Research Article   **Open Access**

Judea Pearl*

Causation and Decision: On Dawid's "Decision Theoretic Foundation of Statistical Causality"

Abstract: In a recent issue of this Journal, Philip Dawid (2021) proposes a framework for causal inference that is based on statistical decision theory and that is, in many aspects, compatible with the familiar framework of causal graphs (e.g., Directed Acyclic Graph (DAGs)). This editorial compares the methodological features of the two frameworks as well as their epistemological basis.

Keywords: directed acyclic graphs, conditional independence, potential outcome, ladder of causation, causal Bayesian network, decision theory, structural causal models, *do*-calculus

1 Introduction

I've followed the works of Professor Dawid since the early 1980's, when I discovered his seminal paper, "Conditional Independence in Statistical Theory" [1]. In that paper, Dawid boldly protests statistics' stalemate over causality, and declares: "Causal inference is one of the most important, most subtle, and most neglected of all the problems of statistics." In the past four decades, a period that saw a revolutionary progress in causal inference [2], Dawid has contributed substantially to this progress, but has consistently demanded that our understanding of causality be grounded in the tradition of statistical decision theory (DT), both conceptually and notationally. This paper [3], is a culmination of Dawid's efforts, and shows vividly what portions of the causal revolution can be re-formulated in the statistical DT paradigm, and what the costs and benefits are of imposing this re-formulation.

2 The statistical DT paradigm

The main thrust of the DT paradigm is to view causal inference as a decision aiding exercise, and to avoid whenever possible any concept or assumption that is not absolutely necessary for that exercise, especially those expressed in a vocabulary alien to traditional statistics. The outcome of this stat-exclusive strategy can be seen in the way Dawid articulates the assumptions that are needed for a DT task to commence. Equations (2)–(5), for example, convey conditional independence (CI) relations among observed or observable variables, but do not involve any counterfactual variables (e.g., Y_x) or do-operators, nor any notational device outside the vocabulary of traditional probability theory. Even the structure of the graph and the directionality of its arrows can be ignored once we accept this set of independencies. The only extra-statistical object is the "regime indicator variable," F_X, which acts as an intervention instrument.

In summary, the constraints implied by the DT framework amount to translating all causal knowledge (usually encoded in a causal graph) to a set of independencies involving regime indicators, allowing no other notational device, and pursuing the analysis as if it were an exercise of prediction, rather than intervention.

*__Corresponding Author: Judea Pearl:__ Department of Computer Science, University of California, Los Angeles, CA 90095; E-mail: judea@cs.ucla.edu

3 What we gain and what we lose

Readers familiar with the ladder of causation [2] and its logical foundations [4] would recognize immediately that, by forbidding counterfactual expressions, the entirety of Rung-3 of the hierarchy (also labeled "counterfactual" or "imagination") would become inaccessible to researchers adhering to the DT paradigm. This means that questions related to probabilities of causation [5, 6], Personalized Decision Making [7], Causes of Effects [8, 9] and large segments of Mediation Analysis [10] would be excluded from the analysis.

On the other hand, once F_X adequately simulates the *do*-operator, it immediately lifts the model from Rung-1 of the ladder to Rung-2 (intervention) level, and endows Dawid's DT framework with all the capabilities of the *do*-calculus and its associated applications, including covariate selection, identifiability results, transportability analysis, missing data, and more.[1]

Dawid's paper gives us in fact a comprehensive account of those tasks that do not rely on Rung-3 information but could be accomplished entirely within Rung-2 (equivalently, within DT), including tasks that were first formulated and solved using Structural Causal Model (SCM) as a starting point. (The analysis of non-compliance [12] is a good example).

It thus appears that, for the restricted set of applications limited to Rung-2 of the ladder, the DT paradigm offers a coherent, self-contained framework for causal inference and analysis. We will soon see, however, that, even in this restricted set of applications, DT cannot live up to its promise of avoiding concepts that are alien to traditional statistics. This will be shown both from basic principles and in the specific model of Figure 2.

4 Can the DT paradigm deliver on its promises?

From basic principles we know that in order to guarantee that F_X adequately simulates an intervention on X, the starting graph (also called "observational" or "unaugmented") must already carry some causal information, in addition to its conditional independencies. Absent such information would leave us on Rung-1 of the ladder of causation, unable to construct the conditional independencies associated with F_X, hence unable to reason with interventions.

