Proceedings of 2012 4th International Conference on Machine Learning and Computing
IPCSIT vol. 25 (2012) © (2012) IACSIT Press, Singapore

The Inference of Contrast for Probabilistic Theories of Causality

Joonsung Kim[1 +]

Department of Philosophy, Myongji University, Seoul, Korea (ROK)

Abstract. Hitchcock [8] argues that the ternary probabilistic theory of causality alone conveys the information about complex relations of causal relevance. In this paper, I show that the unanimity probabilistic theory of causality, which Hitchcock criticizes, also reveals complex relations of causal relevance. I conclude that the two probabilistic theories of causality carve up the same causal structure in two formally different and conceptually consistent ways.

Keywords: binary, causality, contrast, disjunctive factors, ternary, population, probabilistic theory, unanimity

1. Introduction

Several leading philosophers of science (Cartwright [1]) have developed probabilistic theories of causality in 40 years. The kernel of the probabilistic theories of causality is that the presence of a cause factor over the absence of the cause factor raises the probability of its effect factor. A formally intriguing but uneasy criticism has been raised of the probabilistic theories of causality. This criticism, which might be seemingly devastating to the probabilistic theories of causality, is due to the problem of disjunctive factors. In order to assess the causal significance of a medicine for patients' recovery, a medical team divides the study group of patients into three treatment groups, and in turn provides the first group with placebo *A*, the second group with a moderate dose *B* and the third group with a strong dose *C*. Suppose that the probability of recovery *Y* given each of *A, B, C* is as follows[1]: $Pr\ (Y \mid A) = 0.2$, $Pr\ (Y \mid B) = 0.4$, $Pr\ (Y \mid C) = 0.9$. The medical team wants to know whether the moderate dose *B* is a positive causal factor for the patients' recovery *Y*. According to the probabilistic theories of causality, the medical team compares the probability of *Y* in the presence of *B*, *i.e.*, $Pr(Y \mid B)$, with the probability of *Y* in the absence of *B*, *i.e.*, $Pr(Y \mid -B)$ in which '–' refers to 'negation'. Since $-B$, the absence of *B*, is equivalent to a disjunctive factor *A* v *C* (in which 'v' refers to 'or'), the medical team needs to assess the probability of *Y* in the presence of *A* v *C*, *i.e.*, $Pr(Y \mid A \text{ v } C)$. In computing $Pr(Y \mid A \text{ v } C)$, each disjunct, *A* and *C*, of the disjunctive factor *A* v *C* confers *different* probabilities on the factor *Y*. The problem then arises of how one identifies a single causally significant probability of the factor *Y* in the presence of the disjunctive factor *A* v *C*. Hitchcock [9] introduces two problems due to the problem of disjunctive factors, and argues that the probabilistic theories of causality cannot meet them.[2] Suppose that $Pr\ (A)$ is equal to $Pr(C)$. Then, $Pr(Y \mid -B) = [Pr(A)\ Pr(Y \mid A) + Pr(C)\ Pr(Y \mid C)] / [Pr(A) + Pr(C)] = [(0.5)\ (0.2) + (0.5)\ (0.9)] / [(0.5) + (0.5)] = 0.55$. So $Pr(Y \mid B) < Pr(Y \mid -B)$, which tells us that *B* is a *negative* causal factor for *Y*. Hitchcock (1993, p.341) claims that this casual claim conflicts with our intuition that the moderate dose has *positive* causal significance for patients' recovery. This is the first problem the probabilistic theories of causality confront. Again, suppose instead that $Pr(A)$ is

+ Joonsung Kim. Tel.: + 82-2-300-0634; fax: + 82-2-300-0504.
E-mail address: jkim30@mju.ac.kr

1 This example is introduced by Hitchcock [8] (pp. 340-343). Hitchcock slightly modifies the original example Humphreys [10] (pp. 41-42) first presented.

2 Humphreys [10], as far as I know, first posed the problem of disjunctive factors and the two problems due to it for the probabilistic theories of causality.

