Abstracting Probabilistic Relational Models

Vaishak Belle
University of Edinburgh &
Alan Turing Institute
vaishak@ed.ac.uk

Abstract

Abstraction is a powerful idea widely used in science to model, reason and explain the behavior of systems in a more tractable search space, by omitting irrelevant details. While notions of abstraction have matured for deterministic systems, the case for abstracting probabilistic models is not yet fully understood.

In this paper, we develop a foundational framework for abstraction in probabilistic relational models from first principles. These models borrow syntactic devices from first-order logic and are very expressive, thus naturally allowing for relational and hierarchical constructs with stochastic primitives. We motivate a definition of consistency between a high-level model and its low-level counterpart, but also treat the case when the high-level model is missing critical information present in the low-level model. We prove properties of abstractions, both at the level of the parameter as well as the structure of the models.

1 Introduction

Abstraction is a powerful idea widely used in science to explain phenomena at the required granularity. Think of explaining a heart disease in terms of its anatomical components versus its molecular composition. In computer science, it is often understood as the process of mapping one representation onto a simpler representation by suppressing irrelevant information. The motivation is three-fold:

(a) When representing complex pieces of knowledge, abstraction can provide a way to structure that knowledge, hierarchically or otherwise, so as to yield descriptive clarity and modularity.

(b) Reasoning over large graphs, programs, and other structures is almost always computationally challenging, and so abstracting the problem domain to a smaller search space is attractive. Even in the case of tractable representations, such as arithmetic circuits [Darwiche and Marquis, 2002], reasoning is polynomial in the circuit size, so clearly a smaller circuit is more effective.

(c) Lastly, and perhaps most significantly, abstraction features pervasively in commonsense reasoning, and there is much discussion in the fields of cognitive science and philosophy on the role of abstractions for explanations [Jorland, 1994, Dedre and Christian, 2017]. For example, [Gartinkel, 1981] argues that concrete explanations containing too much detail are sensitive to perturbations and are impractical for understanding physical phenomena. Thus, abstractions will likely be critical for explainable AI [Gunning, 2016], and indeed, much of that literature focuses on extracting high-level symbolic and/or programmatic representations from low-level data (e.g., [Penkov and Ramamoorthy, 2017, Sreedharan et al., 2018]).

However, the formal analysis of abstraction has largely focused on categorical (e.g., deterministic, non-probabilistic) domains [Giunchiglia and Walsh, 1992, Milner, 1989], and so is not immediately applicable to the field of statistical machine learning. We do not yet have a full understanding of which aspects of one probabilistic model, representing some low-level phenomena, can be omitted when building a less granular model, possibly standing for a high-level understanding of the domain.

In this paper, we develop a foundational framework for abstraction in probabilistic relational models from first principles. Probabilistic relational models (PRMs) generalize standard (propositional) probabilistic models in borrowing syntactic constructs from first-order logic [Heckerman et al., 2004, Getoor and Taskar, 2007, Richardson and Domingos, 2006]. Thus, our results are
Bayesian network:
The model instantiates constraints for a (parameterised) on entity-relationships for a university database. For example, consider the PRM from [Heckerman et al., 2004] 

To motivate that using an example, consider the PRM from [Heckerman et al., 2004] on entity-relationships for a university database \( \mathcal{U} \). The model instantiates constraints for a (parameterised) Bayesian network:

\[
\text{Difficulty} \rightarrow \text{Grades} \leftarrow \text{IQ}
\]

as follows, referred to as the low-level theory \( \mathcal{U}_l \) in the sequel:

\[
.7 \ \text{diff}(x, E)
\]

\[
.1 \ \text{diff}(x, M)
\]

\[
.2 \ \text{diff}(x, H)
\]

\[
.25 \ \text{iq}(x, L) \land \text{diff}(y, E) \land \text{takes}(x, y) \supset \text{grades}(x, y, u)
\quad \text{for } u \in \{7, 8, 9, 10\}
\]

\[
.25 \ \text{iq}(x, L) \land \lnot\text{diff}(y, E) \land \text{takes}(x, y) \supset \text{grades}(x, y, u)
\quad \text{for } u \in \{5, 6, 7, 8\}
\]

where the constants \( E, M, H, L \) stand for easy, medium, hard, low respectively. (A precise encoding will be presented in a subsequent section.)

The first constraint says that for given any course, say \( E \), the probability that its difficulty level is easy is .7. More generally, this theory says that courses come in three levels of difficulty, and when a low IQ student takes an easy course, his grades can be modeled as a uniform distribution on \{7, 8, 9, 10\}, and when he does not take an easy course, it is a uniform distribution on \{5, 6, 7, 8\}.

A simple yet powerful type of abstraction to apply here is domain abstraction. Assuming the above sentences are the only ones of interest to us, we can lump the constants \( \{M, H\} \) as \( N \), standing for not easy, and lump the mentioned grade values together as \{5, 6\}, \{7, 8\}, \{9, 10\} and denote them as \( B, O, G \), standing for bad, ok and good respectively. Then, we would obtain the following model, referred to as the high-level theory \( \mathcal{U}_h \) in the sequel:

\[
.7 \ \text{diff}(x, E)
\]

\[
.3 \ \text{diff}(x, N)
\]

\[
.5 \ \text{iq}(x, L) \land \text{diff}(y, E) \land \text{takes}(x, y) \supset \text{grades}(x, y, u)
\quad \text{for } u \in \{O, G\}
\]

\[
.5 \ \text{iq}(x, L) \land \text{diff}(y, N) \land \text{takes}(x, y) \supset \text{grades}(x, y, u)
\quad \text{for } u \in \{B, O\}
\]

On closer inspection, the reader may observe that \( \mathcal{U}_h \) is, in fact, a very faithful abstraction of \( \mathcal{U}_l \), in terms of accurately grouping together probabilistic events. Indeed, we will formally show that the two models agree on a large class of probabilistic queries. The benefit, of course, is that \( \mathcal{U}_h \) is defined over a smaller set of random variables.

However, such a faithful alignment may not always be needed, or even feasible. Consider a case of predicate abstraction, where one groups definitions and complex formulas using new predicates. Suppose we had a course list database \( C \). Let \( C_l \) be a low-level theory:

\[
.9 \ \text{CS}(x) \supset \text{diff}(x, H)
\]

\[
.8 \ \text{Physics}(x) \supset \text{diff}(x, E)
\]

\[
1 \ (\text{AI}(x) \supset \text{CS}(x)) \land (\text{Astronomy}(x) \supset \text{Physics}(x))
\]

We may want to define a high-level theory \( C_h \) that simply uses \( \text{Science}(x) \) in place of \( \text{CS}(x) \) and \( \text{Physics}(x) \). But then the weight on rules such as \( \text{Science}(x) \supset \text{diff}(x, H) \) or \( \text{Science}(x) \supset \text{diff}(x, E) \) may not be immediate to derive, in general. Predicate abstraction can also be used as a strategy to check for probabilistically significant events. For example, an administrator may only be interested in ensuring that all low IQ students enroll in an easy course: 

\[
\text{alert} \equiv \lnot[\exists x, \exists y \ (\text{iq}(x, L) \supset (\text{takes}(x, y) \land \text{diff}(y, E)))]
\]

and specifically, whether that atom ever obtains a non-zero probability. Indeed, the literature on verification and security often approach the reasoning of complex systems by distinguishing bad states (e.g., invalid paths.

\[1\] Although the abstraction uses the same predicates as \( \mathcal{U}_h \), note that some of these are essentially new predicates, with different domains. For example, in \( \mathcal{U}_h \), the difficulty ranges over \( \{E, M, H\} \) whereas in \( \mathcal{U}_h \), it ranges over \( \{E, N\} \). The context will make clear whether the predicates and constants are from \( \mathcal{U}_h \) or from \( \mathcal{U}_h \), and so we do not distinguish symbols from \( \mathcal{U}_h \) by means of superscripts and such.
Understand as the underlying probabilistic semantic constraints for analyzing abstractions. Thus, we assume a first-order language with finitely many relations, constants, and a set of relations \( \{ P_1(x), \ldots \} \) and constants \( D \). The domain is fixed to a finite set and \( x, y, z \). In particular, \( x, y, z \) are \( \Delta \), the domain of discourse. The model \( M \) is to be abstracted.

