**m-Transportability: Transportability of a Causal Effect from Multiple Environments**

Sanghack Lee and Vasant Honavar  
Artificial Intelligence Research Laboratory  
Department of Computer Science  
Iowa State University  
Ames, IA 50011  
{shlee, honavar} at iastate.edu

**Abstract**

We study *m-transportability*, a generalization of transportability, which offers a license to use causal information elicited from experiments and observations in *m* ≥ 1 source environments to estimate a causal effect in a given target environment. We provide a novel characterization of *m*-transportability that directly exploits the completeness of do-calculus to obtain the necessary and sufficient conditions for *m*-transportability. We provide an algorithm for deciding *m*-transportability that determines whether a causal relation is *m*-transportable; and if it is, produces a transport formula, that is, a recipe for estimating the desired causal effect by combining experimental information from *m* source environments with observational information from the target environment.

**1 Introduction**

Eliciting causal effects from observations and experiments is a hallmark of intelligence. Pearl (1995; 2000) introduced causal diagrams, a formal representation for combining data with causal information, and do-calculus (Pearl 1995; 2000; 2012), a set of three rules that constitutes a sound (Pearl 1995) and complete (Shpitser and Pearl 2006b; Huang and Valtorta 2006) inferential machinery for causal inference. The resulting framework has been used to decide the identifiability of causal effects (Tian and Pearl 2002; Shpitser and Pearl 2006; Tian 2004), i.e., determine whether a given set of causal assumptions is sufficient for determining causal effects from observations.

The practical need to transfer causal effects elicited in one domain (setting, environment, population) e.g., a controlled laboratory setting, to a different setting presents us with the problem of transporting causal information from a source environment to a possibly different target environment. For example, one might want to know if causal relation between teaching strategies and student learning obtained through a randomized trial in a public school in Chicago can be transported to a public school in Minneapolis that has an admittedly different population of students. Pearl and Bareinboim (2011) introduced selection diagrams which provide a formal representation for expressing knowledge about differences and commonalities between the source and target environments. They used the selection diagrams to provide a formal definition of transportability of causal effects elicited from experimental studies in a source environment to a target environment in which only an observational study is possible. Bareinboim and Pearl (2012b) described procedures to (i) decide transportability between a given source and target environment, and (ii) when the answer is affirmative, determining the set of experimental and observational studies that must be carried out to license the desired transport. Against this background, we consider a setting where we have access to experiments and observations from *m* ≥ 1 source environments and a target environment; and causal information is not transportable between each individual source environment and the target environment. In such a setting, we examine the conditions under which it is possible to transfer causal information learned collectively from the *m* source environments to the target environment.

The main contributions of this paper are as follows: (i) We study *m*-transportability, a generalization of transportability (Pearl and Bareinboim 2011), which offers a license to transfer causal information learned collectively from experiments and observations from *m* source environments to a given target environment where only observational information can be obtained. (ii) We introduce a novel technique that directly exploits the completeness of do-calculus (Shpitser and Pearl 2006b; Huang and Valtorta 2006) to obtain the necessary and sufficient conditions for *m*-transportability. Our results directly apply to transportability (*m*-transportability where *m* = 1) as well as identifiability which is a special case of transportability. (iii) We provide, based on a modification from Bareinboim and Pearl’s algorithm for deciding transportability (Bareinboim and Pearl 2012b), a sound and complete algorithm for deciding *m*-transportability that determines whether the causal effect is *m*-transportable; and if it is, produces a transport formula, that is, a recipe for estimating the causal effect by combining the experimental information from the *m* source environments with observational information from the target environment.

This work was carried out independently of Bareinboim and Pearl (2013) which introduces μ-transportability, which, despite some differences in the formulation, upon...
closer examination, turns out to be essentially identical to \( m \)-transportability. Not surprisingly, the algorithm for deciding \( \mu \)-transportability is essentially identical to our algorithm for deciding \( m \)-transportability: Both algorithms are simple extensions of \textsc{sd}, the algorithm for deciding transportability (a special case of \( m \)-transportability where \( n=1 \)) introduced by Bareinboim and Pearl (2012b); and both algorithms rely on the same graphical criterion to decide whether a causal relation is \( \mu \)-transportable (\( m \)-transportable). However, the work described in this paper differs from that of Bareinboim and Pearl (2013) with respect to the technique used to derive the graphical criterion for lack of \( m \)-transportability (\( \mu \)-transportability). Bareinboim and Pearl (2013) construct a certificate that serves as counter-example that establishes lack of \( \mu \)-transportability. In contrast, we demonstrate the nonexistence of a transport formula (and hence lack of \( m \)-transportability) by directly exploiting the completeness of \textsc{do-calculus} (Shpitser and Pearl 2006b; Huang and Valtorta 2006) and standard probability manipulations.

