆反推論的有效基礎” (On the rationale of reverse inference in neuroscience)，這個主題非常有趣，探討神經科學中，從核磁共振照影推論到認知狀態的推論是一個標準的無效論證，但這個推論形式卻快速增加我們對認知狀態的知識，也具有很大的實用性。在這個議題中，最核心的問題，在於如何對逆反推理加以適當限制，以免做出太離譜的推論。我個人對逆反推論有多年的研究，對神經科學很有興趣卻苦無切入點，所以這個演講提供了一個我對神經科學哲學議題的絕佳切入點。(在網路搜尋時，用逆反推論 (reverse inference) 及 神經科學(neuroscience) 去搜尋，可以找到很多相關資料。)

五、會議心得

（一）會議中，對於因果及機率推論有大量討論。從人類語言中複雜而多樣的動詞中，我們可以看出，因果關係只是多種事件關連性 (event relation) 的一種。其他的事件關連性，如防止、預防、停止、延緩等，在文獻及本此會議中都很少有人討論，這些因果關連性外的事件關連性，應是值得發展的研究主題。

（二）在文獻及會議中，關於因果模型的討論集中在針對個別模型的性質加以討論。而對於模型的比較，則幾乎無人觸及。模型間的比較，在個個不同領域的問題中，當涉及到因果結構如何因行動或事件的發生而改變，模型比較應有重要的運用。例如，從訊息中做出決策，涉及到訊息的因果模型與因訊息而決策的因果模型間的差異。當人們基於訊息做出干預行為，訊息的因果結構及呈現干預行為對因果結構的影響，需要「跨時間性」因果模型比較。

（三）關於逆反推論神經科學及神經心理學中的使用，將成為個人研究的一個方向，很難得可以在參與會議過程中，發現對未來研究有所助益的主題。

（四）歐文等人 (Oven et al. 2007)在認知心理學期刊(Cognitive Psychology) 有一

篇關於反事實條件句的心理學文獻。

六、個人報告及交流

本人會議報告論文題目為「Paradigmatic Causation and Multitudes of Trumping」（中文譯名：典範因果關係與壓制性特例的多面性），論文中文摘要如下(論文全文請見附錄 2)：

摘要：本論文的目標是對西區考克 (Hitchcock) 的兩個論點提出一個整合性理論。西區考克的第一個論點認為，當甲是乙的典範因 (paradigmatic cause)，代表乙反事實的依賴在甲上，且甲對乙提供了一個令人滿意的解釋。第二個論點，西區考克認為認為，文獻中的壓迫性特例 (trumping)，並不單純是一個關於干擾性因果關係(preemption)的例子，而具有多面性，可以是一個單純的非多餘(non-redundant)的因果關係，也可以是所謂的過度決定(over-determination)因果關係，端賴如何選擇對比事件。為了完成這兩個論點的整合，我們論證，不論是在對比式 (contrastive) 及非對比式 (non-contrastive) 的因果架構中，西區考克以自我包含 (self-containment)的概念來補捉令人滿意的解釋，都是不完全洽當的。更進一步，我們發展了一個用來建構利人滿意說明的新概念，也就是完整性(integrity)概念，我們論證這個新概念在對比式與非對已是架構下都是洽當的。

在本次論文的口頭報告中，與會學者對論文提出兩個主要的疑問。

第一個問題來自許邦 (Wolfgan Spohn)教授。他認為，除了有一階的初始值/偏差值 (default-deviant) 的區分，還有高階的初始值/偏差值的區分。任一 n 階的初始值/偏差值區分，是基於 n-1 階初始值/偏差值的區分而成立。許邦教授心中很顯然有一些例子，但時間很短，他也沒多提。因為我從沒想過這個問題，我只

能據實說我沒想過。這個問題視後續要進一步細究的問題，感謝許邦教授的提問。(會後，許邦有向我提另一個看法。雖然我用因果模型來解決壓制性特例的問題，但許邦教授認為，壓制性特例的問題在他的排序理論 (ranking thoery) 中可以很簡單的被解決，並向我仔細的說明了一下。幾天後許邦教授寄了一封信給我，附上了他關於初始值/偏差值的新論文。)

第二個問題來自漢斯教授，她同樣關心初始值/偏差值區分的問題。他指出，這個區份似乎會因法律的規範而備被有所影響，也舉了一個例子。我回應她，在西區考克(2005) 年一篇發表在哲學雜誌 (Journal of Philosophy) 的文章中，對這個問題有一些有趣的研究與討論，感謝漢斯教授的提問。

七、會後心得

「因果與機率推理國際會議」有許多哲學與社會科學領域重要的專家學者與會，這些學者的參與，讓本次會議在問題討論時，能夠進行實質且重要的意見交流，同時也提升了本次會議的學術重要性。本次會議的核心議題，關切因果與機率推理，但它的重要性在於墊基在因果模型的相關理論上。因果模型不只在理論上有其突出特性及是一一個前瞻議題，更重要的是，它是一個生物醫學及社會科學科學家實務上運用的理論，且其影響力正逐步擴大中，這個會議的舉辦顯然是這個影響力的一個顯現，我因個人學術興趣而參與其中，臺灣關於此類研究的學者甚少，希望臺灣能迎頭趕上。目前我也再規劃一些相關的會議，希望能成功辦理這些會議，對臺灣哲學與生物醫學社會學的前瞻議題上有所助益。

八、建議事項

就哲學相關領域來說，關於因果推理及機率推理的研究這相對是少數，而且，使用因果模型來作為研究工具及概念架構的學者更少，本次在會議中見到了許多相關的人。就前瞻研究議題來說，就以下的面向來說，這個議題有無可取代

的重要性。首先，這個議題有廣大的科學集團在背後撐腰，不斷的進行實證上的運用 (有很多套關於因果模型的軟體)，而使其有經驗基礎，以及為了實證需求的改良，使理論的豐富度快速爭加。第二，因果模型是提供了人類一個有史以來，第一個可用來思考因果關連性的數學模型，同時這個數學模型可以用來清晰的分析科學實驗的構作。這兩個特點，使因果理論得以突破所謂的休姆-達因屏障(Hume-Quine barrier)，讓經驗主義者第一次可以跨出觀察與概念運作的極限。在哲學上，其深刻意涵無法比擬。臺灣哲學學者基於數學工具上較西方學者為優，相信這個前瞻議題值得臺灣學者發展。

附錄 1：會議議程。

附錄 2：會議論文「Paradigmatic Causation and Multitudes of Trumping」全文。

Program

Room Arrangement

Date	Time	Address, Room
18 June	09:00 - 12:15	Geschwister-Scholl-Platz 1, Room M010
	13:45 - 18:00	Prof. Huber Platz 2, Room W201
19 June	10:00 - 17:45	Prof. Huber Platz 2, Room V005
20 June	09:00 - 17:45	Prof. Huber Platz 2, Room W101

18 June

Time	Topic
08:30 - 09:00	Registration
09:00 - 09:15	Welcome
09:15 - 10:30	**Keynote Lecture:** Spohn, Wolfgang: Fifteen Dimensions of Evaluating Theories of Causation. A Case Study of the Structural Model and the Ranking Theoretic Approach to Causation Chair: Gregory Wheeler
10:30 - 10:45	Coffee Break
10:45 - 11:30	Leuridan, Bert/Beilaen, Mathieu: A Logic for the Discovery of Causal Regularities
11:30 - 12:15	Fan, Da/Wang, Linton: Paradigmatic Causation and Multitudes of Trumping
12:15 - 13:45	Lunch
13:45 - 14:30	Danks, David: Graphical Models, Cognitive Representations, and Semantic Heterogeneity
14:30 - 15:15	Poellinger, Roland/Hubert, Mario: Bell's Theorem and Non-Markovian Network Models
15:15 - 15:30	Coffee Break
15:30 - 16:15	Mayrhofer, Ralf/Waldmann, Michael: Agents and Causes: Dispositional Intuitions as a Guide to Causal Structure
16:15 - 17:00	Näger, Paul: The Causal Markov Condition and Non-Screening-Off

