A Multivalued Conditional Logic with Probabilistic Interpretation for Causal Decision Theory

François Lepage
Département de philosophie
Université de Montréal

Jian-Yun Nie
Dept. d'informatique et R. O.
Université de Montréal

Causal decision theory has been developed to deal with rational decision making in the case where making a decision may be seen as an indication that some outcome will obtain. It has been suggested that counterfactual conditionals should be used in such a decision situation. The crucial problem posed is that of calculating the probability of a counterfactual conditional. The only technique available until now is *imaging* which, however, only applies to a restricted structure of possible worlds. In order to make the causal decision theory applicable to the general case, in this paper, we extend the imaging technique to the general structure. This results in a logic in which counterfactual conditionals are multivalued. We further provide a probabilistic interpretation to this logic.

1. Introduction

Counterfactual conditionals (or conditionals for short) are studied in order to provide a suitable framework for nonmonotonic reasoning, in particular, that based on causality. A conditional is a statement such as "if A were true, then B would be true", denoted as $A \Rightarrow B$. In contrast with the material implication $A \supset B$, the counterfactual conditional $A \Rightarrow B$ does not systematically result in the truth of the conditional when the antecedent A is actually false, as is the case for material implication. Rather, the truth of the conditional depends on the possible connections between the antecedent and the consequent. The difference between the two logical connectors may be shown by the following examples:

1) If the bus drivers hadn't gone on strike, I would not have been late.
2) If the bus drivers hadn't gone on strike, Paris would be in Spain.

Suppose the actual situation is that the bus drivers have gone on strike. Then according to the semantics of material implication, both sentences above are true. According to our intuition, however, although the first sentence seems reasonable, the second will certainly be rejected by everyone because there is no connection between the bus drivers' strike and Paris' being in Spain. Counterfactual conditionals try to mirror the way people think. They allow a better modeling of notions such as causality. Their importance in AI has been underlined in [1]. Another area in which counterfactuals have a great potential value is decision theory [2].

To begin with, let us describe decision problems as follows: Given a set of acts $\{..., a_i, ...\}$ open to an agent and a set of mutually exclusive states of the world $\{..., s_j, ...\}$. For each act-state pair (a_i, s_j), only one outcome O_{ij} obtains and is assigned a *desirability* $\mathcal{D}O_{ij}$ by the agent. The decision problem is that of choosing to perform the act a_i that gives the highest *expected desirability* expressed as follows:

$$\text{utility}(a_i) = \Sigma_j \; prob(\text{if I perform } a_i \text{ then } s_j \text{ obtains})\mathcal{D}O_{ij}$$

Let us denote the proposition "I perform a_i" by A_i, and "s_j obtains" by S_j. In traditional Bayesian decision making [3], the probability of "if I do a_i then s_j obtains" is formulated as the conditional probability $P(S_j/A_i)$, i.e. the probability of obtaining the state on the assumption that the act is about to be performed. This formulation results in the following $\mathcal{V}$-utility:

$$\mathcal{V}(A_i) = \Sigma_{S_j} P(S_j/A_i)\mathcal{D}O_{ij}$$

This formulation, however, fails to deal with decision making in a situation where choosing to perform an act may be seen as an indication that some state will be obtained. Newcomb's

problem [4] is typically in such a situation:

Suppose that there are two boxes, one transparent containing a thousand dollars and one opaque which may contain a million dollars or nothing. An agent can take either the opaque box only (A_1) or both boxes (A_2). A predictor has already placed a million dollars in the opaque box if he predicted that the agent would take the opaque box only, and nothing in the box if he predicted that the agent would take both boxes. The agent knows all this and he also knows that the predictor is highly reliable.

Using the $\mathcal{V}$-utility, the agent would take only the opaque box (A_1) because A_1 is an indication that the predictor has placed a million dollars in the box and A_2 is an indication that there is nothing. So, both P("the box contains a million dollars"/A_1) and P("the box contains nothing"/A_2) are very high (near 1). Thus, the agent will decline in taking only the opaque box in order to fit the good indication.

We can see that the decision is irrational: since the content of the opaque box has already been decided and will not be affected by the agent's decision, taking both boxes will give him a thousand dollars more in any case.