2 Preliminaries

A causal diagram (Pearl 2000) \( G \) is a semi-Markovian graph or an acyclic directed mixed graph, (i.e., a graph with directed as well as bidirected edges that does not have directed cycles) which encodes a set of causal assumptions. We denote by \( G[Y] \), a subgraph of \( G \) containing nodes in \( Y \) and all arrows between the corresponding nodes in \( G \). Following (Pearl 2000), we denote by \( G_X \), the edge subgraph of \( G \) where all incoming arrows into nodes in \( X \) are deleted; by \( G_Y \), the edge subgraph of \( G \) where all outgoing arrows from nodes in \( Y \) are deleted; and by \( G_{XY} \), the edge subgraph of \( G \) where all incoming arrows into nodes in \( X \) and all outgoing arrows from nodes in \( Y \) are deleted.

A causal model (Pearl 2000) is a tuple \( (U, V, F) \) where \( U \) is a set of background or hidden variables that cannot be observed or experimented on but which can influence the rest of the model; \( V \) is a set of observed variables \( \{V_1, \ldots, V_n\} \) that are determined by variables in the model, i.e., variables in \( U \cup V \); \( F \) is a set of deterministic functions \( \{f_1, \ldots, f_n\} \) where each \( f_i \) specifies the value of the observed variable \( V_i \) given the values of observable parents of \( V_i \) and the values of hidden causes of \( V_i \). A probabilistic causal model (Pearl 2000) \( \text{PCM} \) is a tuple \( M = \langle U, V, F, P(U) \rangle \) where \( P(U) \) is a joint distribution over \( U \).

Intervention (Pearl 2000) on a set of variables \( X \subseteq V \) of \( \text{PCM} M = \langle U, V, F, P(U) \rangle \) involves setting to \( X = \bot \) and is denoted by \textsc{do}-operation \( \text{do}(X=x) \) or simply \( \text{do}(x) \). A causal effect of \( X \) on a disjoint set of variables \( Y \subseteq V \) is written as \( P(y|do(x)) \) or simply \( P_x(y) \). Intervention on a set of variables \( X \subseteq V \) creates a submodel (Pearl 2000) \( M_X \) of \( M \) defined as follows: \( M_X = \langle U, V, F_X, P(U) \rangle \) where \( F_X \) is obtained by taking a set of distinct copies of functions in \( F \) and replacing the functions that determine the value of variables in \( X \) by constant functions setting the variables to values \( x \). It is easy to see that a causal diagram \( G \) that encodes the causal assumptions of model \( M \) is modified to \( G_X \) by intervention on \( X \). The causal effect of \( X \) on a disjoint set of variables \( Y \subseteq V \) is said to be identifiable from \( P \) in \( G \) if \( P_x(y) \) can be computed uniquely from the joint distribution \( P(V) \) of the observed variables in any PCM that induces \( G \) (Pearl 2000).

Do-calculus (Pearl 1995) offers a sound and complete (Shpitser and Pearl 2006b; Huang and Valtorta 2006) inferential machinery for deciding identifiability (Tian and Pearl 2002; Shpitser and Pearl 2006a; Tian 2004) in the sense that, if a causal effect is identifiable, there exists a sequence of applications of the rules of do-calculus that transforms the causal effect formula into one that includes only observational quantities. Let \( G \) be a causal diagram and \( P \) be a distribution on \( G \). Let \( W, X, Y, Z \) be disjoint sets of variables in \( G \). Then, the three rules of do-calculus are (Pearl 1995):

- (Rule 1) Insertion/deletion of observations: \( P_x(y | z, w) = P_x(y | w) \) if \( Y \perp Z | X, W \) in \( G_X \).
- (Rule 2) Intervention/observation exchange: \( P_{x,z}(y | w) = P_x(y | z, w) \) if \( Y \perp Z | X, W \) in \( G_X \).
- (Rule 3) Insertion/deletion of interventions: \( P_{x,z}(y | w) = P_x(y | w) \) if \( Y \perp Z | X, W \) in \( G_X \).

where \( Z(W) \) represents \( Z \setminus \text{An}(W) \) in \( G_X \).