	Common Causes
17:00 - 17:15	Coffee Break

19 June

Time	Topic
10:00 - 11:15	**Keynote Lecture:** Hahn, Ulrike: Causal Argument Chair: Karolina Krzyżanowska
11:15 - 11:30	Coffee Break
11:30 - 12:15	Sydow, Momme von: Logical Inclusion Fallacies Within One Polytomous Dimension
12:15 - 13:45	Lunch
13:45 - 14:30	Lukits, Stefan: A Natural Generalization of Jeffrey Conditioning
14:30 - 15:15	Betz, Gregor: The Veritistic Merit of Probabilistic Degrees of Justification and Doxastic Conservativism in Belief Revision
15:15 - 15:30	Coffee Break
15:30 - 16:15	Horne, Zachary/Livengood, Jonathan: Ordering Effects, Updating Effects, and the Specter of Global Skepticism
16:15 - 17:00	Gyenis, Balazs: Do Genuine Ordering Effects Exist?
17:00 - 17:15	Coffee Break
17:15 - 18:00	Stenning, Keith/Martignon, Laura: A Qualitative Intensional Approach to Causal Reasoning and Decision in Uncertainty
18:00 - 18:45	Stenning, Keith/Varga, Alexandra: Normative Gloves That Fit Agents' Hands: Consequences of a Multiple Logics-Framework for Reasoning
20:00	Conference Dinner (Cafe Reitschule)

20 June

Time	Topic
09:00 - 10:15	**Keynote Lecture:** Hertwig, Ralph: Navigating the Twilight of Uncertainty: Decisions from Experience Chair: Michael Waldmann
10:15 - 10:30	Coffee Break
10:30 - 11:15	Carter, Sam: Probabilistic Judgement about Indicative Conditionals: A

	Formal Model-Based Theory
11:15 - 12:00	Pfeifer, Niki/Stöckle-Schobel Richard: The Probability of Indicative and Counterfactual Conditionals in Causal and Non-causal Settings
12:00 - 13:30	Lunch
13:30 - 14:15	Meder, Bjoern/Mayrhofer, Ralf/Waldmann, Michael R.: Structure Induction in Diagnostic Causal Reasoning
14:15 - 15:00	Brössel, Peter: On the Rationale of Reverse Inference in Neuroscience
15:00 - 15:15	Coffee Break
15:15 - 16:00	Sprenger, Jan/Colombo Matteo: Graded Causation and Explanatory Power, Explicated Probabilistically
16:00 - 16:45	Weinberger, Naftali: Probabilistic Causality, Structural Equations, and Causal Intermediaries
16:45 - 17:00	Coffee Break
17:00 - 17:45	Harinen, Totte: On the Need to Model Proportionality

Paradigmatic Causation and Multitudes of Trumping

Da Fan & Linton Wang

Abstract. The objective of this paper is to propose and defend a synthesis of Hitchcock's two theses of token causation: first, C is a paradigmatic cause of E just in case E counterfactually depends on C and C provides a satisfactory explanation for E; second, from the perspective of contrastive causation, trumping cases in the literature are not simply cases of preemption, but are cases of non-redundancy or overdetermination depending on the choice of contrastive pairs. To accomplish the objective of the synthesis, we argue that, in both the contrastive and non-contrastive frameworks, Hitchcock's proposal for how an event C provides a satisfactory explanation of E, based on his notion of self-containment, is not yet fully adequate. Furthermore, based on Hitchcock's proposal, we develop an alternative notion of satisfactory explanation – the notion of integrity – which is suitable for both the contrastive and non-contrastive framework.

Keywords. Paradigmatic Causation; Trumping; Counterfactual Dependence; Parasitic Dependence; Causal Model; Overdetermination

1 Introduction

The objective of this paper is to propose and defend a synthesis of Hitchcock's two theses of token causation.

> **(Paradigmatic Causes)** C is a paradigmatic cause of E just in case (a) E counterfactually depends on C, and (b) C provides a satisfactory explanation for E (cf. Hitchcock 2007).
>
> **(Multitudes of Trumping)** From the perspective of contrastive causation, the trumping cases in the literature are not simply cases of preemption, but are cases of either non-redundancy or overdetermination, depending on the choice of contrastive pairs (cf. Hitchcock 2011).

Though Hitchcock does not attempt to synthesize the thesis of paradigmatic cases in Hitchcock (2007) and the thesis of multitudes of trumping in Hitchcock (2011), we will attempt to do so. To accomplish this objective, we argue that, in both the contrastive and non-contrastive framework, Hitchcock's proposal for how an event C provides a satisfactory explanation of

E is not yet fully adequate. Then, based on Hitchcock's proposal, we develop an alternative notion of satisfactory explanation that is suitable for both the contrastive and the non-contrastive framework.

In Hitchcock (2007), he attempts to qualify the sort of situations in which our judgements about token causation, i.e., the paradigmatic cases of causation, strongly align with counterfactual dependence. The situations are identified by his "principle of sufficient reason" (PSR henceforth): for a deviant event (which is roughly either a non-typical or an unexpected event) in a given story, there is another deviant event which provides a satisfactory explanation for it. Moreover specifically, Hitchcock formulates PSR by using the notion of *self-containment*, in that, within a given story (which Hitchcock calls a *causal network*) including how the cause C is "connected" to the effect E in question, if a particular event is causally influenced by other events which are all in their default (non-deviant) states, the particular event is at its default state.[1] If the story on how an event C is connected to an event E is not self-contained, Hitchcock identifies the counterfactual dependence of E on C as a *parasitic* dependence to signify that the explanation of E by C is not satisfactory. Hitchcock shows that the challenges to the simple counterfactual analysis for causation from cases of prevention and omission are cases of parasitic dependence, and finds himself agnostic on the question of whether those cases of parasitic dependence are genuine cases of causation. Hitchcock (2007)'s proposal will be recapitulated in section 3.

Nonetheless, in section 4, we will argue that there are cases to show that self-containment is not fully adequate for one to pick up a satisfactory explanation relation between events, meaning that the distinction between paradigmatic causation and parasitic dependence based on self-containment is not yet fully adequate with respect to its intended purpose. We argue that, in those cases, there are events such that counterfactual dependence and self-containment are obtained, but the events are explanatorily incomplete. A new line to remedy the challenge from the case is not far away from Hitchcock's self-containment. In section 5, the distinction between paradigmatic causes and parasitic dependence is alternatively drawn based on the notion of *integrity*, in that, roughly speaking, the counterfactual relationships in a given story of how events are connected are *not disturbed by the deviant events outside the story*. In other words, the counterfactual dependence in cases of paradigmatic causation are those that are not disturbed by the deviant events outside the story for the counterfactual dependence. Under the restriction of integrity, since deviant events outside of the story have no impact on the standing of counterfactual dependence inside of the story, we can say that counterfactual dependence provides a satisfactory explanation between events in the story. A contrastive version of integrity will be provided in section 7, which will be shown to be a proper extension of Hitchcock (2011)'s contrastive analysis of trumping cases, which is to be introduced in section 6.

[1]This pragmatically oriented default/deviant distinction is elaborated in Hitchcock (2007: 506-507). Hall (2007), Halpern (2008), Hitchcock and Knobe (2009), and Halpern and Hitchcock (2013, forthcoming) also adopt this distinction in order to explain actual causation.

