In Proceedings of the 27th AAAI Conference on Artificial Intelligence, pp. 95-101, 2013.

TECHNICAL REPORT
R-408
April 2013

Causal Transportability with Limited Experiments

Elias Bareinboim and Judea Pearl
Cognitive Systems Laboratory
Department of Computer Science
University of California, Los Angeles
Los Angeles, CA. 90095
{eb,judea} at cs.ucla.edu

Abstract

We address the problem of transferring causal knowledge learned in one environment to another, potentially different environment, when only limited experiments may be conducted at the source. This generalizes the treatment of transportability introduced in [Pearl and Bareinboim, 2011; Bareinboim and Pearl, 2012b], which deals with transferring causal information when any experiment can be conducted at the source. Given that it is not always feasible to conduct certain controlled experiments, we consider the decision problem whether experiments on a selected subset Z of variables together with qualitative assumptions encoded in a diagram may render causal effects in the target environment computable from the available data. This problem, which we call z-transportability, reduces to ordinary transportability when Z is all-inclusive, and, like the latter, can be given syntactic characterization using the do-calculus [Pearl, 1995; 2000]. This paper establishes a necessary and sufficient condition for causal effects in the target domain to be estimable from both the non-experimental information available and the limited experimental information transferred from the source. We further provides a complete algorithm for computing the transport formula, that is, a way of fusing experimental and observational information to synthesize an unbiased estimate of the desired causal relation.

Introduction

The challenge of transporting experimental knowledge across heterogeneous settings is pervasive in science. Conclusions that are obtained in a laboratory setting are transported and applied elsewhere, in an environment that differs in many aspects from that of the laboratory [Pearl, 2012]. Similarly, when a robot is trained in a simulated environment, the question arises whether it could put the acquired knowledge into use in a new environment where relationships among agents, objects and features are different.

AI is in a unique position to tackle this challenge formally. First, the distinction between statistical and causal knowledge has received syntactic representation through causal diagrams [Pearl, 1995; Spirtes, Glymour, and Scheines, 2000; Pearl, 2000]. Second, graphical models provide a language for representing differences and commonalities among domains, environments, and populations [Pearl and Bareinboim, 2011] (henceforth, PB). Finally, the inferential machinery provided by the do-calculus [Pearl, 1995; 2000; Koller and Friedman, 2009] is particularly suitable for combining these two features into a coherent framework and developing effective algorithms for knowledge transfer.

In this line of research, the *transportability* problem [PB, 2011] deals with transferring causal knowledge between two environments Π and Π^*. In environment Π, (randomized) experiments can be performed and causal knowledge gathered. In Π^*, potentially different from Π, only passive observations can be collected but no experiments conducted. The problem is to infer a causal relationship R in Π^* using the gathered knowledge. Clearly, if nothing is known about the relationship between Π and Π^*, the problem is unsolvable. [1] Using a graphical representation called *selection diagrams* to encode commonalities and differences between environments [PB, 2011], a complete graphical and algorithmic characterization was provided in [Bareinboim and Pearl, 2012b] (henceforth, BP), which decides if and how transportability is feasible.

In real world applications, however, it may happen that certain controlled experiments cannot be conducted in the source environment (for financial, ethical, or technical reasons), so only a limited amount of experimental information can be gathered. A natural question arises whether the investigator in possession of a limited set of experiments would still be able to estimate the desired effects at the target.

This problem is called here "z-transportability" and generalizes ordinary transportability. Whenever any experiment may be conducted in the source, the two problems coincide. More formally, the z-transportability problem concerns the transfer of causal knowledge from a source domain Π to a target domain Π^*. In Π, experiments over the elements of a set $Z \subset V$ may be conducted (where V represent all variables in the system), so the set I_z contains the causal knowledge derived from the experiments $P(v|do(z'))$ [2], for all $Z' \subseteq Z$. In Π^*, potentially different from Π, only passive

[1] Unsolvable in the sense that for any estimation strategy, examples can be presented where the estimated value and the true effect are divergent even when sample size goes to infinity.

[2] We use $P_x(y)$ interchangeably with $P(y|do(x))$.

observations can be collected but no experiments conducted.

The goal of this paper is to provide a systematic analysis of the z-transportability problem, taking as input *any* arbitrary selection diagram together with an arbitrary set of experiments Z. Our contributions are summarized below:

- We provide a necessary and sufficient graphical condition for solving the z-transportability problem when Z is a set of variables under experimental control. We show that z-transportability is feasible if and only if the selection diagram does not contain a subgraph with certain properties.
- We then construct a complete algorithm for deciding z-transportability of causal effects, which returns a transport formula whenever those effects are z-transportable.
- We further show that the do-calculus is complete for the task of deciding z-transportability.

Motivating Examples

Consider Fig. 1(a) in which the node S represents factors that produce differences between source and target populations. Assume that we conduct a randomized trial in Los Angeles (LA) and estimate the causal effect of treatment X on outcome Y for every age group $Z = z$, denoted $P(y|do(x), z)$. We now wish to generalize the results to the population of New York City (NYC), but we find the distribution $P(x, y, z)$ in LA to be different from the one in the NYC (call the latter $P^*(x, y, z)$). In particular, the average age in NYC is significantly higher than that in LA. How are we to estimate the causal effect of X on Y in NYC, denoted $R = P^*(y|do(x))$? [3] [4]

The selection diagram for this example (Fig. 1(a)) conveys the assumption that the *only* difference between the two populations are factors determining age distributions, shown as $S \to Z$, while age-specific effects $P^*(y|do(x), Z = z)$ are invariant across cities. Difference-generating factors are represented by a special set of variables called *selection variables* S (or simply S-variables), which are graphically depicted as square nodes (■). From this assumption, the overall causal effect in NYC can be derived as follows:

$$\begin{aligned} R &= \sum_z P^*(y|do(x), z)P^*(z) \\ &= \sum_z P(y|do(x), z)P^*(z) \qquad (1) \end{aligned}$$

The last line is the *transport formula* for R. It combines experimental results obtained in LA, $P(y|do(x), z)$, with observational aspects of NYC population, $P^*(z)$, to obtain an experimental claim $P^*(y|do(x))$ about NYC. In this trivial example, the transport formula amounts to a simple recalibration (or re-weighting) of the age-specific effects to account for the new age distribution. In general, however, a more involved mixture of experimental and observational findings would be necessary to obtain a bias-free estimate of the target relation R, a full characterization of which is given in [BP, 2012b]. Interestingly, Fig. 1(b) is an example where bias-free transport of $P^*(y|do(x))$ is not feasible.