3 \( m \)-Transportability

We proceed to formalize the notion of \( m \)-transportability, which offers a license to transfer causal information learned collectively from experiments and observations from \( m \) source environments to a target environment where only observational information can be obtained. The definitions of selection diagrams and \( m \)-transportability and the \( m \)-transportability theorem are direct adaptations of their counterparts introduced by Bareinboim and Pearl (2012b) for transportability (1-transportability). Let \( J \) denote an index set \( \{1, \ldots, m\} \) for \( m \) source environments:

**Definition 1 (Selection Diagram (adapted from Bareinboim and Pearl 2012b)).** Let \( M = \{M_j\}_{j \in J} \) and \( M^\ast \) be a set of causal models relative to domains \( \Pi = \{\Pi_j\}_{j \in J} \) and \( \Pi^\ast \), sharing a causal diagram \( G \). The set \( M \cup \{M^\ast\} \) is said to induce a selection diagram \( D \) if \( D \) is constructed as follows:

1. Every edge in \( G \) is also an edge in \( D \).
2. \( D \) contains an extra edge \( S_i \rightarrow V_i \) if there might exist a discrepancy \( f_i \neq f_i^* \) or \( P^\ast(U^i) \neq P(U^i) \) between \( M^\ast \) and some model \( M_j \in M \).

We denote by \( S \), the set of selection variables. We associate with each domain \( \Pi^i \), a set of domain-specific selection variables \( S_i \subseteq S \) where: \( S_i = \{S_i | f_i \neq f_i^* \} \)

*Given a DAG \( G \), \( \text{an}(W)_G \) denotes the set of ancestors of a node \( W \). Moreover, \( \text{An}(W)_G = \text{an}(W)_G \cup \{W\} \), \( \text{an}(W)_G = \bigcup_{W \in W} \text{an}(W)_G \), and \( \text{An}(W)_G = \bigcup_{W \in W} \text{An}(W)_G \).
two unmeasured variables, gender and ethnicity.

terms of their gender and ethnicity distributions respectively.

The following example illustrates the power of m-transportability.

**Definition 2 (Causal Effects m-Transportability (adapted from (Bareinboim and Pearl 2012b)))**. Let D be a selection diagram and S a collection of sets of domain-specific selection variables relative to a set of domains Π∗ ∪ Π. Let I be the information set \( I(Π^\ast) \) \( \cup \) \( I(Π^\ast) \) where \( I(Π^\ast) \) = \( \{ P^\ast(\mathbf{V} \setminus \mathbf{V}^\prime \mid \mathbf{do} (\mathbf{v}^\prime) \} \setminus \mathbf{V}^\prime \), a set of interventional distributions from Π∗ for every \( j \) ∈ J, and \( I(Π^\ast) \) = \( \{ P^\ast(\mathbf{V}) \} \), an observational distribution from the target domain Π∗. The causal effect \( R = P^\ast_x(\mathbf{y}) \) is said to be m-transportable from Π to Π∗ in D if \( P^\ast_x(\mathbf{y}) \) is uniquely computable from I in any model that induces D and S.

The following example illustrates the power of m-transportability.

**The Power of m-Transportability** Consider the selection diagram (adapted from Figure 2(c) in (Pearl and Bareinboim 2011)) shown in Figure 1 which includes three observed variables corresponding to treatment \( D \), disease \( D \), and a biological marker \( B \); two confounders, gender (between \( T \) and \( B \)) and ethnicity (between \( T \) and \( D \)). Suppose we are interested in estimating a causal effect of treatment on disease \( P^\ast_x(d) \) in a population Π∗ given observational study from Π∗ and experimental studies on two different populations Π1 and Π2. Suppose Π∗ differs from Π1 and Π2 in terms of their gender and ethnicity distributions respectively. These differences are denoted by a set of selection variables \( S = \{ S_T, S_B, S_D \} \) and domain-specific selection variables \( S^1 = \{ S_T, S_B \} \) and \( S^2 = \{ S_T, S_D \} \). It follows from Theorems 4 and 5 of (Bareinboim and Pearl 2012b) that the causal effect \( P^\ast_x(d) \) in the population Π∗ is not transportable from Π1 or Π2 alone. However, it is easy to see that \( P^\ast_x(d) \) can be computed by combining the observational and experimental data from Π1 and Π2 and applying the standard rules of probability and rule 1 of do-calculus:

\[
P^\ast_x(d) = \sum_b P^\ast(d \mid b) P^\ast(b)
\]

\[
= \sum_b P^\ast(d \mid b, s_D = s) P^\ast(b \mid s_B = s)
\]

\[
= \sum_b P^\ast(d \mid b) P^\ast(b).
\]

The necessary and sufficient conditions for m-transportability follow immediately from the completeness (Shpitser and Pearl 2006b; Huang and Valtorta 2006; Pearl and Bareinboim 2011) of do-calculus.