In order to make rational decision in such a situation, Gibbard and Harper [2] suggest that the conditional probability should be replaced by the probability of conditional. This results in the following $\mathcal{U}$-utility:

$$\mathcal{U}(A_i) = \sum_{S_j} P(A_i \Rightarrow S_j)\mathcal{D}O_{ij}$$

$\mathcal{U}$-utility results in a higher value for A_2: since the actual state of the boxes is causally independent from the agent's act, $P(A_i \Rightarrow S_j) = P(S_j)$ for both A_1 and A_2. As choosing A_2 whatever the state is results in a thousand dollars more, the agent would assign desirability such that $\mathcal{D}O_{2j} > \mathcal{D}O_{1j}$). Thus, $\mathcal{U}(A_2) > \mathcal{U}(A_1)$. This choice is rational.

In Gibbard and Harper's approach, the probability of conditional $P(A_i \Rightarrow S_j)$ is taken as grounded. However, the computation of the probability of a conditional is still problematic. To this end, Lewis developed a technique within possible-world semantics called *imaging* [5-7]. To make use of this technique, it is assumed, however, that, given any state (or world in possible-world semantics) and any sentence which is the antecedent (i.e. performing some act), there is at most one *nearest* alternative state in which the sentence is satisfied. This assumption is too restrictive. Indeed, there may be several possible “near” alternatives. For example, if the antecedent is a disjunction ($A_1 \vee A_2$) and A_1 and A_2 are equally plausible, then there may be two alternatives (one for each element in the disjunction), each being as "near" as the other. This situation may occur in decision making. For example, for a party which is to be organized by a family, the organizer selector may need to estimate the utility of the act "the family Robinson organizes the party" (R), but there may be several states near the actual state, for example, "Mary Robinson organizes the party alone" (R_1), "Robert Robinson organizes the party alone" (R_2), "Mary and Robert Robinson organize the party together" (R_3), and so on. Thus, R corresponds to the disjunction ($R_1 \vee R_2 \vee R_3 \vee ...$). It may be the case that R_1, R_2, ... are equally plausible, thus equally near to the actual state of the world. Lewis' imaging fails to cope with such a situation.

How then should we define a probabilistic function for conditionals in this general case? Until now, no one had successfully answered the question. Nute [8] (p. 118) even suggests that this is not feasible:

... there does not appear to be any way to apply the imaging technique to SOS-models [System-Of-Spheres-models, that is the models in which there may be multiple nearest alternatives] *and thereby provide an account of subjunctive* [counterfactual] *probabilities for that logic by this means.*

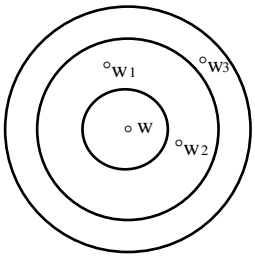
We think the problem underlying this failure is that a conditional has always been considered as a two-valued sentence. In this paper, we develop an alternative logic in which a conditional is multivalued. This will enable us to define a compatible probabilistic function for conditionals using an extended imaging technique.

The remainder of the paper is organized as follows: We first describe some previous development on conditional logic in section 2. After defining a new model for conditional logic in section 3, we will provide the model with a probabilistic interpretation in section 4. Finally, some concluding remarks will be given in section 5.

2. Previous Development on Conditional Logic

In this section, we describe one of the systems of conditional logic, the system of spheres developed by Lewis [9].

Roughly speaking, a system of spheres is built as follows: In possible-world semantics, from a given world *w* (which we will call the actual world), the worlds around it are accessible but their *distance* to the actual world varies. According to this distance, a set of nested spheres around the actual world may be constructed so as to create concentric spherical regions, as shown in the following figure. Worlds having the same distance from the actual world are in the same spherical region. The larger the spherical region, the more distant the worlds in it from the actual world.



In the above figure, the world *w* is the actual world. The worlds w_1 and w_2 are at equal distance from the actual world *w*, but they are closer to *w* than is the world w_3.

The distance may be seen as being determined by a set of ordered constraints $\{C_1, C_2, ..., C_i, ... \}$ such that $C_i \supset C_j$ for any $i < j$. Each sphere then consists of the worlds in which some constraint C_i is satisfied. In the smallest sphere $\{w\}$, the strongest constraint C_1 is satisfied. In a world in other spherical regions, less strong constraints are satisfied. A world w_i is nearer to the actual world *w* than another world w_j is iff w_i satisfies some constraint C_k that w_j does not. Thus, any set of ordered constraints induces a definition of distance.