0.6 and $Pr(C)$ is 0.1. Then, contrary to the previous case, B is now a *positive* causal factor for Y since $Pr(Y \mid B) > Pr(Y \mid -B)$. Thus what causal significance B has for Y depends on the ratio of $Pr(A)$ to $Pr(C)$. Hitchcock [8] (p.341) finds it odd that the *objective* causal significance of B for Y depends on the ratio of $Pr(A)$ to $Pr(C)$. This is the second problem he raises for the probabilistic theories of causality. Hitchcock [8] (pp. 350-351) further argues that the unanimity theory, which is founded on the binary contrast, cannot convey information about complex relations of causal relevance such that, as doses of medicine change, so does the probability of patients' recovery change. Hitchcock claims that *ternary* contrast instead should be considered not only to meet the problem of disjunctive factors but also to explicate the complex relations of causal relevance. Inspired by Holland [9], Hitchcock claims that B is, with regard to Y, contrasted not with $-B$ but with a specific alternative to B, which is the ternary contrast. He compares $Pr(Y \mid B)$ and $Pr(Y \mid A)$, and $Pr(Y \mid B)$ and $Pr(Y \mid C)$, which are the ternary relations. B is a positive causal factor for Y relative to A, $Pr(Y \mid B) > Pr(Y \mid A)$, whereas B is a negative causal factor for Y relative to C, $Pr(Y \mid B) < Pr(Y \mid A)$. We also see that as doses of medicine change, so does the probability of patients' recovery change: $Pr(Y \mid A) < Pr(Y \mid B) < Pr(Y \mid C)$. Hitchcock formally generalizes these complex relations of causal relevance in terms of a conditional probability distribution function $f(\mathrm{x}) = Pr(\boldsymbol{Y} \mid \boldsymbol{X} = x)$ where $\boldsymbol{X}$ is a random variable standing for doses of medicine.[3] The ternary theory[4] introduces a conditional probability distribution function $f_i(\mathrm{x}) = Pr(\boldsymbol{Y} = y \mid \boldsymbol{X} = x \; \& \; K_i)$, and contrasts different values of the random variable $\boldsymbol{X}$ with regard to values of a random variable $\boldsymbol{Y}$. "i "of $f_i(\mathrm{x})$ represents each i of the background contexts K_i, so that the function $f_i(\mathrm{x})$ may have different shapes, depending on what background context it is relative to. Suppose that the values of the random variable $\boldsymbol{X}$ are doses of medicine determined as the result of a random experiment, values of a random variable $\boldsymbol{Y}$ are recovery or non-recovery, and the random experiment is relative to, for example, a background context K_2. If the probability of $\boldsymbol{Y}$ given $\boldsymbol{X} = 2$ (e.g., a moderate dose of medicine) is greater than the probability of $\boldsymbol{Y}$ given $\boldsymbol{X} = 1$ (e.g., a placebo), then $\boldsymbol{X} = 2$ tends to cause $\boldsymbol{Y}$ *when compared with* $\boldsymbol{X} = 1$. If the probability of $\boldsymbol{Y}$ given $\boldsymbol{X} = 3$ (e.g., a strong dose of medicine) is greater than the probability of $\boldsymbol{Y}$ given $\boldsymbol{X} = 2$, then $\boldsymbol{X} = 3$ tends to cause $\boldsymbol{Y}$ *when compared with* $\boldsymbol{X} = 2$. Thus the ternary theory meets the two problems by showing that the moderate dose of medicine is a positive causal factor for the patients' recovery without depending on $Pr(X_1)$ and $Pr(X_3)$. Hitchcock goes further. The function $f_i(\mathrm{x})$ has a shape of probability increasing relative to the background context K_2 such that $f_2(\mathrm{l}) < f_2(2)$ and $f_2(2) < f_2(3)$, assuming that $f_2(1) = Pr(\boldsymbol{Y} \mid \boldsymbol{X} = 1 \; \& \; K_2) = 0.2$, $f_2(2) = Pr(\boldsymbol{Y} \mid \boldsymbol{X} = 2 \; \& \; K_2) = 0.4$ and $f_2(3) = Pr(\boldsymbol{Y} \mid \boldsymbol{X} = 3 \; \& \; K_2) = 0.9$.[5] If the above random experiment is relative to another background context, then $f_i(\mathrm{x})$ may have a shape of probability decreasing or a shape of probability not changing. The relations, $f_2(\mathrm{l}) < f_2(2)$ and $f_2(2) < f_2(3)$, convey the information about the function $f_i(\mathrm{x})$ such that the probability of $\boldsymbol{Y}$ increases from $\boldsymbol{X} = 1$ through $\boldsymbol{X} = 2$ to $\boldsymbol{X} = 3$. Hitchcock [8] (pp. 350-351) claims that only the ternary theory conveys the information about the function $f_i(\mathrm{x})$, and is superior to the unanimity theory.