**Theorem 1 (m-Transportability Theorem (adapted from Theorem 1 in (Pearl and Bareinboim 2011)))**. Let D be the selection diagram characterizing Π and Π∗, and S be a collection of sets of domain-specific selection variables in D. The effect \( P^\ast_x(\mathbf{y} \mid \mathbf{z}) \) is m-transportable from Π to Π∗ if and only if the expression \( P^\ast_x(\mathbf{y} \mid \mathbf{z}, s) \) is reducible, using the rules of do-calculus, to an expression where terms are either do-free or \( S^\ast \)-free (selection variables that appear as conditioning variable in do-terms are a subset of \( S \setminus S^\ast \) for some \( \Pi \in \Pi^\ast \)).

**Proof.** The proof follows immediately from the completeness of do-calculus (Shpitser and Pearl 2006b; Huang and Valtorta 2006) and the do-calculus characterization lemma (Lemma 1 in (Pearl and Bareinboim 2011)) with the one minor modification with respect to when terms with do-operators are estimable from Π; Since Π1 and Π2 differ with respect to selection variables \( S^\ast \), every do-term that does not contain \( S^\ast \) is estimable from Π1.

By applying rule 2 of do-calculus and marginalization (Shpitser and Pearl 2006a), we can reduce estimation of a conditional causal effect to the estimation of unconditional causal effect. Hence, without loss of generality, we focus on m-transportability of unconditional causal effects.

**Corollary 1.** If no finite sequence of rewritings of a causal effect using standard rules of probability and the rules of do-calculus yields a transport formula, then the effect is not m-transportable from source domains to the target domain.

### 4 Characterizing m-Transportability

We proceed to characterize m-transportability in terms of properties of selection diagrams. We first extend the definitions of \( s^\ast \)-tree and s-hedge (Bareinboim and Pearl 2012b) to the setting with m source domains instead of a single source domain. We say that a selection diagram is an ms*-tree if it is an s*-tree for every source domain. We define an ms-hedge by replacing \( s^\ast \)-trees by ms*-trees in the definition of an s-hedge (Bareinboim and Pearl 2012b).

**4.1 Analysis of Hedges**

A hedge (Shpitser and Pearl 2006b) is an essential structure which hinders identification of a causal effect. Hedge structures were used to generate counter-examples (Shpitser and Pearl 2006b) to establish non-identifiability of causal effects in specific settings. Unlike (Shpitser and Pearl 2006b; Bareinboim and Pearl 2012b), we exploit the completeness of do-calculus to show the non-m-transportability of a causal
effect. Here, we examine hedge, s-hedge, and ms-hedge more closely to reveal some formulae that are not uniquely computable on them.

**Definition 3 (Bare-Hedge).** Let $D$ be a selection diagram and $X$, $Y$ be nonempty disjoint sets of $V$. Then a triple $(D, X, Y)$ is called a bare-hedge if $D$ and $D \setminus X$ are $Y$-rooted $m^*$-trees forming an ms-hedge for $P_{x}(y)$ in $D$.

It follows from Corollary 1 that if $P_{x}(y)$ is not m-transportable, then any sequence of rewritings of $P_{x}(y | s)$ must yield a formula with at least one do-term which is not $S^j$-free. This implies that non-m-transportable causal effects must contain a term that cannot be eliminated to yield a transport formula. When such a term is rewritten using other terms, at least one of them must contain such a term as well. We say that the probability terms that satisfy this condition are of the definite form (defined below):

**Definition 4 (Definite Form).** Given a bare-hedge $(D, X, Y)$, a probability term is said to be of the definite form if and only if it is of the form $P_{x}(v' | v'', s^j)$ where $V'$, $V'' \subset V$ and the following hold: (i) $X \subseteq X$ and $X \not= \emptyset$, (ii) $V' \cup V'' \supseteq Y$, (iii) $V' \setminus X \not= \emptyset$, and (iv) $S_{v',X} \subseteq S'' \subseteq S$. We will use $\Psi^{D}_{X,Y}$ to denote the set of all probability terms of the definite form and $\Psi^{D}_{X,Y}$ to denote its complement.

A set of definitions on a selection diagram is of the form $P_{v_1} (v_2 | v_3, s')$ where $V_1$, $V_2$, and $V_3$ are disjoint subsets of observed variables, $V_2$ is nonempty, and a subset $S'$ of selection variables satisfies $S' = S$ or $(S \setminus S') \subseteq V_3 | V_1$. We choose $s' = s$ or $s' \subseteq S' = j$ to select a population from among $\Pi^*$ and $\Pi^j$.

**Hedges and Do-calculus** Because hedges play an important role in the identifiability of causal effects, it is instructive to examine the interplay between a hedge and the rules of do-calculus. We will show a closure property of $\Psi^{D}_{X,Y}$ with respect to do-calculus. Towards this end, we define mutually exclusive ancestors of a DAG $G$ with respect to a set of nodes.