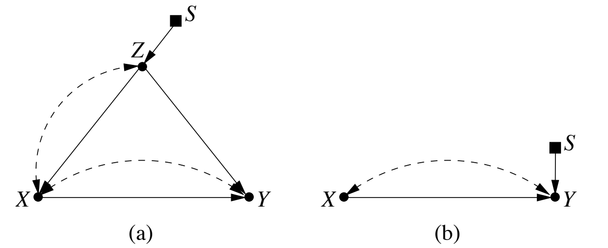


Figure 1: (a) Selection diagram illustrating when transportability among two domains is trivially solved through a simple recalibration. (b) Smallest selection diagram in which a causal relation is not transportable.

To illustrate a z-transportability problem, consider Fig. 2(a) and assume we wish, again, to estimate $P^*(y|do(x))$ but, now, X cannot be randomized. Instead, variable Z can be randomized, and we ask whether we can still estimate $P^*(y|do(x))$ despite this constraint and despite the fact that the two populations differ in the prior probabilities of Z (as shown by the variable S). [5]

Fortunately, in this case, the problem has a positive solution as can be seen from the following derivation. First apply Rule 3 of the do-calculus to add $do(z)$ to the expression,

$$P^*(y|do(x)) = P^*(y|do(x), do(z)) \text{ since } (Y \perp\!\!\!\perp Z|X)_{G_{\overline{X}\overline{Z}}}$$

Then apply Rule 2 to exchange $do(x)$ with x:

$$P^*(y|do(x), do(z)) = P^*(y|x, do(z)) \text{ since } (Y \perp\!\!\!\perp X|Z)_{G_{\underline{X}\overline{Z}}}$$

This last expression can be rewritten as,

$$P^*(y|x, do(z)) = P(y|x, do(z), s) = \frac{P(y, x|do(z))}{P(x|do(z))}, \quad (2)$$

where the first equality follows from the definition of selection diagram and the second using the separation of S from $\{X, Y\}$ after intervening on Z. Therefore, performing an experiment on Z in Π suffices to estimate the causal effect of X on Y in Π^* (without resorting to experimentation on X.)

There are subtle features of this problem that are worth illustrating. Whereas the graph in Fig. 2(a) permits the effect to be z-transportable, the graph in Fig. 2(b) does not. One is tempted to explain this difference by noting that in the mutilated graph from which the edges incoming to Z are cut (to simulate intervention), the causal effect of X on Y is identifiable in Fig. 2(a) but not in (b). That this is not the case is shown in the graph in Fig. 2(c). The resulting mutilated graph in this case entails both the identifiability and transportability of $P^*(y|do(x))$, but this effect is neither identifiable, nor transportable, nor z-transportable (shown later).

In a more involved manner, one might surmise that the solution for the z-identification problem [BP, 2012a] could yield the solution for z-transportability – z-identification

[3] We use the structural interpretation of causal diagrams as described in [Pearl, 2000, pp. 205–208]; see also Appendix 1.

[4] Following standard notation [Pearl, 2000], the dashed bidirected arrows in a graph stands for latent variables.

[5] A typical example is whether we can estimate the effect of cholesterol (X) on heart failure (Y) by experiments on diet (Z) given that cholesterol levels cannot be randomized [Pearl 2000, pp. 88–89].

asks for expressing the causal relation $R = P(y|do(x))$ in terms of experiments on Z (in a fixed domain Π) – however, this too turns out to not be the case. To witness, consider the diagram G in Fig. 3(a), and note that even though R is z-identifiable in Π, it is not the case that R is z-transportable.

Furthermore, consider the same task in regard to Fig. 3(b), a simple analysis for z-identification in the source would yield expression similar to the one in Fig. 2(a),

$$P(y|do(x)) = \frac{P(y, x|do(z))}{P(x|do(z))}, \tag{3}$$

but in this case, the availability of the ratio in eq. (3) is not sufficient for estimating the target quantity $R = P^*(y|do(x))$ in Π^*. Interestingly enough, the quantity R is z-transportable through the transport formula

$$P^*(y|do(x)) = \sum_w P(y|x, w, do(z))P^*(w|x, z), \tag{4}$$

which combines experimental results over Z obtained in the source Π, $P(y|x, w, do(z))$, with observational aspects of the target domain, $P^*(w|x, z)$, to obtain an experimental claim $P^*(y|do(x))$ about the target. (The derivation of this expression is shown more explicitly later on.)

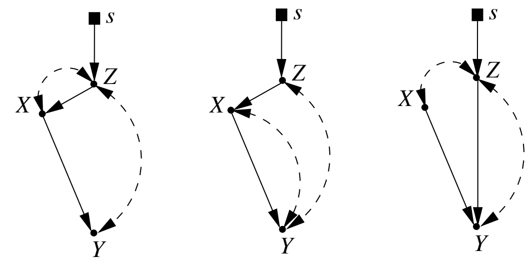
We note that the z-transportability problem reduces neither to transportability nor to z-identifiability, which leaves open the question of how to algorithmically characterize z-transportability. Our goal next is to get a better understanding of this problem and provide formal conditions for deciding whether a given quantity is (or is not) z-transportable from the available information at hand.

