Towards a Computationally Viable Framework for Semantic Representation

Shalom Lappin
University of Gothenburg

Symposium on Logic and Algorithms in Computational Linguistics 2018
University of Stockholm

August 29, 2018

Outline

Possible Worlds in Formal Semantics

- Since Montague (1974) a mainstream view among formal semanticists has depended on possible worlds to model the meanings of natural language expressions.
- Montague imported possible worlds into his model theory through his use of Kripke frame semantics (Kripke(1959,1963)) for modal logic.
- This approach is anticipated in Carnap's (1947) characterisation of intensions as functions from state descriptions to extensions.

Possible Worlds in Formal Semantics

- Since Montague (1974) a mainstream view among formal semanticists has depended on possible worlds to model the meanings of natural language expressions.
- Montague imported possible worlds into his model theory through his use of Kripke frame semantics (Kripke(1959,1963)) for modal logic.
- This approach is anticipated in Carnap's (1947) characterisation of intensions as functions from state descriptions to extensions.

Possible Worlds in Formal Semantics

- Since Montague (1974) a mainstream view among formal semanticists has depended on possible worlds to model the meanings of natural language expressions.
- Montague imported possible worlds into his model theory through his use of Kripke frame semantics (Kripke(1959,1963)) for modal logic.
- This approach is anticipated in Carnap's (1947) characterisation of intensions as functions from state descriptions to extensions.

Possible Worlds in Epistemic Reasoning

- Kripke frame semantics has also been influential in the related field of epistemic reasoning (Halperin (1995)).
- More recent formal semantic approaches, such as Dynamic semantics (Groenendijk and Stokhof (1990, 1991)), and Inquisitive Semantics (Ciardelli, Groenendijk, and Roelofsen (2013), Ciardelli, Roelofsen, and Theiler (2017)) use possible worlds to incorporate epistemic elements into formal semantics.
- They characterise sentence meanings as functions from discourse contexts to discourse contexts.
- Speakers use sentences to communicate information by modifying their hearers' representation of a discourse context.

Possible Worlds in Epistemic Reasoning

- Kripke frame semantics has also been influential in the related field of epistemic reasoning (Halperin (1995)).
- More recent formal semantic approaches, such as Dynamic semantics (Groenendijk and Stokhof (1990, 1991)), and Inquisitive Semantics (Ciardelli, Groenendijk, and Roelofsen (2013), Ciardelli, Roelofsen, and Theiler (2017)) use possible worlds to incorporate epistemic elements into formal semantics.
- They characterise sentence meanings as functions from discourse contexts to discourse contexts.
- Speakers use sentences to communicate information by modifying their hearers' representation of a discourse context.

Possible Worlds in Epistemic Reasoning

- Kripke frame semantics has also been influential in the related field of epistemic reasoning (Halperin (1995)).
- More recent formal semantic approaches, such as Dynamic semantics (Groenendijk and Stokhof (1990, 1991)), and Inquisitive Semantics (Ciardelli, Groenendijk, and Roelofsen (2013), Ciardelli, Roelofsen, and Theiler (2017)) use possible worlds to incorporate epistemic elements into formal semantics.
- They characterise sentence meanings as functions from discourse contexts to discourse contexts.
- Speakers use sentences to communicate information by modifying their hearers' representation of a discourse context.

Possible Worlds in Epistemic Reasoning

- Kripke frame semantics has also been influential in the related field of epistemic reasoning (Halperin (1995)).
- More recent formal semantic approaches, such as Dynamic semantics (Groenendijk and Stokhof (1990, 1991)), and Inquisitive Semantics (Ciardelli, Groenendijk, and Roelofsen (2013), Ciardelli, Roelofsen, and Theiler (2017)) use possible worlds to incorporate epistemic elements into formal semantics.
- They characterise sentence meanings as functions from discourse contexts to discourse contexts.
- Speakers use sentences to communicate information by modifying their hearers' representation of a discourse context.

Kripke Frame Semantics

- A model $M = \langle D, W, F, R \rangle$, where D is a non-empty set of individuals, W is a non-empty set of worlds, F is an interpretation function that assigns intensions to the constants of a language, and R is an accessibility relation on W.
- Formal semanticists have expanded M to include additional indices representing elements of context, such as sets of points in time, and sets of speakers.
- The elements of W are points at which a maximal consistent set of propositions are satisfied.
- In fact Carnap (1947), Jonsson and Tarski (1951), and Kripke(1959) originally characterised worlds as maximal consistent sets of propositions.

Kripke Frame Semantics

- A model $M = \langle D, W, F, R \rangle$, where D is a non-empty set of individuals, W is a non-empty set of worlds, F is an interpretation function that assigns intensions to the constants of a language, and R is an accessibility relation on W.
- Formal semanticists have expanded M to include additional indices representing elements of context, such as sets of points in time, and sets of speakers.
- The elements of W are points at which a maximal consistent set of propositions are satisfied.
- In fact Carnap (1947), Jonsson and Tarski (1951), and Kripke(1959) originally characterised worlds as maximal consistent sets of propositions.

Kripke Frame Semantics

- A model $M = \langle D, W, F, R \rangle$, where D is a non-empty set of individuals, W is a non-empty set of worlds, F is an interpretation function that assigns intensions to the constants of a language, and R is an accessibility relation on W.
- Formal semanticists have expanded M to include additional indices representing elements of context, such as sets of points in time, and sets of speakers.
- The elements of W are points at which a maximal consistent set of propositions are satisfied.
- In fact Carnap (1947), Jonsson and Tarski (1951), and Kripke(1959) originally characterised worlds as maximal consistent sets of propositions.

Kripke Frame Semantics

- A model $M = \langle D, W, F, R \rangle$, where D is a non-empty set of individuals, W is a non-empty set of worlds, F is an interpretation function that assigns intensions to the constants of a language, and R is an accessibility relation on W.
- Formal semanticists have expanded M to include additional indices representing elements of context, such as sets of points in time, and sets of speakers.
- The elements of W are points at which a maximal consistent set of propositions are satisfied.
- In fact Carnap (1947), Jonsson and Tarski (1951), and Kripke(1959) originally characterised worlds as maximal consistent sets of propositions.

Worlds as Ultrafilters of Propositions

- There is a one to one correspondence between the elements of W and the elements of the set of maximal consistent sets of propositions.
- Fox et al. (2002), Fox and Lappin (2005), and Pollard (2008) use this correspondence to formally represent worlds as the set U of ultrafilters in the prelattice of propositions.
- A proposition p holds at a world w_i iff $p \in u_i$, where $u_i \in U$.
- The question of how to represent W reduces to the representability of U.

Worlds as Ultrafilters of Propositions

- There is a one to one correspondence between the elements of W and the elements of the set of maximal consistent sets of propositions.
- Fox et al. (2002), Fox and Lappin (2005), and Pollard (2008) use this correspondence to formally represent worlds as the set U of ultrafilters in the prelattice of propositions.
- A proposition p holds at a world w_i iff $p \in u_i$, where $u_i \in U$.
- The question of how to represent W reduces to the representability of U.

Worlds as Ultrafilters of Propositions

- There is a one to one correspondence between the elements of W and the elements of the set of maximal consistent sets of propositions.
- Fox et al. (2002), Fox and Lappin (2005), and Pollard (2008) use this correspondence to formally represent worlds as the set U of ultrafilters in the prelattice of propositions.
- A proposition p holds at a world w_i iff $p \in u_i$, where $u_i \in U$.
- The question of how to represent W reduces to the representability of U.

Worlds as Ultrafilters of Propositions

- There is a one to one correspondence between the elements of W and the elements of the set of maximal consistent sets of propositions.
- Fox et al. (2002), Fox and Lappin (2005), and Pollard (2008) use this correspondence to formally represent worlds as the set U of ultrafilters in the prelattice of propositions.
- A proposition p holds at a world w_i iff $p \in u_i$, where $u_i \in U$.
- The question of how to represent W reduces to the representability of U.

A Simplified Version of the Representation Problem

- Assume that the the prelattice on which the elements of U are defined encodes classical Boolean propositional logic.
- This system is complete and decidable, and so minimal in expressive power.
- To identify any $u_i \in U$ we need to specify all and only the propositions that hold at u_i (an infinite set of propostions).
- We can enumerate the elements of an infinite set if there is an effective procedure (a finite set of rules, an algorithm, a recursive definition, etc.) for recognising its members.
- It is not clear what an effective procedure for enumerating the propositions of u_i would consist in.

A Simplified Version of the Representation Problem

- Assume that the the prelattice on which the elements of U are defined encodes classical Boolean propositional logic.
- This system is complete and decidable, and so minimal in expressive power.
- To identify any $u_i \in U$ we need to specify all and only the propositions that hold at u_i (an infinite set of propostions).
- We can enumerate the elements of an infinite set if there is an effective procedure (a finite set of rules, an algorithm, a recursive definition, etc.) for recognising its members.
- It is not clear what an effective procedure for enumerating the propositions of u_i would consist in.

A Simplified Version of the Representation Problem

- Assume that the the prelattice on which the elements of U are defined encodes classical Boolean propositional logic.
- This system is complete and decidable, and so minimal in expressive power.
- To identify any $u_i \in U$ we need to specify all and only the propositions that hold at u_i (an infinite set of propostions).
- We can enumerate the elements of an infinite set if there is an effective procedure (a finite set of rules, an algorithm, a recursive definition, etc.) for recognising its members.
- It is not clear what an effective procedure for enumerating the propositions of u_i would consist in.

A Simplified Version of the Representation Problem

- Assume that the the prelattice on which the elements of U are defined encodes classical Boolean propositional logic.
- This system is complete and decidable, and so minimal in expressive power.
- To identify any $u_i \in U$ we need to specify all and only the propositions that hold at u_i (an infinite set of propostions).
- We can enumerate the elements of an infinite set if there is an effective procedure (a finite set of rules, an algorithm, a recursive definition, etc.) for recognising its members.
- It is not clear what an effective procedure for enumerating the propositions of u_i would consist in.

A Simplified Version of the Representation Problem

- Assume that the the prelattice on which the elements of U are defined encodes classical Boolean propositional logic.
- This system is complete and decidable, and so minimal in expressive power.
- To identify any $u_i \in U$ we need to specify all and only the propositions that hold at u_i (an infinite set of propostions).
- We can enumerate the elements of an infinite set if there is an effective procedure (a finite set of rules, an algorithm, a recursive definition, etc.) for recognising its members.
- It is not clear what an effective procedure for enumerating the propositions of u_i would consist in.