**Definition 5 (Mutually Exclusive Ancestors).** The mutually exclusive ancestors of a DAG $G$ with respect to a subset $W$ of its nodes is defined as

$$\Upsilon^G(W) = \{An(W) \setminus An(W \cap \text{\emph{do}}(W)G) \}_{W \in W}$$

We use $\mathcal{F}_{W}$ to denote a family of sets $\Upsilon^G(W)$ indexed by $W$. Given $W' \subseteq W$, we say that $\mathcal{F}_{W'}$ is a subfamily of $\mathcal{F}_{W}$. Given $W \in W$, we use $\mathcal{F}_{W}$ to denote the corresponding element of the family $\mathcal{F}_{W}$. Followings are its properties which are tightly related to the d-separation criterion (Pearl 1988).

**Proposition 1.** Let $\mathcal{F}$ be a family $\Upsilon^G(A) (B)$ given a bare-hedge $(G, X, Y)$ where $B \subseteq A$ and $A \subseteq V$. Let $A \in A$ and $B \in B$. Following properties hold:

(a) The union of the family is $An(B)_{G[A]}$;
(b) The only common descendant of all elements in $\mathcal{F}_B$ is $B$.
(c) The family is pairwise disjoint.

(d) Each vertex in $\mathcal{F}_{B}$ is $d$-connected to $B$ in $G [\mathcal{F}_B]$.

(e) Each pair of vertices in $\mathcal{F}_B$ is $d$-connected in $G [\mathcal{F}_B]$ conditioned on $B$.

Suppose, further that $C (A) = \{G[A]\}$, $\text{An}(B)_{G[A]} = A$, and $B$ is the union of two disjoint nonempty sets $C$ and $D$. Then we have:

(f) Each set in $\mathcal{F}_D$ is $d$-connected to some set in $\mathcal{F}_C$ through a bidirected path in $G [\mathcal{F}_A]$ conditioned on $D$.

(g) Let $C$ be the union of two disjoint nonempty sets $C'$ and $C''$. Then $\mathcal{F}_{C'}$ and $\mathcal{F}_{C''}$ are $d$-connected via bidirected edges in $G [\mathcal{F}_A]$ conditioned on $D$.

**Proof.** Omitted.

We denote the syntactic symmetric difference between two probability terms $R_1 = P_{w_1}(v_2 | v_3)$ and $R_2 = P_{w_2}(w_2 | w_3)$ by $R_1 \ominus R_2$ which is given by $\bigcup_{V_1 \cup W_1}$. If two probability terms are equal according to do-calculus, then $V_2 = w_2$.

**Lemma 1.** Let $(D, X, Y)$ be a bare-hedge, $\psi$ be a probability term in $\Psi^{D}_{X,Y}$, and $\psi'$ be a probability term defined on $D$. If $\psi \ominus \psi'$ is a nonempty subset of $S_{v',X} \cup V \setminus X$, then $\psi$ is $d$-connected to $\psi'$ in $G [\mathcal{F}_A]$.

**Proof.** Since $\psi$ does not contain interventions on $S_{v',X} \cup V \setminus X$, we do not consider deletion or change of interventions. In the case of the selection variables, since they appear only as observations and $S_{v',X} \subseteq S''$, it suffices to consider only removal of selection variables. Suppose that $P_{x}(v' | v'', s'')$ and let $W$ be $\psi \ominus \psi'$. In the following, we prove conditional dependencies related to rules of do-calculus by showing the existence of a $d$-connection path in a subgraph of $D$.

(a) Since $(V' \subseteq W | V'', X, S'')_{\mathcal{F}_W}$ insertion of $W$ as observations on $\psi$ is impossible. Let $\mathcal{F}$ be $\Upsilon^{D[V \setminus X]}(V' \cup V'' | X)$. From Proposition 1(a), it follows that either $W \in \mathcal{F}_{V' \setminus X}$ or $W \in \mathcal{F}_{V'' \setminus X}$. When $W \in \mathcal{F}_{V' \setminus X}$, there exists a variable $V'$ such that $W \in \mathcal{F}_{V'}, \in \mathcal{F}_{V' \setminus X}$. From Proposition 1(d), it follows that there must be a $d$-connection path between $W$ and $V'$. When $W \in \mathcal{F}_{V'' \setminus X}$, there exists a variable $V''$ such that $W \in \mathcal{F}_{V''}, \in \mathcal{F}_{V'' \setminus X}$. Because $\Upsilon(D [V \setminus X]) = \{D [V \setminus X] \} \text{An}(V' \cup V'' \setminus X)_{D[V \setminus X]} = V \setminus X$ from Proposition 1(f), $A \in \mathcal{F}_{V'}$, and $B \in \mathcal{F}_{V''}$. From Proposition 1(e) and Proposition 1(d) respectively imply the existence of $d$-connection paths between $A$ and $V'$ and between $B$ and $V''$. Hence, $W$ and $V'$ must be $d$-connected in $D[V \setminus X]$.