Preliminary Results

The basic semantical framework in our analysis rests on *probabilistic causal models* as defined in [Pearl, 2000, pp. 205], which are also called structural causal models. In the structural causal framework [Pearl, 2000, Ch. 7], actions are modifications of functional relationships, and each atomic action $do(\mathbf{X} = \mathbf{x})$ on a causal model M produces a new model $M_\mathbf{x} = \langle \mathbf{U}, \mathbf{V}, \mathbf{F}_\mathbf{x}, P(\mathbf{U})\rangle$, where $F_\mathbf{x}$ is obtained after replacing $f_X \in \mathbf{F}$ for every $X \in \mathbf{X}$ with a new function that outputs a constant value x given by $do(\mathbf{X} = \mathbf{x})$.

We follow the conventions given in [Pearl, 2000]. We denote variables by capital letters and their values by lower case. Similarly, sets of variables are denoted by bold capital letters, sets of values by bold letters. We will use graph-theoretic terminology with the typical kinship relationships (e.g., parents, ancestors). We usually omit the graph subscript whenever the graph in question is unambiguous. A graph $G_\mathbf{Y}$ will denote the induced subgraph G containing nodes in $\mathbf{Y}$ and all arrows between such nodes. Finally, $G_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}$ stands for the edge subgraph of G where all incoming arrows into $\mathbf{X}$ and all outgoing arrows from $\mathbf{Z}$ are removed.

Key to the analysis of z-transportability is the notion of "identifiability", defined in [Pearl, 2000, pp. 77], which expresses the requirement that causal effects be *computable* from a combination of passive data P and the assumptions embodied in a causal graph G (no experimental information is invoked). In identifiability problems, causal models and their induced graphs are associated with one domain (also called setting, study, population, environment).



(a) (b) (c)

Figure 2: Selection diagrams illustrating z-transportability of the causal effect $R = P^*(y|\hat{x})$. R can be z-transported with experiments on Z in model (a), but not in (b) and (c).

In transportability analysis, this representation was extended to capture properties of two domains simultaneously, using selection diagrams, to be defined next: [5]

Definition 1 (Selection Diagram [BP, 2012b]). *Let $\langle M, M^*\rangle$ be a pair of structural causal models [Pearl, 2000, pp. 205] relative to domains $\langle \Pi, \Pi^*\rangle$, sharing a causal diagram G. $\langle M, M^*\rangle$ 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 $S_i \to V_i$ whenever there might exist a discrepancy $f_i \neq f_i^*$ or $P(U_i) \neq P^*(U_i)$ between M and M^*.*

In words, the S-variables locate the *mechanisms* where structural discrepancies between the two domains are suspected to take place.[6] Alternatively, the absence of a selection node pointing to a variable represents the assumption that the mechanism responsible for assigning value to that variable is identical in both domains.

Armed with the concepts of identifiability and selection diagrams, the problem of z-transportability of causal effects can be defined as follows:

Definition 2 (z-Transportability). *Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ be disjoint sets of variables, and let D be a selection diagram relative to domains $\langle \Pi, \Pi^*\rangle$. Let $\langle P, I_z\rangle$ be the pair of observational and interventional distributions available in Π, where $I_z = \bigcup_{\mathbf{Z}' \subseteq \mathbf{Z}} P(\mathbf{v}|do(\mathbf{z}'))$, and P^* be the observational distribution of Π^*. The causal effect $R = P^*_\mathbf{x}(\mathbf{y})$ is z-transportable from Π to Π^* in D if $P^*_\mathbf{x}(\mathbf{y})$ is uniquely computable from $\langle P, I_z, P^*\rangle$ in any model that induces D.* [7]

The requirement that R be uniquely computable from $\langle P, I_z, P^*\rangle$ has a syntactic image in do-calculus, which is captured by the following Theorem.

[5] The assumption of no structural changes can be easily relaxed [BP, 2012b].

[6] Transportability assumes that enough structural knowledge about both domains is known in order to substantiate the production of their respective causal diagrams. In the absence of such knowledge, *causal discovery* algorithms might be used to infer the diagrams from data [Pearl and Verma, 1991; Pearl, 2000; Spirtes, Glymour, and Scheines, 2000].

[7] Henceforth, "z-transportability" will assume a specified set $\mathbf{Z}$.

Theorem 1. *Let D be the selection diagram characterizing Π and Π^*, and $\mathbf{S}$ a set of selection variables in D. The relation $R = P^*(\mathbf{y}|do(\mathbf{x}), \mathbf{z})$ is z-transportable from Π to Π^* in D if the expression $P(\mathbf{y}|do(\mathbf{x}), \mathbf{z}, \mathbf{s})$ is reducible, using the rules of do-calculus, to an expression in which all do-operators apply to subsets of Z, and the S-variables are separated from these do-operators.*

Proof. The result follows from the definition of z-transportability and soundness of the do-calculus. □

While it is not immediately obvious whether a sequence of rules exist that achieves the reduction required by the theorem, two elementary cases exist which are easily recognizable and can help in answering this question in general:

Definition 3. *(Trivial Transportability)*
A causal relation R is said to be trivially transportable *from Π to Π^*, if $R(\Pi^*)$ is identifiable from the data in Π^*.*

Definition 4. *(Direct z-Transportability)*
A causal relation $R = P^*(\mathbf{y}|do(\mathbf{z}), \mathbf{w})$ is said to be directly z-transportable *from Π to Π^* with $do(\mathbf{Z})$, if experiments $do(\mathbf{Z})$ are available in Π and $R(\Pi) = R(\Pi^*)$.*

A graphical test for direct z-transportability of $R = P^*(\mathbf{y}|do(\mathbf{z}), \mathbf{w})$ follows from the do-calculus and reads: $(S \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z}, \mathbf{W})_{D_{\overline{\mathbf{Z}}}}$; in words, $\mathbf{Z}$ blocks all paths from S to $\mathbf{Y}$ once we remove all arrows pointing to $\mathbf{Z}$ and condition on $\{\mathbf{Z}, \mathbf{W}\}$ in the selection diagram D. As an example, the X-specific causal effects in Fig. 2(a), $P^*(y|x, do(z))$, are directly z-transportable from Π to Π^*. These two cases (i.e., trivial and direct) will act as a basis to decompose the problem of z-transportability into smaller and more manageable subproblems.