A more formal description is as follows:

The alphabet of the language of the logic is $\{(,), \neg, \wedge, \Rightarrow, p_1, p_2, \ldots, p_n\}$, where *n* is an arbitrary natural number. The set $P_n = \{p_i : 1 \leq i \leq n\}$ is the set of atomic sentences (which we will also call propositions). The set of well-formed-formulas (or sentences) *Wff* is defined as follows:

i) $P_n \subset Wff$

ii) if $A, B \in Wff$, then $\neg A$, $(A \wedge B)$, $(A \Rightarrow B) \in Wff$.

The other connectives are introduced by the defined ones, in particular: $\Box A =_{\text{def}} (\neg A \Rightarrow A)$.

A **model** of the logic is a triple $<W, V, \$>$ such that:

(a) $W = \{0,1\}^{P_n}$ is the set of possible worlds. Each world $w \in W$ is the characteristic function of the set of atomic sentences which are true in *w*. Equivalently, $w(\text{-})$ is a function which gives the truth value ($\in \{0,1\}$) of any atomic sentence.

(b) $\$$ is a function $\$: W \rightarrow \{0,1\}^{(\{0,1\}^W)}$ which maps every world to a set of nested spheres such that:
- for any $w \in W$, $\{w\} \in \$(w)$; and
- for any $w \in W$, if $X, Y \in \$(w)$, then $X \subseteq Y$ or $Y \subseteq X$.

(c) V is a valuation function such that for any $w \in W$, $V_w : Wff \rightarrow [0,1]$ verifies the following conditions:

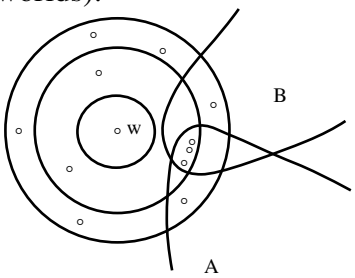
(1) $V_w(p_i) = w(p_i)$;

(2) $V_w(\neg A) = 1 - V_w(A)$;

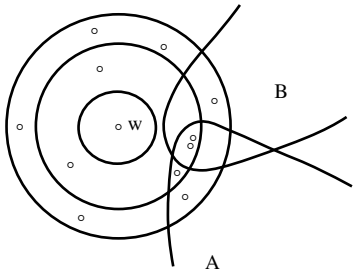
(3) $V_w(A \wedge B) = 1$ if $V_w(A) = 1$ and $V_w(B) = 1$,
$= 0$, otherwise;

(4) $V_w(A \Rightarrow B) = 1$, if either there is no world w such that $V_w(A) = 1$, or in the smallest sphere $X \in \$(w)$ such that there is a $w' \in X$ and $V_{w'}(A)=1$, for any $w'' \in X$ we have $V_{w''}(A \supset B) = 1$;
$= 0$, otherwise.

For any A and any w, we will call w an "A-world" if $V_w(A) = 1$, and those A-worlds included in the smallest sphere the "nearest A-worlds". In fact, the conditional $(A \Rightarrow B)$ is true if all the nearest A-worlds are also B-worlds. The two figures below show respectively the case of $A \Rightarrow B$ being true and false (where the regions A and B represent respectively regions of A-worlds and B-worlds):



$A \Rightarrow B$



$\neg(A \Rightarrow B)$

To give an example, let us consider the two sentences given earlier:

1) If the bus drivers hadn't gone on strike, I would not have been late.
2) If the bus drivers hadn't gone on strike, Paris would be in Spain.

The actual world is supposed to be as follows: the bus drivers have gone on strike, I was late, and the only transportation means I had is buses and I did not want to be late. For the first sentence, there may be several alternative worlds around the actual world in which the bus drivers hadn't gone on strike. But the nearest one(s) is(are) that in which anything other than the bus drivers' strike is not changed. In any of them, the buses travel regularly as usual, I can take buses as usual, and I will not be late. So the first sentence is true. For the second sentence, we have the same nearest world(s) as for the first sentence. However, Paris will remain in France despite the hypothesis concerning the bus drivers' strike. So the sentence is false.

As we said earlier, there is no probabilistic interpretation available now with the model described in this section. In order to obtain a compatible probability function $P(A \Rightarrow B)$, in the next section, we will define a new model for counterfactuals. We will show later that this model allows a compatible probabilistic interpretation.