(b) Since $(V' \subseteq W | V'', X, S'')_{\mathcal{F}_W}$ deletion of $W$ in observation of $\psi$ is impossible. Let $W_1 = W \setminus V''$ and $W_2 = W \setminus S''$. When $W_1 \not= \emptyset$.

\[^4\]The symmetric difference between two sets $C$ and $D$ is denoted by $C \ominus D$ which is given by $(C \cup D) \setminus (C \cap D)$. 586
The preceding results lead to the following lemma:

**Lemma 4.** Given a bare-hedge \(\langle D, X, Y \rangle\), a set \(\Psi_{D, X, Y}^D\) is closed under do-calculus.

**Proof.** By Lemma 1–3, if a probability \(\psi\) in \(\Psi_{D, X, Y}^D\) is equal to a probability \(\psi'\) by do-calculus, then the syntactic symmetric difference between \(\psi\) and \(\psi'\) is a subset of \(X \cup S_X\) and \(\psi'\) is also in \(\Psi_{D, X, Y}^D\).

This directly leads to the following corollary.

**Corollary 2.** Given a bare-hedge \(\langle D, X, Y \rangle\), the complement of \(\Psi_{D, X, Y}^D\) is closed under do-calculus.

**Hedges and Standard Probability Manipulations** We examine the behavior of definite form probability associated with a bare-hedge under the standard probability manipulations.

**Lemma 5.** Given a bare-hedge \(\langle D, X, Y \rangle\), if a formula is derived from a probability term in \(\Psi_{D, X, Y}^D\) by standard probability manipulations, then the formula contains at least one probability term in \(\Psi_{D, X, Y}^D\).

**Proof.** Let \(\psi \in \Psi_{D, X, Y}^D\) be expressed as \(P_X(v', v'', s'')\). By the definition of conditional probability, \(\psi = P_X(v', v'', s'') / P_X(v'', s'')\) if \(P_X(v'' | s'') \neq 0\) and the dividend is of the definite form. From the definition of marginal probability (or law of total probability), we have:

\[ \psi = \sum_w P_X(v', w | v'', s''). \]
The sum and product rules are sufficient for probabilistic inference (Jaynes and Bretthorst 2003, pg. 35). From the sum rule, we have: \( \psi_Y = 1 - \sum_{w \in \mathcal{W}_Y} P_X(w | v', s') \) where \( \mathcal{W}_Y \) is the complement of event \( v' \). It is easy to see that this does not yield a transport formula for \( \psi_Y \). From the product rule, we have: \( \psi = P_X(v' \setminus w | v', s') P_X(w | v', s') \) where \( \mathcal{W} \subset \mathcal{W}' \). If \( \mathcal{W}' \setminus \mathcal{W} \neq \emptyset \), then the first term of \( \psi \) is in \( \Psi_{S, Y}^D \). Otherwise, the second term of \( \psi \) is in \( \Psi_{Y, Y}^D \). We can show similar results for other probability manipulations e.g., Bayes’ theorem.

**Hedges and Transport Formula**

**Lemma 6.** Given a bare-hedge \( \langle D, X, Y \rangle \), none of probability terms in \( \Psi_{S, Y}^D \) can be expressed in terms of probability terms only in \( \mathcal{W}_Y \).

**Proof.** This follows from the closure property of \( \Psi_{S, Y}^D \) under do-calculus (Lemma 4) and the behavior of probability terms in \( \Psi_{S, Y}^D \) under standard probability manipulations (Lemma 5).

The preceding results lead to the following theorem which establishes a crucial connection between bare hedge structures and (lack of) m-transportability.

**Theorem 2.** Let \( D, S \), and \( I \) be as defined in the definition of causal effects m-transportability. Let \( \langle D, X, Y \rangle \) be a bare-hedge. Then, none of causal effect of the target domain \( v \) and Pearl 2012b). The line 6 of \( \Psi_{S, Y}^D \) to account for the use of experimental studies from \( \mathcal{W}_Y \). From the sum and product rules are sufficient for probabilistic information by the completeness of \( \Psi_{S, Y}^D \).