From this basic fact we can conclude immediately that the arrows in the starting graph are endowed with causal information, that the author of that information chose to encode it in the form of causal symbols (nodes and arrows), and that the translation to the CI representation requires an understanding of those causal symbols.[2] In other words, the analyst must be versed in the calculus of symbols that are alien to statistics, contrary to the DT agenda.

More generally, this means that every DT researcher, even when restricted to decision making tasks, must carry in mind a mental representation of causal information and must be endowed with the logic of translating this information to conditional independence statements. How then is this information stored in the researcher's mind?

Going to the specific model of Figure 2, the insufficiency of CI can be shown by assuming that Z is unobserved, so that the starting DAG consists of just three variables: $X \to Y \leftarrow U \to X$, and Equation (2) would then read $F_X \!\perp\!\!\!\perp U$. Now reverse all arrows and let the starting DAG be $X \leftarrow Y \to U \leftarrow X$. All conditional independencies in the original (unextended) graph would remain the same as before (i.e., no CI exists) but, upon adding F_X to the graph, Equation (2) would be violated. In order for F_X to adequately represent a regime indicator for X, Equation (2) needs to be revised to read $F_X \!\perp\!\!\!\perp Y$.

---

1 A complete formal charactgerization of Rung-2 inferences, also known as "Causal Bayesian Networks," is given in [11, pp. 23–24]. Dawid refers to it as Pearlian DAGs, with pure intentions, I hope.

2 The *do*-operator teaches us that the causal information carried by the starting DAG can be reduced to the identities of the parents of X.

We again conclude, as before, that the starting graph must be treated as a carrier of causal information, not merely statistical information, and that at least part of this causal information must be judgmental, provided by a domain expert, since data alone is insufficient to lift us from Rung-1 to Rung-2. This, again, makes it impossible for a DT analyst to avoid the language of nodes and arrows.

This brings us the major criticism I have of the DT framework, its consistency and its ontological basis.

5 Is the DT paradigm worthy of pursuit?

I'll start by questioning the very purpose of DT, which I understand to be: Liberating analysts from dependence on foreign languages, contaminated with non-statistical objects such as causal arrows or counterfactual variables, and shielding them from the dangers that may loom from judgmental assumptions involving such objects.

I perfectly understand Dawid's apprehension of these objects because I've been there with him in the early 1990's, when causality was in its embryonic stage, viewed with suspicion, and when statistical entities were universally judged to be more principled, more understood, and certainly more trustworthy and respectable. Naturally, as an active member of this culture of suspicion, I found it safer to introduce the *do*-operator and the backdoor criterion using an intervention variable called F_i, with three regimes [13], precisely the way it is defined in Dawid's paper. I've quickly abandoned this notation for reasons that Dawid now finds to be "obscure," but which I've found to be unassailable: (1) If the causal information needed to define F_X comes from the unaugmented DAG, what is the point of cluttering the DAG with redundant F_X variables when we can go directly to the DAG and extract whatever information is needed using graphical algorithms? (2) If the causal information needed to define F_X comes from the unaugmented DAG, then the directionality of the arrows in that DAG comes not from mentally simulated interventions but from deeper, more reliable modes of judgment.[3] What are those modes? And why not harness them to guide all levels of causal reasoning, from Rung-1 up to Rung-3?

6 The fundamental question: How do people store causal knowledge?

This fundamental question is rarely asked by causal analysts or even by philosophers, because it is deemed to be psychological, rather than methodological. I disagree. The question of how scientists store scientific knowledge or how children store toy-world information is of fundamental importance in any framework, because whenever we rely on that knowledge, we must extract it from its very source, in its natural habitat, with minimum distortion, so as to preserve its veracity; the quality of our decisions depends on this veracity.

---

3 Another option exists, which Dawid would probably prefer. Instead of concluding that the causal information carried by the unaugmented DAG is encoded as nodes and arrows, one can postulate that this information is already available to the analyst in the form of conditional independence judgements like Equations (2)–(5), relative to every decision variable X needed for the DT task.

I rule out this option on the grounds that judgments of this sort require formidable cognitive efforts, hence they are unlikely to be stored explicitly in the analyst mind, and are bound to be unreliable if elicited.

First, people are bad in handling independence assumptions void of causal interpretation. (Recall how difficult it is for people to handle colliders [2].)

Second, such assumptions must to be checked, judgmentally, for every decision variable in the graph, and ensure that every combination of the corresponding F_i variables adequately simulates its instrumental function.