2 Causal Models and Default/Deviant Distinction

A took kit including causal modeling semantics and default/deviant distinction will be exploited throughout the rest of this paper. We will pack the tool kit as compactly as possible. For a more detailed discussion of the formalism we use, of causal modeling semantics, or of the default/deviant distinction, the reader is directed to the relevant literature.[2]

2.1 Causal Modeling Semantics

Consider a causal model $M =< U, V, E_1, E_2 >$ where U is a set of exogenous variables in which every variable is associated with a set of values, V is a set of endogenous variables in which every variable is associated with a set of values, E_1 is a set of equations of the form $U_i = \kappa_i$ for every $U_i \in U$ with a value κ_i associated to it, and E_2 is a set of equations such that for every $V_i \in V$ of the form $V_i = f_{V_i}(W_1, ..., W_n)$, where f_{V_i} is a function that determines the value of V_i via the variables $W_1, ..., W_n \in U \cup V$.[3] For the equations in E_2, the equations expressed by Boolean connectives over variables associated with only two values $\{0, 1\}$ take their usual truth-functional meanings. That is, the equation $Z = X \vee Y$ (or $Z = f_{\vee}(X, Y)$) means that Z takes the maximal value between X and Y; $Z = X \wedge Y$ (or $Z = f_{\wedge}(X, Y)$) means that Z takes the minimal value between X and Y; $Z = \neg X$ (or $Z = f_{\neg}(X, Y)$) means that $Z = 1 - X$. In this paper, we consider only causal models in which every exogenous variable is assigned with only one value in E_1 and the equations in E_2 are so designed that only one value for every $V_i \in V$ can be calculated based on the values given for the variables in E_1.

A causal graph for a causal model $M =< U, V, E_1, E_2 >$ is a directed graph $G_M =< U \cup V, E >$ such that $E = \{< W_i, W_j > | f_{W_j} = (..., W_i, ...)$ and $f_{W_j} \in E_2\}$, i.e., W_i is a variable that contributes to the determination function of the value of W_j. The terminology of kinship, e.g., *parents*, *children*, *ancestors*, and *descendants*, can be defined accordingly. For example, for the parents of a node X, $Pa(X) = \{Y| < Y, X >\in E\}$.

The formulas with respect to a causal model M include (a) atomic formulas of the form '$W_i = \kappa_i$' where $W_i \in U \cup V$ and κ_i is a value associated with W_i, (b) Boolean combinations of atomic formulas, and (c) counterfactuals of the form '$\varphi > \psi$', in which φ is a conjunction of atomic formulas $W_1 = \kappa_1 \wedge ... \wedge W_i = \kappa_i$ (represented by $\bigwedge_{1 \leq i \leq n} W_i = \kappa_i$), and ψ is a Boolean formula. The truth definition for counterfactuals requires the notion of submodels.

> **(Submodel)** Let Σ be a set of atomic formulas and $M_\Sigma =< U^*, V^*, E_1^*, E_2^* >$ be the submodel of $M =< U, V, E_1, E_2 >$ such that
>
> 1. $U^* = U \cup \{W_i | W_i = \kappa_i \in \Sigma\}$,

[2] For causal modeling semantics, see, among others, Halpern & Pearl (2005), Hitchcock (2001, 2007), and Pearl (2000). For the default/deviant distinction, see references in fn. 1 in this paper.

[3] The formalism of causal models in this paper basically follows that of Pearl (2000).

2. $V^* = V - \{W_i | W_i = \kappa_i \in \Sigma\}$,
3. $E_1^* = (E_1 - \{W_i = \kappa_j \in E_1 | W_i = \kappa_i \in \Sigma\}) \cup \{W_i = \kappa_i | W_i = \kappa_i \in \Sigma\}$, and
4. $E_2^* = E_2 - \{f_{W_i} | W_i = \kappa_i \in \Sigma\}$.

Informally, a submodel M_Σ of M is generated by the set of *interventions* Σ such that every intervention $W_i = \kappa_i \in \Sigma$ functions to turn the variable $W_i \in U \cup V$ into an exogenous variable and set its value to κ_i.

The truth conditions for formulas are as follows.

> **(Truth Conditions, Pearl 2000)** A formula φ is true in a model M – i.e., $M \models \varphi$ – given that
>
> 1. $M \models W_i = \kappa_i$, if and only if the value of W_i calculated from equations in M is κ_i,
> 2. $M \models \varphi \wedge \psi$, if and only if $M \models \varphi$ and $M \models \psi$,
> 3. $M \models \varphi \vee \psi$, if and only if $M \models \varphi$ or $M \models \psi$,
> 4. $M \models \neg\varphi$, if and only if $M \not\models \varphi$,
> 5. $M \models \varphi > \psi$, if and only if $M_{\varphi^*} \models \psi$, where $\varphi^* = \{W_i = \kappa_i | 1 \leq i \leq n$ and $\varphi = \bigwedge_{1 \leq i \leq n} W_i = \kappa_i\}$.

Truth conditions for Boolean formulas are defined as usual in propositional logic, and the truth condition for counterfactuals is defined by using submodels.

We use the case OM to elaborate the causal modeling semantics.

> **(OM)** Assassin shoots Victim, Bodyguard does not push Victim away, and Victim dies. If Assassin did not shoot, or, if Bodyguard pushed Victim away, Victim would not die.

The following causal model OM is to represent the OM case, where three bi-valued variables A, B, and D are to represent, respectively, whether Assassin shoots ($A = 1$ if he does and $A = 0$ if not), whether Bodyguard pushes Victim away ($B = 1$ if he does and $B = 0$ if not), and whether Victim dies ($D = 1$ if he does and $D = 0$ if not); Fig. 1 is the causal graph for OM.

OM

- $A = 1$,
- $B = 0$,
- $D = A \wedge \neg B$

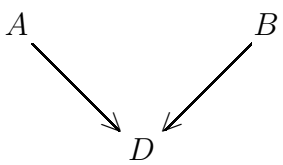


Fig. 1

By the causal modeling semantics, the causal model OM satisfies three atomic sentences: $OM \models A = 1$ and $OM \models B = 0$, for the first two equations assign 1 and 0 to A and B respectively, and $OM \models D = 1$ since 1 is the result of calculating D's value according to the third equation and values of A and B.[4] Moreover, $OM \models B = 1 > D = 0$, as the consequent $D = 0$ is satisfied by the submodel $OM_{B=1}$ which resets the value of B to 1 according to the antecedent $B = 1$. In other words, 0 is the result of recalculating the value of D according to its equation $D = A \wedge \neg B$, with B's value reset to 1 and A's value intact.

The following list of definitions will be handy for later use.

> **Directed Paths (DP).** Let G_M be the causal graph of the causal model M. The sequence of variables $< X_1, ..., X_n >$ is a directed path, a path from the variable X_1 to the variable X_n, in G_M just in case, for any $1 \leq i \leq n$, the variable X_i is a parent of the variable X_{i+1}. $Node(< X_1, ..., X_n >)$ is the set of nodes on the path $< X_1, ..., X_n >$ (cf. Hitchcock 2007: 509).
>
> **Causal Networks (CN).** Let G_M be the causal graph of the causal model M. N is a causal network connecting the variable X to the variable Y in G_M just in case $N = \bigcup\{Node(Pt) | Pt$ is a directed path from X to $Y\}$ (cf. Hitchcock 2007: 509).

Later, DP and CN will be used to define self-containment and integrity.

2.2 Default/Deviant Distinction

For the default/deviant distinction, according to Hitchcock (2007),

> As the name suggests, the default value of a variable is the one that we would expect in the absence of any information about intervening causes. More specifically, there are certain states of a system that are self-sustaining, that will persist in the absence of any causes other than the presence of the state itself: the default assumption is that a system, once it is in such a state, will persist in such a state.
>
> (Hitchcock 2007: 506)

Some pragmatically oriented rules of thumb to identify default values of variables are as follows (cf. Hitchcock 2007: 507).

- Actions or events which do not last long are typically deviant.
- Intentional actions, or bodily movements requiring volition, are typically deviant.
- Actions or events (e.g. causes and effects) in need of explanation are typically deviant.

[4] The third equation can be reformulated more mathematically as '$D = min(A, 1 - B)$', but the expression utilizing Boolean connectives should be understandable.

- Positive events are typically deviant.