In that spirit, we show that abstraction can be understood both from the viewpoint of the parameters (i.e., weights and/or probabilities) and structure (i.e., the logical sentences). To do this, we consider the case of aligning probabilities exactly between the high-level and low-level models, where one obtains an alignment between the probable and improbable events. When it comes to abstracting structure, we show that one wants to ensure that the high-level model is consistent, and perhaps additionally that it is not missing critical information present at the low-level model. This then motivates a definition of soundness and completeness.

We reiterate that our focus here is primarily about the semantic constraints for analyzing abstractions. Thus, at the outset, we assume that we are given a high-level theory, capturing the more abstract probabilistic model, and a low-level theory, understood as the underlying probabilistic model that is to be abstracted.

### 3 Preliminaries

#### 3.1 Logical Language

We assume a first-order language with finitely many relational symbols \( \{ P_1(x), \ldots, P_2(x, y), \ldots, P_3(x, y, z), \ldots \} \), variables \( \{ x, y, z, \ldots \} \), connectives \( \lor, \land, \lor, \land, v \) and a finite set of constants \( D \), serving as the domain of discourse for quantification. Usual abbreviations hold for connectives: we write \( \alpha \supset \beta \) (material implication) to mean \( \neg \alpha \lor \beta \), \( \alpha \equiv \beta \) (equivalence) to mean \( (\alpha \supset \beta) \land (\beta \supset \alpha) \), and \( \exists \alpha \) (existential quantification) to mean \( \forall x \neg \alpha \). In particular, then the domain is fixed to a finite set \( D \), we write \( \forall x \alpha(x) \) to mean \( \forall c \in D \alpha(c) \). Moreover, \( \alpha \land \beta \) is equivalent to \( \neg (\neg \alpha \lor \neg \beta) \), so in proofs, we only consider the connectives \( \land, \lor, \neg \).

The set of (ground) atoms is defined as: \( \{ P(c_1, \ldots, c_k) \mid \exists c \in D, c_i \in D \} \). The set of literals is obtained from the set of atoms, and their negations.

A model \( M \) is a \( \{ 0, 1 \} \) assignment to the set of atoms. Using \( \models \) to denote satisfaction, the semantics for a formula \( \phi \) is defined inductively: \( M \models P(\bar{c}) \) for atom \( P(\bar{c}) \) if \( M[P(\bar{c})] = 1 \); \( M \models \neg \phi \) if \( M \models \phi \) does not hold (also written \( M \models \phi \)); \( M \models \phi \lor \psi \) iff \( M \models \phi \) or \( M \models \psi \); and \( M \models \phi \land \psi \) iff \( M \models \phi \) and \( M \models \psi \). We write \( I \) to mean \( M \models I \) for literal \( I \).

We say a formula \( \phi \) is satisfiable iff there is a model \( M \) such that \( M \models \phi \). We say that \( \phi \) is valid, written \( \models \phi \), iff for every model \( M, M \models \phi \). Moreover, we write \( \phi \models \alpha \) to mean that in every model \( M \) such that \( M \models \phi \), it is also the case that \( M \models \alpha \).

To prepare for our technical discussion, we discuss some notational conventions. Given a formula \( \Delta \), we write \( \text{Lang}(\Delta) \) to mean the logical sub-language implicit in \( \Delta \): that is, the set of well-formed formulas constructed from relations relations \( \{ P_1(x), \ldots \} \) and constants \( D \) mentioned in \( \Delta \). We can then write \( \alpha \in \text{Lang}(\Delta) \) to mean such as well-formed formula. Analogously, we write \( \text{Lits}(\Delta) \) to mean the set of literals obtained from \( \text{Lang}(\Delta) \). For example, if \( \Delta = P(c) \lor Q(c, a) \), then \( \neg P(a) \in \text{Lang}(\Delta), Q(a, a) \in \text{Lang}(\Delta), P(a) \in \text{Lits}(\Delta), \neg Q(a, c) \in \text{Lits}(\Delta) \), and so on. We often abuse notation and write \( \bar{c} \in D \) to mean that each of the constants mentioned in \( \Delta \) is taken from \( D \). Finally, given a \( \Delta \), when we write \( M \models \Delta \), it is implicit here that we take \( M \) to be a model for the language \( \text{Lang}(\Delta) \); that is, it is a \( \{ 0, 1 \} \) assignment to the set of atoms in \( \text{Lang}(\Delta) \). We can make this explicit by writing \( M \in \text{Models}(\text{Lang}(\Delta)) \), or simply \( M \in \text{Models}(\Delta) \) for short.

#### 3.2 Weighted Model Counting

To develop our framework, we appeal to the technical device of weighted model counting (WMC) \cite{Bacchus2009}. WMC is defined over the models of a propositional formula, and serves as an assembly language for a number of heterogeneous representations, including factor graphs, Bayesian networks, probabilistic databases and probabilistic programs \cite{Fierens2011, Bacchus2009, Suciu2011}. WMC enjoys a number of interesting properties that make it particularly well-suited for our endeavor. First, it separates the symbolic representation (i.e., a logical encoding of the probabilistic model) from a weight function denoting the probabilities of variables, which allows us to investigate abstractions both at the level of structures and at the level of parameters. Second, WMC provides a semantic as well as a computational view for probabilistic reasoning. Semantically, the models of propositional formulas map

---

2The reason we go to some length to discuss our notational conventions is this: when we work with a fixed language, the set of relations, literals, and models to consider is immediate. That will no longer be true when we are thinking of different logical languages for high-level and low-level theories, in which case our notation will provide context.
to states in probability spaces (i.e., assignments of values to random variables). Computationally, we are able to reuse SAT technology for building exact and approximate solvers [Gomes et al., 2009], while still leveraging context-specific independences [Boutilier et al., 1996].