In this paper I shall show how the unanimity theory too conveys the information about the complex relations of causal relevance the ternary theory is intended to do. I conclude that the unanimity theory and the ternary theory both carve up the same causal structure in two formally different but conceptually consistent ways, while pointing out that the ternary theory is founded on the unanimity theory.

2. Multiple Ways of Carving up Causal Structure

We should notice that the probabilistic theory of causality, which is called the unanimity theory[6] (Eells [4], Eells and Sober [5]), is also a ternary theory[7]: it says that a factor X is a causal factor for another factor Y

[3] $\boldsymbol{Y}$ is of course also a random variable. In this example, Hitchcock considers only two cases of effect, i.e., recovery and none-recovery, so that $\boldsymbol{Y}$ has only two values 1 and 0.

[4] The theory (Hitchcock [8], pp. 349-353) is a generalization of Holland's [9] (p. 946) interpretation of causal relevance such that X is a positive, negative, or neutral cause of Y with respect to an alternative to X.

[5] y of $\boldsymbol{Y}=y$, which represents different non-negative values of the random variable $\boldsymbol{Y}$, will not appear in $f_2(\mathrm{x})$. For, in this example, Hitchcock considers only two cases of effect, *i.e.*, recovery and none- recovery, so that $\boldsymbol{Y}$ has only two values 1 and 0.

[6] The name 'unanimity' is first dubbed by Dupre' [2].

[7] Strictly speaking, the unanimity theory is a quandary theory which says that a factor X is a causal factor for another factor Y relative to a population P exemplifying a kind, or type Q. Hitchcock's ternary theory too is the same as the unanimity theory in that it too carves up causal structure relative to a population P exemplifying a kind, or type Q. Hitchcock did not notice how versatile the relativity of causal roles to population would be while I do.

relative to a population P. The causal significance of X for Y depends on which population we are considering. This is understood in two ways. First, a population P always exemplifies a population type Q. The causal significance of X for Y depends on which population type Q the population P is taken to exemplify.[8] For example, smoking may have a positive causal significance for lung cancer in a population of middle-aged human beings. But smoking may not have positive causal significance for lung cancer in a population of teen-aged human beings. Second, a population P, in which a factor is a causal factor for another factor, is basically taken as a homogeneous subpopulation. Causal role may be different, depending on which subpopulation we are considering. For example, if X is a positive causal factor for Y in a homogeneous subpopulation, then X may be a negative causal factor for Y in another homogeneous subpopulation. If this causal information is true, then X is causally mixed for Y in a subpopulation into which the two subpopulations are combined. Let us see how this feature of the probabilistic theory of causality conveys information about the function $f_i(\mathrm{x})$ the ternary theory alone allegedly does.

Consider, for brevity, only the three cases of the experiment relative to a background context K_2 in terms of Hitchcock's ternary theory introduced in the previous section, $f_2(1) = Pr(\boldsymbol{Y} \mid \boldsymbol{X} = 1 \ \&\ K_2) = 0.2$, $f_2(2) = Pr(\boldsymbol{Y} \mid \boldsymbol{X} = 2 \ \&\ K_2) = 0.4$ and $f_2(3) = Pr(\boldsymbol{Y} \mid \boldsymbol{X} = 3 \ \&\ K_2) = 0.9$. Let $\boldsymbol{X} = 1$, $\boldsymbol{X} = 2$, $\boldsymbol{X} = 3$ be in turn X_1, X_2, X_3, which constitute a partition of doses of medicine. See Figure 1.