The default/deviant distinction is made on the pragmatic level: the verdict of the default is to be varied according to the theory we adopt, the level of analysis, etc. (cf. Hitchcock 2007: 506). The default value of a variable is a value that represents the expected state which is *thought* to be normal, according to some theoretical or pre-theoretical principles, with the absence of other *information*, rather than a value that represents the "genuine" normal state of a system.

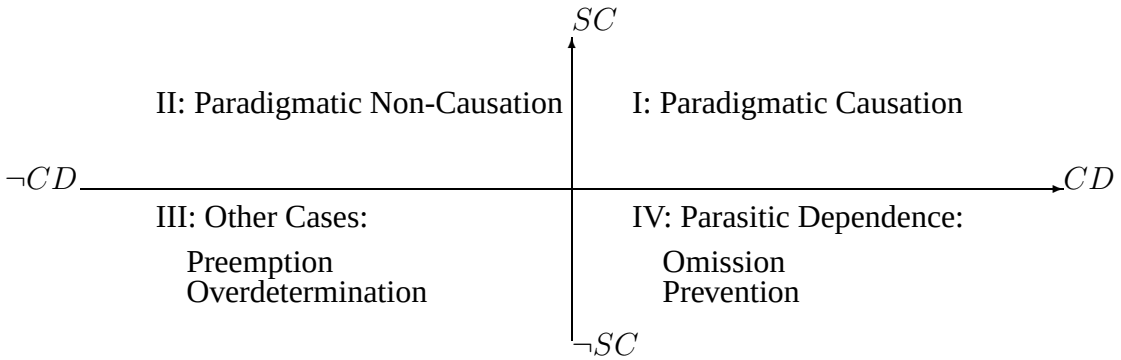
3 Self-Containment

It has long been observed that causation seems strongly connected with counterfactual dependence, and in various cases the two kinds of relations go hand in hand. Nonetheless, the *simple counterfactual analysis* (SCA henceforth) is a too hasty over-generalization of the observation.

> **(SCA)** C and E are two distinct actual events. C is a cause of E if and only if E counterfactually depends on C (i.e., E would have occurred had C occurred, and E would not have occurred had C not occurred).

SCA is rejected because of the existence of counterexamples. On the one hand, cases of omission and prevention illustrate that counterfactual dependence is not sufficient for causation of two actual events; cases of preemption and overdetermination show that it is not necessary, on the other hand.[5]

To remedy the problems of SCA, Hitchcock (2007) adopts a very different route. Instead of providing a comprehensive theory of causation which covers cases including counterfactual dependence or not, he focuses on identifying the condition, which he finds to be self-containment (to be detailed shortly), under which counterfactual dependence between events is necessary and sufficient for the events to be causes and effects.[6] With the two factors self-containment (SC) and counterfactual dependence (CD), the cases for causal concern in question are divided into four quadrants.



[5]For counter-examples to SCA, see, among others, Lewis (1973, 2000) and Hitchcock (2001, 2007).

[6]The attempts for a comprehensive theory of causation based on counterfactual dependence can be found, for example, in Lewis (1973, 2000), Halpern & Pearl (2005), and Hitchcock (2001).

Fig. 2

As in Fig. 2, when we are enquiring whether C causes E, we should firstly consider the self-containment of the story connecting C to E. If the story is self-contained, the case falls in either quadrant I or quadrant II, depending on the counterfactual dependence of E on C. If otherwise, the case falls in one of the other two quadrants. For cases in quadrant IV, such as omission and prevention, the given story is not self-contained but there is counterfactual dependence between the two events. Hitchcock calls this kind of counterfactual dependence *parasitic dependence* and refrains from judging whether it is genuine causation. For cases in quadrant III, such as preemption and overdetermination, Hitchcock suggests that we should make judgements about causation by appealing to theories such as those proposed by Hitchcock (2001) or Halpern and Pearl (2005).

To introduce Hitchcock's notion of self-containment, take the case of omission OM in section 2 as an illustration. Given OM, two objections to SCA can be raised. First, $D = 1$ does counterfactually depend on $B = 0$ ($OM \models B = 1 > D = 0$), but $B = 0$ may not seem to be a cause of $D = 1$. Someone may insist that $B = 0$ is a genuine cause, but in various cases with the same causal structure, such omissions are clearly not causes. Think about the queen case: John did not water his flowers, and so they withered; but the queen of the United Kingdom did not water his flowers either. The queen's not watering his flowers should not be a cause of his flowers' withering, even though whether his flowers withered counterfactually depends on whether the queen watered them.

Second, according to Hitchcock, even if it is acceptable that $B = 0$ is a genuine cause, there is still some intuitive difference between $A = 1$ as a cause and $B = 0$ as a cause. It is the case that $D = 1$ counterfactually depends on both $A = 1$ and $B = 0$. However, on the one hand, without mentioning B, the story involving only A and D strikes us as a self-contained story including enough information to explain itself: Assassin shoots and Victim dies; if Assassin did not shoot, Victim would not die. On the other hand, the story only involving B and D seems explanatorily incomplete: Bodyguard does not push Victim away, Victim dies; if he did, Victim would not die. The second story, without mentioning the fact that Assassin shoots, seems to leave why Bodyguard's action could make the difference on Victim's survival unexplained. In Hitchcock's words, the counterfactual dependence of $D = 1$ on $B = 0$ is *parasitic* upon $A = 1$ (cf. Hitchcock 2007: 504-505). But this intuitive difference is not explained merely in terms of counterfactual dependence. Hitchcock's diagnosis indicates that underlying the intuitive difference is their difference with respect to self-containment: the story connecting A to D is self-contained, so $A = 1$ is a paradigmatic cause of $D = 1$ based on the counterfactual dependence; on the contrary, the story connecting B to D is not self-contained, so no matter whether or not $B = 0$ is a cause of $D = 1$ *simpliciter*, it is not a paradigmatic cause.

Clearly, the causal model OM *per se* does not discriminate between the counterfactual dependence of $D = 1$ on $A = 1$ and the counterfactual dependence of $D = 1$ on $B = 0$. Hitchcock

identifies self-contained stories based on the *default/deviant* distinction between values of each variable (cf. Hitchcock 2007: 510).

> **(Self-Containment)** Given a causal model M and two variables X and Y in M, the causal network N connecting X to Y in G_M is self-contained if and only if for any $Z \in N$ and all its parents $Z_1, ..., Z_n \in N$, $M \models (Z_1 = def(Z_1) \wedge, ..., Z_n = def(Z_n)) > Z = def(Z)$, where def is a function assigning default values to variables.

Informally, the notion of self-containment is meant to capture a special version of the *principle of sufficient reason* adopted by Hitchcock, which claims that any deviant event must be brought about by events including at least one deviant event, rather than by all default events (cf. Hitchcock 2007: 507-508).

To apply the definition to OM, consider the distribution of default and deviant events according to Hitchcock's rules of thumb: that Assassin does not shoot, that Bodyguard does not push, and that Victim does not die are default states ($def(A) = def(B) = def(C) = 0$). In contrast, shooting, pushing, and dying are all deviant ($dev(A) = dev(B) = dev(C) = 1$). (In the remainder of this paper, we just make it implicit that 0 is default and other values are deviant for each variable.) Applying the definition of self-containment to OM, the causal network connecting A to D is self-contained, since $OM \models A = def(A) > D = def(D)$, where $def(A) = def(D) = 0$. On the contrary, the causal network connecting B to D is not self-contained, since $OM \not\models B = def(B) > D = def(D)$, where $def(B) = def(D) = 0$.

Generally, for two events C and E such that E counterfactually depends on C, there are two cases: if the causal network connecting C to E is self-contained, C is a paradigmatic cause of E; if the causal network is not self-contained, the counterfactual dependence is parasitic. For the latter, though Hitchcock refuses to judge whether C is a cause of E *simpliciter*, he claims that C is not a paradigmatic cause of E. The intuitive difference between $A = 1$ as a cause and $B = 0$ as a cause in OM can now be explained based on the difference in self-containment between the two networks. As the causal network connecting A to D is self-contained and $D = 1$ counterfactually depends on $A = 1$, $A = 1$ is a paradigmatic cause of $D = 1$. On the contrary, as the causal network connecting B to D is not self-contained, though $D = 1$ counterfactually depends on $B = 0$, the counterfactual dependence is *parasitic* upon the fact that $A = 1$, and thus that $B = 0$ is not a paradigmatic cause of $D = 1$.