3. A New conditional Logic

The language of our logic is exactly the same as that in section 2. A model of our logic is a triple $<W, v, \$>$ such that:

- $W = \{0,1\}^{P_n}$ is the set of possible worlds as before;
- $\$: W \rightarrow \{0,1\}^{(\{0,1\}^W)}$ is the function as defined before;
- v is a valuation function such that for any $w \in W$, $v_w : Wff \rightarrow [0,1]$ verifies the following conditions:

(1) $v_w(p_i) = w(p_i)$;

(2) $v_w(\neg A) = 1 - v_w(A)$;

(3) $v_w(A \Rightarrow B) = \b\lc\{(\a\al(\o(\sum_{w' \in N_{w,A}};)\f(v_{w'}(B) ; |N_{w,A}|) \quad$ if A is not a contradiction; ; 1 if there is no such sphere))

where $N_{w,A}$ is the set of A-worlds w'' (i.e. $v_{w''}(A) = 1$) and w'' belongs to the smallest sphere around w containing at least one A-world; and $|N_{w,A}|$ is its cardinality;

(4) $v_w(A \wedge B) = \sqrt{v_w(A) \cdot v_w(B) \cdot v_w(A \Rightarrow B) \cdot v_w(B \Rightarrow A)}$

We have a multivalued evaluation (within [0,1]) for conditionals. In particular, if B is a classical formula (i.e., not containing '$\Rightarrow$'), the truth value of $(A \Rightarrow B)$ in the world w is the proportion of B-worlds among the nearest A-worlds.

For the evaluation of a conjunction, the rationale is to take into account the interconnection between the two elements in conjunction, and the two counterfactuals $(A \Rightarrow B)$ and $(B \Rightarrow A)$ are used to cope with the interconnection. This evaluation is radically different from those used in other multivalued logic, such as $v_w(A \wedge B) = min(v_w(A), v_w(B))$ and $v_w(A \wedge B) = v_w(A) \cdot v_w(B)$. These latter are all extensional: a conjunction of related elements are evaluated in the same way as a conjunction of independent elements. Extensional approaches lead to the loss of some interesting properties. For example, in the first extensional evaluation, we do not have $v_w(A \wedge \neg A) = 0$; in the second, we do not have $v_w(A \wedge A) = v_w(A)$. Our evaluation of conjunction, which is not extensional, verifies most classical properties of conjunction, except the associatively in the non-classical case, i.e. $v_w(A \wedge (B \wedge C)) \not\equiv v_w(A \wedge B) \wedge C)$ when A, B or C has a non-classical valuation in w. However, the valuation function preserves all the classical properties in the classical cases.

Proposition 1

If $v_w(A)$ and $v_w(B)$ have classical Boolean evaluation, so do $v_w(\neg A)$ and $v_w(A \wedge B)$.

The proof is straightforward. This property provides our valuation function with classical behavior when their arguments are classical values. Let us now define the following notion of (causal) independence.

Definition

Two sentences A and B are said to be independent if $((A \Rightarrow B) \equiv B) \wedge (B \Rightarrow A) \equiv A))$ is a logical truth.

The following properties, which are incompatible together in any *extensional* multivalued logic, hold in our logic.

Proposition 2

(a) For any A and any w, $v_w(A \wedge A) = v_w(A)$

(b) For any A and any w, $v_w(A \wedge \neg A) = 0$ and $v_w(A \vee \neg A) = 1$

(c) For any A, B and any w, if A and B are independent, then

$$v_w(A \wedge B) = v_w(A) \cdot v_w(B).$$

It is worth noticing that our logic fits the widely shared views about conditional logic. The following formulas and rules commonly rejected by conditional logic are also rejected in our logic.

(1) $((A \Rightarrow B) \wedge (B \Rightarrow C)) \supset (A \Rightarrow C)$ (transitivity)

Let us show a counterexample. Let w be such that $w(p_0) = w(p_1) = 0$. So, $v_w(p_0 \Rightarrow \neg p_1) = 1$ because in the nearest p_0-world, p_1 is still false. Then the following is false in w:

$$((p_0 \wedge p_1) \Rightarrow p_0) \wedge (p_0 \Rightarrow \neg p_1)) \supset ((p_0 \wedge p_1) \Rightarrow \neg p_1).$$

In a similar way, we can also reject the following:

(2) $(A \Rightarrow B) \supset ((A \wedge C) \Rightarrow B)$ (monotonicity)

(3) $(A \Rightarrow B) \supset (\neg B \Rightarrow \neg A)$ (contraposition)

On the other hand, the following commonly accepted rules are also valid in our logic.