We consider below conditions for when a quantity *cannot* be z-transported even when reduced to one of the elementary forms discussed above. The next lemma provides an auxiliary tool to prove that a quantity is not z-transportable based on the refutation of the uniqueness property required by Def. 2, which will be instrumental for proving completeness.

Lemma 1. *Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ be subsets of disjoint variables in domains Π and Π^*, and let D be the respective selection diagram. $R = P^*_{\mathbf{x}}(\mathbf{y})$ is not z-transportable from Π to Π^* in D if there exist two structural causal models M^1 and M^2 compatible with D such that $P_{M_1}(\mathbf{v}) = P_{M_2}(\mathbf{v})$, $P^*_{M_1}(\mathbf{v}) = P^*_{M_2}(\mathbf{v})$, $P_{M_1}(\mathbf{v}|do(\mathbf{z}')) = P_{M_2}(\mathbf{v}|do(\mathbf{z}'))$, for all $\mathbf{Z}' \subseteq \mathbf{Z}$, and $P^*_{M_1}(\mathbf{y}|do(\mathbf{x})) \neq P^*_{M_2}(\mathbf{y}|do(\mathbf{x}))$.*

Proof. Let $\mathcal{I}_z = \bigcup_{\mathbf{Z}' \subseteq \mathbf{Z}} P(\mathbf{v}|do(\mathbf{z}'))$, the collection of all interventional distributions in domain Π. The latter inequality of the Lemma rules out the existence of a function from $\langle \mathcal{P}, \mathcal{I}_z, P^* \rangle$ to $P^*_{\mathbf{x}}(\mathbf{y})$. □

While Lemma 1 might appear convoluted, it is nothing more than a formalization of the statement "R cannot be computed from information set IS alone." Naturally, when IS has three components, $\langle P, I_z, P^* \rangle$, it becomes lengthy. In turn, we use this lemma to show the non-z-transportability of $P^*(y|do(x))$ in the graphs in Fig. 2(b) and (c).

Theorem 2. *$P^*(y|do(x))$ is not z-transportable in the selection diagrams in Fig. 2(b) and (c).*

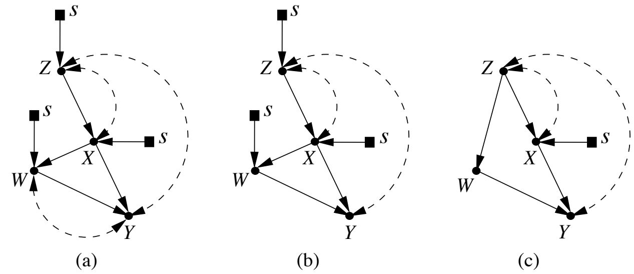


Figure 3: Selection diagrams illustrating the non-trivial relationship among the problems of z-identifiability, transportability, and z-transportability.

Proof. Consider the diagram G in Fig. 2(b). The existence of a hedge $\mathcal{F}', \mathcal{F}$ for $R = P^*(y|do(x))$ in $G_{\overline{Z}}$ implies that Z cannot help in the z-identifiability of R in Π [BP, 2012a]. Assume that R is z-transportable. Note that Z does not participate in the hedge $\mathcal{F}', \mathcal{F}$ since there is no bidirected edge going towards any of its elements in $G_{\overline{Z}}$. Further, consider a parametrization such that Z is a fair coin, so disconnected from $\mathbf{V} \backslash Z$ in G. We can use Lemma 1 and the witness $\mathcal{F}', \mathcal{F}$ for non-z-identifiability to show non-z-transportability in G. The inequality of R between the two models is obvious, and the agreement of the interventional distributions $do(Z)$ follows since Z is disconnected from $\mathbf{V} \setminus Z$ by construction. Contradiction. Since z-transportability has to be valid for any parametrization compatible with G, our claim follows.

In the diagram G in Fig. 2(c), the result is direct. First note that there exists a s-hedge $\mathcal{F}', \mathcal{F}$ for R in G, and by Theorem 5 in [BP, 2012b], R is not transportable. The equality of the interventional distributions $do(Z)$ follows by construction. Therefore, the same two models that witness non-transportability based on the s-hedge $\mathcal{F}', \mathcal{F}$ together with Lemma 1 entail the result. □

Characterizing z-Transportable Relations

We have seen positive as well as negative special cases of z-transportability. In the sequel, we build on the analysis of these cases to give a general characterization of z-transportability for *any* arbitrary selection diagram. A key concept in this characterization will be that of sC-components [BP, 2012b], which is essentially Tian's C-components over selection diagrams [Tian and Pearl, 2002].

Definition 5 (sC-component)**.** *Let D be a selection diagram such that a subset of its bidirected arcs forms a spanning tree over all vertices in D. Then D is an sC-component (selection confounded component).*

If D is not an sC-component itself, it can be uniquely partitioned into a set $\mathcal{C}(D)$ of sC-components. For instance, in Fig. 3(b), there are sC-components $C_1 = \{Z, X, Y\}$ and $C_2 = \{W\}$, since those are the two clusters of variables connected through bidirected edges. Each graph induces a unique sC-component decomposition, which is important since it will provide a way to re-express the target quantity into smaller pieces so as to allow the test of z-transportability in each of these pieces independently.

A special subset of sC-components that embraces the ancestral set of Y turns out to play an important role in deciding transportability (and first noted by Shpitser in the context of the identifiability problem), which will also be useful to z-transportability, as shown in turn.