**Algorithm for m-Transportability**

We present below, msID, a sound and complete algorithm for deciding m-transportability that generalizes sID, an algorithm for deciding transportability (or 1-transportability) (Bareinboim and Pearl 2012b), which is a modification of the causal effects identifiability algorithm (Shpitser and Pearl 2006b). Instead of a single source domain as in the case of sID, the arguments of msID include \( m \) source domains; A second key difference between sID and msID has to do with the substitution of line “if (S ⊥ Y | X) \[ D \] , return \( P_X(y) \)” (line 6 of sID) with “if \( \exists_j \langle S \setminus Y | X \rangle \[ D \] , return \( P_X(y) \)” (line 6 in Figure 2) to account for the use of experimental studies from \( m \) source domains instead of a single source domain.

---

6 As noted earlier, msID is, modulo some notational differences, essentially equivalent to \( \mu \text{SID} \) (Bareinboim and Pearl 2013). We include msID here to make the paper self-contained.

**Function msID (\( y, x, P, D \))**

1. if \( x = \emptyset \), return \( P(y) \)
2. if \( V \setminus \mathcal{A}_0(Y) \neq \emptyset \), return msID(\( y, x \cap \mathcal{A}_0(Y) \), \( P(\mathcal{A}_0(Y)) \), \( D(\mathcal{A}_0(Y)) \))
3. if \( W \not\leftarrow V \setminus \mathcal{A}_0(Y) \), msID(\( y, \cup w, P, D \)
4. if \( C(D \setminus X) = \{ C_1, \ldots, C_k \} \), return \( \sum_{s \in \mathcal{C}_s} \mathcal{ID}(c, v, c_i, P, D) \)
5. if \( C(D \setminus X) = \{ C \} \)
6. if \( \exists_j \langle S \setminus Y | X \rangle \[ D \] , return \( P_x(y) \)
7. if \( C(D) = \{ D \} \), throw FAIL(\( D, C \))
8. if \( C \in \mathcal{C}(D) \), return \( \sum_{v \in C} P(v) \big| v_{D}^{-1} \big) \)
9. if \( \{ C' \} C \subset C \in \mathcal{C}(D) \), return msID(\( y, x \cap C' \), \( P(V_{D} \big| V_{D}^{-1} \cap C', v_{D}^{-1} \big) \), \( C' \))

Figure 2: An m-transportability algorithm msID to identify \( P_x^*(y) \) by calling msID(\( y, x, P^*, D \)). A set of s-components in a graph \( D \) is denoted by \( C(D) \) (Bareinboim and Pearl 2012b). We assume that \( I \) and \( S \) are globally defined.

**Theorem 3 (Soundness).** Whenever msID(\( y, x, P^*, D \)) outputs an expression for \( P_x^*(y) \), the expression correctly estimates \( P_x^*(y) \).

**Proof.** The correctness of steps except the line 6 of msID are shown in (Shpitser and Pearl 2006b; Bareinboim and Pearl 2012b). The line 6 of msID parallelizes the line 6 of sID. Hence, the correctness follows from sID which is due to S-admissibility of \( X \) along with Corollary 1 in (Pearl and Bareinboim 2011).

We proceed to show that whenever msID returns FAIL with an ms-hedge, the causal effect in question is indeed not m-transportable.

**Lemma 7 (adapted from Theorem 5 in (Bareinboim and Pearl 2012b)).** Let \( D \) be a selection diagram induced from \( \Pi \) and \( \Pi^* \). Assume that there exists \( F, F' \) that form an ms-hedge for \( P_x(y) \) in \( \Pi \) and \( \Pi^* \). Then \( P_x(y) \) is not m-transportable from \( \Pi \) to \( \Pi^* \).

**Proof.** The assumptions of the lemma, together with the definitions of bare-hedge and ms-hedge (Section 4) imply that there exists a bare-hedge \( \langle D', X, R \rangle \) where \( R \subset \mathcal{A}_0(Y) \) and \( D' \) is a subgraph of \( D \). Then, from Lemma 6, \( P_x^*(y) \) is not m-transportable from \( \Pi \) to \( \Pi^* \). From Lemma 14 in (Bareinboim and Pearl 2012b) it follows that \( P_x^*(y) \) is not m-transportable from \( \Pi \) to \( \Pi^* \).

**Lemma 8 (adapted from Theorem 7 in (Bareinboim and Pearl 2012b)).** Whenever msID(\( y, x, P^*, D \)) returns FAIL (i.e., fails to transport \( P_x^*(y) \) from \( \Pi \) to \( \Pi^* \)), there must exist \( X' \not\subseteq X, Y' \subseteq Y \) such that the graph pair \( D, C \) returned by the FAIL condition of msID contains as edge subgraphs ms*-trees \( F, F' \) that form an ms-hedge for \( P_x^*(y') \).
Proof. Since msID generalizes sID, the proof follows from arguments along the lines of those used to prove Theorem 7 in (Bareinboim and Pearl 2012b).