Essentially, model counting is the task of counting the models of a propositional formula [Gomes et al., 2009]. WMC extends that problem setting in additionally according weights to literals, and summing the weights of the models, defined in terms of the product of the literal weights. Formally,

**Definition 1.** Suppose $\Delta$ is a ground first-order sentence. Suppose $\mathbf{w}$ is a function that maps the elements of Lits($\Delta$) to $\mathbb{R}^{[0,\infty)}$. Then the WMC of $\Delta$ is defined as:

$$\text{WMC}(\Delta, \mathbf{w}) = \sum_{M \models \Delta} \prod_{e \in M} w(l)$$

Given a formula $\phi \in \text{Lang}(\Delta)$, we can query $\phi$ wrt evidence $e$ for theory $(\Delta, \mathbf{w})$ using:

$$\text{Pr}(\phi | e, \Delta, \mathbf{w}) = \frac{\text{Pr}(\phi \land e, \Delta, \mathbf{w})}{\text{WMC}(\phi \land e \land \Delta, \mathbf{w})}$$

When $e = \text{true}$, we simply write $\text{Pr}(\phi, \Delta, \mathbf{w})$. We remark for $\text{Pr}(\phi, \Delta, \mathbf{w})$ to be well-defined, which is assumed, WMC($\Delta, \mathbf{w}) \neq 0$. (Thus, it is assumed that $\Delta$ is satisfiable, and that $\mathbf{w}$ does not map all the corresponding literals to 0.) If the context is clear, we often refer to $\Delta$ as the theory, and to $\phi$ as the query or event.

We immediately observe the following property from the definition of WMC. (Proofs are provided in the supplementary material.)

**Theorem 2.** If $\Delta \models \phi$, then $\text{Pr}(\phi, \Delta, \mathbf{w}) = 1$. If $\Delta \not\models \phi$ is not satisfiable, then $\text{Pr}(\phi, \Delta, \mathbf{w}) = 0$.

**Example 3.** We illustrate a WMC encoding for $\mathcal{U}_h$ based on the university PRM; the encoding for others considered in this work are analogous. First, note that in atoms such as diff($x$, $y$), the logical variable $y$ captures the possible values of a random variable. Thus, they are to behave like logical functions. Formally, let $\mathcal{U}_h$ be the union of the following, the free variables being implicitly universally quantified from the outside:

- $\text{diff}(y, E) \lor \text{diff}(y, M) \lor \text{diff}(y, H)$
- $f_1(x, y, u) \equiv [\text{iq}(x, L) \land \text{diff}(y, E) \land \text{takes}(x, y) \lor \text{grades}(x, y, u)]$ for $u \in \{7, 8, 9, 10\}$
- $f_2(x, y, u) \equiv [\text{iq}(x, L) \land \neg \text{diff}(y, E) \land \text{takes}(x, y) \lor \text{grades}(x, y, u)]$ for $u \in \{5, 6, 7, 8\}$

The reason we need to introduce auxiliary predicates $f_1$ and $f_2$ is because WMC only allows weights on (ground) literals.

We also need the following hard constraints for capturing the logical functions:

$$\exists u(\text{diff}(y, u), \text{diff}(y, u) \land \text{diff}(y, v) \lor u = v)$$

$$\exists u(\text{grades}(x, y, u), \text{grades}(x, y, u) \land \text{grades}(x, y, v) \lor u = v)$$

Suppose the domain of quantification for the students is only $\{A\}$ and for courses is only $\{B\}$. We then obtain atoms such as:

$$\text{diff}(B, E), \text{diff}(B, M), \text{diff}(B, H), \text{iq}(A, L), \text{diff}(B, E), \text{takes}(A, B), \text{grades}(A, B, 7), \ldots$$

with a weight function $w_l$ for positive atoms derived from the parametric specification in an obvious fashion:

$$w_l(\text{diff}(B, E)) = .7, \ldots, w_l(f_1(A, B, 7)) = .25, \ldots$$

We let the weight of a negated atom $w_l(\neg a)$ to be $1 - w_l(a)$. Moreover, the ground instances $f_1$ and $f_2$ obtain the weights discussed in the parameterized version. The weights of all atoms not mentioning predicates diff, $f_1, f_2$ is taken to be 1. It then follows that $\text{Pr}(\text{diff}(B, E), \mathcal{U}_h, w) = .7$, and $\text{Pr}(\text{grades}(A, B, 7) | e, \mathcal{U}_h, w) = .25$, where $e = \text{takes}(A, B) \land \text{iq}(A, L) \land \text{diff}(A, E)$.

### 4 Abstraction Framework

We assume that the abstraction framework is realized in terms of two types of representations: a high-level/abstract theory that is mapped to a pre-existing low-level/concrete theory. Essentially, the logical symbols (predicates and constants) may differ arbitrarily between the two theories. In terms of notation, we use the subscript $h$ to refer to components of the high-level theory, and $l$ to refer to that of the low-level theory.

The first step is to formally establish the construct of a refinement mapping between the two theories: the mapping associates each high-level atom to a low-level formula, which may be arbitrarily complex.

**Definition 4.** Suppose $\Delta_h$ and $\Delta_l$ are two theories. We say $m$ is a refinement mapping from $\Delta_h$ to $\Delta_l$ iff for all high-
level atoms \(P(\vec{c}) \in \text{Lang}(\Delta_b), m(P(\vec{c})) = \theta_{P(\vec{c})}\) for some \(\theta_{P(\vec{c})} \in \text{Lang}(\Delta_l)\).

The mapping \(m\) is assumed to extend to complex formulas \(\phi \in \text{Lang}(\Delta_l)\) inductively: for atoms \(\phi = P(\vec{c})\), \(m(\phi)\) is as above; \(m(\neg \phi) = \neg m(\phi)\); \(m(\phi \land \psi) = m(\phi) \land m(\psi)\).

It is worth noting that a mapping is deliberately asymmetrical in the sense that its range need not include all the atoms of the low-level theory. That is, there may be atoms \(d \in \text{Lang}(\Delta_l)\), and consequently, also constants and relations, that do not appear in \(m(P(c))\) for every \(P(c) \in \text{Lang}(\Delta_h)\). After all, abstractions are about omitting irrelevant details.

In general, we will want to use these mappings to discuss model-theoretic properties of the two theories, so we introduce the notion of an isomorphism:

**Definition 5.** Given a refinement mapping \(m\) as above, we say that \(M_h \in \text{Models}(\Delta_h)\) is \(m\)-isomorphic to \(M_l \in \text{Models}(\Delta_l)\) i f f for all \(P(\vec{c}) \in \text{Lang}(\Delta_h)\), we have \(M_h \models P(\vec{c}) \iff M_l \models m(P(\vec{c}))\). We write this as \(M_h \sim_m M_l\).

Thus, isomorphism provides a way to align the truth values between high-level atom and low-level formulas. In particular, because of how refinement mappings can be defined for complex formulas, we obtain the following property:

**Theorem 6.** Suppose \(M_h \sim_m M_l\). Then for all \(\phi \in \text{Lang}(\Delta_h)\), \(M_h \models \phi \iff M_l \models m(\phi)\).

**Example 7.** For the university PRM, we provide a mapping \(m_U\) below. When free variables appear, we take it mean that the mapping applies to all substitutions. So, let \(m_U\) map \(\text{diff}(x, E)\), \(\text{takes}(x, y)\), \(i_q(x, l)\) from \(\mathcal{U}_l\) to the same atoms in \(\mathcal{U}_h\), \(m_U(\text{diff}(x, N)) = \text{diff}(x, M) \lor \text{diff}(x, H)\), \(m_U(\text{grades}(x, y, B)) = \text{grades}(x, y, 5) \lor \text{grades}(x, y, 6)\), \(m_U(\text{grades}(x, y, O)) = \text{grades}(x, y, 7) \lor \text{grades}(x, y, 8)\), and \(m_U(\text{grades}(x, y, G)) = \text{grades}(x, y, 9) \lor \text{grades}(x, y, 10)\).