The Repesentation of a World as a SAT Problem

- Simplifying further, assume that we are able to generate u_i from a finite set P_{u_i} of propositions, all of which are in *Conjunctive Normal Form* (CNF).
- A proposition in CNF is a conjunction of disjunctions of literals (elementary propositional variables or their negations).
- The propositions in P_{u_i} can be conjoined in a single formula p_{u_i} that is itself in CNF.
- For p_{u_i} to hold it is necessary to determine a distribution of truth-values for its literals that renders the entire formula true.

The Repesentation of a World as a SAT Problem

- Simplifying further, assume that we are able to generate u_i from a finite set P_{u_i} of propositions, all of which are in *Conjunctive Normal Form* (CNF).
- A proposition in CNF is a conjunction of disjunctions of literals (elementary propositional variables or their negations).
- The propositions in P_{u_i} can be conjoined in a single formula p_{u_i} that is itself in CNF.
- For p_{u_i} to hold it is necessary to determine a distribution of truth-values for its literals that renders the entire formula true.

The Repesentation of a World as a SAT Problem

- Simplifying further, assume that we are able to generate u_i from a finite set P_{u_i} of propositions, all of which are in *Conjunctive Normal Form* (CNF).
- A proposition in CNF is a conjunction of disjunctions of literals (elementary propositional variables or their negations).
- The propositions in P_{u_i} can be conjoined in a single formula p_{u_i} that is itself in CNF.
- For p_{u_i} to hold it is necessary to determine a distribution of truth-values for its literals that renders the entire formula true.

The Repesentation of a World as a SAT Problem

- Simplifying further, assume that we are able to generate u_i from a finite set P_{u_i} of propositions, all of which are in *Conjunctive Normal Form* (CNF).
- A proposition in CNF is a conjunction of disjunctions of literals (elementary propositional variables or their negations).
- The propositions in P_{u_i} can be conjoined in a single formula p_{u_i} that is itself in CNF.
- For p_{u_i} to hold it is necessary to determine a distribution of truth-values for its literals that renders the entire formula true.

The Complexity of the SAT Problem

- Determining the complexity of this satisfaction problem is an instance of the kSAT problem, where k is the number of literals in p_{u_i}.
- If $3 \leq k$, then the satisfiability problem for p_{u_i} is, in the general case, NP-complete, and so intractable (Papadimitriou (1995)).
- Given that this formula is intended to express the finite core of propositions from which the entire ultrafilter u_i is derived, it is reasonable to allow it to contain a large number of distinct elementary propositional constituents, each corresponding to a "core" fact that holds in u_i.

The Complexity of the SAT Problem

- Determining the complexity of this satisfaction problem is an instance of the kSAT problem, where k is the number of literals in p_{u_i}.
- If $3 \leq k$, then the satisfiability problem for p_{u_i} is, in the general case, NP-complete, and so intractable (Papadimitriou (1995)).
- Given that this formula is intended to express the finite core of propositions from which the entire ultrafilter u_i is derived, it is reasonable to allow it to contain a large number of distinct elementary propositional constituents, each corresponding to a "core" fact that holds in u_i.

The Complexity of the SAT Problem

- Determining the complexity of this satisfaction problem is an instance of the kSAT problem, where k is the number of literals in p_{u_i}.
- If $3 \leq k$, then the satisfiability problem for p_{u_i} is, in the general case, NP-complete, and so intractable (Papadimitriou (1995)).
- Given that this formula is intended to express the finite core of propositions from which the entire ultrafilter u_i is derived, it is reasonable to allow it to contain a large number of distinct elementary propositional constituents, each corresponding to a "core" fact that holds in u_i.

The Intractability of the Representation Problem

- It will also be necessary to include law like statements expressing regular relations among events that hold in a world (such as the laws of physics).
- These will be expressed as conditionals $A \rightarrow B$, which are encoded in a CNF formula by disjunctions of the form $\neg A \vee B$.
- Even given the generous simplifying assumptions concerning the enumeration of u_i, specifying the ultrafilter of propositions that corresponds to an individual world is, in general, a computationally intractable problem.
- It follows that it is not possible to compute W efficiently.

The Intractability of the Representation Problem

- It will also be necessary to include law like statements expressing regular relations among events that hold in a world (such as the laws of physics).
- These will be expressed as conditionals $A \rightarrow B$, which are encoded in a CNF formula by disjunctions of the form $\neg A \vee B$.
- Even given the generous simplifying assumptions concerning the enumeration of u_i, specifying the ultrafilter of propositions that corresponds to an individual world is, in general, a computationally intractable problem.
- It follows that it is not possible to compute W efficiently.

The Intractability of the Representation Problem

- It will also be necessary to include law like statements expressing regular relations among events that hold in a world (such as the laws of physics).
- These will be expressed as conditionals $A \rightarrow B$, which are encoded in a CNF formula by disjunctions of the form $\neg A \vee B$.
- Even given the generous simplifying assumptions concerning the enumeration of u_i, specifying the ultrafilter of propositions that corresponds to an individual world is, in general, a computationally intractable problem.
- It follows that it is not possible to compute W efficiently.

The Intractability of the Representation Problem

- It will also be necessary to include law like statements expressing regular relations among events that hold in a world (such as the laws of physics).
- These will be expressed as conditionals $A \rightarrow B$, which are encoded in a CNF formula by disjunctions of the form $\neg A \vee B$.
- Even given the generous simplifying assumptions concerning the enumeration of u_i, specifying the ultrafilter of propositions that corresponds to an individual world is, in general, a computationally intractable problem.
- It follows that it is not possible to compute W efficiently.

Three Possible Escape Moves which Do Not Work: Move 1

- We could follow Montague in claiming that formal semantics is a branch of mathematics rather than psychology.
- Questions of efficient computability and representability are not relevant to the theoretical constructions that it employs.
- This move raises the obvious question of what formal semantics is explaining.
- If it seeks to account for the way in which people interpret the expressions of a natural language, then one cannot simply discard the cognitive aspect of meaning.
- To do so would eliminate the empirical basis for assessing semantic theories.

Three Possible Escape Moves which Do Not Work: Move 1

- We could follow Montague in claiming that formal semantics is a branch of mathematics rather than psychology.
- Questions of efficient computability and representability are not relevant to the theoretical constructions that it employs.
- This move raises the obvious question of what formal semantics is explaining.
- If it seeks to account for the way in which people interpret the expressions of a natural language, then one cannot simply discard the cognitive aspect of meaning.
- To do so would eliminate the empirical basis for assessing semantic theories.

Three Possible Escape Moves which Do Not Work: Move 1

- We could follow Montague in claiming that formal semantics is a branch of mathematics rather than psychology.
- Questions of efficient computability and representability are not relevant to the theoretical constructions that it employs.
- This move raises the obvious question of what formal semantics is explaining.
- If it seeks to account for the way in which people interpret the expressions of a natural language, then one cannot simply discard the cognitive aspect of meaning.
- To do so would eliminate the empirical basis for assessing semantic theories.

Three Possible Escape Moves which Do Not Work: Move 1

- We could follow Montague in claiming that formal semantics is a branch of mathematics rather than psychology.
- Questions of efficient computability and representability are not relevant to the theoretical constructions that it employs.
- This move raises the obvious question of what formal semantics is explaining.
- If it seeks to account for the way in which people interpret the expressions of a natural language, then one cannot simply discard the cognitive aspect of meaning.
- To do so would eliminate the empirical basis for assessing semantic theories.

Three Possible Escape Moves which Do Not Work: Move 1

- We could follow Montague in claiming that formal semantics is a branch of mathematics rather than psychology.
- Questions of efficient computability and representability are not relevant to the theoretical constructions that it employs.
- This move raises the obvious question of what formal semantics is explaining.
- If it seeks to account for the way in which people interpret the expressions of a natural language, then one cannot simply discard the cognitive aspect of meaning.
- To do so would eliminate the empirical basis for assessing semantic theories.

A Weaker Version of Move 1

- We could acknowledge that using and interpreting natural language is indeed a cognitive process, but invoke the competence-performance distinction to insulate formal semantic theory from computational and processing concerns.
- On this view formal semantics offers a theory of semantic competence, which underlies speakers' linguistic performance.
- Unless one provides an explicit account of the way in which this competence drives processing and behaviour, then the notion of competence remains devoid of explanatory content (Lau, Clark, and Lappin (2016)).
- We cannot simply set aside questions of effective computability if we are interested in semantic theories that are grounded on sound cognitive foundations.

A Weaker Version of Move 1

- We could acknowledge that using and interpreting natural language is indeed a cognitive process, but invoke the competence-performance distinction to insulate formal semantic theory from computational and processing concerns.
- On this view formal semantics offers a theory of semantic competence, which underlies speakers' linguistic performance.
- Unless one provides an explicit account of the way in which this competence drives processing and behaviour, then the notion of competence remains devoid of explanatory content (Lau, Clark, and Lappin (2016)).
- We cannot simply set aside questions of effective computability if we are interested in semantic theories that are grounded on sound cognitive foundations.

A Weaker Version of Move 1

- We could acknowledge that using and interpreting natural language is indeed a cognitive process, but invoke the competence-performance distinction to insulate formal semantic theory from computational and processing concerns.
- On this view formal semantics offers a theory of semantic competence, which underlies speakers' linguistic performance.
- Unless one provides an explicit account of the way in which this competence drives processing and behaviour, then the notion of competence remains devoid of explanatory content (Lau, Clark, and Lappin (2016)).
- We cannot simply set aside questions of effective computability if we are interested in semantic theories that are grounded on sound cognitive foundations.

A Weaker Version of Move 1

- We could acknowledge that using and interpreting natural language is indeed a cognitive process, but invoke the competence-performance distinction to insulate formal semantic theory from computational and processing concerns.
- On this view formal semantics offers a theory of semantic competence, which underlies speakers' linguistic performance.
- Unless one provides an explicit account of the way in which this competence drives processing and behaviour, then the notion of competence remains devoid of explanatory content (Lau, Clark, and Lappin (2016)).
- We cannot simply set aside questions of effective computability if we are interested in semantic theories that are grounded on sound cognitive foundations.

Move 2: Stratification

- This technique stratifies a class of intractable problems into subclasses in order to identify the largest subsets of tractable tasks within the larger set (Clark and Lappin (2011)).
- So, for example, work on the tractable subclasses of kSAT problems is an active area of research.
- Similarly, first-order logic is undecidable, but many efficient theorem provers have been developed for subsets of first-order logic that are tractably decidable.
- We could focus on identifying the largest subsets of each $u_i \in U$ that can be tractably specified.

