

A Causal Markov Condition for Value

Olav Benjamin Vassend¹

¹University of Inland Norway

Abstract

This paper proposes a causal independence principle for *value*—the *value Causal Markov Condition* (v-CMC)—and develops the conceptual and mathematical foundations of a “causal value theory” linking causality and utility. After motivating a local formulation of the v-CMC, we introduce a probability–value duality that translates standard causal-inference results into the value setting. In particular, we formulate local, global, and decomposition versions of the v-CMC and prove their equivalence. We also define *v-separation* and show that it is sound and complete for conditional value independence. Furthermore, we derive a Bellman-type recursion as a special case of the v-CMC, thereby generalizing standard Bellman recursion from linear chains to causal DAGs. Finally, we show how the v-CMC supports modular transfer and updating of utility information across causal contexts and develop algorithms for causally structured utility elicitation and canonical influence-diagram construction.

1 INTRODUCTION

In the foundations of causal inference and decision theory, probability and value or utility occupy strikingly different positions.¹ On the probability side, there is a deep theory relating causal graphs to probability distributions, whose central pillar is the Causal Markov Condition (CMC) (Pearl, 2009; Spirtes et al., 2000). According to the CMC, any probability distribution compatible with a directed acyclic graph (DAG) encoding causal relations must respect the conditional independence relations implied by the graph.

On the value side, however, there is no analogous principle.

Standard Expected Utility Theory (EUT) treats the utility function as a descriptive representation whose role is to encode an agent’s subjective preferences (Anscombe and Aumann, 1963; Savage, 1954; von Neumann and Morgenstern, 1944; Jeffrey, 1983). A familiar slogan is “*de gustibus non est disputandum*” (there is no arguing about taste). On this view, if an agent’s preferences exhibit complex interdependencies, then a correct value function faithfully captures all this complexity. Contrary to this standard view, the central thesis of this paper is that causal assumptions often can and should guide judgments about utility dependencies. To motivate the idea with a simple example, consider a headache pill. The utility of taking the pill plausibly depends on whether one has a headache, because one believes that the pill alleviates headaches. If one were to discover that the pill has no causal effect, one’s utility of taking it would (and should) become independent of whether one has a headache.

Aside from being intuitively plausible, the idea that causal information can guide utility judgments is practically significant. The probabilistic CMC plays a central role in enabling predictions to be transferred across causal systems, since it enables modular decompositions of probability distributions that track causal structure. But decisions depend on both probabilities and utilities. Thus, extrapolating decisions from one context to another requires not only transferring probabilistic information, but also determining which aspects of a utility function should remain stable across contexts. This, in turn, calls for a way of decomposing utility functions modularly in light of causal assumptions. The aim of this paper is to formulate such a principle.

More concretely, we propose the *value Causal Markov Condition* (v-CMC), which relates causal graphs to utility functions, and formulate a causal compatibility criterion for DAG–utility pairs (G, u) . We argue that this criterion is a natural normative constraint in many decision-making and learning settings. Furthermore, we argue that there is a conceptual duality in how causal information relates to probability and value: probability flows downstream (from

¹We use “value” and “utility” interchangeably in this paper.

causes to effects), while value flows upstream (from effects to causes). This duality allows us to formulate a translation scheme from causal-inference results to the value setting. Using it, we (i) define local, global, and decomposition versions of the v-CMC and prove they are equivalent; (ii) define *v*-separation and show it is sound and complete for conditional value independence; and (iii) derive the Bellman recursion (Bellman, 1957) as a special case, thereby clarifying when Bellman-style decomposition holds and yielding a principled elicitation procedure over causal DAGs. Importantly, we show how the resulting decomposition identifies which local utility terms must be revised, and which can be transferred unchanged, as causal assumptions change. Finally, we propose an algorithm for automated influence diagram construction.

1.1 RELATED WORK

Graphs for representing preference dependencies have been studied in Multi-Attribute Utility Theory (MAUT) (Keeney and Raiffa, 1993; Fishburn, 1974) and graphical extensions such as CP-nets (Boutilier et al., 2004), UCP-nets (Boutilier et al., 2001), and GAI networks (Bacchus and Grove, 1995). Related formalisms—such as Expected Utility Networks and conditional-utility-independence (CUI) networks—also study graphical representations of utility (or expected-utility) independencies (Mura and Shoham, 1999; Engel and Wellman, 2008; Leonelli and Smith, 2017). Most directly, Brafman and Engel (2009, 2010) introduce the subtractive notion of conditional value used here and show how it supports DAG-based value decompositions. These decompositions differ from ours in three key ways. First, in the existing literature DAGs encode *preference* rather than *causal* structure, aiming at a modular representation of whatever dependencies an agent happens to have. The v-CMC is complementary: it specifies which value independencies an agent’s preferences *should* exhibit to be compatible with causal information. Second, whereas existing (directed) graphical utility decompositions rely on parent-based “screening-off” criteria, the v-CMC maintains that children (in a causal graph) perform the screening-off role. Third, and crucially, the v-CMC provides a principled method for *updating* utility functions modularly and locally as our causal assumptions change. Finally, although the v-CMC is complementary to, and provides a causal basis for, MAUT elicitation (as we explain in Section 7), its treatment and development here are intentionally framework-neutral: the v-CMC is compatible with utility frameworks other than MAUT, some of which we mention later.

A closely related tradition is influence diagrams (IDs) (Howard and Matheson, 2005; Shachter, 1986; Everitt and Hutter, 2021; Everitt et al., 2021), where diamond-shaped value nodes encode which variables contribute to overall value, often assuming an additive decomposition for con-

venience. Our framework provides a more principled basis for these choices in that the local v-CMC guides which value dependencies should (or should not) be represented, and the decomposition v-CMC specifies when additive decomposition of value is warranted. Indeed, in Section 7, we propose an algorithm that uses the v-CMC to automatically construct influence diagrams that respect causal information. In sum, the v-CMC provides a common causal framework that can be applied across Multi-Attribute Utility Theory, reinforcement learning, and influence diagrams.

2 SETUP, NOTATION, AND BASIC ASSUMPTIONS

Let $G = (V, E)$ be a causal DAG with vertices $V = \{X_1, \dots, X_n\}$. We use uppercase letters such as X_i for variables and lowercase letters such as x_i for their realizations. When no confusion can arise, however, we will often write X_i rather than x_i for a particular realized value of X_i . For any $X_i \in V$, let $\text{Pa}(X_i)$ denote the parents of X_i in G , and let $\text{Ch}(X_i)$ denote its children. Let $\text{NA}(X_i)$ denote the set of nodes that are *not* ancestors of X_i , excluding X_i itself and its children $\text{Ch}(X_i)$.

We assume that we have, or are interested in constructing, a utility function u associated with the variables in G . Our treatment is framework-neutral: u may be interpreted in different ways depending on the decision-theoretic framework, and nothing in the formal development below depends on choosing one such interpretation. What matters for our purposes is only the domain and scale of u .

Let $\mathcal{P}(V)$ be the power set of V , i.e., the set of all subsets of V . The utility function need not be defined on every subset of variables in V . Rather, we assume that there is some collection $S \subseteq \mathcal{P}(V)$ of variable sets for which utilities are defined. For example, S might contain all subsets of V , or only those subsets relevant to a particular decision problem. For each $s \in S$, u assigns a real number to each joint realization of the variables in s . We also assume that all these utilities are measured on a common scale, unique up to positive affine transformations. Thus we define:

Definition 1 (Utility function). *Let $G = (V, E)$ be a causal DAG, and let $S \subseteq \mathcal{P}(V)$ be a collection of subsets of V . For each $s \in S$, let $\mathcal{X}(s)$ be the set of joint realizations of the variables in s . A utility function over S is a function*

$$u : \bigcup_{s \in S} (\{s\} \times \mathcal{X}(s)) \rightarrow \mathbb{R}, \quad (1)$$

unique up to positive affine transformations on a common scale across all $s \in S$.

The domain in Definition 1 may look needlessly cumbersome. The explicit $\{s\}$ -tag in the domain keeps track of

which variables are included in the utility assessment. For example, evaluating $X = x$ alone can be different from evaluating $X = x$ together with $Y = y$. Thus $u(s, x)$ records both the variable set s being evaluated and the realization x of that set. If $s \in S$ and $x \in \mathcal{X}(s)$, we write $u(x)$ rather than $u(s, x)$ whenever s is clear from context. When no confusion can arise, we also write $u(X)$ for the utility of a realization of the variable or variable set X .

The probabilistic CMC is stated in terms of conditional probabilities. To formulate an analogous principle for value, we therefore need a notion of conditional value. Following Brafman and Engel (2009, 2010); Bradley (2017), we define the value of x conditional on y as the additional value contributed by x once y is already given:

$$u(x \mid y) := u(x, y) - u(y). \quad (2)$$

Equivalently, $u(x, y) = u(y) + u(x \mid y)$. Thus, joint value can always be written as the value of y plus the conditional contribution of x given y . Note that this definition does not assume additive separability. Additive separability is obtained only in the special case where $u(x \mid y) = u(x)$, in which case $u(x, y) = u(x) + u(y)$. Note also that the subtractive definition generalizes the familiar notion of marginal utility. If the underlying variable y is continuous and x is interpreted as a small increment Δy , then $u(x \mid y) = u(y + \Delta y) - u(y) \approx u'(y)\Delta y$.

This definition is provided in terms of variables x and y , but it is straightforward to extend it to sets of variables. For disjoint sets of variables S, S' and realizations s, s' , define:

$$u(s \mid s') := u(s \cup s') - u(s'). \quad (3)$$

By convention, $u(s \mid \emptyset) := u(s)$ (so that $u(\emptyset) = 0$). This yields a natural notion of conditional value independence (Brafman and Engel, 2010):

Definition 2 (Conditional value independence). *Let A, B , and C be disjoint sets of variables. If for all realizations a, b, c we have $u(a \mid b, c) = u(a \mid c)$, then we write $(A \perp\!\!\!\perp_u B \mid C)$, and say that A is value independent of B given C .*

Finally, we impose a weak consistency constraint:

Definition 3 (Conditional consistency). *Let A, B, B' , and C be sets of variables with $B' \subseteq B$. Then*

$$(A \perp\!\!\!\perp_u B \mid C) \Rightarrow (A \perp\!\!\!\perp_u B' \mid C).$$

Conditional consistency says that if conditioning on C renders an entire collection of variables B irrelevant to the conditional value of A , then it should also render every subset of B irrelevant. If conditional consistency were false, then the whole of B could carry no value-relevant information about A , while some part of B nevertheless does.

This is arguably inconsistent with the intended meaning of screening off. Standard decision theoretic frameworks like the ones we discuss below automatically satisfy conditional consistency, but in our framework-neutral setup, we must impose it as an additional constraint. Conditional consistency ensures that conditional value independence is a semi-graphoid in the sense of Pearl and Paz (1987) (proof in appendix):

Theorem 1 (Conditional value independence obeys semi-graphoid axioms). *Let u be a value function defined as in (1), let conditional value be defined as in (3), and let conditional value independence be defined as in Definition 2. If conditional value independence obeys conditional consistency (Definition 3), then it obeys all the semi-graphoid axioms.*

The discussion up to now has been rather abstract, and some assumptions may appear strong. However, several existing theories in decision theory and reinforcement learning are in fact instantiations of the above framework:

A. In the Jeffrey–Bolker–Bradley framework (Bolker, 1967; Jeffrey, 1965; Bradley, 2017), utility (or desirability) is defined on a full algebra of propositions, i.e., on the full power set $\mathcal{P}(V)$. Bradley (2017) defines conditional utility as in (2) and gives a money pump argument for why this definition is uniquely correct.