Definition 6 (s^*-tree)**.** *Let D be a selection diagram where $\mathbf{Y}$ is the maximal root set. Then D is a $\mathbf{Y}$-rooted s^*-tree if D is an sC-component and all observable nodes have at most one child.*

There exist two interesting structures that stem from s^*-trees. When a pair of s^*-trees shares the same root set, and one encompasses the nodes in $\mathbf{X}$, and the other does not, this structure witnesses non-identifiability and is called *hedge* (i.e., identifiability is infeasible when a hedge is present as an edge subgraph of the inputted causal diagram) [Shpitser and Pearl, 2006]. If there exists a S-node pointing to the s^*-tree that does not intersect with $\mathbf{X}$, this structure witnesses non-transportability and is called s-hedge (i.e., transportability is infeasible when a *s-hedge* is present as an edge subgraph of the inputted selection diagram) [BP, 2012b]. The latter generalizes the former. [8]

Unfortunately, it is not the case that the s-hedge structure characterizes the set of z-transportable relations. To witness, note that there is no s-hedge in Fig. 2(b) and 3(c), so the effects $R = P^*(y|do(x))$ are transportable, but they are not z-transportable. For instance, in Fig. 2(b), there is no s-hedge because, even though there are s^*-trees $F' = \{Y\}$, $F = \{X, Z\} \cup F'$, there is not a selection node pointing F'.

Clearly, if a quantity R is not transportable, R is also not z-transportable, since z-transportability requires more stringent conditions to hold than transportability. The converse does not hold, and we have seen examples in which R is transportable but not z-transportable. In other words, when R is transportable, it is the case that R might be either z-transportable (e.g., Fig. 3(b)), or not z-transportable (e.g., Fig. 2(b) and 3(c)). Furthermore, the fact that any z-transportable quantity is transportable also follows.

Based on these observations, there exists a structure that generalizes s-hedges and will be shown to characterize z-transportable relations, which is defined below.

Definition 7 (zs-hedge)**.** *Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ be subsets of variables in the selection diagram D. Let F, F' be $\mathbf{R}$-rooted s^*-trees such that $F \cap \mathbf{X} \neq \emptyset$, $F' \cap \mathbf{X} = \emptyset$, $F' \subset F$, $\mathbf{R} \subset An(\mathbf{Y})_{D_{\overline{\mathbf{X}}}}$. Then F and F' form a zs-hedge for $P^*_{\mathbf{x}}(\mathbf{y})$ in Π and Π^* relative to $\mathbf{Z}$ if one of the following conditions hold:*

1. *There exists a S-node pointing to some node in F', or*
2. *For any $\mathbf{Z}' \subseteq \mathbf{Z} \cap F$: if all directed paths from $\mathbf{Z}'$ to $\mathbf{Y}$ in D are blocked by $\mathbf{X}$, $F \setminus \mathbf{Z}'$ is also a zs-hedge for $P^*_{\mathbf{x}}(\mathbf{y})$ in Π and Π^* relative to $\mathbf{Z} \setminus \mathbf{Z}'$; otherwise, F is also a zs-hedge for $P^*_{\mathbf{x}}(\mathbf{y})$ in Π and Π^* relative to $\mathbf{Z} \setminus \mathbf{Z}'$, or*
3. *$\mathbf{Z}$ is an empty set.*

We can see that zs-hedge captures the known cases of z-transportability. For example, if there is no S-node in the diagram, there must exist an experiment Z in the source such that the target quantity is rewritten so as to make use of this experiment (i.e., the second condition should fail). If the set Z is empty, the problem is unsolvable since there is no experiment that might yield z-transportability. If there is a Z-node that has a directed open path to $\mathbf{Y}$ (in the d-separation sense), this implies that the expression cannot be rewritten to make use of experimental data over Z.

Finally, we make the formal connection between the existence of a zs-hedge and the impossibility of z-transporting a certain causal relation.

Theorem 3. *Assume there exist s^*-trees F, F' that form a zs-hedge for $P^*_x(y)$ in Π and Π^* relative to $\mathbf{Z}$. Then $R = P^*_x(y)$ is not z-transportable from Π to Π^* in D.*

Proof sketch. Based on Lemma 1, we construct two models that agree on the observables $\langle P, I_z, P^* \rangle$, but disagree on the target relation R. These models extend the construction given in Thm. 3 [BP, 2012a] and Thm. 5 [BP, 2012b], being more involved due to the additional requirements imposed by z-transportability. See the full technical report for the explicit construction [Bareinboim and Pearl, 2013a]. □

While this result establishes the fact that zs-hedges precludes z-transportability, Theorem 3 shows neither how to locate a zs-hedge given a specific selection diagram, nor whether zs-hedges characterizes z-transportability (i.e., whether the converse holds). In the next section, we construct an algorithm which z-transports any causal effects in a diagram which does not contain a zs-hedge.

PROCEDURE $\mathbf{TR^z}(\mathbf{y}, \mathbf{x}, P_{\mathcal{I}}, \mathbf{Z}, \mathcal{I}, D)$
INPUT: $\mathbf{x}, \mathbf{y}$ value assignments, $P_{\mathcal{I}}$ observational distribution in Π^* (if $\mathcal{I} = \emptyset$), and interventional distribution in Π (if $\mathcal{I} \neq \emptyset$), $\mathbf{Z}$ set of variables with interventional distributions available in Π, $\mathcal{I}$ set of active variables in $\mathbf{Z}$, D a selection diagram, $\mathbf{S}$ set of selection nodes. [$P^*, P_{\mathbf{Z}}$ are globally available, and $P_{\mathcal{I}}$ represents the distribution given active $\mathbf{Z}$, removing the nodes after conditioning in the top. order.]
OUTPUT: $P^*_{\mathbf{x}}(\mathbf{y})$ in terms of P^*, $P_{\mathbf{Z}}$ or $FAIL(D, C_0)$.