Move 2: Stratification

- This technique stratifies a class of intractable problems into subclasses in order to identify the largest subsets of tractable tasks within the larger set (Clark and Lappin (2011)).
- So, for example, work on the tractable subclasses of kSAT problems is an active area of research.
- Similarly, first-order logic is undecidable, but many efficient theorem provers have been developed for subsets of first-order logic that are tractably decidable.
- We could focus on identifying the largest subsets of each $u_i \in U$ that can be tractably specified.

Move 2: Stratification

- This technique stratifies a class of intractable problems into subclasses in order to identify the largest subsets of tractable tasks within the larger set (Clark and Lappin (2011)).
- So, for example, work on the tractable subclasses of kSAT problems is an active area of research.
- Similarly, first-order logic is undecidable, but many efficient theorem provers have been developed for subsets of first-order logic that are tractably decidable.
- We could focus on identifying the largest subsets of each $u_i \in U$ that can be tractably specified.

Move 2: Stratification

- This technique stratifies a class of intractable problems into subclasses in order to identify the largest subsets of tractable tasks within the larger set (Clark and Lappin (2011)).
- So, for example, work on the tractable subclasses of kSAT problems is an active area of research.
- Similarly, first-order logic is undecidable, but many efficient theorem provers have been developed for subsets of first-order logic that are tractably decidable.
- We could focus on identifying the largest subsets of each $u_i \in U$ that can be tractably specified.

Why Stratification won't Work for The World Representation Problem

- By definition, a world is (corresponds to) a maximal set of consistent propositions, an ultrafilter in a prelattice.
- If we specify only a proper subset of such an ultrafilter (a non-maximal filter), then it is no longer identified by all and only the propositions that hold at that world.
- In principle, several distinct worlds could share the same set of efficiently representable subsets of propositions, in which case they would not be efficiently distinguishable.

Why Stratification won't Work for The World Representation Problem

- By definition, a world is (corresponds to) a maximal set of consistent propositions, an ultrafilter in a prelattice.
- If we specify only a proper subset of such an ultrafilter (a non-maximal filter), then it is no longer identified by all and only the propositions that hold at that world.
- In principle, several distinct worlds could share the same set of efficiently representable subsets of propositions, in which case they would not be efficiently distinguishable.

Why Stratification won't Work for The World Representation Problem

- By definition, a world is (corresponds to) a maximal set of consistent propositions, an ultrafilter in a prelattice.
- If we specify only a proper subset of such an ultrafilter (a non-maximal filter), then it is no longer identified by all and only the propositions that hold at that world.
- In principle, several distinct worlds could share the same set of efficiently representable subsets of propositions, in which case they would not be efficiently distinguishable.

Move 3: Possible Situations

- We could substitute the set of possible situations for the set of possible worlds, where situations are partial worlds (Heim (1990), Lappin (2000), Kratzer (2014)).
- It is indeed the case that some non-maximal individual situations, and certain sets of such situations are easier to represent than worlds (Barwise and Perry (1983)).
- However, the representability problem for the entire set of possible situations is even more severe than the one that we encounter for the set of possible worlds.

Move 3: Possible Situations

- We could substitute the set of possible situations for the set of possible worlds, where situations are partial worlds (Heim (1990), Lappin (2000), Kratzer (2014)).
- It is indeed the case that some non-maximal individual situations, and certain sets of such situations are easier to represent than worlds (Barwise and Perry (1983)).
- However, the representability problem for the entire set of possible situations is even more severe than the one that we encounter for the set of possible worlds.

Move 3: Possible Situations

- We could substitute the set of possible situations for the set of possible worlds, where situations are partial worlds (Heim (1990), Lappin (2000), Kratzer (2014)).
- It is indeed the case that some non-maximal individual situations, and certain sets of such situations are easier to represent than worlds (Barwise and Perry (1983)).
- However, the representability problem for the entire set of possible situations is even more severe than the one that we encounter for the set of possible worlds.

The Representability Problem for the Set of Possible Situations

- For any given u_i corresponding to a world w_i, a situation $s_i \subseteq u_i$.
- The set of situations S_i for u_i is $\mathcal{P}(u_i)$, the power set of u_i.
- If $|u_i| = \aleph_0$, by Cantor's theorem on the cardinality of power sets, $|S_i|$ is uncountably infinite.

The Representability Problem for the Set of Possible Situations

- For any given u_i corresponding to a world w_i, a situation $s_i \subseteq u_i$.
- The set of situations S_i for u_i is $\mathcal{P}(u_i)$, the power set of u_i.
- If $|u_i| = \aleph_0$, by Cantor's theorem on the cardinality of power sets, $|S_i|$ is uncountably infinite.

The Representability Problem for the Set of Possible Situations

- For any given u_i corresponding to a world w_i, a situation $s_i \subseteq u_i$.
- The set of situations S_i for u_i is $\mathcal{P}(u_i)$, the power set of u_i.
- If $|u_i| = \aleph_0$, by Cantor's theorem on the cardinality of power sets, $|S_i|$ is uncountably infinite.

Modifying Move 3

- It is possible to avoid this difficulty if we limit ourselves to subsets of situations that we can specify effectively, as we require them for particular analyses.
- This is, in effect, a form of stratification.
- But as situations are not maximal in the way that worlds are, it is a viable method when applied to situations.
- In order for stratification to work, it is necessary to show that we do, in fact, have effective procedures for representing the situations that we need for our theories.

Modifying Move 3

- It is possible to avoid this difficulty if we limit ourselves to subsets of situations that we can specify effectively, as we require them for particular analyses.
- This is, in effect, a form of stratification.
- But as situations are not maximal in the way that worlds are, it is a viable method when applied to situations.
- In order for stratification to work, it is necessary to show that we do, in fact, have effective procedures for representing the situations that we need for our theories.

Modifying Move 3

- It is possible to avoid this difficulty if we limit ourselves to subsets of situations that we can specify effectively, as we require them for particular analyses.
- This is, in effect, a form of stratification.
- But as situations are not maximal in the way that worlds are, it is a viable method when applied to situations.
- In order for stratification to work, it is necessary to show that we do, in fact, have effective procedures for representing the situations that we need for our theories.

Modifying Move 3

- It is possible to avoid this difficulty if we limit ourselves to subsets of situations that we can specify effectively, as we require them for particular analyses.
- This is, in effect, a form of stratification.
- But as situations are not maximal in the way that worlds are, it is a viable method when applied to situations.
- In order for stratification to work, it is necessary to show that we do, in fact, have effective procedures for representing the situations that we need for our theories.

Operational and Denotational Semantics of Programming Languages

- It is common to distinguish between the operational and the denotational semantics of a program (Stump (2013)).
- Operational meaning corresponds (roughly) to the sequence of state transitions that occur when a program is executed.
- It can be identified with the computational process through which the program produces an output for a specified input.
- The denotational meaning of a program is the mathematical object that represents the output which it generates for a given input.
- Operational and denotational semantics can be understood compositionally in terms of their contributions to the state transitions of the program, and the value that it yields, respectively.

Operational and Denotational Semantics of Programming Languages

- It is common to distinguish between the operational and the denotational semantics of a program (Stump (2013)).
- Operational meaning corresponds (roughly) to the sequence of state transitions that occur when a program is executed.
- It can be identified with the computational process through which the program produces an output for a specified input.
- The denotational meaning of a program is the mathematical object that represents the output which it generates for a given input.
- Operational and denotational semantics can be understood compositionally in terms of their contributions to the state transitions of the program, and the value that it yields, respectively.

Operational and Denotational Semantics of Programming Languages

- It is common to distinguish between the operational and the denotational semantics of a program (Stump (2013)).
- Operational meaning corresponds (roughly) to the sequence of state transitions that occur when a program is executed.
- It can be identified with the computational process through which the program produces an output for a specified input.
- The denotational meaning of a program is the mathematical object that represents the output which it generates for a given input.
- Operational and denotational semantics can be understood compositionally in terms of their contributions to the state transitions of the program, and the value that it yields, respectively.

Operational and Denotational Semantics of Programming Languages

- It is common to distinguish between the operational and the denotational semantics of a program (Stump (2013)).
- Operational meaning corresponds (roughly) to the sequence of state transitions that occur when a program is executed.
- It can be identified with the computational process through which the program produces an output for a specified input.
- The denotational meaning of a program is the mathematical object that represents the output which it generates for a given input.
- Operational and denotational semantics can be understood compositionally in terms of their contributions to the state transitions of the program, and the value that it yields, respectively.

Operational and Denotational Semantics of Programming Languages

- It is common to distinguish between the operational and the denotational semantics of a program (Stump (2013)).
- Operational meaning corresponds (roughly) to the sequence of state transitions that occur when a program is executed.
- It can be identified with the computational process through which the program produces an output for a specified input.
- The denotational meaning of a program is the mathematical object that represents the output which it generates for a given input.
- Operational and denotational semantics can be understood compositionally in terms of their contributions to the state transitions of the program, and the value that it yields, respectively.

Example 1

- It is possible to construct a theorem prover for first-order logic using either semantic tableaux or resolution (Blackburn and Bos (2003)).
- Both theorem provers use proof by contradiction, but they employ alternative formal methods, and they are implemented as different computational procedures.
- They exhibit distinct efficiency and complexity properties.
- The two classifier predicates $theorem_{tableaux}$ and $theorem_{resolution}$ are operationally distinct, but they are provably equivalent in their denotations.

Example 1

- It is possible to construct a theorem prover for first-order logic using either semantic tableaux or resolution (Blackburn and Bos (2003)).
- Both theorem provers use proof by contradiction, but they employ alternative formal methods, and they are implemented as different computational procedures.
- They exhibit distinct efficiency and complexity properties.
- The two classifier predicates $theorem_{tableaux}$ and $theorem_{resolution}$ are operationally distinct, but they are provably equivalent in their denotations.

Example 1

- It is possible to construct a theorem prover for first-order logic using either semantic tableaux or resolution (Blackburn and Bos (2003)).
- Both theorem provers use proof by contradiction, but they employ alternative formal methods, and they are implemented as different computational procedures.
- They exhibit distinct efficiency and complexity properties.
- The two classifier predicates $theorem_{tableaux}$ and $theorem_{resolution}$ are operationally distinct, but they are provably equivalent in their denotations.