B. In Multi-Attribute Utility Theory (MAUT) (Fishburn, 1974; Keeney and Raiffa, 1993), one specifies local von Neumann–Morgenstern utility functions on subsets $V' \subseteq V$ and calibrates them to a common scale; in this tradition, Brafman and Engel (2010) propose the subtractive definition (2) and the conditional independence relation (Definition 2), and show that this relation is semi-graphoid.

C. In reinforcement learning (RL), the value function is defined on nested subsets V_i , yielding a Bellman recursion $u(V_i) = r(v_i) + u(V_{i-1})$. The advantage function $Q(a, s) - V(s)$ can be viewed as a special case of the subtractive definition (Definition 2) under the obvious identification $Q(a, s) = u(a, s)$ and $V(s) = u(s)$. See Section 7 for more discussion of the relationship to RL.

Thus, our definition of conditional value (Definition 2) is not arbitrary, and can be justified in several different frameworks. Here we give an additional, framework-neutral justification. Since u is assumed unique only up to positive affine transformations, any definition of conditional value must be affine invariant. The following result shows that, under plausible conditions, the subtractive definition is the only definition of conditional value that is affine invariant.

Theorem 2 (Affine Invariance of Conditional Value). *Assume that $u(x, y)$ and $u(y)$ are utility functions that are unique up to arbitrary positive affine transformations. Assume, furthermore, that conditional utility is a function f of $u(x, y)$ and $u(y)$, i.e., $u(x \mid y) = f(u(x, y), u(y))$, such*

that f is differentiable at $(0, 0)$ and f is increasing in $u(x, y)$ when holding fixed $u(y)$. Finally, suppose f is invariant to all positive affine transformations of the utility scale, so that, for all $a > 0, b \in \mathbb{R}$:

$$f(aJ + b, aM + b) = af(J, M) \quad (4)$$

where $J = u(x, y)$ and $M = u(y)$. Then f must be the arithmetic difference: $f(J, M) = k(J - M)$, where k is a scaling factor that can be set to 1 without loss of generality.

3 STATEMENT AND EXPLANATION OF THE (LOCAL) v-CMC

We are now in a position to state the local version of the v-CMC, which states that a node’s causal children screen off our evaluation of the value of the node from the node’s non-ancestors:

Definition 4 (v-CMC (local formulation)). *Let G be a causal DAG with vertices $V = X_1, \dots, X_n$. Let u be a utility function defined on some set of subsets $S \subset \mathcal{P}(V)$. Then, for any set of non-ancestors $N \subset \text{NA}(X_i)$:*

$$X_i \perp\!\!\!\perp_u N \mid \text{Ch}(X_i)$$

whenever u is defined on the relevant subsets.

Since the v-CMC only applies to subsets on which u is defined, we will omit this qualification henceforth. However, all our results depend on the assumption that u is defined (or can be constructed) on the relevant subsets of interest.

The probabilistic CMC holds only for DAGs that are “causally sufficient” in the sense that, for any two nodes in the graph, any common cause of the nodes is also included. We need a similar condition in order to formulate our compatibility criterion:

Definition 5 (Value sufficiency). *Let G be a causal DAG with vertices $V = X_1, \dots, X_n$. Then G is value sufficient iff, for any two variables X_i and X_j , all shared causal effects of X_i and X_j are included in G .*

Value sufficiency is the exact dual of the classic definition of causal sufficiency given by Spirtes et al. (2000). It is arguably a principle that all decision-making implicitly assumes, whether or not it relies on an explicit causal principle such as the v-CMC. For example, you might judge that the utility of drinking orange juice is independent of the medication you are taking because you assume they are causally independent. However, if you learn (e.g., from medical evidence) that orange juice and the medication causally interact to produce a nasty side effect (an omitted joint causal effect), your utilities will change, as they should.

Thus, in practice, value sufficiency is a tentative modeling assumption on which all decision making relies, and its

adequacy depends on empirical and scientific facts about the underlying causal system. Typically, one will begin with a set of variables deemed relevant to the decision problem; value sufficiency then requires drawing arrows from any two variables into any shared colliders already in this set. If it turns out—on the basis of improved causal knowledge or revised modeling assumptions—that an important common effect has been omitted from the original variable set, both common sense and value sufficiency require adding it to the set and extending the utility function over the new variable. A key benefit of the v-CMC is that it enables this sort of extension to be carried out in a modular and local way, as we will see below. With these preliminaries in place, we can now formulate our causal compatibility criterion:

Definition 6 (Compatibility criterion). *Let G be a causal DAG with vertices $V = X_1, \dots, X_n$. Let u be a utility function defined on some set of subsets $S \subset \mathcal{P}(V)$. If G is value sufficient, then the pair (G, u) is compatible only if u obeys the v-CMC with respect to the subsets of G on which it is defined.*

Our contention is that rational decision-making should often be based on pairs (G, u) that are compatible in this sense. There are multiple arguments for this claim, which will be detailed below: first, it arguably accords with reasonable judgments in concrete cases; second, it follows from widely applicable premises; third, as we will see in Section 7, the v-CMC and Definition 6 provide a natural foundation and generalization of methods already widely applied in decision theory and reinforcement learning. We recognize that the v-CMC is likely to raise questions, and have therefore included a set of clarificatory notes in the appendix. However, two remarks are worth making already here.

First, the v-CMC, like its probabilistic dual, concerns variables that can stand in causal relationships and can therefore be meaningfully represented in a causal DAG. Many utility dependencies and independencies thus fall outside its scope. For example, it is natural to think that the disvalue of lying may depend on broader ethical considerations, such as whether Utilitarianism is true. However, an abstract proposition such as “Utilitarianism is true” is plausibly not a variable that can be meaningfully represented in a causal DAG. Thus, while such commitments may matter to an agent’s evaluative outlook, they fall outside the scope of a principle like the v-CMC.

Second, the v-CMC says that the children of a variable screen off its non-ancestors; it does not say that its *ancestors* are screened off. A strict consequentialist would probably argue that causal ancestors should also be screened off, since once some state B has occurred, its history should be irrelevant to how we assess its value, but the v-CMC does not make this demand. The v-CMC is therefore weaker than it may at first look: it does not rule out letting the causal history of a state influence how we value that state.

4 APPLICATIONS AND JUSTIFICATIONS OF THE COMPATIBILITY CRITERION

We begin with two simple applications of the v-CMC, to illustrate how it delivers reasonable verdicts in particular cases.

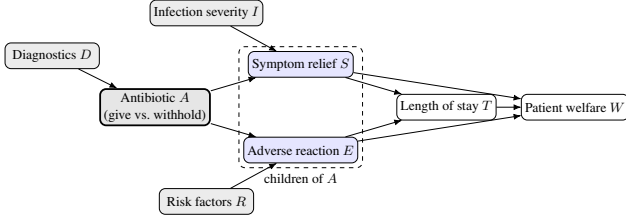


Figure 1: Antibiotic example: A affects S, E , which affect outcomes T, W ; D, I, R are exogenous.

Consider A in Figure 1, and suppose the graph is value sufficient. Since A 's causal children are S and E , the v-CMC implies $u(A \mid S, E, R) = u(A \mid S, E)$. That is, once we condition on symptom relief and an adverse reaction, the risk factor R provides no further information relevant to assessing the utility of administering the antibiotic. This seems reasonable: if we assume the subject *will* have an adverse reaction, it does not matter whether they have R ; likewise, if we assume they *will not* have an adverse reaction, R again makes no difference. Hence R is *value irrelevant* to administering the antibiotic once we condition on A 's children, as entailed by the local v-CMC.

A dynamic example sharpens the point. In Figure 1 it is reasonable that A and R are value dependent—e.g. $u(A=1 \mid R=1) < u(A=1 \mid R=0)$; the v-CMC permits, but does not force, such dependence. Suppose, however, that we discover R is not in fact a cause of E , while the remainder of the DAG is correct. In this modified graph, the v-CMC entails that A and R are value independent. This again seems appropriate: if R has no causal pathway to any variable affected by A , then R is *causally isolated* and should not influence whether to administer the antibiotic. Figure 1 also clarifies why value sufficiency is necessary for v-CMC: if E were erroneously omitted, the graph would be value insufficient since R and A jointly cause E . In that case, R would still be relevant to the value of administering the antibiotic, $u(A=1 \mid R=1) \neq u(A=1 \mid R=0)$; but because the only causal pathway making R relevant to A (namely via E) would be omitted, the causal children of A in the value insufficient graph would fail to screen off R from A .

We now give a more general defense of the Compatibility criterion in Definition 6 from three widely applicable assumptions. To motivate the first, note that the value of a node depends (in part) on its causally downstream effects: value “propagates upwards” from effects to causes, since a cause is more valuable insofar as it produces better effects. In a value-sufficient DAG, every causal path from a node

W to any descendant set D' must pass through its children $\text{Ch}(W)$, so $\text{Ch}(W)$ is a bottleneck: holding $\text{Ch}(W)$ fixed blocks every causal route by which W could affect D' . If the value of W still changes when one then learns downstream details D' , one is effectively treating D' as an additional channel of influence not represented in the graph. Thus, if the DAG is value sufficient, it is reasonable to require:

Definition 7 (Child mediation). *Let $G = (V, E)$ be a causal DAG and let u be a value function defined on $S \subseteq \mathcal{P}(V)$. Then u is child-mediated with respect to G iff, for every node W and for every set D' of causal descendants of W that are not children of W ,*

$$W \perp\!\!\!\perp_u D' \mid \text{Ch}(W). \quad (5)$$

In other words—recalling our definition of conditional utility— u satisfies child mediation iff, to evaluate the conditional value contribution of W over and above its causal effects D' , it suffices to consider the conditional value contribution of W over and above its most immediate effects $\text{Ch}(W)$. This does not mean that non-child descendants are irrelevant to total utility. Indeed, below we define a decomposition version of the v-CMC, which specifies how total utility decomposes over a causal DAG. Child mediation says only that, to determine the incremental contribution of W to total value, it suffices to consider the utility of W conditional on its children. This is reasonable in a value-sufficient DAG, since all of W 's causal effects must pass through $\text{Ch}(W)$.

This point is analogous to the distinction, familiar from the probabilistic CMC, between unconditional relevance and conditional screening-off. The local probabilistic CMC does not say that a node's non-descendants are irrelevant to its unconditional probability; it says that they are irrelevant to its conditional probability once the node's parents are fixed. Similarly, child mediation concerns conditional utility rather than total utility or the expected utility of an action—distal descendants can affect both of the latter even if they do not affect conditional utility.