(1) $\vdash (A \Rightarrow B) \vee (A \Rightarrow \neg B)$

(2) $\vdash (A \Rightarrow \neg B) \equiv \neg(A \Rightarrow B)$

(3) $\Box (A \supset B) \vdash (A \Rightarrow B)$

(4) $(A \Rightarrow B) \vdash (A \supset B)$

(5) $((A \Rightarrow B) \wedge (A \Rightarrow C)) \vdash (A \Rightarrow (B \wedge C))$

The validity of these rules may be easily checked using the definition of v. The first rule is called the law of the conditional excluded middle. It may seem odd to obtain such a law in a multivalued logic. In fact, this law does not mean that it is always the case that either $(A \Rightarrow B)$ or $(A \Rightarrow \neg B)$ is true; it just says that $v_w(A \Rightarrow B) + v_w(A \Rightarrow \neg B) = 1$ for any w. It is also interesting to point out that, when any p_i is equally plausible for w, we have:

$$v_w((p_1 \vee p_2) \Rightarrow p_1) = v_w((p_1 \vee p_2) \Rightarrow p_2) = 1/2$$

$$v_w((p_1 \vee p_2) \Rightarrow p_1) \vee ((p_1 \vee p_2) \Rightarrow p_2)) = 1$$

4. A Probabilistic Interpretation

One could be tempted to interpret the fractional truth value of a conditional as a probability as is usually done in AI. This, we think, would be misleading. The truth and the probability of a sentence are two different notions. However, a logic may be endowed with a probabilistic interpretation. This is what we do for our logic in what follows.

Let us first suppose that we have a probability measure P defined on the Boolean algebra of the power set of the set of worlds, which satisfies the following usual properties :

(1) $P : \{0,1\}^W \rightarrow [0,1]$; (2) $P(W) = 1$; (3) $P(X \cup Y) = P(X) + P(Y)$, if $X \cap Y = \varnothing$

To define the probability of every sentence not containing the connector '$\Rightarrow$', the following formula is used as usual (note: $P(w)$ below is an abbreviation for $P(\{w\})$):

$$P(A) = \sum_{w \in W} v_w(A) \cdot P(w)$$

This formula expresses the idea that the probability of any classical sentence is the sum of the probabilities of the set the worlds in which the sentence is true.

It is not straightforward to extend the above formula to sentences containing the

conditional connector. To calculate the probability of these sentences, Lewis [5] developed the imaging technique. This technique only applies to a restricted structure such that for any sentence A and any world w, there is at most one nearest A-world. Let us denote this latter as $s(w,A)$. The restricted structure for imaging is equivalent to a trivial system of spheres: each concentric spherical region contains exactly one world (except when A is a contradiction). In such a system $v_w(A \Rightarrow B)$ always has a Boolean value: it is 1 if the nearest A-world is a B-world and 0 if not. The imaging technique is as follows:

Given a probability function P, to calculate $P(A \Rightarrow B)$, one first has to define a new probability function P_A with respect to the antecedent A. P_A is obtained by shifting the probability of a world onto its nearest A-world, i.e.:

$$P_A(w') = \sum_{w \in W} P(w) \cdot \delta(w', s(w,A))$$

where $\delta(x,y)$ is the Kronecker delta function such that $\delta(x,y) = 1$ if $x=y$, and $\delta(x,y) = 0$ otherwise. It may be shown that when A is not a contradiction, P_A is also a probability function. Then, we have:

$$P_A(B) = \sum_{w \in W} v_w(B) \cdot P_A(w)$$

It has been shown [5] that with the function P_A obtained in this way, we can prove the following equation for the case where A is not a contradiction:

$$P(A \Rightarrow B) = \sum_{w \in W} v_w(A \Rightarrow B) \cdot P(w) = P_A(B).$$

This is the formal way to state the idea that the probability of a conditional is the probability of the consequent after having imaged on the antecedent, i.e. after having created a situation in which the antecedent has the complete probability (note that $P_A(A) = 1$).

As has been noted [6, 8], the restriction made in this imaging is too strong. Nute [8] then developed the notion of *reasonableness* to replace the function $s(w,A)$ in order to spread the probability of a world to all the near reasonable A-worlds. However, he fails to extend Lewis' imaging technique to the general case, and even suggests that such an extension is not feasible (see introduction). Gärdenfors [6, 7] developed a similar approach.