if $\mathbf{x} = \emptyset$, **return** $\sum_{\mathbf{V} \setminus \mathbf{Y}} P_{\mathcal{I}}(\mathbf{V})$
if $\mathbf{V} \setminus An(\mathbf{Y})_D \neq \emptyset$, **return** $\mathbf{TR^z}(\mathbf{y}, \mathbf{x} \cap An(\mathbf{Y})_D,$
  $\sum_{\mathbf{V} \setminus An(\mathbf{Y})_D} P_{\mathcal{I}}, \mathbf{Z}, \mathcal{I}, An(\mathbf{Y})_D)$
Set $\mathbf{W} = (\mathbf{V} \setminus \mathbf{X}) \setminus An(\mathbf{Y})_{D_{\overline{\mathbf{X}}}}$.
  if $\mathbf{W} \neq \emptyset$, **return** $\mathbf{TR^z}(\mathbf{y}, \mathbf{x} \cup \mathbf{w}, P_{\mathcal{I}}, \mathbf{Z}, \mathcal{I}, D)$
if $\mathcal{C}(D \setminus \mathbf{X}) = \{C_0, C_1, ..., C_k\}$,
  return $\sum_{\mathbf{V} \setminus \{\mathbf{Y}, \mathbf{X}\}} \prod_i \mathbf{TR^z}(c_i, \mathbf{V} \setminus c_i, P_{\mathcal{I}}, \mathbf{Z}, \mathcal{I}, D)$
if $\mathcal{C}(D \setminus \mathbf{X}) = \{C_0\}$,
  if $\mathcal{C}(D) \neq \{D\}$,
    if $C_0 \in \mathcal{C}(D)$, **return** $\sum_{s \setminus \mathbf{Y}} \prod_{i|V_i \in C_0} P_{\mathcal{I}}(v_i|V_D^{(i-1)})$
    if $(\exists C') C_0 \subset C' \in \mathcal{C}(D)$, **return** $\mathbf{TR^z}(\mathbf{y}, \mathbf{x} \cap C',$
    $\prod_{i|V_i \in C'} P_{\mathcal{I}}(V_i|V_D^{(i-1)} \cap C', v_D^{(i-1)} \setminus C'), \mathbf{Z}, \mathcal{I}, C')$
  else,
    if $\left((\mathbf{S} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X})_{D_{\overline{\mathbf{X}}}} \wedge (\mathbf{Z} \cap \mathbf{X} \neq \emptyset)\right)$,
    return $\mathbf{TR^z}(\mathbf{y}, \mathbf{x} \setminus \mathbf{z}, P_{\mathcal{I}}, \mathbf{Z} \setminus \mathbf{X}, \mathbf{Z} \cap \mathbf{X}, D \setminus \{\mathbf{Z} \cap \mathbf{X}\})$
    else, $\mathbf{FAIL}(D, C_0)$

Figure 4: Modified version of transportability algorithm capable of recognizing z-transportable relations.

[8] For a more detailed discussion on the relationship between these two structures, refer to [BP, 2012b].

A Complete Algorithm for z-Transportability

Some of the previous analyses of identifiability, z-identifiability, and transportability [Kuroki and Miyakawa, 1999; Tian and Pearl, 2002; Shpitser and Pearl, 2006; Huang and Valtorta, 2006; BP, 2012b; 2012a] will be useful in the algorithmization of z-transportability which generalizes these problems. We construct an algorithm called $\mathbf{TR^z}$ (see Fig. 4) based on the algorithm **sID** algorithm introduced in [BP, 2012b] (a variant of **ID** [Shpitser and Pearl, 2006]), which explicitly employed the s-hedge structure that will show to be instrumental to prove completeness.

In the sequel, we explain the general strategy undertaken by $\mathbf{TR^z}$, which builds on two observations developed so far:

(i) z-transportability (sufficiency): Causal relations can be z-transported if trivially transportable (def. 3) or directly z-transportable (def. 4), which relies on the experiments performed over $\mathbf{Z}$. The current algorithms already operate on the first part, proceeding through a sequence of equalities in do-calculus based on the sC-component decomposition. The idea is to apply a divide-and-conquer strategy breaking the problem into smaller, more manageable pieces, and then to assemble them back when this is possible. In each base case, we have to evaluate these pieces, checking whether they are z-transportable based on the definitions 3 or 4.

(ii) Non-z-transportability (necessity): The algorithm proceeds until it is not able to resolve a certain subproblem, which implies the existence of a zs-hedge. It is not immediately obvious that failure of the algorithm implies the existence of this zs-hedge. Assuming that this is the case, it is not difficult to see that Theorem 3 can be used to generate a counterexample to non-z-transportability based on the refutation of the uniqueness property using Lemma 1.

Before showing the more formal properties of $\mathbf{TR}^z$, we demonstrate how $\mathbf{TR}^z$ works through the z-transportability of $R = P^*(y|do(x))$ using $\{Z\}$ in the graph D in Fig. 3(b).

The process starts with $\mathbf{TR}^z(\mathbf{y}, \mathbf{x}, P_{\mathcal{I}}, \{Z\}, \{\}, D, 1)$, and after failure in the tests in lines 1 and 2, $\mathbf{TR}^z$ succeed in line 3, setting $\mathbf{W} = \{Z\}$ (since $An(\mathbf{Y})_{D_{\overline{\mathbf{X}}}} = \{X, Y, W\}$, which does not include $\{Z\}$). Thus, the original problem is reduced to the call $\mathbf{TR}^z(\mathbf{y}, \{x\} \cup \{z\}, P_{\mathcal{I}}, \mathbf{Z}, \{\}, D)$.

After failure in the previous tests, $\mathbf{TR}^z$ invokes line 4. Since $D = An(Y)$ and $\mathcal{C}(D \setminus \{X, Z\}) = (C_0, C_1)$, where $C_0 = D(\{W\})$ and $C_1 = D(\{Y\})$, the original problem is reduced to z-transporting respectively $Q_0 = P^*_{x,z,y}(w)$ and $Q_1 = P^*_{x,w}(y, z)$, which implies that $R = \sum_w Q_0\, Q_1$.