Example 1

- It is possible to construct a theorem prover for first-order logic using either semantic tableaux or resolution (Blackburn and Bos (2003)).
- Both theorem provers use proof by contradiction, but they employ alternative formal methods, and they are implemented as different computational procedures.
- They exhibit distinct efficiency and complexity properties.
- The two classifier predicates $theorem_{tableaux}$ and $theorem_{resolution}$ are operationally distinct, but they are provably equivalent in their denotations.

Example 2

- Consider two functions from fundamental sound frequencies to the letters indicating musical notes and half tones.
- The first takes as its arguments the pitch frequency waves of the electronic sensor in a chromatic tuner, and the second the pitch frequency graphs of a spectrogram.
- Assume that both functions can recognise notes and half tones in the same range of octaves, to the same level of accuracy.
- Again, their operational semantics are distinct, but they are denotationally equivalent.

Example 2

- Consider two functions from fundamental sound frequencies to the letters indicating musical notes and half tones.
- The first takes as its arguments the pitch frequency waves of the electronic sensor in a chromatic tuner, and the second the pitch frequency graphs of a spectrogram.
- Assume that both functions can recognise notes and half tones in the same range of octaves, to the same level of accuracy.
- Again, their operational semantics are distinct, but they are denotationally equivalent.

Example 2

- Consider two functions from fundamental sound frequencies to the letters indicating musical notes and half tones.
- The first takes as its arguments the pitch frequency waves of the electronic sensor in a chromatic tuner, and the second the pitch frequency graphs of a spectrogram.
- Assume that both functions can recognise notes and half tones in the same range of octaves, to the same level of accuracy.
- Again, their operational semantics are distinct, but they are denotationally equivalent.

Example 2

- Consider two functions from fundamental sound frequencies to the letters indicating musical notes and half tones.
- The first takes as its arguments the pitch frequency waves of the electronic sensor in a chromatic tuner, and the second the pitch frequency graphs of a spectrogram.
- Assume that both functions can recognise notes and half tones in the same range of octaves, to the same level of accuracy.
- Again, their operational semantics are distinct, but they are denotationally equivalent.

An Operational View of Intensions

- We take the operational meaning of an expression to be the computational process through which speakers compute its extension.
- Its denotational meaning is the extension that it generates for a given argument.
- Intensions are computable functions.
- This view of intension avoids the intractability of representation problem that arises with possible worlds.

An Operational View of Intensions

- We take the operational meaning of an expression to be the computational process through which speakers compute its extension.
- Its denotational meaning is the extension that it generates for a given argument.
- Intensions are computable functions.
- This view of intension avoids the intractability of representation problem that arises with possible worlds.

An Operational View of Intensions

- We take the operational meaning of an expression to be the computational process through which speakers compute its extension.
- Its denotational meaning is the extension that it generates for a given argument.
- Intensions are computable functions.
- This view of intension avoids the intractability of representation problem that arises with possible worlds.

An Operational View of Intensions

- We take the operational meaning of an expression to be the computational process through which speakers compute its extension.
- Its denotational meaning is the extension that it generates for a given argument.
- Intensions are computable functions.
- This view of intension avoids the intractability of representation problem that arises with possible worlds.

The Problem of Hyperintensionality

If logically equivalent expressions have the same denotations in all possible worlds and intensions are functions from worlds to denotations, then these expressions are identical in intension.

(1) a. If $A \subseteq B$ and $B \subseteq A$, then $A = B$. $\Leftrightarrow$

b. A prime number is divisible only by itself and 1.

(2) a. Mary believes that if $A \subseteq B$ and $B \subseteq A$, then $A = B$. $\not\Leftrightarrow$

b. Mary believes that a prime number is divisible only by itself and 1.

An Operational Solution to Hyperintensionality

- If we identify intensions with operational meaning, then (1)a and b are intensionally distinct.
- (1)a is a theorem of set theory, while (1)b is a theorem of number theory.
- Their proofs are entirely different, and so they encode distinct objects of belief.
- The operational notion of intension permits us to individuate objects of propositional attitude with the necessary degree of fine-grained meaning.

An Operational Solution to Hyperintensionality

- If we identify intensions with operational meaning, then (1)a and b are intensionally distinct.
- (1)a is a theorem of set theory, while (1)b is a theorem of number theory.
- Their proofs are entirely different, and so they encode distinct objects of belief.
- The operational notion of intension permits us to individuate objects of propositional attitude with the necessary degree of fine-grained meaning.

An Operational Solution to Hyperintensionality

- If we identify intensions with operational meaning, then (1)a and b are intensionally distinct.
- (1)a is a theorem of set theory, while (1)b is a theorem of number theory.
- Their proofs are entirely different, and so they encode distinct objects of belief.
- The operational notion of intension permits us to individuate objects of propositional attitude with the necessary degree of fine-grained meaning.

An Operational Solution to Hyperintensionality

- If we identify intensions with operational meaning, then (1)a and b are intensionally distinct.
- (1)a is a theorem of set theory, while (1)b is a theorem of number theory.
- Their proofs are entirely different, and so they encode distinct objects of belief.
- The operational notion of intension permits us to individuate objects of propositional attitude with the necessary degree of fine-grained meaning.

Modality

(3) a. Necessarily if $A \subseteq B$ and $B \subseteq A$, then $A = B$.

b. Possibly interest rates will rise in the next quarter.

c. It is likely that the Social Democrats will win the next election in Sweden.

The Classical View

- In possible worlds semantics modal operators are generalised quantifiers (GQs) on worlds.
- Necessity is a universal quantifier.
- Possibility an existential quantifier.
- *Likely* is a variant of the second-order GQ *most*.

The Classical View

- In possible worlds semantics modal operators are generalised quantifiers (GQs) on worlds.
- Necessity is a universal quantifier.
- Possibility an existential quantifier.
- *Likely* is a variant of the second-order GQ *most*.

The Classical View

- In possible worlds semantics modal operators are generalised quantifiers (GQs) on worlds.
- Necessity is a universal quantifier.
- Possibility an existential quantifier.
- *Likely* is a variant of the second-order GQ *most*.

The Classical View

- In possible worlds semantics modal operators are generalised quantifiers (GQs) on worlds.
- Necessity is a universal quantifier.
- Possibility an existential quantifier.
- *Likely* is a variant of the second-order GQ *most*.

Classical Truth Conditions for Modal Statements

1. $||\Box\alpha||^{M,w_i} = t$ iff $\forall w_{\in W} ||\alpha||^{M,w} = t$.
2. $||\Diamond\beta||^{M,w_i} = t$ iff $\exists w_{\in W} ||\beta||^{M,w} = t$.
3. $||\textit{Likely}\ \gamma||^{M,w_i} = t$ iff for an appropriately defined $W' \subseteq W$, $|\{w_j \in W' : ||\gamma||^{M,w_j} = t\}| \geq \epsilon$, where ϵ is a parametric cardinality value that is greater than 50% of W'.

An Alternative Probabilistic View of Modality

- We can reformulate modal statements as types of probability judgments.
- A probability model M consists of a sample space of events with all possible outcomes given, and a probability distribution over these outcomes, specified by a function p (Halpern (2003)).
- A model of the throws of a die assigns probabilities to each of its six sides landing up.
- If the die is not biased towards one or more sides, the probability function will assign equal probability to each of these outcomes, with the values of the sides summing to 1.

An Alternative Probabilistic View of Modality

- We can reformulate modal statements as types of probability judgments.
- A probability model M consists of a sample space of events with all possible outcomes given, and a probability distribution over these outcomes, specified by a function p (Halpern (2003)).
- A model of the throws of a die assigns probabilities to each of its six sides landing up.
- If the die is not biased towards one or more sides, the probability function will assign equal probability to each of these outcomes, with the values of the sides summing to 1.

An Alternative Probabilistic View of Modality

- We can reformulate modal statements as types of probability judgments.
- A probability model M consists of a sample space of events with all possible outcomes given, and a probability distribution over these outcomes, specified by a function p (Halpern (2003)).
- A model of the throws of a die assigns probabilities to each of its six sides landing up.
- If the die is not biased towards one or more sides, the probability function will assign equal probability to each of these outcomes, with the values of the sides summing to 1.

An Alternative Probabilistic View of Modality

- We can reformulate modal statements as types of probability judgments.
- A probability model M consists of a sample space of events with all possible outcomes given, and a probability distribution over these outcomes, specified by a function p (Halpern (2003)).
- A model of the throws of a die assigns probabilities to each of its six sides landing up.
- If the die is not biased towards one or more sides, the probability function will assign equal probability to each of these outcomes, with the values of the sides summing to 1.

Worlds and Sample Spaces

- Probability theorists often refer to the set of possible outcomes in a sample space as possible worlds, but this is misleading.
- Unlike worlds in Kripke frame semantics, outcomes are non-maximal.
- They are more naturally described as situations, which can be as large or as small as required by the sample space of a model.
- In specifying a sample space it is not necessary to distribute probability over the set of all possible situations (even of a certain type).
- We estimate the likelihood of an event of a particular type on the basis of observed occurrences of events, either of this type, or of others that might condition it.

Worlds and Sample Spaces

- Probability theorists often refer to the set of possible outcomes in a sample space as possible worlds, but this is misleading.
- Unlike worlds in Kripke frame semantics, outcomes are non-maximal.
- They are more naturally described as situations, which can be as large or as small as required by the sample space of a model.
- In specifying a sample space it is not necessary to distribute probability over the set of all possible situations (even of a certain type).
- We estimate the likelihood of an event of a particular type on the basis of observed occurrences of events, either of this type, or of others that might condition it.

Worlds and Sample Spaces

- Probability theorists often refer to the set of possible outcomes in a sample space as possible worlds, but this is misleading.
- Unlike worlds in Kripke frame semantics, outcomes are non-maximal.
- They are more naturally described as situations, which can be as large or as small as required by the sample space of a model.
- In specifying a sample space it is not necessary to distribute probability over the set of all possible situations (even of a certain type).
- We estimate the likelihood of an event of a particular type on the basis of observed occurrences of events, either of this type, or of others that might condition it.

Worlds and Sample Spaces

- Probability theorists often refer to the set of possible outcomes in a sample space as possible worlds, but this is misleading.
- Unlike worlds in Kripke frame semantics, outcomes are non-maximal.
- They are more naturally described as situations, which can be as large or as small as required by the sample space of a model.
- In specifying a sample space it is not necessary to distribute probability over the set of all possible situations (even of a certain type).
- We estimate the likelihood of an event of a particular type on the basis of observed occurrences of events, either of this type, or of others that might condition it.