In Figure 1, for example, child mediation says that the utility of administering the antibiotic A , *conditional* on its children S and E , does not depend on further causal descendants such as T . But T may still be relevant to the *expected utility* of administering A , since the probability of T may depend on A through S and E . For more on this point, see Section A of the Appendix.

The second and third assumptions draw on Causal Decision Theory (CDT) (Lewis, 1981; Joyce, 1999; Gibbard and Harper, 1978; Pearl, 2009; Stern, 2017), whose basic premise is that decisions should be based on the causal effects of the acts under consideration. Thus, the value of an intervention should not depend on causally irrelevant factors (i.e., non-descendants), a sensible requirement in many

economic, political, and medical decision problems (as in Figure 1).

Let $G_{do(W)}$ be obtained from G by severing all arrows into W . Let $u_{do(W)}$ be the (possibly updated) utility function associated with $G_{do(W)}$. As we will see below (Section 7), $u_{do(W)}$ will sometimes differ from u , and the v-CMC gives a principled method for how to discover these differences. However, a basic principle consistent with CDT is that if G' is a subgraph of G that is not affected by the intervention, then $u_{do(W)}$ should be identical to u on G' . More precisely, we assume:

Definition 8 (Utility modularity). *If $V' \subseteq V$ is such that the induced subgraph on V' is identical in G and $G_{do(W)}$, i.e. $G[V'] = G_{do(W)}[V']$, then for every $S \subseteq V'$ on which both are defined, $u_{do(W)}(S) = u(S)$.*

Finally, we require:

Definition 9 (CDT property). *Let $G = (V, E)$ be a causal DAG and let u be a value function defined on some set of subsets $S \subseteq \mathcal{P}(V)$. Let $W \in V$, and let $G_{do(W)}$ and $u_{do(W)}$ be defined as above. Let D be the set of all descendants of W in $G_{do(W)}$ and let ND be the set of non-descendants of W in $G_{do(W)}$. Then u has the CDT property with respect to G iff for any $ND' \subset ND$:*

$$W \perp\!\!\!\perp_{u_{do(W)}} ND' \mid D. \quad (6)$$

Although the CDT property may look formally complicated, it simply requires that the utility of an intervention be assessed on the basis of its causal effects alone, with other variables relevant only insofar as they causally influence those effects. We now have:

Theorem 3 (CDT-based justification of v-CMC). *Suppose G is a DAG and u is a value function on subsets of G satisfying:*

1. *u has the child-mediation property with respect to G .*
2. *u obeys modularity with respect to G .*
3. *u has the CDT property with respect to G .*

Then u obeys the local v-CMC on G .

This theorem provides widely applicable *sufficient* conditions for the v-CMC. Child mediation is reasonable to require of any utility function over a value-sufficient DAG, and utility modularity and the CDT property are, as noted, appealing in many important decision-making contexts. However, we do not claim that the v-CMC applies *only* under these conditions; a full discussion of its scope is beyond this paper.²

²See Section A of the Appendix for more discussion of the scope of the v-CMC.

5 THE CAUSAL PROBABILITY-VALUE DUALITY PRINCIPLE

Our discussion in the preceding section highlighted a striking asymmetry: in a causal graph, probability “flows downwards” whereas value “flows upwards.” In particular, the local v-CMC is the exact dual of the probabilistic CMC: whereas the CMC says that causal *parents* probabilistically screen off variables from their *non-descendants*, the v-CMC says that *children* screen off (in the value sense) variables from their *non-ancestors*. Taking this duality seriously suggests a systematic recipe for translating principles about DAGs and probability into corresponding principles about DAGs and value.

Table 1: The Causal Probability-Value Translation Key

Probability concept	Value concept
Probabilistic independence	Value independence
Parents ($\text{Pa}(\cdot)$)	Children ($\text{Ch}(\cdot)$)
Non-descendants	Non-ancestors
Common cause	Common effect
Parent-closed	Child-closed
$p(A \mid B) = \frac{p(A,B)}{p(B)}$	$u(A \mid B) = u(A, B) - u(B)$

Using Table 1, we can dualize any principle relating causal DAGs and probability into a corresponding principle relating DAGs and value. While the existence of a dual does not by itself guarantee its truth, we show in the next two sections that several such utility-theoretic duals of familiar causal-inference principles can in fact be proven.

6 EQUIVALENT FORMULATIONS AND GRAPHICAL CRITERION

As is well known, the probabilistic CMC admits three equivalent formulations: local, global, and decomposition (factorization). By the translation key in Table 1, each has a value dual. We first dualize the factorization formulation:

Definition 10 (v-CMC (decomposition)). *Let $G = (V, E)$ be a causal DAG. A subgraph $G' = (V', E')$ is child-closed if $X_i \in V'$ and $X_j \in \text{Ch}_G(X_i)$ implies $X_j \in V'$. We say that G and u jointly obey the decomposition v-CMC if for every child-closed G' with vertex set $V' = \{X_{i_1}, \dots, X_{i_m}\}$,*

$$u(X_{i_1}, \dots, X_{i_m}) = \sum_{k=1}^m u(X_{i_k} \mid \text{Ch}_G(X_{i_k})).$$

Readers may note that Definition 10 is somewhat more complex than the standard factorization version of the probabilistic CMC. This is because utility functions inherently have

less structure than probability functions; this extra structure must therefore be imposed in the definition. Readers familiar with the literature on generalized additive independence (GAI) utility measures may also note that Definition 10 essentially yields a causally based GAI decomposition.

Next, we formulate a global version via the value dual of d -separation:

Definition 11 (*v*-separation). *Let X and Y be nodes (or sets of nodes) in a DAG. We say that X and Y are v -separated by Z iff every path between them is blocked by Z , where a path is blocked iff: (1) it contains a chain $X \rightarrow M \rightarrow Y$ (or $Y \rightarrow M \rightarrow X$) with $M \in Z$; (2) it contains a common effect $X \rightarrow M \leftarrow Y$ with $M \in Z$; or (3) it contains a common cause $X \leftarrow M \rightarrow Y$ with neither M nor any ancestor of M in Z .*

For example, in Figure 1, the only path between A and R goes through the common effect E . Hence, E v -separates R from A .

By construction, v -separation is the exact dual of d -separation. Hence, if G^{Rev} is identical to G except that all the arrows are reversed, then X and Y are v -separated by Z on G if and only if X and Y are d -separated by Z on G^{Rev} —this fact will be useful below.

We are now in a position to formulate the global formulation of the v -CMC:

Definition 12 (v -CMC (global version)). *Let G be a causal DAG and u a value function. We say that G and u jointly obey the global v -CMC if whenever X and Y are v -separated by Z , then $X \perp\!\!\!\perp_u Y \mid Z$.*

In the appendix, we show that all three formulations of the v -CMC are equivalent:

Theorem 4 (v -CMC Equivalence Theorem). *Let G be a DAG and u a value function. Then:*

1. *Local v -CMC $\iff$ Global v -CMC.*
2. *Local v -CMC $\iff$ Decomposition v -CMC.*

Proof. Proof sketch: (i) **Local v -CMC $\iff$ Global v -CMC**: we first show that v -separation on a DAG G is equivalent to d -separation on the DAG G^{Rev} that is identical to G except that all the arrows are reversed (the fact we stated earlier). We then use this equivalence together with the fact that, for a semi-graphoid independence relation, the Global and Local versions of the CMC are equivalent (see Theorem 2.44 of Lauritzen (2019)) in order to conclude Local v -CMC $\iff$ Global v -CMC. (ii) **Local v -CMC $\iff$ Decomposition v -CMC**: The fact that the utility functions have less mathematical structure than probability functions (in particular, they cannot be marginalized) means we cannot easily use the relevant analogous proof from the causal inference

literature. Instead, we prove the $\leftarrow$ direction directly and the $\rightarrow$ direction by induction on the number of nodes in G . $\square$

Finally, v -separation provides a sound and complete graphical criterion for the class of value functions that obey v -CMC on G :

Theorem 5 (Soundness and Completeness). *Let $G = (V, E)$ be a causal DAG and let $X, Y, Z \subset V$ be disjoint.*

1. **Soundness**: *If X and Y are v -separated by Z in G , then $X \perp\!\!\!\perp_u Y \mid Z$ for all value functions u that obey the decomposition v -CMC on G .*
2. **Completeness**: *If X and Y are not v -separated by Z in G , then there exists a value function u that obeys the decomposition v -CMC on G such that $X \not\perp\!\!\!\perp_u Y \mid Z$.*

Proof. (i) **Completeness** (proof sketch; complete proof in the appendix): If X and Y are not v -separated in G , then they are not d -separated in G^{Rev} . Since d -separation is complete (Geiger et al., 1990b) and the set of unfaithful distributions has measure 0 (Meek, 1995), there exists a strictly positive probability function P satisfying the factorization property with respect to G^{Rev} such that $X \not\perp\!\!\!\perp_P Y \mid Z$. We then map P to the value function $u = \log P$, and show that conditional value independence (with respect to u) is a semi-graphoid, and that u obeys the decomposition property with respect to G . Finally, we show that $X \not\perp\!\!\!\perp_u Y \mid Z$, and hence u acts as a witness showing that if X and Y are not v -separated in G , then there exists a u such that $X \not\perp\!\!\!\perp_u Y \mid Z$.

(ii) **Soundness**: Suppose X and Y are v -separated by Z in G and let u be a value function that obeys the decomposition v -CMC on G . By Theorem 4, u obeys the global v -CMC on G , and therefore (by Definition 12), $X \perp\!\!\!\perp_u Y \mid Z$. $\square$

7 IMPLICATIONS AND APPLICATIONS

We now discuss several implications of the v -CMC, and present constructions it licenses. We begin by noting that the v -CMC generalizes the Bellman equation (Bellman, 1957) to causal DAGs.

7.1 A BELLMAN-TYPE RECURSION ON CAUSAL DAGS

Define the *height* of a node as the length of the longest directed path from that node to any leaf (a node with no children). Let $G_i := (V_i, E_i)$ be the subgraph of G consisting of all nodes of height at most i (so G_0 contains exactly the leaf nodes). By construction, each G_i is child-closed. We then obtain:

Theorem 6 (Bellman-type decomposition). *Let $G = (V, E)$ be a causal DAG and let u be a value function over V satisfying the v-CMC. Let $P_i := V_i \setminus V_{i-1}$. Then*

$$u(V_i) = \sum_{W \in P_i} u(W \mid \text{Ch}(W)) + u(V_{i-1}). \quad (7)$$

The proof is in the appendix. To see the connection with the standard Bellman equation, consider the special case of a linear order $W_0 \leftarrow W_1 \leftarrow \dots \leftarrow W_n$ of temporally ordered states, and assume that $u(W \mid \text{Ch}(W))$ is invariant to the values of $\text{Ch}(W)$. Then $u(W \mid \text{Ch}(W))$ depends only on W and can be interpreted as a reward function $r(W)$, in which case Theorem 6 reduces to the finite-horizon Bellman recursion

$$u(V_i) = r(W_i) + u(V_{i-1}).$$

Thus, the v-CMC provides a normative justification for Bellman recursion, while also clarifying both how it can be generalized and when the corresponding decomposition may fail: (1) if $u(W \mid \text{Ch}(W))$ depends on $\text{Ch}(W)$, the resulting decomposition is a strict generalization of the standard Bellman recursion; and (2) if the DAG is not value sufficient, the corresponding Bellman-style decomposition need not hold.