Evaluating the first factor $Q_0 = P^*_{x,z,y}(w)$, $\mathbf{TR}^z$ triggers line 2, noting that nodes which are not ancestors of W can be ignored. This implies that $P^*_{x,z,y}(w) = P^*_{x,z}(w)$ with induced subgraph $C_0 = \{Z \to X, X \to W, Z \leftarrow U_{zx} \to X\}$, where U_{zx} stands for the hidden variable between Z and X. In the new call, $\mathbf{TR}^z$ goes to line 5, where locally $\mathcal{C}(D \setminus \{X, Z\}) = \{C_1\}$, for $C_1 = \{W\}$. Given that $C_0 \neq C_1$, the test in line 6 succeed, and so the test in line 7, noting that C_1 is an sC-component itself (there is no bidirected edge connecting $\{W\}$ and $\{Z, X\}$). So, $\mathbf{TR}^z$ can trivially z-transport this factor and returns $P^*(w|z, x)$.

Evaluating the second factor $Q_1 = P^*_{x,z,w}(y)$, $\mathbf{TR}^z$ fails until the tests in lines 5 and 6, where the local induced subgraph is $C_0 = \{Y\}$. $\mathbf{TR}^z$ fails in line 7 since C_0 is not an sC-component itself (just a part of another sC-component). In the sequel, the test in line 8 comes true, where $C' = \{Z \to X, X \to Y, Z \leftarrow U_{zx} \to X, Z \leftarrow U_{zy} \to Y\}$, so the original call is reduced through the removal of $\{W\}$, which is not part of the sC-component (there are no bidirected edges between $\{W\}$ and C'). In the new call, $\mathbf{TR}^z$ succeed in the test in line 5, but fails in line 6. In the sequel, both tests in line 10 come true, and $\{X, Z\} \cap \{Z\} = \{Z\} = \mathcal{I}$, which induces the graph $C' \setminus \{Z\} = \{X \to Y\} = C_2$. Finally, $\mathbf{TR}^z$ fails until line 6, and then triggers line 7 since $\{Y\}$ is a component itself in C_2, so returning $P(y|w, x, do(z))$ (in the source since $\mathcal{I} \neq \emptyset$). This result coincides with Eq. (4).

We prove next soundness and completeness of $\mathbf{TR^z}$.

Theorem 4 (soundness)**.** *Whenever* $\mathbf{TR^z}$ *returns an expression for* $P^*_{\mathbf{x}}(\mathbf{y})$*, it is correct.*

Proof. The result partly follows from the soundness of **sID** shown in Thm. 6 [BP, 2012b], which is inherited by $\mathbf{TR^z}$ by construction. Note that the process of identification of the target relation without the $\mathbf{Z}$-nodes, that were considered in line 10, is allowed since, by assumption, the interventional distribution $do(\mathbf{Z})$ can be used after testing for direct z-transportability in the respective local call. ☐

Theorem 5. *Assume* $\mathbf{TR^z}$ *fails to z-transport* $P^*_{\mathbf{x}}(\mathbf{y})$ *(executes line 11). Then, there exists* $\mathbf{X}' \subseteq \mathbf{X}$*,* $\mathbf{Y}' \subseteq \mathbf{Y}$*,* $\mathbf{Z}' \subseteq \mathbf{Z}$ *such that the graph pair* D, C_0 *returned by the fail condition of* $\mathbf{TR^z}$ *contains as edge subgraphs* s^**-trees* F*,* F' *that form a* zs*-hedge for* $P^*_{\mathbf{x}'}(\mathbf{y}')$ *in* Π *and* Π^* *relative to* $\mathbf{Z}'$*.*

Proof sketch. We can use the specific topological relation between the graphs D, C_0, remove non-essential edges, and show that the remaining structure matches the definition of a zs-hedge. See [Bareinboim and Pearl, 2013a]. ☐

The following results are now immediate.

Corollary 1 (completeness)**.** $\mathbf{TR^z}$ *is complete.*

Corollary 2 (do-calculus characterization)**.** *The rules of do-calculus, together with standard probability manipulations are complete for determining z-transportability of* $P^*_{\mathbf{x}}(\mathbf{y})$*.*

Corollary 3 (zs-hedge criterion)**.** $P^*_{\mathbf{x}}(\mathbf{y})$ *is z-transportable from* Π *to* Π^* *in D if and only if there does not exist a zs-hedge for* $P^*_{\mathbf{x}'}(\mathbf{y}')$ *in D, for any* $\mathbf{X}' \subseteq \mathbf{X}$*,* $\mathbf{Y}' \subseteq \mathbf{Y}$*,* $\mathbf{Z}' \subseteq \mathbf{Z}$*.*

Conclusions

This paper treats transportability problems in which experiments can be conducted only over limited sets of variables $\mathbf{Z}$. We provide a necessary and sufficient graphical condition under which causal effects in a target environment can be estimated from experimental information transported from the source environment, potentially different from the former. We further provide a complete algorithm for computing the resulting mapping, that is, a formula for fusing available observational and experimental data to synthesize an estimate of the desired causal effects. We show that the do-calculus is complete for characterizing the z-transportability class. While practical applications of these results are predicated on the availability of problem-specific selection diagrams, the general understanding of why some problems permit information transfer and other do not has scientific merit on its own. It informs investigators what kind of disparities between environments would make transportability theoretically impossible, and what disparities can be circumvented by clever information fusion strategies. Even though the construction of a selection diagram might be a demanding task, the completeness result makes such construction unavoidable if one seeks theoretical guarantees for a given method of information transfer. Fortunately, the knowledge necessary to construct a diagram is not much different than that required for ordinary causal diagrams as used, for example, to establish internal validity (i.e., identifiability). This paper complements a recent work on transportability called meta-transportability [Bareinboim and Pearl, 2013b], which deals with transferring causal information from multiple heterogeneous domains.

Acknowledgement

We thank the reviewers for their comments. This research was supported in parts by grants from NSF #IIS-1249822, and ONR #N00014-13-1-0153 and #N00014-10-1-0933.

Appendix 1

The do-calculus [Pearl, 1995] consists of three rules that permit us to transform expressions involving do-operators into other expressions of this type, whenever certain conditions hold in the causal diagram G. We consider a DAG G in which each child-parent family represents a deterministic function $x_i = f_i(pa_i, \epsilon_i), i = 1, \ldots, n$, where pa_i are the parents of variables X_i in G; and $\epsilon_i, i = 1, \ldots, n$ are arbitrarily distributed random disturbances, representing background factors that the investigator chooses not to include in the analysis.