Worlds and Sample Spaces

- Probability theorists often refer to the set of possible outcomes in a sample space as possible worlds, but this is misleading.
- Unlike worlds in Kripke frame semantics, outcomes are non-maximal.
- They are more naturally described as situations, which can be as large or as small as required by the sample space of a model.
- In specifying a sample space it is not necessary to distribute probability over the set of all possible situations (even of a certain type).
- We estimate the likelihood of an event of a particular type on the basis of observed occurrences of events, either of this type, or of others that might condition it.

Bayesian Probability

- In Bayesian models we compute the posterior probability of an event A (the hypothesis) given observed events B (the evidence) with Bayes' Rule, where $p(B) \neq 0$.
- $p(A|B) = \dfrac{p(B|A)p(A)}{p(B)}$
- $p(A)$ is the prior probability that the model assigns to the hyothesis that A will occur, and the denominator $p(B)$ normalises the value of the numerator so that all probabilities in the sample space sum to 1.

Bayesian Probability

- In Bayesian models we compute the posterior probability of an event A (the hypothesis) given observed events B (the evidence) with Bayes' Rule, where $p(B) \neq 0$.
- $p(A|B) = \dfrac{p(B|A)p(A)}{p(B)}$
- $p(A)$ is the prior probability that the model assigns to the hyothesis that A will occur, and the denominator $p(B)$ normalises the value of the numerator so that all probabilities in the sample space sum to 1.

Bayesian Probability

- In Bayesian models we compute the posterior probability of an event A (the hypothesis) given observed events B (the evidence) with Bayes' Rule, where $p(B) \neq 0$.
- $p(A|B) = \dfrac{p(B|A)p(A)}{p(B)}$
- $p(A)$ is the prior probability that the model assigns to the hyothesis that A will occur, and the denominator $p(B)$ normalises the value of the numerator so that all probabilities in the sample space sum to 1.

Conditional Probability

- Assume that the probability of A is conditioned by several event types $V_1, \ldots V_k$, where these are random variables.
- Each such V_i contains a set of probability assignments for different outcomes with respect to an event of that type.
- All assignments for events in V_i sum to 1.

Conditional Probability

- Assume that the probability of A is conditioned by several event types $V_1, \ldots V_k$, where these are random variables.
- Each such V_i contains a set of probability assignments for different outcomes with respect to an event of that type.
- All assignments for events in V_i sum to 1.

Conditional Probability

- Assume that the probability of A is conditioned by several event types $V_1, \ldots V_k$, where these are random variables.
- Each such V_i contains a set of probability assignments for different outcomes with respect to an event of that type.
- All assignments for events in V_i sum to 1.

An Example

- Let A be the event of John arriving home on time.
- Let the random variables that A depends on be whether his meeting ends on time (T), if he leaves work immediately after the meeting (W), and whether his bus is running on schedule (B).
- Assume that T includes probabilities for John's meeting ending on time (t_1), for the meeting ending late (t_2), and for it ending early (t_3).
- If these are the only event instances for the random variable T, then $p(t_1) + p(t_2) + p(t_3) = 1$.
- The other random variables, W and B, have similar distributions of probability values for their instances.

An Example

- Let A be the event of John arriving home on time.
- Let the random variables that A depends on be whether his meeting ends on time (T), if he leaves work immediately after the meeting (W), and whether his bus is running on schedule (B).
- Assume that T includes probabilities for John's meeting ending on time (t_1), for the meeting ending late (t_2), and for it ending early (t_3).
- If these are the only event instances for the random variable T, then $p(t_1) + p(t_2) + p(t_3) = 1$.
- The other random variables, W and B, have similar distributions of probability values for their instances.

An Example

- Let A be the event of John arriving home on time.
- Let the random variables that A depends on be whether his meeting ends on time (T), if he leaves work immediately after the meeting (W), and whether his bus is running on schedule (B).
- Assume that T includes probabilities for John's meeting ending on time (t_1), for the meeting ending late (t_2), and for it ending early (t_3).
- If these are the only event instances for the random variable T, then $p(t_1) + p(t_2) + p(t_3) = 1$.
- The other random variables, W and B, have similar distributions of probability values for their instances.

An Example

- Let A be the event of John arriving home on time.
- Let the random variables that A depends on be whether his meeting ends on time (T), if he leaves work immediately after the meeting (W), and whether his bus is running on schedule (B).
- Assume that T includes probabilities for John's meeting ending on time (t_1), for the meeting ending late (t_2), and for it ending early (t_3).
- If these are the only event instances for the random variable T, then $p(t_1) + p(t_2) + p(t_3) = 1$.
- The other random variables, W and B, have similar distributions of probability values for their instances.

An Example

- Let A be the event of John arriving home on time.
- Let the random variables that A depends on be whether his meeting ends on time (T), if he leaves work immediately after the meeting (W), and whether his bus is running on schedule (B).
- Assume that T includes probabilities for John's meeting ending on time (t_1), for the meeting ending late (t_2), and for it ending early (t_3).
- If these are the only event instances for the random variable T, then $p(t_1) + p(t_2) + p(t_3) = 1$.
- The other random variables, W and B, have similar distributions of probability values for their instances.

Marginalising out Conditional Probabilities

- We can compute a non-conditional probability for A by marginalising out the probabilities of T, W, B.
- This involves summing across the joint probability values for A and all instances of the random variables T, W, B.
- $p(A) = \sum_{t_{\in T}, w_{\in W}, b_{\in B}} p(A, t, w, b)$
- Joint probabilities of this kind are equivalent to the probabilities of a conjunction of events, and we can compute these through the chain rule for conjunction, which treats it as a product of conditional probabilities.
- $p(A, T, W, B) = p(A|T, W, B) \times p(T|W, B) \times p(W|B) \times p(B)$

Marginalising out Conditional Probabilities

- We can compute a non-conditional probability for A by marginalising out the probabilities of T, W, B.
- This involves summing across the joint probability values for A and all instances of the random variables T, W, B.
- $p(A) = \sum_{t_{\in T}, w_{\in W}, b_{\in B}} p(A, t, w, b)$
- Joint probabilities of this kind are equivalent to the probabilities of a conjunction of events, and we can compute these through the chain rule for conjunction, which treats it as a product of conditional probabilities.
- $p(A, T, W, B) = p(A|T, W, B) \times p(T|W, B) \times p(W|B) \times p(B)$

Marginalising out Conditional Probabilities

- We can compute a non-conditional probability for A by marginalising out the probabilities of T, W, B.
- This involves summing across the joint probability values for A and all instances of the random variables T, W, B.
- $p(A) = \sum_{t_{\in T}, w_{\in W}, b_{\in B}} p(A, t, w, b)$
- Joint probabilities of this kind are equivalent to the probabilities of a conjunction of events, and we can compute these through the chain rule for conjunction, which treats it as a product of conditional probabilities.
- $p(A, T, W, B) = p(A|T, W, B) \times p(T|W, B) \times p(W|B) \times p(B)$

Marginalising out Conditional Probabilities

- We can compute a non-conditional probability for A by marginalising out the probabilities of T, W, B.
- This involves summing across the joint probability values for A and all instances of the random variables T, W, B.
- $p(A) = \sum_{t_{\in T}, w_{\in W}, b_{\in B}} p(A, t, w, b)$
- Joint probabilities of this kind are equivalent to the probabilities of a conjunction of events, and we can compute these through the chain rule for conjunction, which treats it as a product of conditional probabilities.
- $p(A, T, W, B) = p(A|T, W, B) \times p(T|W, B) \times p(W|B) \times p(B)$

Tractability of the Representation Problem for Bayesian Probability Models

- Computing the full set of such joint probability assignments is, in the general case, intractable.
- However, there are efficient ways of estimating or approximating them within a Bayesian network (Pearl (1990), Murphy (2001), Halpern (2003), Koski and Noble (2009)).
- It is, then, possible to efficiently represent a large subset of probability models, and to compute probability distributions for the possible events in their sample spaces.
- The maximality of worlds and the absence of any apparent procedure for generating their representations excludes the application of these methods to possible worlds of the kind that figure in classical formal semantics.

Tractability of the Representation Problem for Bayesian Probability Models

- Computing the full set of such joint probability assignments is, in the general case, intractable.
- However, there are efficient ways of estimating or approximating them within a Bayesian network (Pearl (1990), Murphy (2001), Halpern (2003), Koski and Noble (2009)).
- It is, then, possible to efficiently represent a large subset of probability models, and to compute probability distributions for the possible events in their sample spaces.
- The maximality of worlds and the absence of any apparent procedure for generating their representations excludes the application of these methods to possible worlds of the kind that figure in classical formal semantics.

Tractability of the Representation Problem for Bayesian Probability Models

- Computing the full set of such joint probability assignments is, in the general case, intractable.
- However, there are efficient ways of estimating or approximating them within a Bayesian network (Pearl (1990), Murphy (2001), Halpern (2003), Koski and Noble (2009)).
- It is, then, possible to efficiently represent a large subset of probability models, and to compute probability distributions for the possible events in their sample spaces.
- The maximality of worlds and the absence of any apparent procedure for generating their representations excludes the application of these methods to possible worlds of the kind that figure in classical formal semantics.

Tractability of the Representation Problem for Bayesian Probability Models

- Computing the full set of such joint probability assignments is, in the general case, intractable.
- However, there are efficient ways of estimating or approximating them within a Bayesian network (Pearl (1990), Murphy (2001), Halpern (2003), Koski and Noble (2009)).
- It is, then, possible to efficiently represent a large subset of probability models, and to compute probability distributions for the possible events in their sample spaces.
- The maximality of worlds and the absence of any apparent procedure for generating their representations excludes the application of these methods to possible worlds of the kind that figure in classical formal semantics.

Using Probability Models to Characterise Modality

Let M be a probability model, and p the probability function in M.

1'. $||\textit{Necessarily } \alpha||^{M,p} = t$ iff for all models $M' \in R, p_{\in M'}(\alpha) = 1$, where R is a suitably restricted subset of probability models.

2'. $||\textit{Possibly } \beta||^{M,p} = t$ iff $p(\beta) > 0$.

3'. $||\textit{Likely } \gamma||^{M,p} = t$ iff $p(\gamma) > \epsilon$, where ϵ is a parametric probability value that is greater than 0.5.