Suppose the domain includes a single student \(A\), who takes course \(B\). Suppose \(M_h\) is a model of \(\mathcal{U}_h\) where \(\{i_q(A, L), \text{takes}(A, B), \text{diff}(B, E), \text{grades}(A, B, O)\}\) holds. Now consider the model \(M_l\) of \(\mathcal{U}_l\) where \(\{i_q(A, L), \text{takes}(A, B), \text{diff}(B, E), \text{grades}(A, B, 7)\}\) holds. It is easy to verify that \(M_h \sim_m M_l\), because the main question is whether \(M_l\) satisfies \(m_U(\text{grades}(A, B, O)) = \text{grades}(A, B, 7) \lor \text{grades}(A, B, 8)\), which it does.

---

In the following sections, we will discuss the properties of abstractions based on mappings and isomorphisms.

### 5 Unweighted Abstractions

To obtain intuitions about the properties of abstract models from first principles, we will consider a fundamental type of abstraction: the absence of probabilities. In so much as probabilistic assertions quantify the likelihood of worlds, omitting probabilities still informs us about the possible and the certain, thus allowing us to test whether \(\Delta_h\) is consistent with \(\Delta_l\).

**Definition 8.** Given a weighted theory \((\Delta, w)\), the unweighted setting refers to the case when for all \(P(\vec{c}) \in \text{Lang}(\Delta)\), we have \(w(P(\vec{c})) = w(\neg P(\vec{c})) = 1\).

Since probabilities do not occur in the setting, we can establish consistency by checking whether all conclusions by \(\Delta_h\) (that is, certain events) are also conclusions by \(\Delta_l\): in other words, are the conclusions sound? We define:

**Definition 9.** The theory \(\Delta_h\) is a sound abstraction of \(\Delta_l\) relative to refinement mapping \(m\) i f f for all \(M_l \in \text{Models}(\Delta_l)\), there is a \(M_h \in \text{Models}(\Delta_h)\) such that \(M_h \sim_m M_l\).

**Theorem 10.** Suppose \(\Delta_h\) is a sound abstraction of \(\Delta_l\) relative to \(m\). Then for all \(\phi \in \text{Lang}(\Delta_h)\): (a) if \(Pr(m(\phi), \Delta_l, w_l) > 0\) then \(Pr(\phi, \Delta_h, w_h) > 0\); and (b) if \(Pr(\phi, \Delta_h, w_h) = 1\) then \(Pr(m(\phi), \Delta_l, w_l) = 1\).

**Example 11.** It is easy to check that for the university PRM, \(U_h\) is a sound abstraction of \(U_l\) wrt \(m_U\).

It is fairly straightforward to construct trivially unsound abstractions. To see a less obvious example, consider \(C_i\) from before, and suppose it also included: \(CS(x) \lor \text{Programming}(x)\) and \(\text{Physics}(x) \lor \text{Fieldwork}(x)\). And as discussed, let \(C_h\) be a high-level theory with the predicate \(\text{Science}(x)\), but not \(CS(x)\) and \(\text{Physics}(x)\).

Suppose \(B\) is a CS-course. Suppose \(m_C\) is a mapping that replaces \(\text{Science}(x)\) by \(CS(x) \lor \text{Physics}(x)\), but maps every other predicate to itself. Then, we have \(Pr(\phi, C_h, w_h) = 1\) for \(\phi = \text{Science}(B) \land \text{Programming}(B) \land \text{Fieldwork}(B)\), whereas, \(Pr(m(\phi), C_l, w_l) \neq 1\), because there will be possible worlds where \(CS(B) \land \neg \text{Fieldwork}(B)\).

Sound abstractions ascertain that conclusions by \(\Delta_h\) are consistent with \(\Delta_l\). What about events considered possible

---

3 When the high-level and low-level theories are defined over the same domain of discourse \(D\), \(m\) can have a compact specification of the form \(m(P(\vec{x})) = \theta_{P(\vec{x})}\), where \(P(\vec{x})\) is a non-ground predicate, and \(\vec{x}\) are the only free variables in \(\theta_P\). So effectively the mapping works by substitutions: for every instance \(P(\vec{c})\), we have \(m(P(\vec{c})) = \theta_{P(\vec{c})}\).

4 Thus, this section can be seen to establish a framework for abstraction in classical (unweighted) model counting.
by $\Delta_h$? Because we are omitting information when constructing an abstract model, it may be that $\Delta_h$ entertains an event as possible even though $\Delta_l$ does not.

**Definition 12.** The theory $\Delta_h$ is a complete abstraction of $\Delta_l$ relative to $m$ iff for all $M_h \in \text{Models}(\Delta_h)$, there is a $M_l \in \text{Models}(\Delta_l)$ such that $M_h \sim_m M_l$.

**Theorem 13.** Suppose $\Delta_h$ is a complete abstraction of $\Delta_l$ relative to $m$. Then for all $\phi \in \text{Lang}(\Delta_h)$:
(i) if $\Pr(\phi, \Delta_h, w_h) > 0$ then $\Pr(m(\phi), \Delta_l, w_l) > 0$; and (ii) if $\Pr(m(\phi), \Delta_l, w_l) = 1$ then $\Pr(\phi, \Delta_h, w_h) = 1$.

Put differently, an exhaustive (but perhaps impractical) way to verify whether $\Delta_h$ is a sound (or complete) abstraction is to verify that the properties discussed in Theorem 10 (or 13, respectively) hold.

**Example 14.** The university PRM can be seen as a complete abstraction wrt $m_{Q_l}$.

To see a case where it is not complete, consider a variant high-level theory $\mathcal{U}'_l$ where we ignore the difficulty of courses and have only one rule: $\text{iq}(x, L) \land \text{takes}(x, y) \supset \text{grades}(x, y, u)$ where $u \in \{B, O, G\}$. Suppose the low-level theory is $\mathcal{U}'_l = \text{diff}(B, H) \land \mathcal{U}_l$, and $\Delta$ is a low-IQ student who takes $B$. It is easy to see that $\Pr(\phi, \mathcal{U}'_l, w_h) > 0$ for $\phi = \text{iq}(A, L) \land \text{takes}(A, B) \land \text{grades}(A, B, G)$, because $\mathcal{U}'_l$ says that any of the three grades levels are possible. But clearly, $B$ being a hard course means that $\text{diff}(B, H) \land m_{Q_l}(\phi)$ cannot be satisfiable, and so it is a zero-probability event wrt $\mathcal{U}'_l$.

**Definition 15.** The theory $\Delta_h$ is a sound and complete abstraction of $\Delta_l$ relative to $m$ iff $\Delta_h$ is both a sound and a complete abstraction of $\Delta_l$ relative to $m$.

**Theorem 16.** Suppose $\Delta_h$ is a sound and complete abstraction of $\Delta_l$ relative to $m$. Then for every $\phi \in \text{Lang}(\Delta_h)$, (a) $\Pr(\phi, \Delta_h, w_h) > 0$ if $\Pr(m(\phi), \Delta_l, w_l) > 0$; and (b) $\Pr(\phi, \Delta_h, w_h) = 1$ iff $\Pr(m(\phi), \Delta_l, w_l) = 1$.