Let X, Y, and Z be disjoint sets of nodes in a causal DAG G. We denote by $G_{\overline{X}\underline{Z}}$. the graph obtained by deleting from G all arrows pointing to nodes in X and all edges emerging from nodes in Z. The following three rules are valid for every interventional distribution compatible with G.

Rule 1: $P_{\mathbf{x}}(\mathbf{y}|\mathbf{z}, \mathbf{w}) = P_{\mathbf{x}}(\mathbf{y}|\mathbf{w})$ if $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z}|\mathbf{X}, \mathbf{W})_{G_{\overline{\mathbf{X}}}}$.
Rule 2: $P_{\mathbf{x},\mathbf{z}}(\mathbf{y}|\mathbf{w}) = P_{\mathbf{x}}(\mathbf{y}|\mathbf{z}, \mathbf{w})$ if $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z}|\mathbf{X}, \mathbf{W})_{G_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}}$.
Rule 3: $P_{\mathbf{x},\mathbf{z}}(\mathbf{y}|\mathbf{w}) = P_{\mathbf{x}}(\mathbf{y}|\mathbf{w})$ if $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z}|\mathbf{X}, \mathbf{W})_{G_{\overline{\mathbf{X}, \mathbf{Z}^*}}}$, where $\mathbf{Z}^* = \mathbf{Z} \setminus Anc(\mathbf{W})_{G_{\overline{\mathbf{X}}}}$.

In words, rule 1 affirms that the d-separation criterion holds when the system is under an intervention $do(X = x)$; rule 2 gives a condition to exchange the action $do(Z = z)$ with passive observation $(Z = z)$; rule 3 gives a condition under which the action $do(Z = z)$ is irrelevant and can be deleted (similarly to conditional independences when an observation is irrelevant). The do-calculus was proven to be complete for the task of identification of causal effects [Shpitser and Pearl, 2006; Huang and Valtorta, 2006].

Appendix 2

To exemplify the use of do-calculus, we apply it to derive the transport formula for the model of Fig. 3(b) (Eq. 4),

$$P^*(y|do(x)) = P(y|do(x), S) = P(y|do(x), do(z), S)$$
$\big($3rd rule of do-calculus, $(Z \perp\!\!\!\perp Y|X, S)_{G_{\overline{X}}}\big)$
$$= \textstyle\sum_w P(y|do(x), do(z), w, S)P(w|do(x), do(z), S)$$
$$= \textstyle\sum_w P(y|do(x), do(z), w)P(w|do(x), do(z), S)$$
$\big($1st rule of do-calculus, $(S \perp\!\!\!\perp Y|X, Z)_{G_{\overline{X,Z}}}\big)$
$$= \textstyle\sum_w P(y|x, do(z), w)P(w|do(x), do(z), S)$$
$\big($2nd rule of do-calculus, $(X \perp\!\!\!\perp Y|Z)_{G_{\underline{X}\overline{Z}}}\big)$
$$= \textstyle\sum_w P(y|x, do(z), w)P(w|x, z, S)$$
$\big($2nd rule of do-calculus, $(X, Z \perp\!\!\!\perp W)_{G_{\underline{X,Z}}}\big)$
$$= \textstyle\sum_w P(y|x, do(z), w)P^*(w|x, z)$$

References

Bareinboim, E., and Pearl, J. 2012a. Causal inference by surrogate experiments: z-identifiability. In de Freitas, N., and Murphy, K., eds., *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence (UAI)*, 113–120. AUAI Press.

Bareinboim, E., and Pearl, J. 2012b. Transportability of causal effects: Completeness results. In Hoffmann., J., and Selman, B., eds., *Proceedings of The Twenty-Sixth Conference on Artificial Intelligence (AAAI)*, 698–704. AAAI Press.

Bareinboim, E., and Pearl, J. 2013a. Causal transportability with limited experiments. Technical Report R-408, Cognitive Systems Laboratory, Department of Computer Science, UCLA.

Bareinboim, E., and Pearl, J. 2013b. Meta-transportability of causal effects: A formal approach. In *Proceedings of The Sixteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*. JMLR (in press).

Huang, Y., and Valtorta, M. 2006. Identifiability in causal bayesian networks: A sound and complete algorithm. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI)*, 1149–1156. AAAI Press.

Koller, D., and Friedman, N. 2009. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press.

Kuroki, M., and Miyakawa, M. 1999. Identifiability criteria for causal effects of joint interventions. *Journal of the Royal Statistical Society* 29:105–117.

Pearl, J., and Bareinboim, E. 2011. Transportability of causal and statistical relations: A formal approach. In *Proceedings of the Twenty-Fifth National Conference on Artificial Intelligence (AAAI)*, 247–254. AAAI Press.

Pearl, J., and Verma, T. 1991. A theory of inferred causation. In Allen, J.; Fikes, R.; and Sandewall, E., eds., *Proceedings of the Second Conference on Principles of Knowledge Representation and Reasoning (KR)*. San Mateo, CA: Morgan Kaufmann. 441–452.

Pearl, J. 1995. Causal diagrams for empirical research. *Biometrika* 82(4):669–710.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press. Second ed., 2009.

Pearl, J. 2012. Do-calculus revisited. In de Freitas, N., and Murphy, K., eds., *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence (UAI)*, 4–11. AUAI Press.

Shpitser, I., and Pearl, J. 2006. Identification of joint interventional distributions in recursive semi-Markovian causal models. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI)*, 1219–1226. AAAI Press.

Spirtes, P.; Glymour, C.; and Scheines, R. 2000. *Causation, Prediction, and Search*. Cambridge, MA: MIT Press, 2nd edition.

Tian, J., and Pearl, J. 2002. A general identification condition for causal effects. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI)*, 567–573. AAAI Press.