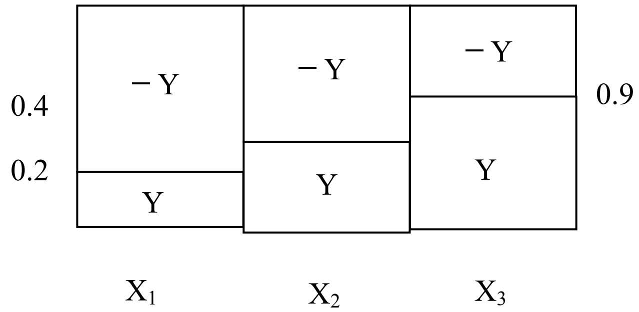

Fig. 1

According to the unanimity theory, the relations between X_1, X_2, X_3 and Y relative to the background context K_2 are $Pr(Y \mid X_1 \& K_2)$, $Pr(Y \mid X_2 \& K_2)$, $Pr(Y \mid X_3 \& K_2)$. These three conditional probabilities are in turn equivalent to $f_2(1) = Pr(\boldsymbol{Y} \mid \boldsymbol{X} =1 \ \&\ K_2) = 0.2$, $f_2(2) = Pr(\boldsymbol{Y} \mid \boldsymbol{X} =2 \ \&\ K_2) = 0.4$, $f_2(3) = Pr(\boldsymbol{Y} \mid \boldsymbol{X} =3 \ \&\ K_2) = 0.9$. Consider, relative to K_2, a subpopulation whose individuals have the property of X_1 or X_2. That is, a population type of the subpopulation is X_1 or X_2. (The shaded parts in Figure 2 are the relations between X_1, X_2 and Y in the subpopulation X_1 or X_2.)

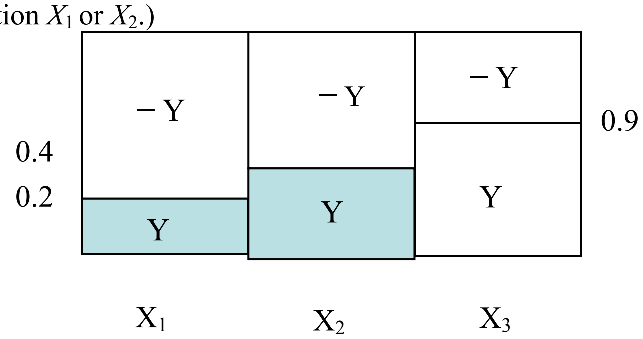

Fig. 2

Let the subpopulation relative to K_2 be K^{1}_{2}. In the subpopulation K^1_2, X_2 is not only the absence of X_1 but also only alternative to X_1. Therefore, in the subpopulation K^1_2, $Pr(Y \mid X_1) < Pr(Y \mid -X_1)$, that is, $Pr(Y \mid X_1) < Pr(Y \mid X_2)$. Again, consider a subpopulation whose individuals have the property of X_2 or X_3.

[8] See Eells [4] (Chapter 1) for details.

(The shaded parts in Figure 3 are the relations between X_1, X_2, and Y in the subpopulation X_2 or X_3.)

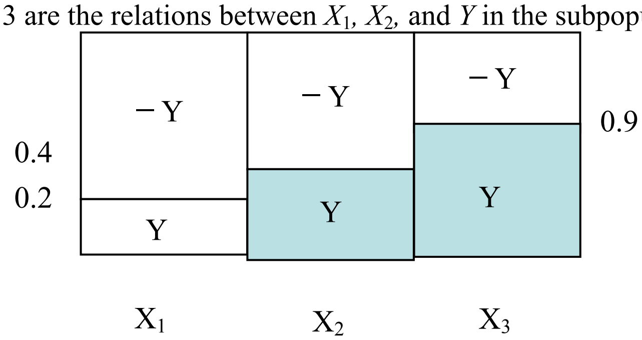

Fig. 3

Let the subpopulation be $K^2{}_2$. In the subpopulation $K^2{}_2$, X_3 is not just the absence of X_2 but also only alternative to X_2. Therefore, in the subpopulation $K^2{}_2$, $Pr(Y \mid X_2) < Pr(Y \mid -X_2)$, that is, $Pr(Y \mid X_2) < Pr(Y \mid X_3)$. Again, consider a subpopulation whose individuals have the property of X_1 or X_3. (The shaded parts in Figure 4 are the relations between X_1, X_3, and Y in the subpopulation X_1 or X_3.)