Similar objections apply of course to the potential outcome framework and the conditional ignorability assumptions with which they commence. Indeed, as noted by generations of investigators, e.g., [14, 15], judging the validity of such assumption is cognitively formidable; they are used primarily for formal justification of desirable results, not because analysts believe they hold true.

To drive the point home, imagine that you are given Equations (2)–(5) instead of the graph, and you are asked to judge whether the four equations adequately represent what you know about the problem domain. I have been working with Dawid's CI notation for 37 years, yet no matter how long I examine these equations, I remain unsure whether I have not left out a symbol or two. This does not happen to me with a DAG; I may doubt whether I have enough knowledge to determine each and every arrow, true, but I never doubt whether the DAG represents the knowledge that I do have (say, that roosters do not cause sunrise or that ice-cream sales does not cause people to drown).

The conclusion I draw is, first, that people do not store causal knowledge in the form of CI assertions and, second, that graphical relationships in the form of 'who listens to whom' is a more promising model of how human knowledge is stored.

The anthropomorphic metaphor of "listening" may drive some purists skeptical but, recall, we are seeking the most rudimentary primitives, or building blocks, with which knowledge is represented in our mind. Such building blocks must be metaphorical, and many of them anthropomorphic, since these are the chunks of expertise we acquire in childhood.[4]

It is not an accident that our everyday language for causation is replete with graphical metaphors (e.g., "causal pathways," "mediate between X and Y"), nor is it a coincident that the first formal representation of causal relations turned out to be "path diagrams" [16].

Still, to pacify the purists, we can replace "listening to" with "sources of variations." In other words, when deciding whether an arrow $X \to Y$ is appropriate in the DAG, the analyst must ask herself "what are the sources of the variations we may observe in Y?" If X qualifies as one of those (direct) sources, an arrow $X \to Y$ is introduced.

Admittedly, super-purists would not buy even the "sources of variation" metaphor as a legitimate basis for capturing human intuition about causation. For them, science has groomed a mathematical object that acts precisely as "listen to" and "sources of variation," yet it is decidedly respectable and fairly common; it is called a "function." Thus, to decide whether an arrow $X \to Y$ is appropriate in the DAG, the analyst must ask whether the relationship between X and Y requires a non-trivial function $y = f(x, u)$ for some experimental unit u. This criterion, though less likely to be understood by rank and file researchers, is a favorite in economics [17, 18] and serves as the basis of Structural Equations Models [19] and Structural Causal Models (SCM) [11, p. 27].

It is the functional, quasi-deterministic nature of SCM's that allows us to derive all counterfactual relationships from any given SCM model [20] and to handle systems with feedback [11, p. 215].[5]

7 The manipulability issue

My insistence on ascribing causal status to arrows in the unaugmented DAG, regardless of the manipulability of the variables involved, has invited a flood of criticism from manipulability-minded researchers, claiming that the *do*-calculus relies on the assumption that all variables are physically manipulable. There is nothing to support this misconception. While I have given causal semantics to the operator $do(X = x)$ with X non-manipulable [21], this does not mean that we have the physical means of directly intervening on X.[6] To suppress such misunderstanding, it is trivial, yet totally unnecessary to attach special labels to

---

4 It is worth noting that advocates of SWIGs, likewise, have failed to define the intuitive basis for constructing the original DAG, from which the SWIG is later configured. The received literature attributes this basis to a "Structured Tree Graph" called FFRCISTG which I have found to be far from intuitive. I certainty don't see how it can help a lay investigator decide whether an arrow $X \to Y$ should be added to the DAG.

5 I should confess at this point that, despite their heretical posture in 1993, I chose to introduce the backdoor criterion using SCMs, rather than intervention-based DAGs [13]. The clarity offered by the former was irresistible, even facing the statistical hegemony of the time.

6 Dawid, for example, writes: Our restriction of regime indicators to nodes where interventions are meaningful and relevant is in contrast with, for example, the approach of Pearl [11], where it is assumed that it is (at least in principle)

such variables in the graph. At the same time, I consider it necessary to be able to accommodate common statements such as "The moon causes tides" without seeking phantom interventions on the moon position. Causal graphs permit us to do so without trepidations.