We would like to supplement Hitchcock's notion of paradigmatic causes with the restriction on applying it only to the causal concern with effects that are deviant events. In general, from the perspective of PSR, a causal concern arises only on events which are deviants. For example, in a variation of OM that Assassin and Backup both actually do nothing and Victim is alive, it is not a causal concern (or a explanatory concern) to ask what causes Victim's being alive. Without this restriction of application, Hitchcock's criterion for paradigmatic causes might, dubiously, lead to identifying that Assassin's not shooting is a paradigmatic cause of Victim's being alive.

4 Parasitism

In OM, though the death of Victim counterfactually depends on both Assassin's shooting and Bodyguard's refraining from action, the different causal roles of the latter two events are identified by self-containment. As Hitchcock (2007) indicates, though whether Bodyguard pushes is a difference-maker of the death of Victim (i.e. Victim's death counterfactually depends on Bodyguard's push), it does not offer a satisfactory explanation for why Victim dies, since the network connecting the two events is not self-contained. In contrast, the self-contained network connecting whether Assassin shoots and whether Victim dies guarantees that the former satisfactorily explains the latter.

But the following *half-dose* case shows that the line drawn by the notion of self-containment between difference-makers which possess satisfactory explanatory power and those do not is not fully adequate. Specifically, the half-dose case illustrates that even in a self-contained network, a particular difference-maker may fail to fully explain its causal consequence.

> **(HD)** The lethal dose of a kind of toxin is 10 grams, meaning that if one took 10 grams or more of the toxin, she would die, and she would survive otherwise. Each of Assassin and Badguy administers 5 grams of the toxin in Victim's coffee. Victim drinks the poisoned coffee and dies. If either of Assassin or Badguy did not administer the poison, Victim would not die.

Let the causal model HD represent the case, with the variable A, B, and D standing for whether Assassin administers ($A = 1$ if Assassin administers and $A = 0$ if not), whether Badguy administers ($B = 1$ if Badguy administers and $B = 0$ if not), and whether Victim dies ($V = 1$ if Victim dies and $V = 0$ if not), respectively.

HD

- $A = 1$,
- $B = 1$,
- $D = A \wedge B$

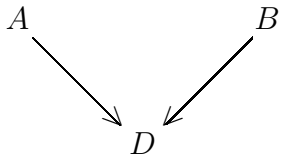


Fig. 3

Applying Hitchcock's theory to HD, both of the causal networks $\{A, D\}$ and $\{B, D\}$ are self-contained, and $D = 1$ counterfactually depends on both $A = 1$ and $B = 1$. As a result, according to Hitchcock (2007), both $A = 1$ and $B = 1$ are paradigmatic causes of $D = 1$.

Nonetheless, the networks $\{A, D\}$ and $\{B, D\}$ come with some parasitic feature which makes neither $A = 1$ nor $B = 1$ provide a satisfactory explanation for $D = 1$. In the HD case, both of the two subplots connecting A or B, respectively, to D strike us as incomplete stories. For both $\{A, D\}$ and $\{B, D\}$, administering a half of the lethal dose does not by itself

lead to Victim's death; Victim's death counterfactually depends on either administration of half of the lethal dose only because of the administration of the other half of the lethal dose. As a result, on their own, neither the fact that Assassin administers half of the lethal dose nor that Badguy administers half the lethal dose can fully explain Victim's death, although both of the two administrations are difference-makers for the death.

If the reason to accept Hitchcock's notion of self-containment is that, by using it, we can secure that each particular paradigmatic cause not only is a difference-maker of its effect but also can provide its effect a satisfactory explanation, the half-dose case constitutes a counterexample showing that the notion of self-containment fails to guarantee the explanatory role of paradigmatic causes: though both $\{A, D\}$ and $\{B, D\}$ are self-contained, neither $A = 1$ nor $B = 1$ provides $D = 1$ a satisfactory explanation. If the line between paradigmatic causation and parasitic dependence is meant to reflect the difference in their explanatory power, Hitchcock's notion of self-containment is not yet fully successful.

5 Integrity

We would like to offer the new notion of *integrity* to avoid the challenge to self-containment. To do that, the notion of integrity should be defined in such a way that (at least) the following four results follow: in OM, $\{A, D\}$ is integral but $\{B, D\}$ is not, and in HD, both $\{A, D\}$ and $\{B, D\}$ are not integral. Some observations about (un)expected counterfactual dependence in OM and HD, stated in an informal manner, may be useful to elaborate the notion first.

> **(O1)** In OM, *if Assassin did not shoot, Victim would not die* seems natural; thus the counterfactual dependence of $D = 1$ on $A = 1$ is in accordance with our expectation and $A = 1$ explains $D = 1$ well. On the other hand, without the additional information about Assassin's shooting, *if Bodyguard pushed, Victim would not die* is somehow confusing. Thus the counterfactual dependence of $D = 1$ on $B = 0$ is not expected and $B = 1$ is not enough to explain $D = 1$.
>
> **(O2)** In HD, when considered separately, neither *if Assassin did not administer his half dose, Victim would not die*, nor *if Badguy did not administer his half dose, Victim would not die* is convincing; thus the counterfactual dependence is not expected and both $A = 1$ and $B = 1$ lack full explanatory power.

For OM and HD, it seems that when counterfactual dependence of an event E on another event C is expected, C has the full power to explain E. If otherwise, although C is a difference-maker of E, C cannot offer satisfactory explanation for E, just as Victim's death cannot be explained well by Bodyguard's refraining from pushing in OM, or by either Assassin's or Badguy's administration in HD.

If the counterfactually dependence of E on C is unexpected, we could say that C influences E in a way that "should" not be the case. For example, in OM, by merely considering the two events about whether Bodyguard pushes and whether Victim dies, it should be the case that Bodyguard's pushing is not a difference-maker for Victim's death at all. In other words, without disturbances imposed by other events, whether Victim dies should not counterfactually depend on whether Bodyguard pushes. Since Assassin's shooting disturbs the counterfactual relationship between Bodyguard's pushing and Victim's death, it *creates* the counterfactual dependence of Victim's death on Bodyguard's pushing. Though there should be no counterfactual dependence between Victim's death and Bodyguard's pushing, there turns out to be such an unexpected counterfactual dependence because of the disturbance.

The notion of integrity is proposed to capture whether the counterfactual relationships in a given causal network are disturbed by other events: if any of them is disturbed by other events, the network is not integral; and if there is no disturbance by other events, the network is integral. Given a causal model and a causal network in it, for a particular variable X in the network, it could have several parents in the model. These parents can be divided into network-internal parents – parents in the given network, and network-external ones – parents outside the network. Each network-external parent is a candidate for a disturbance, since it may influence how X is determined by its network-internal parents. For network-external parents which have default values, they should not be disturbances, since they are in states that we expect, and they contribute nothing to the unexpectedness of the counterfactual relationships, if any, between X and its network-internal parents. However, network-external parents which have deviant values may and may not be disturbances, so integrity is defined based on comparing the actual circumstance with the counterfactual situation where network-external deviants are changed into their default states.

> **(Integrity)** Let M be a causal model, N be a causal network in G_M connecting X to Y. For any variable $Z \in N$ with network-internal parents $I_1, ..., I_n$ and network-external parents $O_1, ..., O_m$, let M_Σ be the submodel of M where $O_1, ..., O_m$ are all altered by intervention to correspond with the default, i.e., set exogenously with their default values. The causal network N is integral if and only if $(*)$ holds for each $Z \in N$ with network-internal parents:
>
> > **$(*)$** If Z has any network-external parents, $M \models (I_1 = i_1 \wedge ... \wedge I_n = i_n) > Z = z$ if and only if $M_\Sigma \models (I_1 = i_1 \wedge ... \wedge I_n = i_n) > Z = z$, for any possible value i_p of I_p $(1 \leq p \leq n)$ and any possible value z of Z.