Note that the recursion in Theorem 6 is purely a recursion for the value function. It does not explicitly include a probability model or policy function, as the standard Bellman equations do, which is why we call it a “Bellman-type” recursion. However, extending the result is straightforward. By taking expectations with respect to a probability distribution and combining the probabilistic CMC with the v-CMC, one obtains a recursion for expected values. Since this extension introduces no essentially new ideas beyond those developed here, we omit the details.

In addition to allowing us to derive a causal version of the Bellman recursion, we note that the v-CMC also has a natural application in *inverse* reinforcement learning. Inverse reinforcement learning (Adams et al., 2022; Arora and Doshi, 2021) is concerned with reconstructing the reward function of an agent based on their behavior. As Arora and Doshi (2021) note, this problem is ill-posed because there is always a large number of reward functions that could potentially fit with a finite set of observed behaviors. The set of potential reward functions therefore needs to be constrained in some way. The v-CMC arguably provides a very natural constraint.

7.2 CAUSALLY STRUCTURED UTILITY ELICITATION

The recursion in the preceding section yields a simple elicitation procedure for any child-closed graph, which is given in

pseudocode in Algorithm 1. The algorithm requires no new elicitation scheme—it is compatible with standard MAUT methods, for example—but it structures elicitation in a principled way around an assumed causal DAG:

Algorithm 1: Elicitation algorithm

Input: Value-sufficient DAG $G = (V, E)$ of height H

Output: Value function u over V

for $i = 0$ **to** H **do**

$u(V_i) \leftarrow u(V_{i-1})$

foreach $W \in P_i$ **do**

 Elicit $u(W \mid \text{Ch}(W))$

$u(V_i) \leftarrow u(V_i) + u(W \mid \text{Ch}(W))$

end

end

Consider Figure 1. On a high level, the procedure then has four steps: **(1)** We begin by eliciting or assigning utilities to the patient’s welfare being high or low, $u(W = 1)$ and $u(W = 0)$. **(2)** Next, we elicit the value of Length of Stay T , conditional on its child W , i.e., $u(T = 1 \mid W)$ and $u(T = 0 \mid W)$. **(3)** We assess Symptom Relief (S) and Adverse Reaction (E). Their children are $\{T, W\}$, so we assess the utility of S and E given T and W , i.e., $u(S \mid T, W)$ and $u(E \mid T, W)$. **(4)** Finally, we assess the antibiotic A given its children S, E , i.e., $u(A \mid S, E)$.

Note that steps 1-4 yield local value functions. If each has been elicited using hypothetical lotteries, then they must be calibrated to a common scale. In an MAUT framework, this can be done by considering hypothetical cross-condition lotteries (see Keeney and Raiffa (1993) for methods).

This algorithm will generally have substantial complexity savings. Even in the simple example with binary variables (as in Figure 1), specifying the joint value function from scratch requires assigning values to all 2^n joint states, with n the number of nodes. By contrast, a rough upper bound on the number of elicitations required by Algorithm 1 is $n \max_{W \in V} 2^{1+|\text{Ch}(W)|}$, i.e., linear in the number of nodes.

7.3 UTILITY TRANSFER ACROSS CAUSAL CONTEXTS

As discussed in the introduction, a central motivation for the v-CMC is that transferring decisions across causal contexts requires not only transferring probabilistic information, but also determining which aspects of a utility function should remain stable when the underlying causal model changes. The causal decomposition provided by the v-CMC makes this possible by allowing an already constructed utility function to be modified locally and modularly as the causal graph changes. Causal graphs may change in several ways: for example, through interventions, node additions, node removals, edge additions, edge removals, or combinations

of these. To illustrate the basic idea, we consider two especially important cases: (1) intervening on a variable, and (2) expanding a DAG by adding newly relevant variables.

Interventions: suppose we intervene on a node X by setting $X = x$, thereby yielding the graph $G_{do(X)}$ where all arrows into X have been severed. This changes the child sets only for the parents of X : for any $P \in \text{Pa}(X)_G$, we have $X \notin \text{Ch}(P)_{G_{do(X)}}$, while all other nodes retain the same children. Hence, under the decomposition v-CMC, updating the decomposed utility representation to remain compatible with $G_{do(X)}$ requires revising only the local terms for those P whose child sets change—namely $u(P \mid \text{Ch}(P)_{G_{do(X)}})$ for $P \in \text{Pa}(X)_G$; all other local terms can be retained unchanged.

DAG expansions: suppose we discover that we have omitted a common causal effect E of variables in our variable set, so that the DAG is not value sufficient. We then expand G to include E , forming the new DAG G_{+E} . Extending u to G_{+E} is straightforward: by the decomposition v-CMC, it suffices to (i) determine the new local term $u(E \mid \text{Ch}(E)_{G_{+E}})$, and (ii) update the local terms $u(W \mid \text{Ch}(W)_{G_{+E}})$ for those W whose child sets changed—namely $W \in \text{Pa}(E)_{G_{+E}}$.

Note that in both cases the cost of updating depends on the size of the subset of affected nodes, not the total size of the graph, which is what makes the v-CMC representation modular in the sense promised above. Section H of the Appendix gives a worked example showing how the decomposition v-CMC can be used to update a joint utility modularly after interventions.

7.4 AUTOMATED INFLUENCE DIAGRAM CONSTRUCTION

Influence diagrams are an effective tool for structuring decision problems and can be used to represent both informational and causal relationships. The v-CMC provides a principled way to construct influence diagrams that respect causal structure. In an influence diagram with causal semantics, an arc from a variable node (i.e., a chance or decision node) to a value node indicates that the variable directly influences the value node. The decomposition formulation of the v-CMC (Definition 10) can therefore be used to construct a canonical influence diagram from a value-sufficient DAG of variable nodes as follows. For each nonzero local term $u(X_i \mid \text{Ch}(X_i))$, introduce a value node representing the local value contributed by X_i , and draw arcs into that value node from X_i and from each member of $\text{Ch}(X_i)$. Variables for which $u(X_i \mid \text{Ch}(X_i)) \equiv 0$ have no local value contribution and may be omitted from the value-node construction, even though they may still matter instrumentally through their causal effects.

Thus, in summary, the procedure is: (i) input a value-

sufficient DAG and the set of variables with nonzero local value contribution; (ii) create one value node for each such variable; and (iii) connect that value node to the variable and its causal children. For example, applying Algorithm 2 to the DAG in Figure 1, with S , E , T , and W identified as the variables with nonzero local value contribution, yields Figure 2.

Algorithm 2: Influence diagram construction algorithm

Input: Value-sufficient DAG $G = (V, E)$; Variable set V'

Output: Canonical influence diagram

for $w \in V'$ **do**

 Construct value node n for w

 Draw arc from w to n

foreach $c \in \text{Ch}(w)$ **do**

 Draw arc from c to n

end

end

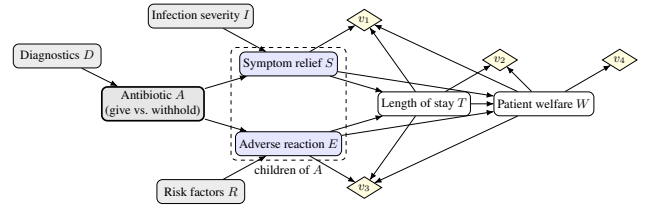


Figure 2: Antibiotic example with value nodes attached to S , E , T , and W .

Note that the value nodes are constructed to agree, in a term-by-term fashion, with the decomposition over the DAG given by the decomposition v-CMC (10). Thus, the value nodes can be added: $v_{total} = v_1 + v_2 + v_3 + v_4$.

8 CONCLUSION

This paper has developed the conceptual and mathematical foundations of how causal assumptions constrain value functions, and we have relied on assumptions it is natural to try to weaken. For example, the v-CMC relates a single assumed DAG to a value function; an important next step is to extend the framework to settings with uncertainty about the DAG, and to causal structures with cycles.

We have also explored only a small set of applications. Further directions include using the v-CMC as a causal constraint in inverse RL (as suggested earlier), developing robust ways of handling possible value insufficiency, and using v -separation as a diagnostic tool for checking compatibility of reported utilities with causal assumptions. Taken together, these directions suggest a broader research program in *causal value theory*, in which causal structure plays a central normative role in shaping rational preferences.

Acknowledgements

Thanks to Steven Diggin, Malcolm Forster, Reuben Stern, Joshua Thong, and several reviewers at UAI who made the paper better than it otherwise would have been. This paper is part of the research project “Towards a Theory of Rational Desires” whose broader goal is to investigate normative constraints on value functions. Work on the project is supported by funding from the European Research Council (ERC) under the European Union’s Horizon Europe research and innovation programme (Grant agreement No. 101164097).

References

- Adams, S., Cody, T., and Beling, P. A. (2022). A survey of inverse reinforcement learning. *Artificial Intelligence Review*, 55:4307–4346.
- Anscombe, F. J. and Aumann, R. J. (1963). A definition of subjective probability. *Annals of Mathematical Statistics*, 34:199–205.
- Arora, S. and Doshi, P. (2021). A survey of inverse reinforcement learning: Challenges, methods and progress. *Artificial Intelligence*, 297:103500.
- Bacchus, F. and Grove, A. J. (1995). Graphical models for preference and utility. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI-95)*, pages 3–10, San Francisco. Morgan Kaufmann.
- Bellman, R. (1957). *Dynamic Programming*. Princeton University Press, Princeton, NJ.
- Bolker, E. D. (1967). A simultaneous axiomatization of utility and subjective probability. *Philosophy of Science*, 34(4):333–340.
- Boutilier, C., Bacchus, F., and Brafman, R. I. (2001). UCP-networks: A directed graphical representation of conditional utilities. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, pages 56–64, San Francisco. Morgan Kaufmann.
- Boutilier, C., Brafman, R. I., Domshlak, C., Hoos, H. H., and Poole, D. (2004). CP-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. *Journal of Artificial Intelligence Research*, 21:135–191.
- Bradley, R. (2017). *Decision Theory with a Human Face*. Cambridge University Press.
- Brafman, R. I. and Engel, Y. (2009). Conditional utility independence: A new notion with applications in multiattribute utility theory. In *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI 2009)*, pages 389–394.
- Brafman, R. I. and Engel, Y. (2010). Decomposed utility functions and graphical models for reasoning about preferences. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, 267–272.
- Engel, Y. and Wellman, M. P. (2008). CUI networks: A graphical representation for conditional utility independence. *Journal of Artificial Intelligence Research*, 31:83–112.
- Everitt, T., Carey, R., Langlois, E. D., Ortega, P. A., and Legg, S. (2021). Agent incentives: A causal perspective. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 11487–11495.
- Everitt, T. and Hutter, M. (2021). Causal influence diagrams for safe and fair AI. *Artificial Intelligence*, 296:103479.
- Fishburn, P. C. (1974). Seven independence concepts and continuous multiattribute utility functions. *Journal of Mathematical Psychology*, 11(3):337–359.
- Geiger, D., Verma, T., and Pearl, J. (1990a). d-separation: From theorems to algorithms. In *Proceedings of the Fifth Conference on Uncertainty in Artificial Intelligence*, pages 139–148. Elsevier.
- Geiger, D., Verma, T. S., and Pearl, J. (1990b). Identifying independence in Bayesian networks. *Networks*, 20(5):507–534.
- Gibbard, A. and Harper, W. L. (1978). Counterfactuals and two kinds of expected utility. In Hooker, C., Leach, J. J., and McClennen, E. F., editors, *Foundations and Applications of Decision Theory*, pages 125–162. Reidel, Dordrecht.
- Howard, R. A. and Matheson, J. E. (2005). Influence diagrams. *Decision Analysis*, 2(3):127–143.
- Jeffrey, R. C. (1965). *The Logic of Decision*. McGraw-Hill.
- Jeffrey, R. C. (1983). *The Logic of Decision*. University of Chicago Press, 2nd edition.
- Joyce, J. M. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press, Cambridge.
- Keeney, R. L. and Raiffa, H. (1993). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Cambridge University Press, Cambridge, 2nd edition.
- Lauritzen, S. L. (2019). Lectures on graphical models. 3rd electronic edition (lecture notes).
- Leonelli, M. and Smith, J. Q. (2017). Directed expected utility networks. *Decision Analysis*, 14(2):108–125.
- Lewis, D. (1981). Causal decision theory. *Australasian Journal of Philosophy*, 59(1):5–30.