Using Probability Models to Characterise Modality

Let M be a probability model, and p the probability function in M.

1'. $||\textit{Necessarily } \alpha||^{M,p} = t$ iff for all models $M' \in R, p_{\in M'}(\alpha) = 1$, where R is a suitably restricted subset of probability models.

2'. $||\textit{Possibly } \beta||^{M,p} = t$ iff $p(\beta) > 0$.

3'. $||\textit{Likely } \gamma||^{M,p} = t$ iff $p(\gamma) > \epsilon$, where ϵ is a parametric probability value that is greater than 0.5.

Epistemic States: the Classical Approach

- Let W_B be the set of worlds (understood as ultrafilters of propositions) compatible with an agent a's beliefs.
- Take F_B to be a possibly non-maximal filter such that $F_B \subseteq \bigcap W_B$, where for every proposition $\phi \in F_B$, a regards ϕ as true.
- Let w_{actual} be the actual world.
- a's knowledge is contained in $F_K \subseteq F_B \cap w_{actual}$ (Halperin (1995)).

Epistemic States: the Classical Approach

- Let W_B be the set of worlds (understood as ultrafilters of propositions) compatible with an agent a's beliefs.
- Take F_B to be a possibly non-maximal filter such that $F_B \subseteq \bigcap W_B$, where for every proposition $\phi \in F_B$, a regards ϕ as true.
- Let w_{actual} be the actual world.
- a's knowledge is contained in $F_K \subseteq F_B \cap w_{actual}$ (Halperin (1995)).

Epistemic States: the Classical Approach

- Let W_B be the set of worlds (understood as ultrafilters of propositions) compatible with an agent a's beliefs.
- Take F_B to be a possibly non-maximal filter such that $F_B \subseteq \bigcap W_B$, where for every proposition $\phi \in F_B$, a regards ϕ as true.
- Let w_{actual} be the actual world.
- a's knowledge is contained in $F_K \subseteq F_B \cap w_{actual}$ (Halperin (1995)).

Epistemic States: the Classical Approach

- Let W_B be the set of worlds (understood as ultrafilters of propositions) compatible with an agent a's beliefs.
- Take F_B to be a possibly non-maximal filter such that $F_B \subseteq \bigcap W_B$, where for every proposition $\phi \in F_B$, a regards ϕ as true.
- Let w_{actual} be the actual world.
- a's knowledge is contained in $F_K \subseteq F_B \cap w_{actual}$ (Halperin (1995)).

Epistemic States: a Probabilistic Approach

- We can use a probability model to encode an agent's beliefs.
- The probability distribution that this model contains expresses the agent's epistemic commitments concerning the likelihood of situations and events.
- One way of articulating the structure of causal dependencies implicit in these beliefs is to use a Bayesian network as a model of belief.

Epistemic States: a Probabilistic Approach

- We can use a probability model to encode an agent's beliefs.
- The probability distribution that this model contains expresses the agent's epistemic commitments concerning the likelihood of situations and events.
- One way of articulating the structure of causal dependencies implicit in these beliefs is to use a Bayesian network as a model of belief.

Epistemic States: a Probabilistic Approach

- We can use a probability model to encode an agent's beliefs.
- The probability distribution that this model contains expresses the agent's epistemic commitments concerning the likelihood of situations and events.
- One way of articulating the structure of causal dependencies implicit in these beliefs is to use a Bayesian network as a model of belief.

Bayesian Networks

- A Bayesian network is a Directed Acyclic Graph (DAG) whose nodes are random variables.
- Each of the values of a random variable is the probability of one of the set of possible states that the variable denotes.
- Its directed edges express dependency relations among the variables.
- When the values of all the variables are specified, the graph describes a complete joint probability distribution (JPD) for its random variables.

Bayesian Networks

- A Bayesian network is a Directed Acyclic Graph (DAG) whose nodes are random variables.
- Each of the values of a random variable is the probability of one of the set of possible states that the variable denotes.
- Its directed edges express dependency relations among the variables.
- When the values of all the variables are specified, the graph describes a complete joint probability distribution (JPD) for its random variables.

Bayesian Networks

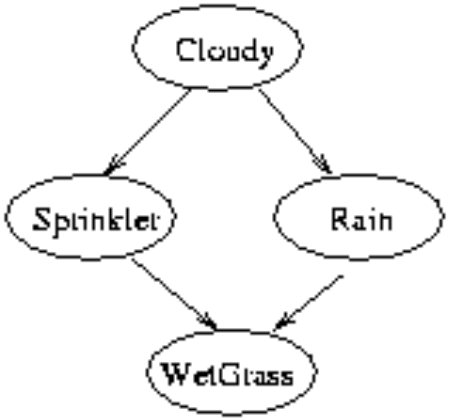
- A Bayesian network is a Directed Acyclic Graph (DAG) whose nodes are random variables.
- Each of the values of a random variable is the probability of one of the set of possible states that the variable denotes.
- Its directed edges express dependency relations among the variables.
- When the values of all the variables are specified, the graph describes a complete joint probability distribution (JPD) for its random variables.

Bayesian Networks

- A Bayesian network is a Directed Acyclic Graph (DAG) whose nodes are random variables.
- Each of the values of a random variable is the probability of one of the set of possible states that the variable denotes.
- Its directed edges express dependency relations among the variables.
- When the values of all the variables are specified, the graph describes a complete joint probability distribution (JPD) for its random variables.

A Bayesian Network (Russell and Norvig(1995))

P(C=F)	P(C=T)
0.5	0.5



C	P(S=F)	P(S=T)
F	0.5	0.5
T	0.9	0.1

C	P(R=F)	P(R=T)
F	0.8	0.2
T	0.2	0.8

S	R	P(W=F)	P(W=T)
F	F	1.0	0.0
T	F	0.1	0.9
F	T	0.1	0.9
T	T	0.01	0.99

Computing the Unconditional Probability of an Event in a Bayesian Network

- We can compute the marginal probability of the grass being wet (W = T) in this network by marginalising out the probabilities of the other variables on which W conditionally depends.
- As we have seen, this involves summing across all the joint probabilities of their instances.
- $p(W = T) = \sum_{s,r,c} p(W = T, S = s, R = r, C = c)$
- As we have a complete JPD for the variables of this network, it is straightforward to compute $p(W = T)$ using the chain rule.

Computing the Unconditional Probability of an Event in a Bayesian Network

- We can compute the marginal probability of the grass being wet (W = T) in this network by marginalising out the probabilities of the other variables on which W conditionally depends.
- As we have seen, this involves summing across all the joint probabilities of their instances.
- $p(W = T) = \sum_{s,r,c} p(W = T, S = s, R = r, C = c)$
- As we have a complete JPD for the variables of this network, it is straightforward to compute $p(W = T)$ using the chain rule.

Computing the Unconditional Probability of an Event in a Bayesian Network

- We can compute the marginal probability of the grass being wet (W = T) in this network by marginalising out the probabilities of the other variables on which W conditionally depends.
- As we have seen, this involves summing across all the joint probabilities of their instances.
- $p(W = T) = \sum_{s,r,c} p(W = T, S = s, R = r, C = c)$
- As we have a complete JPD for the variables of this network, it is straightforward to compute $p(W = T)$ using the chain rule.

Computing the Unconditional Probability of an Event in a Bayesian Network

- We can compute the marginal probability of the grass being wet (W = T) in this network by marginalising out the probabilities of the other variables on which W conditionally depends.
- As we have seen, this involves summing across all the joint probabilities of their instances.
- $p(W = T) = \sum_{s,r,c} p(W = T, S = s, R = r, C = c)$
- As we have a complete JPD for the variables of this network, it is straightforward to compute $p(W = T)$ using the chain rule.

Modelling an Agent's Belief with Bayesian Networks

- In principle we could model an agent's beliefs as a single integrated Bayesian network.
- This would be inefficient, as it would be problematic to determine the dependencies among all of the random variables representing event types that the agent has beliefs about, in a way that sustains consistency and tractability.
- It is more computationally manageable, and more epistemically plausible to construct local Bayesian networks to encode an agent's *a*'s beliefs about a particular domain of situations.
- A complete collection of beliefs for *a* will consist of a set of such local networks, where each element of this set expresses *a*'s beliefs about a specified class of events.

Modelling an Agent's Belief with Bayesian Networks

- In principle we could model an agent's beliefs as a single integrated Bayesian network.
- This would be inefficient, as it would be problematic to determine the dependencies among all of the random variables representing event types that the agent has beliefs about, in a way that sustains consistency and tractability.
- It is more computationally manageable, and more epistemically plausible to construct local Bayesian networks to encode an agent's *a*'s beliefs about a particular domain of situations.
- A complete collection of beliefs for *a* will consist of a set of such local networks, where each element of this set expresses *a*'s beliefs about a specified class of events.

Modelling an Agent's Belief with Bayesian Networks

- In principle we could model an agent's beliefs as a single integrated Bayesian network.
- This would be inefficient, as it would be problematic to determine the dependencies among all of the random variables representing event types that the agent has beliefs about, in a way that sustains consistency and tractability.
- It is more computationally manageable, and more epistemically plausible to construct local Bayesian networks to encode an agent's *a*'s beliefs about a particular domain of situations.
- A complete collection of beliefs for *a* will consist of a set of such local networks, where each element of this set expresses *a*'s beliefs about a specified class of events.

Modelling an Agent's Belief with Bayesian Networks

- In principle we could model an agent's beliefs as a single integrated Bayesian network.
- This would be inefficient, as it would be problematic to determine the dependencies among all of the random variables representing event types that the agent has beliefs about, in a way that sustains consistency and tractability.
- It is more computationally manageable, and more epistemically plausible to construct local Bayesian networks to encode an agent's *a*'s beliefs about a particular domain of situations.
- A complete collection of beliefs for *a* will consist of a set of such local networks, where each element of this set expresses *a*'s beliefs about a specified class of events.

Isomorphic Networks

- Two graphs G_i and G_j are *isomorphic* iff
  1. they contain the same number of vertices,
  2. there is a bijection from the vertices of G_i to the vertices of G_j and vice versa, such that
  3. the same number of edges connect each vertex v_i to G_i and v_j to G_j, through identical corresponding paths.
- For isomorphic DAGs this condition entails that the edges going into v_i and coming from it are of the same directionality as the edges going into and coming out of v_j, and vice versa.