## 6 Weighted Abstractions

Clearly the above theorems would not hold in general when considering non-trivial weights. It is easy to imagine a weight function that redistributes weights such that zero probability events in $\Delta_l$ have high probabilities in $\Delta_h$, and vice versa. So, outside the case of probabilities mapping exactly between $\Delta_h$ and $\Delta_l$ (discussed in the next section), we need to understand how to abstract weighted theories. The previous section provided a recipe for abstractions, from which properties discussed in Theorems 10 and 13 followed. To a first approximation, then, we can motivate a definition for weighted abstractions by requiring that those properties hold categorically, in the form of constraints. But it turns out, we can do better. We can show that if the property about probable events hold as a constraint wrt a sound or complete abstraction, then the corresponding property about certain events follows as a consequence. (Recall that this duality is not about an event and its negation, which would follow from the axioms of probability, but about how the high-level and low-level theories align.)

To prepare for this approach, let us begin with a few properties that follow from the axioms of probability [Fagin and Halpern, 1994], but are established here using WMC:

**Theorem 17.** Suppose $(\Delta, w)$ is a weighted theory. Then the following hold for all $\phi, \psi \in \text{Lang}(\Delta)$:
1. If $\Delta \models \phi$ then $\Pr(\phi, \Delta, w) = 1$.
2. If $\phi \land \Delta$ is not satisfiable, then $\Pr(\phi, \Delta, w) = 0$.
3. $\Pr(\neg \phi, \Delta, w) = 1 - \Pr(\phi, \Delta, w)$.
4. $\Pr(\phi \lor \psi, \Delta, w) = \Pr(\phi, \Delta, w) + \Pr(\psi, \Delta, w) - \Pr(\phi \land \psi, \Delta, w)$.
5. If $\Pr(\phi, \Delta, w) = 0$ then $\Pr(\phi \land \phi, \Delta, w) = 0$.
6. If $\Pr(\phi, \Delta, w) > 0$ then $\Pr(\phi \lor \psi, \Delta, w) > 0$.
7. $\Pr(\phi, \Delta, w) \geq \Pr(\phi \land \psi, \Delta, w)$.

**Definition 18.** The theory $(\Delta_h, w_h)$ is a weighted sound abstraction of $(\Delta_l, w_l)$ relative to refinement mapping $m$ if $\Delta_h$ is a sound abstraction of $\Delta_l$ relative to $m$, and for all $d \in \text{Lits}(\Delta_h)$, if $\Pr(m(d), \Delta_l, w_l) > 0$ then $\Pr(d, \Delta_h, w_h) > 0$.

We will now that this stipulation at the level of literals immediately implies the validity of the constraint for all formulas:

**Theorem 19.** Suppose $(\Delta_h, w_h)$ is a weighted sound abstraction of $(\Delta_l, w_l)$ relative to $m$. Then for all $\phi \in \text{Lang}(\Delta_h)$, if $\Pr(m(\phi), \Delta_l, w_l) > 0$ then $\Pr(\phi, \Delta_h, w_h) > 0$.

The key result of this definition is that the property on certain events, seen in Theorem 10, follows as a consequence:

**Theorem 20.** Suppose $(\Delta_h, w_h)$ is a weighted sound abstraction of $(\Delta_l, w_l)$ relative to $m$. Then for all $\phi \in \text{Lang}(\Delta_h)$, if $\Pr(\phi, \Delta_h, w_h) = 1$ then $\Pr(m(\phi), \Delta_l, w_l) = 1$. 
Example 21. The university PRM can be seen to be a weighted sound abstraction wrt $m_{QL}$. Consider the university PRM with a variant high-level theory $U''_h$, where the third constraint is the following instead:

\[ iq(x, L) \land diff(x, E) \land takes(x, y) \supset grades(x, y, G) \]

Consider the query $\phi = iq(A, L) \land diff(B, E) \land takes(A, B) \supset grades(A, B, O)$. Clearly, the low-level theory accords a non-zero probability to $m_{QL}(\phi)$, but because of the third constraint, $U''_h$ accords a zero probability to $\phi$. Thus, this is not a sound weighted abstraction.

Following these results, extending complete abstractions as well as sound and complete abstractions is analogous, which we state here for the sake of completeness. (The proofs are also analogous and hence omitted.)

Definition 22. The theory $(\Delta_h, w_h)$ is a weighted complete abstraction of $(\Delta_l, w_l)$ relative to refinement mapping $m$ iff $\Delta_h$ is a complete abstraction of $\Delta_l$ relative to $m$, and for all $d \in Lits(\Delta_h)$, if $Pr(d, \Delta_h, w_h) > 0$ then $Pr(m(d), \Delta_l, w_l) > 0$.

Theorem 23. Suppose $(\Delta_h, w_h)$ is a weighted complete abstraction of $(\Delta_l, w_l)$ relative to $m$. Then for all $\phi \in Lang(\Delta_h)$, if $Pr(m(\phi), \Delta_h, w_h) = 1$ then $Pr(\phi, \Delta_h, w_h) = 1$.

Example 24. The university PRM can be seen to be a weighted complete abstraction wrt $m_{QL}$.

Example 15 also applies as an instance of an abstraction that is not weighted complete via:

\[ \exists x \exists y \exists u (iq(x, L) \land takes(x, y) \supset grades(x, y, u)) \]

Mainly because the difficulty of courses is ignored, an event is considered probable by the high-level theory but not by the low-level one.

Definition 25. The theory $(\Delta_h, w_h)$ is a weighted sound and complete abstraction of $(\Delta_l, w_l)$ relative to refinement mapping $m$ iff it is both a weighted sound and a weighted complete abstraction.

Theorem 26. Suppose $(\Delta_h, w_h)$ is a weighted sound and complete abstraction of $(\Delta_l, w_l)$ relative to $m$. Then for all $\phi \in Lang(\Delta_h)$, $Pr(m(\phi), \Delta_l, w_l) = 1$ iff $Pr(\phi, \Delta_h, w_h) = 1$.

7 Exact Abstractions

The most faithful case of aligning the high-level and low-level theories is when the probabilities coincide for all high-level queries.

Definition 27. The theory $(\Delta_h, w_h)$ is a weighted exact abstraction of $(\Delta_l, w_l)$ relative to refinement mapping $m$ iff for all literals $d \in Lits(\Delta_h)$, $Pr(d, \Delta_h, w_h) = Pr(m(d), \Delta_l, w_l)$.

The key property here is that it suffices for the modeler to ascertain the alignment for literals, after which it follows for all formulas.

Theorem 28. Suppose $(\Delta_h, w_h)$ is a weighted exact abstraction of $(\Delta_l, w_l)$ relative to $m$. Then for all $\phi \in Lang(\Delta_h)$, $Pr(\phi, \Delta_h, w_h) = Pr(m(\phi), \Delta_l, w_l)$.

Example 29. The university PRM can be seen to be an instance of a weighted exact abstraction wrt $m_{QL}$.