The notion of integrity so-defined means to indicate that the deviance of network-external variables does not disturb the counterfactual relationships among network-internal variables.

This notion of integrity captures whether the given network is explanatorily complete. On the one hand, in an integral network connecting a variable X to another Y, the specific pattern of

how each variable is determined by its network-internal parents is not disturbed by its network-external parents' being deviant: it depends on its network-internal parents in the way it should do. In this case, the causal network includes sufficient information, and X is properly positioned to offer satisfactory explanation to the value of Y. On the other hand, if the network is not integral, there are some deviant variables outside the network disturbing the way in which some variable depends on its network-internal parents. If that is the case, the information about those deviants outside is ignored by the network and the counterfactual relationships among the network-internal variables are unexpected. Thus a satisfactory explanation for the actual value of Y must include information about those network-external disturbances, and the value of X by itself is not enough.

Therefore, the notion of integrity draws a line between networks connecting two variables in which one cannot satisfactorily explain the other, and networks connecting two variables in which one explains the other well. As a result, integrity lead to adequate results for the cases mentioned in previous sections: for OM, the causal network $\{A, D\}$ is integral, but $\{B, D\}$ is not. For HD, both $\{A, D\}$ and $\{B, D\}$ are not integral. Playing the same role with Hitchcock's notion of self-containment does, integrity classifies $A = 1$ (that Assassin shoots) as a paradigmatic cause of $D = 1$ (that Victim dies) for OM, as the notion of self-containment does. However, disagreeing with self-containment, for HD, it implies that the counterfactual dependence of $D = 1$ (that Victim dies) on $A = 1$ (that Assassin administers a half dose) is parasitic dependence, and the same for the counterfactual dependence of $D = 1$ on $B = 1$ (that Badguy administers a half dose). If the objection to self-containment in section 4 correctly indicates that the reason to distinguish paradigmatic causation from parasitic dependence is that the former provides satisfactory explanations but the latter does not, the new notion of integrity is in the right position to supplant Hitchcock's notion of self-containment.

Further comparison between integrity and self-containment can provide a diagnosis of why integrity does a better job on capturing an event's being satisfactorily explained. To begin with, let a causal model (rather than a causal network) be self-contained just in case if all parents of a variable X in the model take default values, then X takes a default value. Formally, integrity and self-containment are related in the following manner: if M is a self-contained causal model and N is an integral causal network in M, then N is a self-contained causal network.[7] This formal feature means the following: when the causal relationship in a world is constructed in such a way that no deviant event occurs without occurrences of its deviant parents, the observation that integrity implies self-containment means that every anomalous counterfactual relationship in an integral network comes from and thus can be fully explained by deviance *in* the network.

[7]It is not hard to seen why this formal feature holds. Consider a model M which is self-contained, and a given network N that is integral but not self-contained. Given that N is not self-contained, it means that some variable X in N is such that even if all its network-internal parents take default values, X does not take a default value. Since the network is integral, this counterfactual feature for X will hold even if all network-external parents are intervened to take default values. However, this violates the assumption that M is self-contained.

The contrast between integrity and self-containment can also further illuminate two related issues of relevant concern: first, the issue concerning how the default/deviant distinction contributes to how causes provide satisfactory explanation for effects; second, the issue concerning how the default/deviant distinction contributes to our judgement of causation in paradigmatic cases of causation.

For the first issue, observe that self-containment is proposed to capture a special version of the principle of sufficient reason, which requires each could-be anomaly to have at least one abnormal parent in the given network – there must be some abnormal parent explaining the anomaly in question. However, this special version of the principle of sufficient reason allows the explanation from causes to effects to be *partial*. For example, for the self-contained network $\{A, D\}$ in HD, if D takes its deviant value 1, it must be the case that its network-internal parent A is deviant. However, self-containment does not require the network-internal abnormal parent to possess the *full* explanatory power for the network-internal anomaly to arise, for neither A nor B can fully explain D's being deviant. The deviant value of B is required for D's value to be deviant with respect to A's taking a deviant value. On the other hand, integrity requires network-internal parents to possess full explanatory power: D's being deviant counterfactually depends on A, but that dependence is based on B's being deviant, so the network $\{A, D\}$ is not integral and it does not offer enough information to fully explain D's being deviant. Therefore, integrity is a condition more precise than self-containment in terms of capturing the notion of satisfactory explanation.

For the second issue, given the discussion immediately above, in Hitchcock (2007)'s approach to causation by the qualification from self-containment, any paradigmatic cause must not only be a difference-maker for its effect, but also has to possess explanatory power to *some* degree for its effect, but possibly not *completely*. The qualification made by self-containment ensures that the corresponding paradigmatic causes do *not* necessarily provide full explanations for their effects, as the half-dose case shows. Given that all cases of paradigmatic causation and parasitic dependence are cases with counterfactual dependence, the explanatory significance of the distinction between the two categories proposed by Hitchcock (2007) is blurred. Then, in what sense can we say that a paradigmatic causation is *paradigmatic*? It may be argued that the distinction is made to capture whether C possess *any* explanatory power for E when E counterfactually depends on C. However, this point of view presupposes a strong assumption that parasitic dependence possesses *no* explanatory power at all. This assumption seems likely to be incorrect. In OM, it seems that the Bodyguard's refraining from pushing at least explains Victims death to an extent. As self-containment does not require paradigmatic causes to possess full explanatory power, the line drawn by Hitchcock between paradigmatic causation and parasitic dependence fails to mark the difference in their explanatory roles clearly.

In contrast, if the condition about explanatory power is imposed in terms of integrity, each paradigmatic cause must possess the full explanatory power for its effect, and cases of parasitic

dependence are cases that at best provide partial explanations. In this sense, each paradigmatic cause is *the* cause of its effect, for it is a difference-maker with the full explanatory power. On the other hand, for cases of parasitic dependence, each difference-maker in question is at best *a* cause (or just a causal factor), for it cannot fully explain its effect. Since the notion of integrity distinguishes paradigmatic causation with parasitic dependence in accordance with *the* cause and *a* cause, the distinction is more strongly motivated: because they are *the* causes that make differences for and fully explain their effects, paradigmatic causes may be called "paradigmatic".

6 Trumping

For the second part of the synthesis, we turn to the attempt to incorporate integrity into Hitchcock (2011)'s contrastive account for trumping. Before doing that, we first indicate that the notion of self-containment is in tension with Hitchcock's contrastive account for trumping.

The trumping case introduced by Schaffer (2000) constitutes a thorny problem for various accounts of causation. Paraphrasing a realistic version elaborated in Lewis (2000), the trumping case goes as follows.

> **(TP)** Soldiers obey what is ordered by the officer with the highest rank among all officers who issue any order. A sergeant and a major, being the only two officers at the position to order, simultaneously shout 'Advance!'; the soldiers hear both, and advance.

A simple way to model TP is to represent it by the causal model TP_0 with three bi-valued variables J ($J = 0$ for the major ordering nothing, $J = 1$ for the major ordering 'advance'), S ($S = 0$ for the sergeant ordering nothing, $S = 1$ for the sergeant ordering 'advance'), and A ($A = 0$ for no action, $A = 1$ for advancing).