The problem of both Lewis and Nute is that they consider a conditional as always being two-valued. In fact, if we abandon this restriction, Nute's approach is very close to an extension of the imaging technique. Let us pursue in this direction and define the following extended imaging: Given a probability function P over the worlds, and a sentence A, we define the new probability function P_A as the result of sharing the probability of each world uniformly among all the A-worlds in its smallest sphere where there is at least one A-world. Formally:

$$P_A(w') = \sum_{w \in N^*_{w'}} \frac{P(w)}{|N_{w,A}|}$$

where $N^*_{w',A}$ is the dual of $N_{w,A}$, i.e., $w' \in N_{w,A}$ iff $w \in N^*_{w',A}$.
This only expresses that we put on w' a part of the probability of every world w such that w' is one of the nearest A-worlds around w, the part being determined by the fraction $\frac{1}{|N_{w,A}|}$.

We can easily check that this new imaging is compatible with our definition of $v_w(A \Rightarrow B)$. It is also compatible with $P(A \Rightarrow B)$ as expressed in the following proposition.

Proposition 5

Let A be any sentence which is not a contradiction, B any sentence and P any probability

function on the Boolean algebra based on W, then

$$P(A \Rightarrow B) = \sum_{w \in W} v_w(A \Rightarrow B)\, P(w) = P_A(B)$$

If A is a contradiction, then $P(A \Rightarrow B)$ is set to 1.

The proof of the proposition only uses standard technique. We skip it here.

We now have a logic for the general case in which the probability of a conditional is the probability of the consequent after having imaged on A. We can see that the generalized imaging technique may be directly applied to causal decision theory proposed in [2]. The idea is to first construct a set of weakening constraints (as stated in section 2) from the constraints of the agent which vary in strength. This set of constraints determines a unique system of spheres. Then the probability of conditional is calculated using the generalized imaging technique.

5. Final remarks

In this paper, we have defined a new framework for counterfactual conditionals. Contrary to what is usually thought, in the case of multiple "near" worlds, a conditional is multivalued in our logic. In the same case, the imaging technique has been extended. It has been shown that the multivalued conditional logic is compatible with a probabilistic interpretation. One particularity of the logic is that the counterfactual conditional is considered as the most basic notion in reasoning. In this logic, a conjunction is defined by using counterfactual conditionals. This results in a non extensional evaluation of conjunction. As a consequence, most intuitive properties are verified by this logic.

In our probabilistic interpretation, the probability of the conditional is different from the conditional probability. The interpretation of probability of conditional as conditional probability has been extensively examined by Goodman and Nguyen in [10] which shows that this interpretation cannot be extended to the entire Boolean algebra.

With our logic and the generalized imaging technique, the causal decision theory proposed by Gibbard and Harper [2] can be applied to the general case. In this paper, however, for the lack of enough space, we did not present detail of applying our evaluation of $P(A \Rightarrow B)$ to decision making problems. Nevertheless, the application is quite straightforward and readers may imagine the procedure quite well. We will describe the application to decision theory in more detail in another report.

References

[1] M. L. Ginsberg, “Counterfactuals,” *Artificial Intelligence*, vol. 30, pp. 35-79, 1986.

[2] A. Gibbard and W. L. Harper, “Counterfactuals and two Kinds of expected utility,” in *Ifs*, W. L. Harper, R. Stalnaker, and G. Pearce, Eds. Cambridge: Reidel, 1981, pp.153-190.

[3] R. C. Jeffrey, *The logic of decision*. New York: McGraw-Hill, 1965.

[4] R. Nozick, “Newcomb's problem and two principles of choice,” in *Essays in honor of Carl G. Hempel*, N. Rescher, Ed. Dordrecht-Holland: Reidel, 1969.

[5] D. Lewis, “Probability of conditionals and conditional probabilities,” *Philosophical Review*, vol. LXXXV, pp. 297-315, 1976.

[6] P. Gärdenfors, “Imaging and conditionalization,” *The Journal of Philosophy*, vol. 79, pp. 747-760, 1982.

[7] P. Gärdenfors, *Knowledge in Flux: Modeling in Dynamics of Epistemic States*. Cambridge: MIT Press, 1988.

[8] D. Nute, *Topics in Conditional Logic*, vol. 20. Dordrecht, Holland: Reidel, 1980.

[9] D. Lewis, *Counterfactuals*: Harvard University Press, 1973.

[10] I. R. Goodman and H. T. Nguyen, “A Theory of Conditional Information for Probabilistic Inference in Intelligent Systems: III. Mathematical Appendix,” *Information Sciences*, vol. 75, pp. 253-277, 1993.