In contrast, the variant in Example 14 does not belong to this type because the high-level theory accords a probability of 1/3 to a low-IQ student taking a difficult course and still getting a good grade, whereas the low-level theory considers that improbable.

8 Towards Abstracting Evidence

Recall that we can query $\phi$ wrt evidence $e$ for theory $(\Delta, w)$ using \[(\phi).\] We assumed so far that $\phi, e \in Lang(\Delta)$. However, in many applications needing abstraction, it is often the case that observations are low-level (e.g., readings on sensor), whereas the query is at the high-level (e.g., interactions with user). In this section, we discuss some ways to reconcile this issue.

Consider low-level evidence $e \in Lits(\Delta_l)$. For simplicity, let $e$ be a literal. Without loss of generality, let mappings be in conjunctive normal form (CNF). We say a literal is pure in a CNF $\theta$ if its complement does not appear in $\theta$. (E.g., $p$ is pure in $p \lor q$ but not in $\neg p \lor q$; in contrast, $\neg p$ is pure in the latter but not the former.) We observe that, by construction, there may be many high-level atoms that map to formulas involving $e$. So, given a mapping $m$, let us retrieve these by concretization:

\[ m^{-1}(e) = \{ P(\vec{c}) \in Lang(\Delta_h) | e \text{ is mentioned \& pure in } m(P(\vec{c})) \}. \]

(That is, $m(P(\vec{c}))$ is a CNF formula.) Here, $m^{-1}(e)$ is equivalently expressed as a formula: $\lor P_i(\vec{c}_i)$. The idea is that by looking at high-level atoms where $e$ is pure under

\footnote{The distribution on the high-level theory is essentially a “push-forward” measure [Trench, 2003].}

\footnote{It is conceivable that there may be other approaches for this reconciliation, and in our inquiry as well, it will become clear that a number of variants present themselves. We also limit the discussion to exact abstractions for simplicity.}
the mapping, we are essentially finding atoms that agree with the evidence (and not its negation).

We can now retrieve all low-level sentences these map to by re-applying m as follows: \( m(m^{-1}(e)) = \bigvee m(P_i(z_i)) \). (It is easy to see that \( e \) will remain pure in \( m(m^{-1}(e)) \).

An immediate case, then, of conditioning being straightforward is when \( e = m(m^{-1}(e)) \):

**Theorem 30.** Suppose \( (\Delta_h, w_h) \) is a weighted exact abstraction of \( (\Delta_l, w_l) \) relative to \( m \). Suppose \( e \in \text{Lits}(\Delta_l) \) and \( e = m(m^{-1}(e)) \). Then for any \( \phi \in \text{Lang}(\Delta_h) \),

\[
\Pr(\phi | m^{-1}(e), \Delta_h, w_h) = \Pr(m(\phi) | e, \Delta_l, w_l).
\]

A simple example is the case of \( \text{diff}(x, E) \) in the university PRM, as it was mapped to the same atom at both levels.

But beyond this simple case, it is not always possible to reason about low-level events in an exact manner at the high-level. Indeed, as mentioned before, omitting details is the very goal of abstraction. For example, in the university PRM, given any course \( B \), \( \Pr(\text{diff}(B, M), \mathcal{U}_h, w_h) = .1 \), but clearly there is no way to syntactically arrange \( \text{diff}(B, E), \text{diff}(B, N) \) in \( \mathcal{U}_h \) to obtain that number. Of course, it would not be hard to show a more involved property, such as \( \Pr(\text{diff}(B, N), \mathcal{U}_h, w_h) \geq \Pr(\text{diff}(B, M), \mathcal{U}_h, w_h) \).

Rather than treating such properties, we will consider the case where probabilities can correspond exactly. Then, one way to incorporate low-level evidence is to weaken it, in the sense that conditioning wrt the low-level theory would suffer from a loss in detail, which is precisely the problem faced by the high-level theory. We may think of using \( m(m^{-1}(e)) \), for example. However, that is not sufficient for conditioning to be correct, because \( m(m^{-1}(e)) \) can say more and less than \( e \). For example, in the university PRM, suppose we have evidence \( e = \text{diff}(B, M) \) for \( \mathcal{U}_h \). So \( m^{q_1^{-1}}(e) = \text{diff}(B, N) \), and \( m^{q_2^{-1}}(m^{q_1^{-1}}(e)) = \text{diff}(B, M) \vee \text{diff}(B, H) \), which is saying less than \( e \). This is reasonable. But suppose for the sake of the argument, \( m^{q_2^{-1}}(m^{q_1^{-1}}(e)) = (\text{diff}(B, M) \vee \text{diff}(B, H)) \land \text{diff}(C, H) \).

(This is somewhat artificial but well-defined.) The problem is that \( e \) does not imply anything about the difficulty of course \( C \). Thus, if we use \( m^{q_2^{-1}}(m^{q_1^{-1}}(e)) \) as evidence, we will be falsely assuming facts that were not observed.

To get around this, we stipulate this implication formally:

**Definition 31.** Given evidence \( e \) and mapping \( m \), we define the \( m \)-weakening of \( e \) as \( m(m^{-1}(e)) \). It is definable iff \( e \models m(m^{-1}(e)) \).

The most obvious (and reasonable) case where definability follows is when \( m(m^{-1}(e)) \) is a clause, that is, a disjunction of literals. Because \( e \) is pure in \( m(m^{-1}(e)) \), it immediately follows that \( e \models m(m^{-1}(e)) \). (E.g., \( p \) is pure in \( p \vee q \), and of course \( p \models p \vee q \).

**Theorem 32.** Suppose \( (\Delta_h, w_h) \) is a weighted exact abstraction of \( (\Delta_l, w_l) \) relative to \( m \). Suppose \( e \in \text{Lits}(\Delta_l) \) and its \( m \)-weakening is definable. Then, \( \Pr(\phi | m^{-1}(e), \Delta_h, w_h) = \Pr(m(\phi) | m(m^{-1}(e)), \Delta_l, w_l) \).

**Example 33.** For the university PRM and \( e = \text{diff}(B, M) \), its \( m_{q_1} \)-weakening is \( \text{diff}(B, M) \vee \text{diff}(B, H) \). And indeed, \( e \models m_{q_1}(m^{-1}_{q_1}(e)) \). For the query \( \phi = \text{iq}(A, L) \land \text{takes}(A, B) \land \text{grades}(A, B, O) \), its probability given \( m_{q_1}^{-1}(e) = \text{diff}(B, N) \) at the high-level coincides with the probability of \( \text{iq}(A, L) \land \text{takes}(A, B) \land (\text{grades}(A, B, 7) \lor \text{grades}(A, B, 8)) \) given \( m_{q_1}(m_{q_1}^{-1}(e)) \) at the low-level.

## 9 Discussions

Abstraction is a major topic in knowledge representation [Giunchiglia and Walsh, 1992; Erol et al., 1996; Saitta and Zucker, 2013; Banerhashemi et al., 2017]. The idea of establishing mappings between models to yield a semantic theory for abstraction owes its origin to works such as [Milner, 1989]. But formal treatments have been mostly restricted to categorical and non-probabilistic features. Our work is inspired by this body of work, but generalizes it to develop a mathematical framework for stochastic primitives in a rich, relational language.