Isomorphic Networks

- Two graphs G_i and G_j are *isomorphic* iff
  1. they contain the same number of vertices,
  2. there is a bijection from the vertices of G_i to the vertices of G_j and vice versa, such that
  3. the same number of edges connect each vertex v_i to G_i and v_j to G_j, through identical corresponding paths.
- For isomorphic DAGs this condition entails that the edges going into v_i and coming from it are of the same directionality as the edges going into and coming out of v_j, and vice versa.

Knowledge and Belief

- Two subgraphs of two Bayesian networks *match* iff they are isomorphic, and the random variables at their corresponding vertices range over the same event instances, with the same probability values.
- Let BN_B be the Bayesian network that expresses a's beliefs about a given event domain.
- Take BN_R to be the Bayesian network that codifies the actual probabilities and causal dependencies that hold for these events.
- We can identify a's knowledge for this domain as the maximal subgraph BN_K of BN_B that matches a subgraph in BN_R, and which satisfies additional conditions C.
- These conditions enforce constraints like the requirement that the beliefs encoded in BN_B are warranted by appropriate evidence.

Knowledge and Belief

- Two subgraphs of two Bayesian networks *match* iff they are isomorphic, and the random variables at their corresponding vertices range over the same event instances, with the same probability values.
- Let BN_B be the Bayesian network that expresses a's beliefs about a given event domain.
- Take BN_R to be the Bayesian network that codifies the actual probabilities and causal dependencies that hold for these events.
- We can identify a's knowledge for this domain as the maximal subgraph BN_K of BN_B that matches a subgraph in BN_R, and which satisfies additional conditions C.
- These conditions enforce constraints like the requirement that the beliefs encoded in BN_B are warranted by appropriate evidence.

Knowledge and Belief

- Two subgraphs of two Bayesian networks *match* iff they are isomorphic, and the random variables at their corresponding vertices range over the same event instances, with the same probability values.
- Let BN_B be the Bayesian network that expresses a's beliefs about a given event domain.
- Take BN_R to be the Bayesian network that codifies the actual probabilities and causal dependencies that hold for these events.
- We can identify a's knowledge for this domain as the maximal subgraph BN_K of BN_B that matches a subgraph in BN_R, and which satisfies additional conditions C.
- These conditions enforce constraints like the requirement that the beliefs encoded in BN_B are warranted by appropriate evidence.

Knowledge and Belief

- Two subgraphs of two Bayesian networks *match* iff they are isomorphic, and the random variables at their corresponding vertices range over the same event instances, with the same probability values.
- Let BN_B be the Bayesian network that expresses a's beliefs about a given event domain.
- Take BN_R to be the Bayesian network that codifies the actual probabilities and causal dependencies that hold for these events.
- We can identify a's knowledge for this domain as the maximal subgraph BN_K of BN_B that matches a subgraph in BN_R, and which satisfies additional conditions C.
- These conditions enforce constraints like the requirement that the beliefs encoded in BN_B are warranted by appropriate evidence.

Knowledge and Belief

- Two subgraphs of two Bayesian networks *match* iff they are isomorphic, and the random variables at their corresponding vertices range over the same event instances, with the same probability values.
- Let BN_B be the Bayesian network that expresses a's beliefs about a given event domain.
- Take BN_R to be the Bayesian network that codifies the actual probabilities and causal dependencies that hold for these events.
- We can identify a's knowledge for this domain as the maximal subgraph BN_K of BN_B that matches a subgraph in BN_R, and which satisfies additional conditions C.
- These conditions enforce constraints like the requirement that the beliefs encoded in BN_B are warranted by appropriate evidence.

Advantages of the Bayesian Approach

- By modelling knowledge and belief with Bayesian networks we avoid the representability problem that the classical view inherits from possible worlds.
- Belief revision has to be handled by a task specific update function in a classical worlds based model of belief.
- Bayesian networks inherently exhibit the acquisition of beliefs as a dynamic process driven by continual updates in an epistemic agent's observations.
- In a traditional worlds model of epistemic states, inference depends on an epistemic logic, whose rules are added to the model.
- A Bayesian network generates causal inferences directly, through the dependencies that it encodes in its paths.

Advantages of the Bayesian Approach

- By modelling knowledge and belief with Bayesian networks we avoid the representability problem that the classical view inherits from possible worlds.
- Belief revision has to be handled by a task specific update function in a classical worlds based model of belief.
- Bayesian networks inherently exhibit the acquisition of beliefs as a dynamic process driven by continual updates in an epistemic agent's observations.
- In a traditional worlds model of epistemic states, inference depends on an epistemic logic, whose rules are added to the model.
- A Bayesian network generates causal inferences directly, through the dependencies that it encodes in its paths.

Advantages of the Bayesian Approach

- By modelling knowledge and belief with Bayesian networks we avoid the representability problem that the classical view inherits from possible worlds.
- Belief revision has to be handled by a task specific update function in a classical worlds based model of belief.
- Bayesian networks inherently exhibit the acquisition of beliefs as a dynamic process driven by continual updates in an epistemic agent's observations.
- In a traditional worlds model of epistemic states, inference depends on an epistemic logic, whose rules are added to the model.
- A Bayesian network generates causal inferences directly, through the dependencies that it encodes in its paths.

Advantages of the Bayesian Approach

- By modelling knowledge and belief with Bayesian networks we avoid the representability problem that the classical view inherits from possible worlds.
- Belief revision has to be handled by a task specific update function in a classical worlds based model of belief.
- Bayesian networks inherently exhibit the acquisition of beliefs as a dynamic process driven by continual updates in an epistemic agent's observations.
- In a traditional worlds model of epistemic states, inference depends on an epistemic logic, whose rules are added to the model.
- A Bayesian network generates causal inferences directly, through the dependencies that it encodes in its paths.

Advantages of the Bayesian Approach

- By modelling knowledge and belief with Bayesian networks we avoid the representability problem that the classical view inherits from possible worlds.
- Belief revision has to be handled by a task specific update function in a classical worlds based model of belief.
- Bayesian networks inherently exhibit the acquisition of beliefs as a dynamic process driven by continual updates in an epistemic agent's observations.
- In a traditional worlds model of epistemic states, inference depends on an epistemic logic, whose rules are added to the model.
- A Bayesian network generates causal inferences directly, through the dependencies that it encodes in its paths.

Related Work: van Eijck and Lappin (2012)

- vanEijck and Lappin (2012) propose a theory in which the probability of a sentence is the sum of the probability values of the worlds in which it is true.
- If these worlds are construed as maximal in the sense discussed here, then this proposal runs into the representability problem for worlds.

Related Work: van Eijck and Lappin (2012)

- vanEijck and Lappin (2012) propose a theory in which the probability of a sentence is the sum of the probability values of the worlds in which it is true.
- If these worlds are construed as maximal in the sense discussed here, then this proposal runs into the representability problem for worlds.

Related Work: Cooper et al. (2015)

- Cooper, Dobnik, Larsson, and Lappin (2015) develop a compositional semantics in which the probability of a sentence is a judgment on the likelihood that a given situation is of a particular type.
- They specify situation types though their probabilistic type theory (ProbTTR).
- It is not entirely clear how probabilities for sentences are computed in this system.
- The conditions of type membership in ProbTTR may not be efficiently decidable.
- It is not obvious that the type theory is necessary for a viable probabilistic semantics of classifiers.

Related Work: Cooper et al. (2015)

- Cooper, Dobnik, Larsson, and Lappin (2015) develop a compositional semantics in which the probability of a sentence is a judgment on the likelihood that a given situation is of a particular type.
- They specify situation types though their probabilistic type theory (ProbTTR).
- It is not entirely clear how probabilities for sentences are computed in this system.
- The conditions of type membership in ProbTTR may not be efficiently decidable.
- It is not obvious that the type theory is necessary for a viable probabilistic semantics of classifiers.

Related Work: Cooper et al. (2015)

- Cooper, Dobnik, Larsson, and Lappin (2015) develop a compositional semantics in which the probability of a sentence is a judgment on the likelihood that a given situation is of a particular type.
- They specify situation types though their probabilistic type theory (ProbTTR).
- It is not entirely clear how probabilities for sentences are computed in this system.
- The conditions of type membership in ProbTTR may not be efficiently decidable.
- It is not obvious that the type theory is necessary for a viable probabilistic semantics of classifiers.

Related Work: Cooper et al. (2015)

- Cooper, Dobnik, Larsson, and Lappin (2015) develop a compositional semantics in which the probability of a sentence is a judgment on the likelihood that a given situation is of a particular type.
- They specify situation types though their probabilistic type theory (ProbTTR).
- It is not entirely clear how probabilities for sentences are computed in this system.
- The conditions of type membership in ProbTTR may not be efficiently decidable.
- It is not obvious that the type theory is necessary for a viable probabilistic semantics of classifiers.

Related Work: Cooper et al. (2015)

- Cooper, Dobnik, Larsson, and Lappin (2015) develop a compositional semantics in which the probability of a sentence is a judgment on the likelihood that a given situation is of a particular type.
- They specify situation types though their probabilistic type theory (ProbTTR).
- It is not entirely clear how probabilities for sentences are computed in this system.
- The conditions of type membership in ProbTTR may not be efficiently decidable.
- It is not obvious that the type theory is necessary for a viable probabilistic semantics of classifiers.

Related Work: Goodman and Lassiter (2015), Lassiter and Goodman (2017)

- Goodman and Lassiter (2015) and Lassiter and Goodman (2017) take probability to be distributed over partial worlds.
- They implement probabilistic treatments of a scalar adjective, *tall*, and the sorities paradox for nouns like *heap* in the functional probabilistic programming language Church.
- The Goodman-Lassiter account models vagueness by positing the existence of a univocal speaker's meaning that hearers estimate through distributing probability among alternative possible readings.
- They posit a boundary cut off point parameter for graded modifiers, where the value of this parameter is determined in context.

Related Work: Goodman and Lassiter (2015), Lassiter and Goodman (2017)

- Goodman and Lassiter (2015) and Lassiter and Goodman (2017) take probability to be distributed over partial worlds.
- They implement probabilistic treatments of a scalar adjective, *tall*, and the sorities paradox for nouns like *heap* in the functional probabilistic programming language Church.
- The Goodman-Lassiter account models vagueness by positing the existence of a univocal speaker's meaning that hearers estimate through distributing probability among alternative possible readings.
- They posit a boundary cut off point parameter for graded modifiers, where the value of this parameter is determined in context.