- Meek, C. (1995). Strong completeness and faithfulness in Bayesian networks. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 411–418.
- Mura, P. L. and Shoham, Y. (1999). Expected utility networks. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI-99)*, pages 366–373, San Francisco. Morgan Kaufmann.
- Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, 2nd edition.
- Pearl, J. and Paz, A. (1987). Graphoids: A graph-based logic for reasoning about relevance relations. In du Boulay, B., Hogg, D., and Steels, L., editors, *Advances in Artificial Intelligence-II*, pages 357–363. North-Holland.
- Savage, L. J. (1954). *The Foundations of Statistics*. Wiley.
- Shachter, R. D. (1986). Evaluating influence diagrams. *Operations Research*, 34(6):871–882.
- Spirtes, P., Glymour, C. N., and Scheines, R. (2000). *Causation, Prediction, and Search*. MIT Press, Cambridge, MA, 2nd edition.
- Stern, R. (2017). Interventionist decision theory. *Synthese*, 194:4133–4153.
- Verma, T. and Pearl, J. (1988). Causal networks: Semantics of conditional independence. In *Proceedings of the Fourth Conference on Uncertainty in Artificial Intelligence*, pages 69–78.
- von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press.

A Causal Markov Condition for Value (Supplementary Material)

Olav Benjamin Vassend¹

¹University of Inland Norway

A A FEW CLARIFICATORY NOTES ABOUT THE V-CMC

Because the v-CMC is likely to prompt several questions—as well as possibly objections—in readers’ minds, we include here several (hopefully) clarifying notes about its scope.

First, we noted earlier (in Section 4) that the v-CMC is not committed to a causal-decision-theoretic view of decision theory, even though our justification relies on a “CDT property.” In fact, we do not think the plausibility of the v-CMC even depends on having a consequentialist view of value. For example, a non-consequentialist, deontological conception that regards the value of an act as determined by the motive behind the act is also compatible with the v-CMC, provided that the agent’s motive can be represented as a variable in the causal model. In this case, causal non-child descendants are vacuously screened off, since *all* causal consequences are regarded as irrelevant. Furthermore, non-ancestors that are not descendants are *also* irrelevant, since the motive behind an act is a causal ancestor of the act. We thus think the v-CMC is compatible with both consequentialist and non-consequentialist conceptions of value, although a full explanation and defense of this claim is beyond the scope of this paper.

Second, the v-CMC constrains conditional utilities, such as $u(A \mid \text{Ch}(A))$, not expected utilities such as $\text{EU}[A]$. For example, in Figure 1, v-CMC implies that $u(A \mid S, E, I) = u(A \mid S, E)$, but I obviously affects the probability of S , and since S affects the value of A , it is reasonable to expect I to affect $\text{EU}[A]$, even though it is screened off in the conditional value assessment. Indeed, writing (for simplicity) the expected utility of administering the antibiotic as

$$\text{EU}[A \mid I] = \sum_{s,e} u(A \mid s, e) P(s, e \mid A, I),$$

we see that even if $u(A \mid s, e, I) = u(A \mid s, e)$, the quantity $P(s, e \mid A, I)$ may still depend on I , and hence so may $\text{EU}[A \mid I]$.

Third, even though the v-CMC implies that A and B are value independent if neither is a cause of the other and they have no common causal effect, this does not mean that A and B must be value independent *conditionally* on all other variables. Consider Figure 1 again. The utility of the patient’s having an adverse reaction, given that you have administered the antibiotic to them and they experienced symptom relief is plausibly *higher* than the value of the patient’s having an adverse reaction given that you administered the antibiotic and they have *no* symptom relief. Given that you have administered the antibiotic, having symptom relief plausibly compensates, to some degree, for having an adverse reaction. More generally, even if X and Y are value independent, conditioning on a common ancestor may render X and Y conditionally value dependent. This is analogous to how conditioning on a common effect may render X and Y conditionally probabilistically dependent.

B PROOF THAT CONDITIONAL VALUE INDEPENDENCE OBEYS SEMI-GRAPHOID AXIOMS

First, we list the semi-graphoid axioms (Pearl and Paz, 1987):

Definition 13 (The semi-graphoid axioms). *Let X, Y, W, Z be disjoint sets of variables. A conditional independence relation $\perp\!\!\!\perp_u$ is a semi-graphoid if it satisfies:*

1. **Symmetry:**

$$(X \perp\!\!\!\perp_u Y \mid Z) \Rightarrow (Y \perp\!\!\!\perp_u X \mid Z).$$

2. **Semi-graphoid decomposition:**

$$(X \perp\!\!\!\perp_u Y \cup W \mid Z) \Rightarrow (X \perp\!\!\!\perp_u Y \mid Z) \text{ and } (X \perp\!\!\!\perp_u W \mid Z).$$

3. **Weak Union:**

$$(X \perp\!\!\!\perp_u Y \cup W \mid Z) \Rightarrow (X \perp\!\!\!\perp_u Y \mid Z \cup W).$$

4. **Contraction:**

$$(X \perp\!\!\!\perp_u Y \mid Z) \text{ and } (X \perp\!\!\!\perp_u W \mid Z \cup Y) \Rightarrow (X \perp\!\!\!\perp_u Y \cup W \mid Z).$$

We now prove the theorem:

Theorem 7 (Conditional value independence obeys semi-graphoid axioms). *Let u be a value function defined as in (1), let conditional value be defined as in (3), and let conditional value independence be defined as in (2). If conditional value independence has the Consistency property (Definition 3), then it obeys all the semi-graphoid axioms.*

Proof. Throughout, let X, Y, W, Z be disjoint sets of variables. Recall that, by the definition of conditional value (Definition 3),

$$u(X \mid Y, Z) = u(X \cup Y \cup Z) - u(Y \cup Z), \quad u(X \mid Z) = u(X \cup Z) - u(Z).$$

Hence, by Definition 2,

$$(X \perp\!\!\!\perp_u Y \mid Z) \iff u(X \cup Y \cup Z) - u(Y \cup Z) = u(X \cup Z) - u(Z), \quad (8)$$

for all realizations of (X, Y, Z) (we suppress explicit realization notation for readability).

Symmetry. Assume $(X \perp\!\!\!\perp_u Y \mid Z)$. By (8),

$$u(X \cup Y \cup Z) = u(Y \cup Z) + u(X \cup Z) - u(Z).$$

Subtracting $u(X \cup Z)$ from both sides yields

$$u(X \cup Y \cup Z) - u(X \cup Z) = u(Y \cup Z) - u(Z),$$

i.e.,

$$u(Y \mid X, Z) = u(Y \mid Z).$$

Thus $(Y \perp\!\!\!\perp_u X \mid Z)$, proving Symmetry.

Contraction. Assume $(X \perp\!\!\!\perp_u Y \mid Z)$ and $(X \perp\!\!\!\perp_u W \mid Z \cup Y)$. Expanding each by (8) gives:

$$u(X \cup Y \cup Z) - u(Y \cup Z) = u(X \cup Z) - u(Z), \quad (9)$$

$$u(X \cup W \cup Y \cup Z) - u(W \cup Y \cup Z) = u(X \cup Y \cup Z) - u(Y \cup Z). \quad (10)$$

Substituting (9) into the right-hand side of (10) yields

$$u(X \cup W \cup Y \cup Z) - u(W \cup Y \cup Z) = u(X \cup Z) - u(Z),$$

which, by (8), is exactly

$$(X \perp\!\!\!\perp_u Y \cup W \mid Z).$$

Thus Contraction holds.

Semi-graphoid decomposition. Assume $(X \perp\!\!\!\perp_u Y \cup W \mid Z)$. By the Consistency property (Definition 3), independence from a set implies independence from any subset of that set. Since $Y \subseteq Y \cup W$ and $W \subseteq Y \cup W$, we obtain

$$(X \perp\!\!\!\perp_u Y \mid Z) \quad \text{and} \quad (X \perp\!\!\!\perp_u W \mid Z).$$

Thus Decomposition holds.

Weak Union. Assume $(X \perp\!\!\!\perp_u Y \cup W \mid Z)$. We must show $(X \perp\!\!\!\perp_u Y \mid Z \cup W)$, i.e.,

$$u(X \cup Y \cup W \cup Z) - u(Y \cup W \cup Z) = u(X \cup W \cup Z) - u(W \cup Z). \quad (11)$$

From $(X \perp\!\!\!\perp_u Y \cup W \mid Z)$ and (8), we have

$$u(X \cup Y \cup W \cup Z) - u(Y \cup W \cup Z) = u(X \cup Z) - u(Z). \quad (12)$$

By Decomposition (already established, using Consistency) applied to the same assumption $(X \perp\!\!\!\perp_u Y \cup W \mid Z)$, we also have $(X \perp\!\!\!\perp_u W \mid Z)$.

Expanding this by (8) yields

$$u(X \cup W \cup Z) - u(W \cup Z) = u(X \cup Z) - u(Z). \quad (13)$$

Combining (12) and (13) shows that both sides of (11) are equal to $u(X \cup Z) - u(Z)$, and hence (11) holds. Therefore $(X \perp\!\!\!\perp_u Y \mid Z \cup W)$, proving Weak Union.