TP_0

- $J = 1$,
- $S = 1$,
- $A = J \vee S$

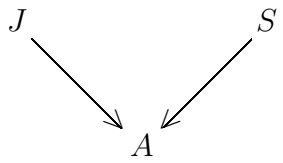


Fig. 4

However, if the trumping case involves any destructive power over and above the overdetermination case, any adequate representation must fully respect the law mentioned in TP – that soldiers obey whatever is ordered by the officer with the highest rank among all officers who issue any order. Unfortunately, the causal model TP_0 represents the trumping case TP by identifying it with the typical overdetermination case, since it accommodates the weaker law that

soldiers advance if *any* officer orders 'advance', which should be the law governing the overdetermination case where $J = 1$ and $S = 1$ share equal approval or disfavor for being a cause of $A = 1$. Therefore, if it is possible that the trumping case is to be appreciated as an issue distinct from overdetermination at all, the TP case requires a more fine-grained representation than the simple bi-valued model TP_0.

To display the full power of the law in TP, the causal model TP properly represents TP by using three multi-valued variables J, S, and A, respectively to represent the order of the major ($J = 0$ for the major ordering nothing, $J = 1$ for the major ordering 'advance', and $J = 2, ..., J = n$ for other orders), the order of the sergeant ($S = 0$ for the sergeant ordering nothing, $S = 1$ for the sergeant ordering 'advance', and $S = 2, ..., S = n$ for other orders, corresponding to those of the major, respectively), and the action of the soldiers ($A = 0$ for no action, $A = 1$ for advancing, and $A = 2, ..., A = n$ for other actions, with respect to the those orders of the major/sergeant, respectively).

TP

- $J = 1$,
- $S = 1$,
- $A = \begin{cases} J, & \text{if } J \neq 0 \\ S, & \text{if } J = 0 \end{cases}$

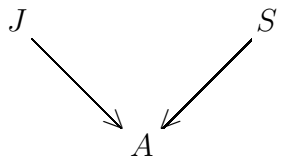


Fig. 5

In TP, the counterfactual structure, with values representing the orders of officers other than 'advance' and actions of the soldiers other than advancing, accommodates the possibility that the major's order conflicts with the sergeant's, and the equation for A respects the full power of the law that soldiers obey the order issued by the highest-ranking officer among those who issue orders.

Hitchcock (2011) advocates contrastivism for causation in analyzing the trumping case. While Schaffer (2000) adopts the trumping case as a preemption case, Hitchcock opposes him by arguing that, when equipped with the contrastivist approach, the trumping case needs not to be seen as a preemption case; rather, it should be analyzed in a fine-grained manner by appealing to contrastivism.[8]

According to Hitchcock (2011), cases of early preemption, late preemption, and overdetermination are all cases called "*redundant causes*", meaning that, in any of these cases, there are two events A and B (or more) such that one or both (singly or collectively) cause some other event C, and either would have brought about C had the other not occurred. In those cases, Hitchcock calls both A and B redundant causes of C, without committing to either of them being a genuine cause – for a redundant cause may or may not be a cause *simpliciter* (cf. Hitchcock 2011: 229). Besides, if some event A is a cause of another event B but not a redundant

[8]For contrastivism of causation, see, among others, Schaffer (2005).

cause, it is called a *non-redundant cause*. A is *causally irrelevant* to B if A is neither a cause nor a redundant cause of B. The main point Hitchcock defends is that, varying with the selection of contrasts, the trumping case may be either a case of a non-redundant cause or a case of overdetermination, but it is definitely not preemption. In particular, Hitchcock (2001) argues for the following claims.

> **(i)** $J = 1$ rather than $J = 2$ is a non-redundant cause, thus, a cause, of $A = 1$ rather than $A = 2$;
>
> **(ii)** $S = 1$ rather than $S = 2$ is causally irrelevant to, thus, not a cause of, $A = 1$ rather than $A = 2$;
>
> **(iii)** $J = 1$ rather than $J = 0$ and $S = 1$ rather than $S = 0$ overdetermine $A = 1$ rather than $A = 0$.[9]

By adopting the contrastivist approach, these results ameliorate the seeming tension between the view that the trumping case is a case of overdetermination, and that it is a case of asymmetry in that J=1 and S=1 can have different causal status.

If the claims (i) – (iii) are collectively acceptable, it is interesting to check whether we can at the same time accept the pragmatically oriented approach presented in Hitchcock (2007). We will argue that when applying the self-containment to the trumping case, the claims mentioned above fail to obtain together.

Consider claim (iii) first. It says that the trumping case is a case of overdetermination, if $J = 0$, $S = 0$, and $A = 0$ are selected to be the contrasts. According to Hitchcock (2007), overdetermination falls into the category of no counterfactual dependence and no self-containment, i.e., the category represented by quadrant III in Fig. 2. We may take the following understanding of counterfactual dependence in the contrastivist style, which seems to be at least reasonable.

> **(Counterfactual Dependence, the contrastive version)** Suppose $X = x$ and $Y = y$ are facts, and $X = x'$ and $Y = y'$ are alternative versions of them. $X = x$ rather than $X = x'$ counterfactually depends on $Y = y$ rather than $Y = y'$ if and only if $Y = y' > X = x'$.

According to this definition, it is easy to see that $A = 1$ rather than $A = 0$ does not counterfactually depend on either $J = 1$ rather than $J = 0$ or $S = 1$ rather than $S = 0$. Moreover, we can see that neither causal networks $\{J, A\}$ and $\{S, A\}$ is self-contained. When events about human actions are involved, the default is the state without any action, so $def(J) = def(S) = def(A) = 0$. Thus, the causal network connecting J to A is not self-contained, since it would

[9] In Hitchcock (2011), the claims are made in a more general form to accommodate the possibility that there are more than two events connecting to the effect.

still be the case that $A = 1$ were it the case that $J = 0$. Similarly, the causal network connecting S to A is not self-contained either. Therefore, by applying the theory in Hitchcock (2007), claim (iii) is comfortably confirmed.

Nonetheless, it is hard for claim (i) to fit in Hitchcock (2007)'s proposal. By referring to $J = 1$ rather than $J = 2$ as a non-redundant cause of $A = 1$ rather than $A = 2$, (i) implies that $J = 1$ rather than $J = 2$ is a genuine cause of $A = 1$ rather than $A = 2$. Thus, there are three possibilities: either it falls into quadrant I, III, or IV, since quadrant II is a category purely for non-causation. First, it must not fall into the category of no counterfactual dependence and no self-containment (quadrant III in Fig. 2), because $A = 1$ rather than $A = 2$ counterfactually depends on $J = 1$ rather than $J = 2$. Second, it must not fall into the category of paradigmatic causation (quadrant I), since, as indicated in the analysis for the claim (iii), the network $\{J, A\}$ is not self-contained. Third, it is also not reasonable to place the case in the category of parasitic dependence (quadrant IV) either. In TP, the counterfactual dependence of the soldiers' advancing rather than shooting on the major's ordering 'advance' rather than 'shoot' is *not* parasitic on the fact that the sergeant orders 'advance'. Disregarding whether the sergeant so orders, the counterfactual dependence obtains anyway.

Hitchcock's claims (i) and (iii) are thus in tension. The root of this tension comes from the fact that bearers of self-containment are causal networks connecting events without being relativized to particular selections of contrasts, so whether the network $\{J, A\}$ is self-contained or not is irrelevant to which selection of contrasts is considered. Moreover, the notion of self-containment does not accomplish the goal of discriminating complete stories and incomplete ones. When we consider the network $\{J, A\}$ by focusing on the particular selection of contrasts, i.e., $J = 1$ rather than $J = 2$ and $A = 1$ rather than $A = 2$, it seems to be complete, for the counterfactual relationships between $J = 1$ and $A = 1$ ($J = 1 > A = 1$) and between $J = 2$ and $A = 2$ ($J = 2 > A = 2$) are just what we would expect without the information about S, and hence $J = 1$ rather $J = 2$ can satisfactorily explain $A = 1$ rather than $A = 2$. In contrast, with respect to the selection of contrasts $J = 1$ rather than $J = 0$ and $A = 1$ rather than $A = 0$, the same network $\{J, A\}$ is incomplete, for the actual situation where whether $J = 1$ or $J = 0$ makes no difference to A is not what we expect; additional information about S is needed in order to explain it.