Theorem 4 (Completeness). Algorithm msID is complete.

Proof. This follows from Lemma 7 and 8 and the completeness of sID (Bareinboim and Pearl 2012b).

It is easy to see that the msID terminates; and that its runtime complexity is polynomial in the number of variables. The number of recursion is $O(|V|^2)$ since (i) each recursion in line 3 (adding interventions on $X$) and 4 (factorization) is called at most once; (ii) line 2 and 9 only reduce the size of $X$; and (iii) line 1, 6, 7, and 8 are terminal. In addition, graphical tests (e.g., tests for conditional independence) can be carried out in time that is polynomial in the number of observed variables.

6 Conclusion

We introduced $m$-transportability which offers a license to transfer causal information learned collectively from experiments and observations from $m$ source environments to a given target environment where only observational information can be obtained. Bareinboim and Pearl (2013) independently introduced $\mu$-transportability, which despite some differences in the formulation, turns out to be essentially identical to $m$-transportability. We established the necessary and sufficient conditions for $m$-transportability by directly exploiting the completeness of do-calculus (Shpitser and Pearl 2006b; Huang and Valtorta 2006). We introduced msID, a sound and complete algorithm for deciding $m$-transportability that outputs (a) evidence of non-$m$-transportability if a causal effect is not $m$-transportable from a given set of $m$ source environments to a specified target environment; and (b) a transport formula, that is, a recipe for estimating the desired causal effect by combining experimental information from $m$ source environments with only observational information from the target environment, otherwise. We note that msID is essentially identical to $\mu$sID, the algorithm for deciding $\mu$-transportability that was independently introduced by Bareinboim and Pearl (2013). This is not surprising in light of the fact that $m$-transportability and $\mu$-transportability are equivalent notions; Both msID and $\mu$sID are simple extensions of sID, the algorithm for deciding transportability (a special case of $m$-transportability where $m=1$) introduced by Bareinboim and Pearl (2012b); Both msID and $\mu$sID rely on the same graphical criterion to decide whether a causal relation is $m$-transportable ($\mu$-transportable). However, the work presented in this paper differs from that of Bareinboim and Pearl (2013) with respect to the technique used to derive the graphical criterion for lack of $m$-transportability ($\mu$-transportability): Bareinboim and Pearl (2013) construct a certificate that serves as counter-example that establishes lack of $\mu$-transportability. In contrast, we demonstrate the nonexistence of a transport formula (and hence lack of $m$-transportability) by directly exploiting the completeness of do-calculus (Shpitser and Pearl 2006b; Huang and Valtorta 2006) and standard probability manipulations.

Causal effects identifiability (Galles and Pearl 1995; Tian 2004; Tian and Pearl 2002; Shpitser and Pearl 2006a; 2006b), transportability (Pearl and Bareinboim 2011; Bareinboim and Pearl 2012b), $z$-identifiability (Bareinboim and Pearl 2012a) (the problem of estimating in a given domain, the causal effect of $X$ on $Y$ from surrogate experiments on $Z$ that are more amenable to experimental manipulation in the domain than $X$) are all special cases of meta-identifiability (Pearl 2012) which has to do with non-parametric identification of causal effects given multiple domains and arbitrary information from each domain. The results presented in this paper expand the subclass of meta-identifiability problems with provably correct and complete algorithmic solutions. There are several additional special cases of meta-identifiability to consider, including in particular, generalizations of $z$-identifiability and transportability to allow causal information from experiments, including surrogate experiments, in multiple source domains to be combined to facilitate the estimation of a causal effect in a target domain.

Acknowledgments

The work on $m$-transportability described in this paper was inspired by the seminal work of Judea Pearl on formal approaches to causal inference. It builds on the key results published by Judea Pearl and his coauthors on causal effects identifiability (Galles and Pearl 1995; Tian 2004; Tian and Pearl 2002; Shpitser and Pearl 2006a; 2006b) and transportability (Pearl and Bareinboim 2011; Bareinboim and Pearl 2012b). The authors are grateful to AAAI 2013 anonymous reviewers for their thorough reviews and AAAI 2013 Program Committee Chairs for giving them the opportunity to publish their results on $m$-transportability and clarify their relationship to those obtained independently by Bareinboim and Pearl (2013). The work of Vasant Honavar while working at the National Science Foundation was supported by the National Science Foundation. Any opinion, finding, and conclusions contained in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References