Since conditional value independence satisfies Symmetry, Decomposition, Weak Union, and Contraction, it obeys all the semi-graphoid axioms. $\square$

C UNIQUENESS OF THE SUBTRACTIVE DEFINITION

Theorem 2 (Affine Invariance of Conditional Value). *Assume that $u(x, y)$ and $u(y)$ are utility functions that are unique up to arbitrary positive affine transformations. Assume, furthermore, that conditional utility is a function f of $u(x, y)$ and $u(y)$, i.e., $u(x \mid y) = f(u(x, y), u(y))$, such that f is differentiable at $(0, 0)$ and f is increasing in $u(x, y)$ when holding fixed $u(y)$. Finally, suppose f is invariant to all positive affine transformations of the utility scale, so that, for all $a > 0, b \in \mathbb{R}$:*

$$f(aJ + b, aM + b) = af(J, M) \quad (4)$$

where $J = u(x, y)$ and $M = u(y)$. Then f must be the arithmetic difference: $f(J, M) = k(J - M)$, where k is a scaling factor that can be set to 1 without loss of generality.

Proof. Let $a = 1$. Then the condition becomes:

$$f(J + b, M + b) = f(J, M).$$

Let $b = -M$. Then $f(J - M, 0) = f(J, M)$. Define $g(z) := f(z, 0)$. Then $f(J, M) = g(J - M)$, so f depends only on the difference $J - M$.

Substituting into the general condition yields

$$g((aJ + b) - (aM + b)) = ag(J - M),$$

so the b terms cancel and we obtain

$$g(a(J - M)) = ag(J - M).$$

Letting $z = J - M$, this is $g(az) = ag(z)$ for all $a > 0$. Hence, for $z > 0$ we may set $a = z$ to get $g(z) = z g(1)$, and for $z < 0$ writing $z = -t$ with $t > 0$ gives $g(z) = g(-t) = t g(-1)$. Let $g(1) = k_+$ and $g(-1) = -k_-$ (equivalently $k_- := -g(-1)$). Then

$$g(z) = \begin{cases} k_+ z & \text{if } z \geq 0, \\ k_- z & \text{if } z \leq 0, \end{cases} \quad \text{with } k_+, k_- > 0.$$

Since f is differentiable at $(0, 0)$, $g(z) = f(z, 0)$ is differentiable at 0, which implies that the left and right derivatives at 0 agree, hence $k_+ = k_- =: k$. Therefore $f(J, M) = k(J - M)$, i.e., $u(x | y) = k(u(x, y) - u(y))$. Since f is increasing in $u(x, y)$ (holding $u(y)$ fixed), we have $k > 0$. Finally, k can be absorbed into the overall utility scale, so we may set $k = 1$ without loss of generality. □

D PROOF OF THE CDT-BASED JUSTIFICATION OF V-CMC

We first restate the relevant definitions:

Definition 7 (Child mediation). *Let $G = (V, E)$ be a causal DAG and let u be a value function defined on $S \subseteq \mathcal{P}(V)$. Then u is child-mediated with respect to G iff, for every node W and for every set D' of causal descendants of W that are not children of W ,*

$$W \perp\!\!\!\perp_u D' \mid \text{Ch}(W). \quad (5)$$

Definition 9 (CDT property). *Let $G = (V, E)$ be a causal DAG and let u be a value function defined on some set of subsets $S \subseteq \mathcal{P}(V)$. Let $W \in V$, and let $G_{do(W)}$ and $u_{do(W)}$ be defined as above. Let D be the set of all descendants of W in $G_{do(W)}$ and let ND be the set of non-descendants of W in $G_{do(W)}$. Then u has the CDT property with respect to G iff for any $ND' \subset ND$:*

$$W \perp\!\!\!\perp_{u_{do(W)}} ND' \mid D. \quad (6)$$

Here is the theorem we wish to prove:

Theorem 3 (CDT-based justification of v-CMC). *Suppose G is a DAG and u is a value function on subsets of G satisfying:*

1. *u has the child-mediation property with respect to G .*
2. *u obeys modularity with respect to G .*
3. *u has the CDT property with respect to G .*

Then u obeys the local v-CMC on G .

Proof. Fix an arbitrary node $W \in V$. We want to show that for every set $N \subseteq \text{NA}(W)_G$,

$$W \perp\!\!\!\perp_u N \mid \text{Ch}(W) \quad (14)$$

Let D denote the set of all descendants of W in G . Note that intervening on W only removes arrows into W , so it does not change which nodes are reachable *from* W . Hence, the set of descendants of W is the same in G and $G_{do(W)}$; in particular, we may use the same D for both graphs.

Now take any $N \subseteq \text{NA}(W)_G$. We first decompose N into its descendant and non-descendant parts:

$$N_D := N \cap (D \setminus \text{Ch}(W)), \quad N_{ND} := N \setminus D.$$

Then $N = N_D \cup N_{ND}$, and these sets are disjoint. Furthermore, $N_D \subseteq D \setminus \text{Ch}(W)$. Using child mediation (5) with $D \setminus \text{Ch}(W)$ yields:

$$W \perp\!\!\!\perp_u (D \setminus \text{Ch}(W)) \mid \text{Ch}(W). \quad (15)$$

Use $\perp\!\!\!\perp_{u_{do(W)}}$ to denote the independence relation induced by $u_{do(W)}$. Since $N_{ND} \subseteq ND$, the CDT property (6) implies

$$W \perp\!\!\!\perp_{u_{do(W)}} N_{ND} \mid D,$$

Note that the graphical relations between W , N_{ND} , and D are not affected by the intervention $do(W)$. Hence, by our assumption that $u_{do(W)}$ and u are identical on all subgraphs unaffected by $do(W)$, we have:

$$W \perp\!\!\!\perp_{u_{do(W)}} N_{ND} \mid D \iff \tag{16}$$

$$W \perp\!\!\!\perp_u N_{ND} \mid D \tag{17}$$

And therefore, since $D = \text{Ch}(W) \cup (D \setminus \text{Ch}(W))$, this can be written as follows:

$$W \perp\!\!\!\perp_u N_{ND} \mid \text{Ch}(W), (D \setminus \text{Ch}(W)) \tag{18}$$

Combining (15) and (18) using Contraction gives:

$$W \perp\!\!\!\perp_u N_{ND} \cup (D \setminus \text{Ch}(W)) \mid \text{Ch}(W).$$

Finally, using conditional consistency (3), since $N_D \subseteq D \setminus \text{Ch}(W)$, this implies:

$$W \perp\!\!\!\perp_u N_{ND} \cup N_D \mid \text{Ch}(W). \tag{19}$$

which is exactly (14). Since W and $N \subseteq \text{NA}(W)_G$ were arbitrary, u obeys the local v-CMC on G .

□

E PROOFS OF CMC EQUIVALENCES

Let us first restate the theorem:

Theorem 4 (v-CMC Equivalence Theorem). *Let G be a DAG and u a value function. Then:*

1. *Local v-CMC $\iff$ Global v-CMC.*
2. *Local v-CMC $\iff$ Decomposition v-CMC.*

To show that the global and local versions of the v-CMC are equivalent, we need several results. First, we need the semi-graphoid version of the CMC. Define:

Definition 14 (Local SG-CMC). *An independence relation $\perp\!\!\!\perp$ and DAG $G = (V, E)$ obey the local semi-graphoid CMC (SG-CMC) if and only if, for every $W \in V$, and every set of nodes ND in the set of non-descendants of W (excluding W 's parents):*

$$W \perp\!\!\!\perp ND \mid \text{Pa}(W) \tag{20}$$

Definition 15 (Global SG-CMC). *An independence relation $\perp\!\!\!\perp_u$ and DAG $G = (V, E)$ obey the global semi-graphoid CMC (SG-CMC) if and only if, for every disjoint set of nodes X, Y , and Z , if Z d-separates X from Y in G , then $X \perp\!\!\!\perp_u Y \mid Z$.*

We need the following result (see Theorem 2.44 of Lauritzen (2019)):

Theorem 8. *Let $G = (V, E)$ be a DAG and let $\perp\!\!\!\perp$ be a semi-graphoid independence relation on V . Then: Local SG-CMC $\iff$ Global SG-CMC*

Next, we need two useful lemmas. Define G^{Rev} as the DAG obtained from G by reversing every edge. We then have the following results:

Lemma 1 (Local v-CMC/CMC duality). *Let G and G^{Rev} be defined as above and let u be a utility function defined on subsets of V , then u and G satisfy the local v-CMC if and only if u and G^{Rev} obey the local SG-CMC.*

Proof. This is immediate by the definition of G^{Rev} since $\text{Ch}(W)$ in G is $\text{Pa}(W)$ in G^{Rev} , and the set of non-descendants of W (excluding W 's parents) in G^{Rev} is the set of non-ancestors of W (excluding W 's children) in G . $\square$

Lemma 2 (v-separation/d-separation duality). *Let G and G^{Rev} be defined as above, then:*

A and B are v-separated by C in $G \iff A$ and B are d-separated by C in G^{Rev} .

Proof. The proof just amounts to verifying that translating the conditions for v-separation using the translation key in 1 gives d-separation on G^{Rev} . By definition, v-separation in G blocks paths based on three conditions:

- Chains $A \rightarrow M \rightarrow B$ where $M \in C$.
- Common Effects $A \rightarrow M \leftarrow B$ where $M \in C$.
- Common Causes $A \leftarrow M \rightarrow B$ where neither M nor its ancestors are in C .

In G^{Rev} , all edges are reversed, so:

- Chains become chains $A \leftarrow M \leftarrow B$, blocked if $M \in C$.
- Common Effects become Common Causes $A \leftarrow M \rightarrow B$, blocked if $M \in C$.
- Common Causes become Common Effects $A \rightarrow M \leftarrow B$, blocked if neither M nor its descendants (which were ancestors in G) are in C .

This mapping shows that v-separation in G corresponds exactly to the definition of d-separation in G^{Rev} , as desired. $\square$

We are now in a position to prove part 1 of Theorem 4.

Global $\leftrightarrow$ Local

Proof. Note that, by Lemma 1, $\perp\!\!\!\perp_u$ and G jointly obey the local v-CMC if and only if $\perp\!\!\!\perp_u$ and G^{Rev} jointly obey the semi-graphoid version of the local CMC for DAGs. Similarly, by Lemma 2 $\perp\!\!\!\perp_u$ and G jointly obey the global v-CMC if and only if $\perp\!\!\!\perp_u$ and G^{Rev} jointly obey the semi-graphoid version of the global CMC for DAGs. Hence, since the local and global versions of the CMC are equivalent by Theorem 8, it follows that the local and global versions of the v-CMC are equivalent as well. $\square$

Next, we prove the second part of Theorem 4:

Local $\rightarrow$ Decomposition

Proof. Let $G = (V, E)$ be the DAG and order the vertices $(Y_1, \dots, Y_n)$ so that if $Y_i \rightarrow Y_j$ then $i < j$ (parents precede children). In particular, Y_1 is a root (it has no parents). For brevity write $Y_{>i} := \{Y_{i+1}, \dots, Y_n\}$. Let $\text{Ch}(Y_i)$ denote the children of Y_i (in G), and let $\text{nCh}(Y_{>i}) := Y_{>i} \setminus \text{Ch}(Y_i)$. Note that under this ordering, $\text{nCh}(Y_{>i})$ contains no ancestors of Y_i . Moreover, $Y_{>i} = \text{Ch}(Y_i) \cup \text{nCh}(Y_{>i})$. Because $\text{nCh}(Y_{>i})$ contains no ancestors of Y_i , the local v-CMC implies that $u(Y_i \mid \text{Ch}(Y_i), \text{nCh}(Y_{>i})) = u(Y_i \mid \text{Ch}(Y_i))$, for all Y_i .