Related Work: Goodman and Lassiter (2015), Lassiter and Goodman (2017)

- Goodman and Lassiter (2015) and Lassiter and Goodman (2017) take probability to be distributed over partial worlds.
- They implement probabilistic treatments of a scalar adjective, *tall*, and the sorities paradox for nouns like *heap* in the functional probabilistic programming language Church.
- The Goodman-Lassiter account models vagueness by positing the existence of a univocal speaker's meaning that hearers estimate through distributing probability among alternative possible readings.
- They posit a boundary cut off point parameter for graded modifiers, where the value of this parameter is determined in context.

Related Work: Goodman and Lassiter (2015), Lassiter and Goodman (2017)

- Goodman and Lassiter (2015) and Lassiter and Goodman (2017) take probability to be distributed over partial worlds.
- They implement probabilistic treatments of a scalar adjective, *tall*, and the sorities paradox for nouns like *heap* in the functional probabilistic programming language Church.
- The Goodman-Lassiter account models vagueness by positing the existence of a univocal speaker's meaning that hearers estimate through distributing probability among alternative possible readings.
- They posit a boundary cut off point parameter for graded modifiers, where the value of this parameter is determined in context.

Related Work: Goodman and Lassiter (2015), Lassiter and Goodman (2017)

- The approach that I am suggesting here does not assume such an inaccessible boundary point for predicates.
- It allows us to interpret the probability value of a sentence as the likelihood that a competent speaker would endorse an assertion, given certain conditions (hypotheses).
- Therefore, predication remains intrinsically vague.
- It consists in applying a classifier to new instances on the basis of supervised training.
- We are not obliged to posit a contextually dependent cut off boundary for graded (or non-graded) predicates.

Related Work: Goodman and Lassiter (2015), Lassiter and Goodman (2017)

- The approach that I am suggesting here does not assume such an inaccessible boundary point for predicates.
- It allows us to interpret the probability value of a sentence as the likelihood that a competent speaker would endorse an assertion, given certain conditions (hypotheses).
- Therefore, predication remains intrinsically vague.
- It consists in applying a classifier to new instances on the basis of supervised training.
- We are not obliged to posit a contextually dependent cut off boundary for graded (or non-graded) predicates.

Related Work: Goodman and Lassiter (2015), Lassiter and Goodman (2017)

- The approach that I am suggesting here does not assume such an inaccessible boundary point for predicates.
- It allows us to interpret the probability value of a sentence as the likelihood that a competent speaker would endorse an assertion, given certain conditions (hypotheses).
- Therefore, predication remains intrinsically vague.
- It consists in applying a classifier to new instances on the basis of supervised training.
- We are not obliged to posit a contextually dependent cut off boundary for graded (or non-graded) predicates.

Related Work: Goodman and Lassiter (2015), Lassiter and Goodman (2017)

- The approach that I am suggesting here does not assume such an inaccessible boundary point for predicates.
- It allows us to interpret the probability value of a sentence as the likelihood that a competent speaker would endorse an assertion, given certain conditions (hypotheses).
- Therefore, predication remains intrinsically vague.
- It consists in applying a classifier to new instances on the basis of supervised training.
- We are not obliged to posit a contextually dependent cut off boundary for graded (or non-graded) predicates.

Related Work: Goodman and Lassiter (2015), Lassiter and Goodman (2017)

- The approach that I am suggesting here does not assume such an inaccessible boundary point for predicates.
- It allows us to interpret the probability value of a sentence as the likelihood that a competent speaker would endorse an assertion, given certain conditions (hypotheses).
- Therefore, predication remains intrinsically vague.
- It consists in applying a classifier to new instances on the basis of supervised training.
- We are not obliged to posit a contextually dependent cut off boundary for graded (or non-graded) predicates.

Related Work: Bernardy et al. (2018)

- Bernardy, Chatzikyriakidis, Blanck, and Lappin (2018) propose a compositional Bayesian semantics that implements the approach proposed here in a functional probabilistic programming language similar to Church.
- It generates probability models that satisfy a set of specified constraints.
- It uses Markov Chain Monte Carlo sampling to estimate the likelihood of a sentence being true in these models.
- It implements a small scale Bayesian paradigm of semantic learning.

Related Work: Bernardy et al. (2018)

- Bernardy, Chatzikyriakidis, Blanck, and Lappin (2018) propose a compositional Bayesian semantics that implements the approach proposed here in a functional probabilistic programming language similar to Church.
- It generates probability models that satisfy a set of specified constraints.
- It uses Markov Chain Monte Carlo sampling to estimate the likelihood of a sentence being true in these models.
- It implements a small scale Bayesian paradigm of semantic learning.

Related Work: Bernardy et al. (2018)

- Bernardy, Chatzikyriakidis, Blanck, and Lappin (2018) propose a compositional Bayesian semantics that implements the approach proposed here in a functional probabilistic programming language similar to Church.
- It generates probability models that satisfy a set of specified constraints.
- It uses Markov Chain Monte Carlo sampling to estimate the likelihood of a sentence being true in these models.
- It implements a small scale Bayesian paradigm of semantic learning.

Related Work: Bernardy et al. (2018)

- Bernardy, Chatzikyriakidis, Blanck, and Lappin (2018) propose a compositional Bayesian semantics that implements the approach proposed here in a functional probabilistic programming language similar to Church.
- It generates probability models that satisfy a set of specified constraints.
- It uses Markov Chain Monte Carlo sampling to estimate the likelihood of a sentence being true in these models.
- It implements a small scale Bayesian paradigm of semantic learning.

Conclusions

- The tradition of formal semantics which uses possible worlds to model intensions, modality, and epistemic states is not built on cognitively viable foundations.
- By adapting the distinction between operational and denotation semantics to natural language it is possible to develop a fine-grained treatment of intensions that dispenses with possible worlds.
- We use probability models to interpret modal expressions, and Bayesian networks to encode knowledge, belief, and inference.
- Stratification, estimation, and approximation techniques allow us to effectively represent significant subclasses of these models.
- Therefore they offer a computationally realistic basis for handling epistemic states and inference.

Conclusions

- The tradition of formal semantics which uses possible worlds to model intensions, modality, and epistemic states is not built on cognitively viable foundations.
- By adapting the distinction between operational and denotation semantics to natural language it is possible to develop a fine-grained treatment of intensions that dispenses with possible worlds.
- We use probability models to interpret modal expressions, and Bayesian networks to encode knowledge, belief, and inference.
- Stratification, estimation, and approximation techniques allow us to effectively represent significant subclasses of these models.
- Therefore they offer a computationally realistic basis for handling epistemic states and inference.

Conclusions

- The tradition of formal semantics which uses possible worlds to model intensions, modality, and epistemic states is not built on cognitively viable foundations.
- By adapting the distinction between operational and denotation semantics to natural language it is possible to develop a fine-grained treatment of intensions that dispenses with possible worlds.
- We use probability models to interpret modal expressions, and Bayesian networks to encode knowledge, belief, and inference.
- Stratification, estimation, and approximation techniques allow us to effectively represent significant subclasses of these models.
- Therefore they offer a computationally realistic basis for handling epistemic states and inference.

Conclusions

- The tradition of formal semantics which uses possible worlds to model intensions, modality, and epistemic states is not built on cognitively viable foundations.
- By adapting the distinction between operational and denotation semantics to natural language it is possible to develop a fine-grained treatment of intensions that dispenses with possible worlds.
- We use probability models to interpret modal expressions, and Bayesian networks to encode knowledge, belief, and inference.
- Stratification, estimation, and approximation techniques allow us to effectively represent significant subclasses of these models.
- Therefore they offer a computationally realistic basis for handling epistemic states and inference.

Conclusions

- The tradition of formal semantics which uses possible worlds to model intensions, modality, and epistemic states is not built on cognitively viable foundations.
- By adapting the distinction between operational and denotation semantics to natural language it is possible to develop a fine-grained treatment of intensions that dispenses with possible worlds.
- We use probability models to interpret modal expressions, and Bayesian networks to encode knowledge, belief, and inference.
- Stratification, estimation, and approximation techniques allow us to effectively represent significant subclasses of these models.
- Therefore they offer a computationally realistic basis for handling epistemic states and inference.

Future Work

- The proposed approach will have to integrate the operational view of intensions into the probabilistic treatment of knowledge and belief.
- It must explain how intensions are acquired by Bayesian learning processes.
- it must develop a wide coverage system that combines a compositional semantics with a procedure for generating probability models in which it is possible to sample a large number of predicates.
- Bernardy et al. (2018) provide an initial prototype for this system.
- Much work remains to be done on both the compositional semantics and the model testing components in order to create a robust Bayesian framework for natural language interpretation.

Future Work

- The proposed approach will have to integrate the operational view of intensions into the probabilistic treatment of knowledge and belief.
- It must explain how intensions are acquired by Bayesian learning processes.
- it must develop a wide coverage system that combines a compositional semantics with a procedure for generating probability models in which it is possible to sample a large number of predicates.
- Bernardy et al. (2018) provide an initial prototype for this system.
- Much work remains to be done on both the compositional semantics and the model testing components in order to create a robust Bayesian framework for natural language interpretation.

Future Work

- The proposed approach will have to integrate the operational view of intensions into the probabilistic treatment of knowledge and belief.
- It must explain how intensions are acquired by Bayesian learning processes.
- it must develop a wide coverage system that combines a compositional semantics with a procedure for generating probability models in which it is possible to sample a large number of predicates.
- Bernardy et al. (2018) provide an initial prototype for this system.
- Much work remains to be done on both the compositional semantics and the model testing components in order to create a robust Bayesian framework for natural language interpretation.

Future Work

- The proposed approach will have to integrate the operational view of intensions into the probabilistic treatment of knowledge and belief.
- It must explain how intensions are acquired by Bayesian learning processes.
- it must develop a wide coverage system that combines a compositional semantics with a procedure for generating probability models in which it is possible to sample a large number of predicates.
- Bernardy et al. (2018) provide an initial prototype for this system.
- Much work remains to be done on both the compositional semantics and the model testing components in order to create a robust Bayesian framework for natural language interpretation.

Future Work

- The proposed approach will have to integrate the operational view of intensions into the probabilistic treatment of knowledge and belief.
- It must explain how intensions are acquired by Bayesian learning processes.
- it must develop a wide coverage system that combines a compositional semantics with a procedure for generating probability models in which it is possible to sample a large number of predicates.
- Bernardy et al. (2018) provide an initial prototype for this system.
- Much work remains to be done on both the compositional semantics and the model testing components in order to create a robust Bayesian framework for natural language interpretation.