In the area of program verification, static analysis and abstraction techniques are commonplace to test the correctness of programs and probabilistic programs. For example, [Sharma et al., 2013] consider statistical properties of program behavior to advocate abstractions. [Zhang et al., 2017] consider verification to the learnability of concepts, [Holzen et al., 2017] study abstract predicates for loop-free probabilistic programs, and [Monniaux, 2001] defines abstract representations for probabilistic program path analysis. A number of additional concerns present themselves in a programmatic setting, including branching in the presence of stochastic primitives, and stochastic transitions between program states. The motivation then is to deduce sound abstractions for verifying correctness (e.g., termination) properties. While some of these works do not consider abstractions themselves to be probabilistic, the developments are related to our goals. Thus, it would be interesting to know how ideas and techniques from the program analysis literature can be carry over to our framework, and vice versa. Abstraction is also a long-standing concern in causal modeling. In recent work, [Rubenstein et al., 2017] study con-
sistency between what they view as micro and macro-level variables via structural equation models.

To summarize, we were motivated in the development of a framework for abstractions in PRMs, based on isomorphisms between models, where atoms in a high-level theory can be mapped to complex formulas at the low-level, thereby exploiting the first-order expressivity. From that, we developed a number of accounts of abstraction, in case exact alignments are not feasible, and how one could handle low-level evidence. Our account was based on WMC, which also serves as an algorithmic account of how to compute and compare probabilities. Given the increasing interest in abstraction for probabilistic planning and statistical explainable AI, we hope our framework is helpful in developing probabilistic abstractions for increased clarity, modularity and tractability.

References

[Bacchus et al., 2009] Bacchus, F., Dalmao, S., and Pitassi, T. (2009). Solving #SAT and Bayesian inference with backtracking search. *J. Artif. Intell. Res. (JAIR)*, 34:391–442.

[Banihashemi et al., 2017] Banihashemi, B., De Giacomomo, G., and Lespérance, Y. (2017). Abstraction in situation calculus action theories. In *AAAI*, pages 1048–1055.

[Boutilier et al., 1996] Boutilier, C., Friedman, N., Goldszmidt, M., and Koller, D. (1996). Context-specific independence in bayesian networks. In *UAI*, pages 1048–1055.

[Boutillier et al., 1996] Boutillier, C., Friedman, N., Goldszmidt, M., and Koller, D. (1996). Context-specific independence in bayesian networks. In *AAAI*, pages 1048–1055.

[Darwiche and Marquis, 2002] Darwiche, A. and Marquis, P. (2002). A knowledge compilation map. *Journal of Artificial Intelligence Research*, 17:229–264.

[Dedre and Christian, 2017] Dedre, G. and Christian, H. (2017). Analogy and abstraction. *Topics in Cognitive Science*, 9(3):672–693.

[Erol et al., 1996] Erol, K., Hendler, J., and Nau, D. S. (1996). Complexity results for htn planning. *Annals of Mathematics and Artificial Intelligence*, 18(1):69–93.

[Fagin and Halpern, 1994] Fagin, R. and Halpern, J. Y. (1994). Reasoning about knowledge and probability. *J. ACM*, 41(2):340–367.

[Fierens et al., 2011] Fierens, D., Van den Broeck, G., Thon, I., Gutmann, B., and De Raedt, L. (2011). Inference in probabilistic logic programs using weighted CNF’s. In *Proceedings of UAI*, pages 211–220.

[Garfinkel, 1981] Garfinkel, A. (1981). *Forms of explanation: Rethinking the questions in social theory*. Yale University Press New Haven.

[Getoor and Taskar, 2007] Getoor, L. and Taskar, B., editors (2007). *An Introduction to Statistical Relational Learning*. MIT Press.

[Giunchiglia and Walsh, 1992] Giunchiglia, F. and Walsh, T. (1992). A theory of abstraction. *Artificial intelligence*, 57(2-3):323–389.

[Gomes et al., 2009] Gomes, C. P., Sabharwal, A., and Selman, B. (2009). Model counting. In *Handbook of Satisfiability*. IOS Press.

[Gunning, 2016] Gunning, D. (2016). Explainable artificial intelligence (xai). Technical report, DARPA/I2O.

[Heckerman et al., 2004] Heckerman, D., Meek, C., and Koller, D. (2004). Probabilistic models for relational data. Technical report, Technical Report MSR-TR-2004-30, Microsoft Research.

[Holtzen et al., 2017] Holtzen, S., Millstein, T., and Van den Broeck, G. (2017). Probabilistic program abstractions. In *Proceedings of the 33rd Conference on Uncertainty in Artificial Intelligence (UAI)*.

[Jorland, 1994] Jorland, G. (1994). Idealization and transformation. *Idealization VI: Idealization in economics*, pages 265–275.

[Milner, 1989] Milner, R. (1989). *Communication and Concurrency. International series in computer science*. Prentice hall Englewood Cliffs.

[Monniaux, 2001] Monniaux, D. (2001). An abstract monte-carlo method for the analysis of probabilistic programs. In *ACM SIGPLAN Notices*, volume 36, pages 93–101. ACM.

[Penkov and Ramamoorthy, 2017] Penkov, S. and Ramamoorthy, S. (2017). Explaining transition systems through program induction. *CoRR*, abs/1705.08320.

[Richardson and Domingos, 2006] Richardson, M. and Domingos, P. (2006). Markov logic networks. *Machine learning*, 62(1):107–136.

[Rubenstein et al., 2017] Rubenstein, P. K., Weichwald, S., Bongers, S., Mooij, J. M., Janzing, D., Grosse-Wentrup, M., and Schölkopf, B. (2017). Causal consistency of structural equation models. *arXiv preprint arXiv:1707.00819*. 

[Garfinkel, 1981] Garfinkel, A. (1981). *Forms of explanation: Rethinking the questions in social theory*. Yale University Press New Haven.

[Getoor and Taskar, 2007] Getoor, L. and Taskar, B., editors (2007). *An Introduction to Statistical Relational Learning*. MIT Press.

[Giunchiglia and Walsh, 1992] Giunchiglia, F. and Walsh, T. (1992). A theory of abstraction. *Artificial intelligence*, 57(2-3):323–389.

[Gomes et al., 2009] Gomes, C. P., Sabharwal, A., and Selman, B. (2009). Model counting. In *Handbook of Satisfiability*. IOS Press.

[Gunning, 2016] Gunning, D. (2016). Explainable artificial intelligence (xai). Technical report, DARPA/I2O.

[Heckerman et al., 2004] Heckerman, D., Meek, C., and Koller, D. (2004). Probabilistic models for relational data. Technical report, Technical Report MSR-TR-2004-30, Microsoft Research.

[Holtzen et al., 2017] Holtzen, S., Millstein, T., and Van den Broeck, G. (2017). Probabilistic program abstractions. In *Proceedings of the 33rd Conference on Uncertainty in Artificial Intelligence (UAI)*.

[Jorland, 1994] Jorland, G. (1994). Idealization and transformation. *Idealization VI: Idealization in economics*, pages 265–275.

[Milner, 1989] Milner, R. (1989). *Communication and Concurrency. International series in computer science*. Prentice hall Englewood Cliffs.

[Monniaux, 2001] Monniaux, D. (2001). An abstract monte-carlo method for the analysis of probabilistic programs. In *ACM SIGPLAN Notices*, volume 36, pages 93–101. ACM.