We now prove the statement by induction. If G has one vertex, then $u(Y_1) = u(Y_1 \mid \emptyset) = u(Y_1 \mid \text{Ch}(Y_1))$, and hence the decomposition property holds (vacuously). Now suppose the decomposition property holds for all graphs with fewer than n vertices. Note that this assumption immediately entails that every proper subgraph of G has the decomposition property. Hence, it suffices to show that G itself can be decomposed into a sum of local conditional value terms. Note that the vertices $\{Y_2, \dots, Y_n\}$ and edges between these vertices form a graph G' . Let us use $\text{Ch}(Y_i)_{G'}$ to denote the children of Y_i in the graph G' . Our ordering guarantees that Y_1 is not a child of any Y_i . Therefore, $\text{Ch}(Y_i)_{G'} = \text{Ch}(Y_i)_G$ for every Y_i , which allows us to drop subscripts and write $\text{Ch}(Y_i)$.

Using the definition of conditional value (3), we can write:

$$u(Y_1, \dots, Y_n) \quad (21)$$

$$= u(Y_1 \mid Y_2, \dots, Y_n) + u(Y_2, \dots, Y_n) \quad (22)$$

$$= u(Y_1 \mid \text{Ch}(Y_1), \text{nCh}(Y_{>1})) + u(Y_2, \dots, Y_n) \quad (23)$$

$$= u(Y_1 \mid \text{Ch}(Y_1)) + u(Y_2, \dots, Y_n) \quad (24)$$

$$= u(Y_1 \mid \text{Ch}(Y_1)) + \sum_{i=2}^n u(Y_i \mid \text{Ch}(Y_i)_{G'}) \quad (25)$$

$$= \sum_{i=1}^n u(Y_i \mid \text{Ch}(Y_i)) \quad (26)$$

Where, in the penultimate step, we used v-CMC, and in the last step we used the induction hypothesis. Since Y_i is just a relabeling of X_i and u is insensitive to arbitrary relabelings, we have:

$$u(X_1, \dots, X_n) = u(Y_1, \dots, Y_n) \quad (27)$$

$$= \sum_{i=1}^n u(Y_i \mid \text{Ch}(Y_i)) = \sum_{i=1}^n u(X_i \mid \text{Ch}(X_i)) \quad (28)$$

This establishes that G itself can be decomposed into local conditional terms, and hence completes the proof. $\square$

Decomposition $\rightarrow$ Local

Proof. Fix a node $X_j \in V$. Let

$$N_j := \text{nAn}X_{jG} \setminus \text{Ch}(X_j)_G,$$

i.e. the set of non-ancestors of X_j in G that are not children of X_j . Let $T := \{X_j\} \cup N_j$ and let S be the (induced) subgraph obtained by closing T under children: starting from T , iteratively add $\text{Ch}(X)$ for every X already included, until no new nodes are added. By construction, S is child-closed.

Moreover, X_j has no ancestors in S : all nodes added to T are descendants (via child-closure) of nodes in T , so no directed path into X_j can be created inside S . Hence $S \setminus \{X_j\}$ is also child-closed, because if some $Y \in S \setminus \{X_j\}$ had X_j as a child then Y would be an ancestor of X_j in S , a contradiction.

Since G has the decomposition property, it applies to the child-closed subgraphs S and $S \setminus \{X_j\}$:

$$u(S) = \sum_{X_i \in S} u(X_i \mid \text{Ch}(X_i)_G), \quad u(S \setminus \{X_j\}) = \sum_{X_i \in S \setminus \{X_j\}} u(X_i \mid \text{Ch}(X_i)_G).$$

Subtracting gives

$$u(S) - u(S \setminus \{X_j\}) = u(X_j \mid \text{Ch}(X_j)_G).$$

By the definition of conditional value, the left-hand side equals $u(X_j \mid S \setminus \{X_j\})$, so

$$u(X_j \mid S \setminus \{X_j\}) = u(X_j \mid \text{Ch}(X_j)_G).$$

Finally, because S is child-closed we have $\text{Ch}(X_j)_S = \text{Ch}(X_j)_G$, and by construction $S \setminus \{X_j\} = \text{Ch}(X_j)_G \cup N_j$. Therefore

$$u(X_j \mid \text{Ch}(X_j)_G \cup N_j) = u(X_j \mid \text{Ch}(X_j)_G).$$

By the semigraphoid decomposition axiom, this implies that for every $n \subseteq N_j$,

$$u(X_j \mid \text{Ch}(X_j)_G \cup n) = u(X_j \mid \text{Ch}(X_j)_G),$$

which is exactly the local v-CMC for X_j .

□

F PROOF OF COMPLETENESS

Here is the theorem we wish to prove:

Theorem 9 (Soundness and Completeness). *Let $G = (V, E)$ be a causal DAG and let $X, Y, Z \subset V$ be disjoint.*

1. **Soundness:** *If X and Y are v -separated by Z in G , then $X \perp\!\!\!\perp_u Y \mid Z$ for all value functions u that obey the decomposition v-CMC on G .*
2. **Completeness:** *If X and Y are not v -separated by Z in G , then there exists a value function u that obeys the decomposition v-CMC on G such that $X \not\perp\!\!\!\perp_u Y \mid Z$.*

Proof. Since the main document contains the proof of soundness, here we just prove completeness.

(i) Completeness. Suppose X and Y are not v -separated by Z in G . Then X and Y are not d -separated by Z in the arrow-reversed graph G^{Rev} . By the completeness of d -separation (Verma and Pearl, 1988; Geiger et al., 1990a), there exists a probability distribution P that factorizes with respect to G^{Rev} such that

$$X \not\perp\!\!\!\perp_P Y \mid Z.$$

Furthermore, P can be chosen strictly positive: the subset yielding unfaithful distributions has Lebesgue measure zero (Meek, 1995), so a strictly positive faithful distribution exists and therefore violates $X \perp\!\!\!\perp Y \mid Z$.

Now define a value function u on (joint) realizations of subsets of V by

$$u(S) := \log P(S), \tag{29}$$

for every set of variables $S \subseteq V$ and every realization s of S (we suppress explicit realization notation), which is well-defined since P is strictly positive. With conditional value defined by subtraction (3), we then have, for all disjoint $S_1, S_2 \subseteq V$,

$$u(S_1 \mid S_2) = u(S_1 \cup S_2) - u(S_2) = \log P(S_1 \cup S_2) - \log P(S_2) = \log P(S_1 \mid S_2). \tag{30}$$

It follows immediately that conditional value independence with respect to u coincides with probabilistic conditional independence with respect to P : for all disjoint $S_1, S_2, S_3 \subseteq V$,

$$S_1 \perp\!\!\!\perp_u S_2 \mid S_3 \iff u(S_1 \mid S_2, S_3) = u(S_1 \mid S_3) \tag{31}$$

$$\iff \log P(S_1 \mid S_2, S_3) = \log P(S_1 \mid S_3) \iff S_1 \perp\!\!\!\perp_P S_2 \mid S_3, \tag{32}$$

where all equalities are understood pointwise for all realizations. In particular, since $X \not\perp\!\!\!\perp_P Y \mid Z$, we obtain

$$X \not\perp\!\!\!\perp_u Y \mid Z.$$

Finally, since P factorizes with respect to G^{Rev} , for every *parent-closed* vertex set $V' \subseteq V$ in G^{Rev} we have

$$P(V') = \prod_{w \in V'} P(w \mid \text{Pa}(w)_{G^{\text{Rev}}}). \quad (33)$$

Taking logs and using (29)–(30) yields

$$u(V') = \sum_{w \in V'} \log P(w \mid \text{Pa}(w)_{G^{\text{Rev}}}) = \sum_{w \in V'} u(w \mid \text{Pa}(w)_{G^{\text{Rev}}}). \quad (34)$$

Now note that $\text{Pa}(w)_{G^{\text{Rev}}} = \text{Ch}(w)_G$ for every node w , and that V' is parent-closed in G^{Rev} iff V' is child-closed in G . Hence (34) is exactly the decomposition v-CMC on G :

$$u(V') = \sum_{w \in V'} u(w \mid \text{Ch}(w)_G),$$

for every child-closed $V' \subseteq V$.

We have thus constructed a value function u that obeys the decomposition v-CMC on G and for which $X \not\perp_u Y \mid Z$. This completes the proof of Completeness. $\square$

G PROOF OF THE BELLMAN TYPE DECOMPOSITION

We start by restating the theorem:

Theorem 6 (Bellman-type decomposition). *Let $G = (V, E)$ be a causal DAG and let u be a value function over V satisfying the v-CMC. Let $P_i := V_i \setminus V_{i-1}$. Then*

$$u(V_i) = \sum_{W \in P_i} u(W \mid \text{Ch}(W)) + u(V_{i-1}). \quad (7)$$

Proof. Note that by the definition of conditional value (3), we can write:

$$u(V_i) = u(P_i \cup V_{i-1}) = u(P_i \mid V_{i-1}) + u(V_{i-1}) \quad (35)$$

Note that, by construction, V_{i-1} contains all the children of P_i , which allows us to write $V_{i-1} = \text{Ch}(P_i) \cup R$, where $\text{Ch}(P_i)$ is the set of all the children of nodes in P_i and R is the remainder of nodes in V_{i-1} that are not in $\text{Ch}(P_i)$. Note that, again by construction, R consists of non-ancestor nodes of P_i . $\text{Ch}(P_i)$ therefore v -separates P_i from R and the (global) v-CMC implies $u(P_i \mid \text{Ch}(P_i) \cup R) = u(P_i \mid \text{Ch}(P_i))$. Thus, we can write (35) as follows:

$$u(V_i) = u(P_i \mid \text{Ch}(P_i)) + u(V_{i-1}) \quad (36)$$

Next, we will show that the following decomposition holds for any set S that has the property that no node in S is an ancestor of any other node in S :

$$u(S \mid \text{Ch}(S)) = \sum_{W \in S} u(W \mid \text{Ch}(W)), \quad (37)$$

which we do by induction on the number of nodes in S . If S has one node, (37) holds vacuously. Now suppose (37) holds for all sets with fewer than n nodes and suppose S has n nodes. Pick one of the nodes $W \in S$ and write $S' := S \setminus \{W\}$. S' has $n - 1$ nodes and hence satisfies (37). So we can write:

$$u(S \mid \text{Ch}(S)) = u(\{W\} \cup S' \mid \text{Ch}(W) \cup \text{Ch}(S')) \quad (38)$$

$$= u(\{W\} \cup \text{Ch}(W) \cup S' \cup \text{Ch}(S')) - u(\text{Ch}(W) \cup \text{Ch}(S')) \quad (39)$$

$$= u(\{W\} \mid \text{Ch}(W) \cup S' \cup \text{Ch}(S')) + u(\text{Ch}(W) \cup S' \cup \text{Ch}(S')) - u(\text{Ch}(W) \cup \text{Ch}(S')) \quad (40)$$