While we do not see an obvious way to relativize self-containment in a contrastivist fashion, and believe that integrity does better than self-containment in articulating paradigmatic causation, the attempt for the synthesis is to relativize integrity to particular selections of contrasts in order to accord with Hitchcock (2011)'s three claims for the trumping case.

7 Relativized Integrity

The unrelativized integrity is inappropriate for a contrastive analysis of causation. Returning to the TP case, the causal network connecting J to A is not integral, since if the network-external parent S of A were to take the default value 0, the counterfactual relations between A and J would be different: if the major were to refrain from issuing any order, the soldiers would do nothing, rather than advance. The claim (i) cannot be achieved by the unrelativized integrity.

It helps to observe that, when we are interested in *how* S disturbs the causal outcome of J on A, it is substantive to compare the specific pattern of the counterfactual relationship between J and A with respect to that S takes its default value and S takes its actual deviant value. On the one hand, when S takes its actual value 1, the counterfactual dependence of $A = 1$ rather than $A = 2$ on $J = 1$ rather than $J = 2$ is not disturbed, for it would be the same were S to take its default: $J = 1 > A = 1$, and $J = 2 > A = 2$. In contrast, the counterfactual dependence of $A = 1$ rather than $A = 0$ on $J = 1$ rather than $J = 0$ would be disturbed, for $J = 0 > A = 1$, unlike what it would be if S were to take its default.

The above observation suggests the idea that the causal network connecting J to A is integral *with respect to* $J = 1$ *rather than* $J = 2$ *and* $A = 1$ *rather than* $A = 2$, since the part of the counterfactual relationship involving the two pairs of contrasts is not disturbed by S's actual value; and the causal network connecting J to A is not integral *with respect to* $J = 1$ *rather than* $J = 0$ *and* $A = 1$ *rather than* $A = 0$, since S's being deviant disturbs the part of counterfactual relationship about the two pairs of contrasts in question. This idea is formally stated as follows.

> **(Integrity, the relativized version)** Let M be a causal model, and N be a causal network in G_M connecting X to Y. Suppose $M \models X = x \wedge Y = y$, and the contrasts in question are $X = x'$ and $Y = y'$ ($x' \neq x$ and $y' \neq y$). For every $Z \in N$, suppose it has network-internal parents $I_1, ..., I_n$ and network-external parents $O_1, ..., O_m$. Furthermore, $M \models Z = z \wedge I_1 = I_1 \wedge ... \wedge I_n = i_n$, and $M \models X = x' > (Z = z' \wedge I_1 = I_1' \wedge ... \wedge I_n = i_n')$. Let M_Σ be the submodel of M where $O_1, ..., O_m$ are all intervened to take their default values. The causal network N is *integral* with respect to $X = x$ rather than $X = x'$ and $Y = y$ rather than $Y = y'$ if and only if both (#) and (##) hold for each $Z \in N$ with network-internal and network-external parents:
>
> > **(#)** $M_\Sigma \models (I_1 = i_1 \wedge ... \wedge I_n = i_n) > Z = z$ if and only if $M \models (I_1 = i_1 \wedge ... \wedge I_n = i_n) > Z = z$;
> >
> > **(##)** $M_\Sigma \models (I_1 = i_1' \wedge ... \wedge I_n = i_n') > Z = z'$ if and only if $M \models (I_1 = i_1' \wedge ... \wedge I_n = i_n') > Z = z'$;

Consider TP again. The actual values of the variables constitute a state of the network $\{J, A\}$: $J = 1$ and $A = 1$. Moreover, the given selection of contrasts offers an alternative state where

$J = 2$ and $A = 2$ (since $J = 2 > A = 2$). If both states remain intact when network-external parents of variables with network-internal parents are intervened to be default, the network is integral with respect to the given selection of contrasts. As a result, for TP, the causal network $\{J, A\}$ is integral with respect to $J = 1$ rather than $J = 2$ and $A = 1$ rather than $A = 2$, for if S were to take 0, it would still be the case that $J = 1 > A = 1$ and $J = 2 > A = 2$, thus S's being deviant does not disturb the counterfactual relationships among the selected contrasts. On the contrary, the network $\{J, A\}$ is not integral with respect to $J = 1$ rather than $J = 0$ and $A = 1$ rather than $A = 0$: if S were to take 0, it would be the case that $J = 0 > A = 0$, but $J = 0 > A = 1$ is actually true. To fit in the contrastive approach, the relativized integrity allows variation across different pairs of selected contrasts.

Since $A = 1$ rather than $A = 2$ counterfactually depends on $J = 1$ rather than $J = 2$ and $\{J, A\}$ is integral with respect to this selection of contrasts, $J = 1$ rather than $J = 2$ is a paradigmatic cause, and hence a non-redundant cause, of $A = 1$ rather than $A = 2$. So (i) is a natural result of this new account. On the other hand, (iii) is confirmed by the following. The causal network connecting J to A is not integral with respect to $J = 1$ rather than $J = 0$ and $A = 1$ rather than $A = 0$; the causal network connecting S to A is not integral with respect to $S = 1$ rather than $S = 0$ and $A = 1$ rather than $A = 0$. Furthermore, $A = 1$ rather than $A = 0$ counterfactually depends on neither $J = 1$ rather than $J = 0$ nor $S = 1$ rather than $S = 0$. This result matches the characterization of overdetermination cases in Hitchcock (2007): they are cases in quadrant III, in which there is no counterfactual dependence and no integrity for the events of interest. The synthesis is done.

8 Concluding Remarks

The synthetic attempt in this paper makes the general framework of Hitchcock's pragmatic-oriented account of causation promising in that the distinction between paradigmatic causation and parasitic dependence is reanalyzed by introducing the notion of integrity. The new line more adequately captures the explanatory roles of causes, and, by reflecting the distinction between *the* cause and *a* cause of an event, it also makes the term "paradigmatic" more sensible. Besides, the notion of integrity is strengthened by relativizing it to particular selections of contrasts. By doing so, the pragmatically oriented framework originated in Hitchcock (2007) fits in well with contrastivism, as the trumping case shows. If the pragmatic framework is acceptable, integrity should be more attractive than the notion of self-containment for proponents of contrastivism such as Hitchcock.

References

Hall, N. (2007). Structural equations and causation. *Philosophical Studies*, 132, 109-136.

Halpern, J. Y. (2008). Defaults and Normality in Causal Structures. In *Principles of Knowledge Representation and Reasoning: Proc. Eleventh International conference (KR '08)*, pp. 198-208.

Halpern, J. Y., and C. Hitchcock (2013). Compact representations of extended causal models. *Cognitive science*, 37(6), 986-1010.

Halpern, J. Y., and C. Hitchcock (forthcoming). Graded causation and defaults. *The British Journal for the Philosophy of Science.*

Halpern, J., and J. Pearl (2005). Causes and explanations: a structural-model approach – Part I: Causes. *British Journal for the Philosophy of Science*, 56, 843-887.

Hitchcock, C. (2001). The intransitivity of causation revealed in equations and graphs. *Journal of Philosophy*, 98, 273-299.

Hitchcock, C. (2007). Prevention, preemption, and the principle of sufficient reason. *Philosophical Review*, 116, 495-532.

Hitchcock, C. (2011). Trumping and contrastive causation. *Synthese*, 181, 227-240.

Hitchcock, C., and J. Knobe (2009). Cause and Norm. *Journal of Philosophy*, 106(11), 587-612.

Lewis, D. (1973). Causation. *Journal of Philosophy*, 70, 556-567.

Lewis, D. (2000). Causation as influence. *Journal of Philosophy*, 97, 182-197.

Pearl, J. (2000). *Causality: models, reasoning and inference.* Cambridge University Press.

Schaffer, J. (2000). Trumping preemption. *Journal of Philosophy*, 97, 165-181.

Schaffer, J. (2005). Contrastive causation. *Philosophical Review*, 114, 327-358.