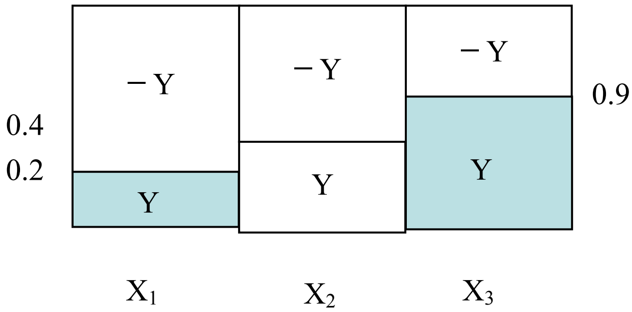

Fig. 4

Let the subpopulation be $K^3{}_2$. In the subpopulation $K^3{}_2$, X_3 is not just the absence of X_1 but also only alternative to X_1. In the subpopulation $K^3{}_2$, $Pr(Y \mid X_1) < Pr(Y \mid -X_1)$, that is, $Pr(Y \mid X_1) < Pr(Y \mid X_3)$. Let us follow the transition of the probability of Y from the subpopulation $K^1{}_2$ through the subpopulation $K^2{}_2$ to the subpopulation $K^3{}_2$. Then it is easy to see that the probability of Y increases from X_1 through X_2 to X_3. This conveys the information exactly about $f_2(x)$ such that $f_2(1) < f_2(2)$ and $f_2(2) < f_2(3)$. By considering *each of the three subpopulations* $K^1{}_2$, $K^2{}_2$, $K^3{}_2$ as the third relatum, the unanimity theory shows that the comparison between $Pr(Y \mid X)$ and $Pr(Y \mid -X)$ conveys the information about $f_2(x)$.

3. Conclusion

Hitchcock[9] does not notice that the unanimity theory is developed in the context of a *three-place* theory of causality: it says that a factor is a causal factor for another factor *relative to a population of a certain type or kind*. Therefore, the theory is, as it should be, sensitive to the mechanism by which subjects are assigned to the three treatment groups. This sensitivity is not an objection to the theory but rather clearly reveals a desirable feature of the unanimity theory. Several philosophers (Hausman [6], [7], Schaffer [12]), which have discussed what should be contrasted as causal relata, argue argue that ternary (or quandary) contrast meets several crucial problems with theories of causation allegedly due to binary contrast. They at least implicitly credit their arguments for quandary contrast to Hitchcock's [8] ternary probabilistic theory of causality. My discussions in this paper may lead them to reconsider the relation between binary contrast and

[9] In the footnote of his paper, Hitchcock too notes Eells' response to his paper. Eells points out that the unanimity theory is three-place theory. It is interesting that Hitchcock disregards the desirable feature of the unanimity theory.

ternary contrast. At least as far as causal relevance is relative to population, the ternary contrast is founded on the binary contrast.

4. References

[1] Cartwright, N. 1979. Causal Laws and Effective Strategies. *Noûs* 13: 419-437. Reprinted in N. Cartwright, *How the Laws of Physics Lie*. Clarendon Press. Oxford. 1983.

[2] Dupre', J. 1984. 'Probabilistic Causality Emancipated', in P. A. French, T. E. Uehling, H. K. Wettstein (eds.), *Midwest Studies in Philosophy IX: Causation and Causal Theories*, University of Minnesota Press, Minneapolis, pp. 169-75.

[3] Eells, E. 1988. 'Probabilistic Causal Interaction and Disjunctive Causal Factors', in J. Fetzer (eds.), *Probability and Causality: Essays In Honor of Wesley C. Salmon*. D. Reidel, Dordrecht, pp. 189-209.

[4] Eells, E. 1991. *Probabilistic Causality*, Cambridge University Press.

[5] Eells, E. and Sober, E. 1983. Probabilistic Causality and the Question of Transitivity. *Philosophy of Science50*: 35-57.

[6] Hausman, D. 1998a. *Causal Asymmetries*, Cambridge University Press.

[7] Hausman, D. 2010. Probabilistic Causation and Practical Causal Generalizations. In E. Eells and J. Fetzer (eds.). *The Place of Probability in Science*.

[8] Hitchcock, C. 1993. A Generalized Probabilistic Theory of Causal Relevance. *Synthese 97*: 335-64.

[9] Holland, P.W. 1986. Statistics and Causal Inference (in Theory and Methods). *Journal of the American Statistical Association,* Vol. 81: 945-960.

[10] Humphreys, P. 1989. *The Chances of Explanation,* Princeton University Press.

[11] Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge Univ. Press.

[12] Schaffer, J. 2005. Contrastive Causation. *Philosophical Review* 114: 297-328.