Since W has no ancestors in S , the v-CMC implies that $u(\{W\} \mid \text{Ch}(W) \cup S' \cup \text{Ch}(S')) = u(\{W\} \mid \text{Ch}(W)) = u(W \mid \text{Ch}(W))$, so (38) becomes:

$$u(S \mid \text{Ch}(S)) = u(W \mid \text{Ch}(W)) \quad (41)$$

$$+ u(\text{Ch}(W) \cup S' \cup \text{Ch}(S')) - u(\text{Ch}(W) \cup \text{Ch}(S')) \quad (42)$$

$$= u(W \mid \text{Ch}(W)) + u(S' \mid \text{Ch}(S') \cup \text{Ch}(W)) \quad (43)$$

Again, by assumption W cannot be an ancestor of any node in S , so the v-CMC implies that $u(S' \mid \text{Ch}(S') \cup \text{Ch}(W)) = u(S' \mid \text{Ch}(S'))$. We can therefore write:

$$u(S \mid \text{Ch}(S)) = u(W \mid \text{Ch}(W)) + u(S' \mid \text{Ch}(S')) \quad (44)$$

$$u(W \mid \text{Ch}(W)) + \sum_{W \in S'} u(W \mid \text{Ch}(W)) \quad (45)$$

$$= \sum_{W \in S} u(W \mid \text{Ch}(W)), \quad (46)$$

which completes the inductive proof of (37). Now note that, by construction, P_i is a set such that no node in P_i is an ancestor of any other node in P_i , because if W is an ancestor of W' , then the height of W will be greater than the height of W' and every node in P_i has height i . Thus, combining (36) and (37) yields the desired Bellman decomposition:

$$u(V_i) = u(P_i \mid \text{Ch}(P_i)) + u(V_{i-1}) = \sum_{W \in P_i} u(W \mid \text{Ch}(W)) + u(V_{i-1}), \quad (47)$$

which completes the proof. □

H A WORKED EXAMPLE OF v -SEPARATION AND V-CMC DECOMPOSITION UNDER INTERVENTIONS

The purpose of this section is to illustrate how v -separation reduces the conditioning set required for local conditional-value assessments, and how the decomposition v-CMC permits utility functions to be updated modularly after interventions.

Consider the antibiotic DAG in Figure 1. Let $G^A := G_{\text{do}(A)}$ denote the intervened-upon graph in which the arrow $D \rightarrow A$ is removed, and let u^A be its associated utility function. For convenience, the graph is shown in Figure 3.

Suppose first that we are interested in assessing the conditional value of administering the antibiotic, $A = 1$, i.e., $u^A(A \mid S, E, D, I, R, T, W)$. In G^A , the variable D is disconnected from A . Moreover, conditioning on S and E blocks the common-effect paths $A \rightarrow S \leftarrow I$ and $A \rightarrow E \leftarrow R$, and blocks every path from A to T or W . Thus, A is v -separated from the remaining variables by its children:

$$A \perp\!\!\!\perp D, I, R, T, W \mid S, E. \quad (48)$$

Thus, by the global v-CMC,

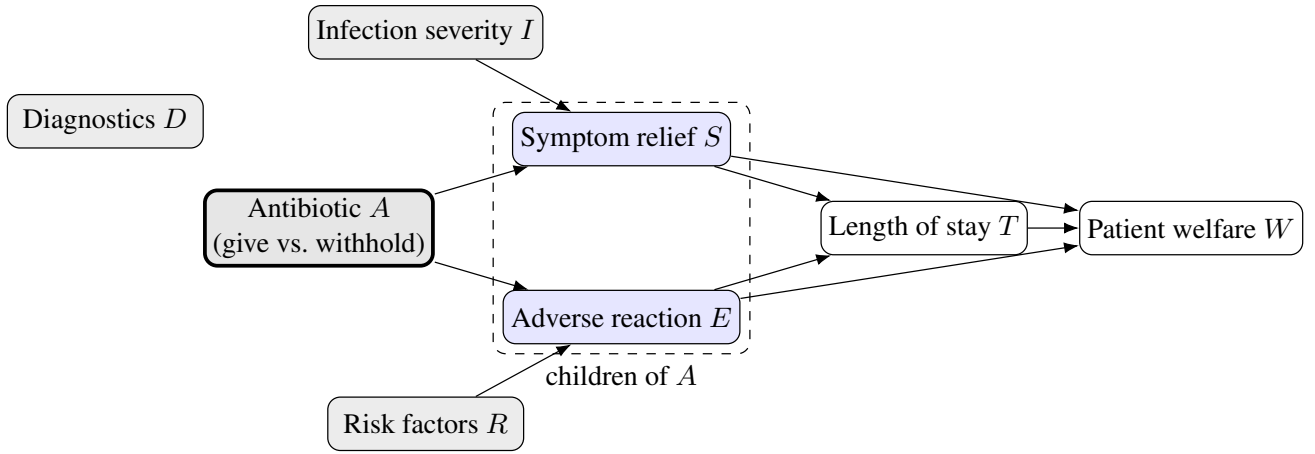


Figure 3: Post-intervention graph G^A : A affects S, E , which affect outcomes T, W ; D is disconnected and I, R are exogenous.

$$u^A(A \mid S, E, D, I, R, T, W) = u^A(A \mid S, E). \quad (49)$$

Hence, although the original graph contains eight variables, the local conditional-value assessment of administering the antibiotic requires only that we condition on the pair (S, E) . This illustrates how v -separation can simplify local conditional value assessments.

Suppose next that we want to evaluate the *total* utility $u^A(A, S, E, T, W)$ on the sub-graph over the variables $\{A, S, E, T, W\}$. Note that this set is child-closed in G^A , and hence the decomposition v-CMC allows us to simplify as follows:

$$\begin{aligned} u^A(A, S, E, T, W) &= u^A(W) + u^A(T \mid W) \\ &\quad + u^A(S \mid T, W) + u^A(E \mid T, W) \\ &\quad + u^A(A \mid S, E). \end{aligned} \quad (50)$$

For a numerical illustration, suppose that

$$\begin{aligned} u^A(W) &= 100W, & u^A(T \mid W) &= -10T, \\ u^A(S \mid T, W) &= 10S, & u^A(E \mid T, W) &= -30E. \end{aligned} \quad (51)$$

Thus, high patient welfare has utility 100 and low welfare has utility 0, a long stay has disutility 10 and a short one has a utility of 0, and so on. Suppose, furthermore, that the local conditional value contribution of administering the antibiotic has the following slightly more complex structure:

$$\begin{array}{c|cccc} (S, E) & (0, 0) & (1, 0) & (0, 1) & (1, 1) \\ \hline u^A(A = 1 \mid S, E) & -8 & 5 & -15 & -5. \end{array} \quad (52)$$

Thus, for example, the conditional utility of administering the antibiotic is 5 when the patient has symptom relief and no adverse reaction, whereas it is -15 when the patient has symptom relief and an adverse reaction.

Given these local utilities, we can calculate the total joint utility for any combination of values of $\{A, S, E, T, W\}$. For example, the total utility of administering the antibiotic when the patient has symptom relief, an adverse reaction, a short length of stay, and ultimately high welfare is:

$$\begin{aligned} u^A(A = 1, S = 1, E = 1, T = 0, W = 1) &= 100 + 0 + 10 - 30 - 5 \\ &= 75. \end{aligned} \quad (53)$$

Next, let us consider how the v -CMC can be used to assess how the total utility changes under a further intervention. Suppose, in particular, that we intervene to prevent adverse reactions, for example by administering a prophylactic. This removes the arrows into E and produces the new graph $G^{AE} := G_{\text{do}(A), \text{do}(E=0)}$, shown in Figure 4.

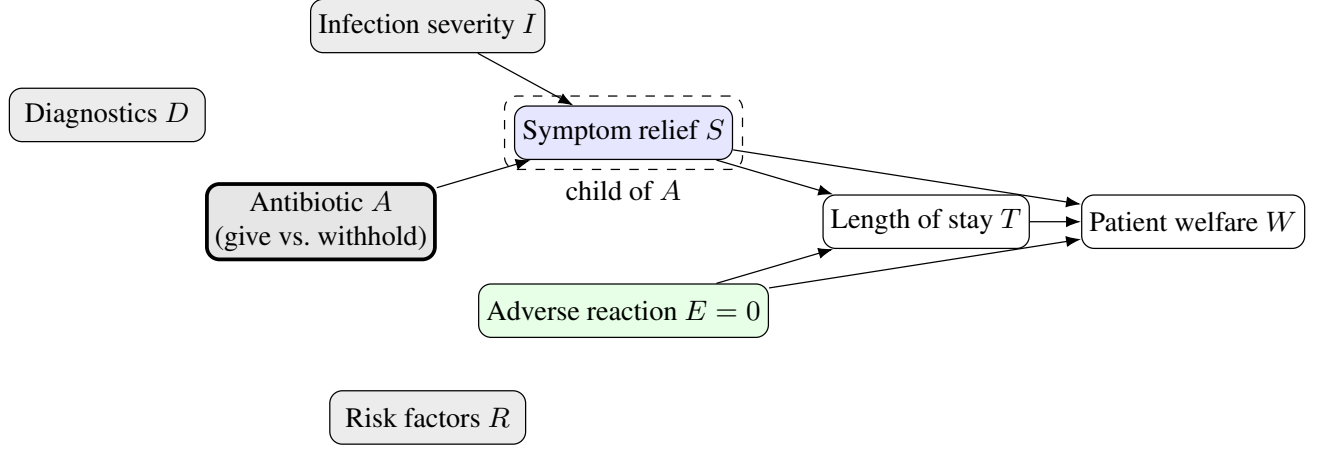


Figure 4: Post-intervention graph G^{AE} : A affects S , while E is fixed at 0 by intervention.

The intervention on E affects the child sets of both R and A . However, R is not part of the reduced child-closed set $\{A, S, E, T, W\}$ used above, and its local term is therefore not needed for the present analysis. Since we are concerned with the downstream utility of A , and E is fixed at 0 and is no longer part of the minimal child-closed set containing A , we may also omit E from the reduced representation. The v -CMC therefore gives:

$$u^{AE}(A, S, T, W) = u^A(W) + u^A(T | W) + u^A(S | T, W) + u^{AE}(A | S). \quad (54)$$

Here, the only local term whose form changes in the reduced representation after intervening on E is $u^{AE}(A | S)$; the remaining local terms in the reduced representation are retained unchanged from u^A . Suppose we have the following conditional values for $u^{AE}(A | S)$:

$$u^{AE}(A = 1 | S = 0) = -8, \quad u^{AE}(A = 1 | S = 1) = 2. \quad (55)$$

Then we can, for example, calculate the total joint utility of $(A = 1, S = 1, T = 0, W = 1)$ by using equation 54:

$$u^{AE}(A = 1, S = 1, T = 0, W = 1) = 100 + 0 + 10 + 2 = 112. \quad (56)$$

Of course, the numerical values are purely illustrative. The substantive point is that under $\text{do}(A)$, v -separation reduces the local assessment of A to $u^A(A | S, E)$; after $\text{do}(E = 0)$, it reduces further to $u^{AE}(A | S)$. The remaining local terms in the reduced representation can be transferred unchanged from u^A .