8 Conclusions

Dawid considers it advantageous to formulate and analyze interventional tasks within the confines of the DT paradigm, as opposed to causal Bayesian Networks or structural equation models. This preference, stated explicitly in [3]:

"Since a Pearlian DAG is just an alternative representation of a particular kind of augmented DAG, its appropriateness must once again depend on the acceptability of the strong assumptions, described in Section 10.2, needed to justify augmentation of an observational DAG."

I hope my comments convince Dawid that the opposite is true; the observational (i.e., unaugmented) DAG comes before its augmentation. Its appropriateness comes directly from domain knowledge independently of any augmentation one may wish to entertain. Moreover, the appropriateness of any augmentation depends not on conditional independence assumptions, but on causal assumptions embedded in the unaugmented DAG. Finally, the causal reading of the arrows in the unaugmented DAG is not based on interventional considerations, but on the more fundamental relation of "listens to," which applies to both manipulable and non-manipulable variables.

References

[1] A.P. Dawid. Conditional independence in statistical theory. *Journal of the Royal Statistical Society, Series B*, 41(1):1–31, 1979.

[2] J. Pearl and D. Mackenzie. *The Book of Why: The New Science of Cause and Effect*. Basic Books, New York, 2018.

[3] A.P. Dawid. Decision-theoretic foundations for statistical causality. *Journal of Causal Inference*, 9:39–77, 2021.

[4] E. Bareinboim, J.D. Correa, D. Ibeling, and T. Icard. On Pearl's hierarchy and the foundations of causal inference. In H. Geffner, R. Decter, and J.Y. Halpern, editors, *Probabilistic and Causal Inference: The Works of Judea Pearl*, volume 36, pages 640–646. Association for Computing Machinery, New York, NY, USA, 2022.

[5] J. Tian and J. Pearl. Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence*, 28:287–313, 2000.

[6] A.P. Dawid, M. Musio, and R. Murtas. The probability of causation. *Law, Probability and Risk*, 16:163–79, 2017.

[7] S. Mueller and J. Pearl. Personalized decision making – a conceptual introduction. Technical Report R-513, Department of Computer Science, University of California, Los Angeles, CA, 2022.

[8] J. Pearl. Causes of effects and effects of causes. *Journal of Sociological Methods and Research*, 44:149–164, 2015.

[9] S. Mueller, A. Li, and J. Pearl. Causes of effects: Learning individual responses from population data. In L. De Raedt, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 2712–2718. International Joint Conferences on Artificial Intelligence Organization, 7 2022.

[10] J. Pearl. Interpretation and identification of causal mediation. *Psychological Methods*, 19:459–481, 2014.

[11] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, 2nd edition, 2009.

[12] A. Balke and J. Pearl. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439):1172–1176, September 1997.

[13] J. Pearl. Comment: Graphical models, causality, and intervention. *Statistical Science*, 8(3):266–269, 1993.

possible to consider interventions at every node in a DAG: while this allows one to interpret every arrow as "causal," that may not be an appropriate representation of the actual problem.

[14] M.M. Joffe, W.P. Yang, and H.I. Feldman. Selective ignorability assumptions in causal inference. *International Journal of Biostatistics*, 2010;6(2): Article 11. doi: 10.2202/1557-4679.1199.

[15] J. Pearl. Myth, confusion, and science in causal analysis. Technical Report R-348, University of California, Los Angeles, CA, 2009.

[16] S. Wright. The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. *Proceedings of the National Academy of Sciences of the United States of America*, 6:320–332, 1920.

[17] T. Haavelmo. The statistical implications of a system of simultaneous equations. *Econometrica*, 11:1–12, 1943. Reprinted in D.F. Hendry and M.S. Morgan (Eds.), *The Foundations of Econometric Analysis*, Cambridge University Press, 477–490, 1995.

[18] J. Pearl. Trygve Haavelmo and the emergence of causal calculus. *Econometric Theory*, 31:152–179, 2015. Special issue on Haavelmo Centennial.

[19] J. Pearl. The causal foundations of structural equation modeling. In R.H. Hoyle, editor, *Handbook of Structural Equation Modeling*, pages 68–91. Guilford Press, New York, 2012. Second edition forthcoming, January 2023.

[20] J. Pearl. On the first law of causal inference, Causal Analysis in Theory and Practice, 29 Nov. 2014, http://causality.cs.ucla.edu/blog/index.php/2014/11/29/on-the-first-law-of-causal-inference/.

[21] J. Pearl. On the interpretation of $do(x)$. *Journal of Causal Inference*, 7(1):20192002, 2019.