[Penkov and Ramamoorthy, 2017] Penkov, S. and Ramamoorthy, S. (2017). Explaining transition systems through program induction. *CoRR*, abs/1705.08320.

[Richardson and Domingos, 2006] Richardson, M. and Domingos, P. (2006). Markov logic networks. *Machine learning*, 62(1):107–136.

[Rubenstein et al., 2017] Rubenstein, P. K., Weichwald, S., Bongers, S., Mooij, J. M., Janzing, D., Grosse-Wentrup, M., and Schölkopf, B. (2017). Causal consistency of structural equation models. *arXiv preprint arXiv:1707.00819*. 

[Garfinkel, 1981] Garfinkel, A. (1981). *Forms of explanation: Rethinking the questions in social theory*. Yale University Press New Haven.
[Saitta and Zucker, 2013] Saitta, L. and Zucker, J.-D. (2013). Abstraction in artificial intelligence and complex systems, volume 456. Springer.

[Sharma et al., 2013] Sharma, R., Gupta, S., Hariharan, B., Aiken, A., and Nori, A. V. (2013). Verification as learning geometric concepts. In International Static Analysis Symposium, pages 388–411. Springer.

[Sreedharan et al., 2018] Sreedharan, S., Srivastava, S., and Kambhampati, S. (2018). Hierarchical expertise level modeling for user specific contrastive explanations. In IJCAI, pages 4829–4836.

[Suciu et al., 2011] Suciu, D., Olteanu, D., Ré, C., and Koch, C. (2011). Probabilistic databases. Synthesis Lectures on Data Management, 3(2):1–180.

[Trench, 2003] Trench, W. F. (2003). Introduction to real analysis. Prentice Hall.

[Zhang et al., 2017] Zhang, X., Si, X., and Naik, M. (2017). Combining the logical and the probabilistic in program analysis. In Proceedings of the 1st ACM SIGPLAN International Workshop on Machine Learning and Programming Languages, MAPL 2017, pages 27–34, New York, NY, USA. ACM.
Appendix: Proofs

Proof of Theorem 2. For the first property, every $M$ such that $M \models \Delta$, $M \models \phi$ also, and so $\text{WMC}(\phi \land \Delta, w) = \text{WMC}(\Delta, w)$. For the second, $\text{WMC}(\phi \land \Delta, w) = 0$.

Proof of Theorem 6. We prove by induction on $\phi$. Base case immediate by definition. Negation: $M_h \models \neg \phi$ iff $M_h \not\models \phi$ iff (by hypothesis) $M_l \not\models m(\phi)$ iff (by semantics) $M_l \models \neg m(\phi)$ iff (by definition) $M_l \models m(\neg \phi)$. Conjunction: $M_h \models \phi \land \psi$ iff $M_h \models \phi$ and $M_h \models \psi$ (by hypothesis) $M_l \models m(\phi)$ and $M_l \models m(\psi)$ iff (by semantics) $M_l \models m(\phi) \land m(\psi)$ iff (by definition) $M_l \models m(\phi \land \psi)$.

Proof of Theorem 10. For (a), suppose the antecedent holds, which means there is an $M_l \in \text{Models}(\Delta_l)$ such that $M_l \models m(\phi)$. By assumption, there is a $M_l \in \text{Models}(\Delta_l)$ such that $M_l \models m(\phi)$, and $\Pr(\phi, \Delta, w_h) \neq 0$. By Theorem 17, it must be that it must be smaller or equal to the other probabilities. (That is, if $a, b, c > 0, c \leq a$ and $c \leq b$, then $a + b - c > 0$.) So, $\Pr(\phi \land \Delta, w_h) > 0$.

Proof of Theorem 19. By induction on $\phi$. The case of atoms and negations is immediate by definition. So we only need an argument for disjunctions. Suppose $\Pr(m(\phi \lor \psi), \Delta, w) > 0$, that is, $\Pr(\phi, \Delta, w_h) > 0$, and $\Pr(\psi, \Delta, w_h) > 0$, and therefore $\Pr(\phi \lor \psi, \Delta, w_h) > 0$.

Proof of Theorem 20. Suppose antecedent but not consequent. Then there is some $M_l \in \text{Models}(\Delta_l)$ such that $M_l \models m(\phi)$ and it has non-zero weight. (If all such $M_l$ have zero weight, then the consequent cannot be falsified because these models do not influence the probability.) By assumption, there is a $M_h \in \text{Models}(\Delta_h)$ such that $M_h \not\models m_l M_l$, and so $M_h \not\models \phi$.

There are now two cases, depending on the weight of the model $M_h$. (And so the proof deviates from that for Theorem 10.)

Case $w_h(M_h) > 0$: The proof follows as in Theorem 19 yielding a contradiction.

Case $w_h(M_h) = 0$: Let $M_1^h$ be a formula denoting the conjunction of the literals true at $M_l$. (Since there are finitely many atoms, such a formula can be obtained.) Because $M_h \models m_l M_l$, $M_l \models m(M_1^h)$. Overloading the notation $M^1$ to mean conjunction and set of literals true at $M_l$, $M^1(M_1^h) \subseteq M^1$, the latter being the set of literals true at $M_l$. But by assumption $M_h$ has non-zero weight, which means $\Pr(M^1(M_1^h), \Delta, w_l) > 0$. It follows that $\Pr(m(M_1^h), \Delta, w_l) > 0$, because otherwise Theorem 17(5) would be contradicted. By Theorem 19, $\Pr(M^1_h, \Delta_h, w_h) > 0$, and so $w_h(M_h) > 0$. Contradiction.
Let $\phi_i$ be $\phi$ but with every occurrence of $d$ replaced by $i$ and every occurrence of $\neg d$ replaced by $1 - i$. Because $\phi_i$ eliminates at least one literal from $\phi$, its length is $\leq k$ and by hypothesis, $\Pr(\phi_i, \Delta_h, w_h) = \Pr(m(\phi_i), \Delta_l, w_l)$.

From Theorem 17 we know that $\Pr(\phi_0 \lor \phi_1) = \Pr(\phi_0) + \Pr(\phi_1) - \Pr(\phi_0 \land \phi_1)$. However, observe that $\phi_0 \land \phi_1 = (\phi \land d = 1) \land (\phi \land d = 0)$, which is inconsistent, so has probability 0. Thus, $\Pr(\phi, \Delta_h, w_h) = \Pr(\phi_0, \Delta_h, w_h) + \Pr(\phi_1, \Delta_h, w_h)$. Analogously, it is not hard to show that $\Pr(m(\phi_0) \lor m(\phi_1), \Delta_l, w_l)$, that is, $\Pr(m(\phi), \Delta_l, w_l) = \Pr(m(\phi_0), \Delta_l, w_l) + \Pr(m(\phi_1), \Delta_l, w_l)$, and therefore, $\Pr(\phi, \Delta_h, w_h) = \Pr(m(\phi), \Delta_l, w_l)$.

**Proof of Theorem 30.** By assumption, the probability of $\phi \land m^{-1}(e)$ wrt $\Delta_h$ must be the same as that of $m(\phi) \land m(m^{-1}(e))$ at the low-level.

**Proof of Theorem 32.** Proof analogous to Theorem 30.