Cognitive Science with Category Theory

Ash Asudeh
University of Oxford & Carleton University
May 6, 2018 | Boğaziçi University

Introduction

- Language can be understood abstractly as a mapping between form and meaning.
- Linguistic theory uses structures (basic elements and relations between them) to represent:
  - Linguistic form (phonology)
  - Linguistic meaning (semantics)
  - The structures mediating them (syntax).
- Mappings between different linguistic representations (interfaces) can therefore be understood as mappings between structures.
- Category theory — a branch of mathematics that has been applied to problems in theoretical physics and theoretical computer science — is concerned with such structural mappings, and is therefore well-placed to help us understand how linguistic structures may, when necessary, be mapped to more complex structures.

Introduction

- In this talk, I present an overview of a seven-year research program with Dr. Gianluca Giorgolo, in which we have explored applications of a particular construction from category theory, the **monad**, to problems in linguistics and cognitive science.

- Monads are a key way to understand **computation** in **computer science**.

- Our work develops in detail some original insights by Ken Shan.

- The capstone work in this collaboration will be our forthcoming Oxford University Press monograph, ***Enriched Meanings: Natural Language Semantics with Category Theory***.

Introduction

- It is noteworthy that this approach reveals interesting solutions to a diverse range of problems in **cognitive science**, all from the perspective of **linguistic theory**, using a formal tool from **computer science**.

    - **Conventional Implicature**, a phenomenon important to linguistic semantics and pragmatics and philosophy of language

    - **Conjunction Fallacies**, a phenomenon important to cognitive psychology, logic, and philosophy of language

    - **Substitution Puzzles**, a phenomenon important to semantics, philosophy of language, and philosophy of mind

- In short, this is a **highly interdisciplinary collaboration** involving a **linguist** (me) and a **logician/computer scientist** (Gianluca) that is arguably at the **leading edge** of symbolic cognitive science, and in fact is a demonstration of the continuing relevance of **highly formalized approaches** to understanding **cognition**, as we see an idea from **computer science** helping to explain problems in **linguistics**, **psychology**, and **philosophy**, all from a **linguistic and logical perspective**.

Overview

- A very little bit of formal background, just focusing on the intuitions
- Conventional Implicatures (**`Writer`** monad)
- Conjunction Fallacies (**`Probability`** monad)
- Substitution Puzzles (**`Reader`** monad)

Formal Background

Category Theory: Intuitions

- We can think of a category as a way to describe a complex ensemble of objects by specifying the ways in which the objects interact.
- This is the approach we take in our project as far as linguistic meanings are concerned.
    - Our objects will be different kinds of meanings (entities, propositions, predicates, relations, quantifiers…).
    - From the category-theoretic perspective, though, instead of focusing on what they *are*, we focus on how they *interact*, and in particular how they can be composed to construct more complex meanings.
- The formal definition: A category C is a collection C_1 of objects together with a collection C_2 of arrows or morphisms such that the objects and arrows satisfy certain conditions, which I omit here. .
- There is an introduction to category theory in our ESSLLI 2015 course notes, available from my website.

Monads: Intuitions

- A construction in category theory that defines a canonical way to map a set of objects and functions that we may consider simple in some sense into a more complex space of objects and functions

- A monad can be thought of as a way to reproduce the structure of a space of values and functions in a richer setting that carries more information, in the sense that we can specify more things about the values and functions.

- The idea is that we can move from the information-poor space to the information-rich space by mapping a value or function in the poor space to an information-enriched counterpart.

- A little more formally, monads are a particular type of functor in category theory that comes with two well-defined natural transformations which encode the notions of *embedding* and *combination/composition*.

Monads: Intuitions

- It's hopefully then intuitively clear why monads might be of particular interest in formal semantics, because embedding and combination are in essence the basic notions of semantic theory, as enshrined in the fundamental semantic principle:

  The Principle of Compositionality
  The meaning of a linguistic expression is determined by its parts (***embedding***) and their syntactic arrangement (***combination***).

- **Key compositional intuition:**
  Monads provide a means to capture certain complex linguistic meanings *without pushing complexity back into the syntax* or *generalizing to the worst case.*

Conventional Implicatures

Conventional implicature

- Conventional implicatures were introduced in the original, foundational work on implicature by Grice, but were subsequently more or less conflated with presupposition and thus received relatively little focused attention.

- This changed with the pioneering modern work of Potts.

- He argued convincingly that CIs are a distinct category from presupposition because their meaning involves a kind of foregrounded but side-lined information rather than information that is taken for granted, as is the case for presupposition.

- Conventional implicature raises challenges for compositionality and for the semantics/pragmatics divide, which it seems to somehow straddle.

Dog poop blues

- Two categories of CI-contributing linguistic expressions are *expressives* (which are well-illustrated by salty language, so please pardon my French), as in (1), and *appositives*, as in (2).

  1. **A:** Most fucking neighbourhood dogs crap on my damn lawn.
     B: No, that's not true.
     ⇒ The neighbourhood dogs don't defecate on your lawn

     ⇏ There's nothing wrong with neighbourhood dogs and/or their defecating on your lawn.

  2. **A:** John Lee Hooker, the bluesman from Tennessee, appeared in *The Blues Brothers*.
     B: No, that's not true.
     ⇒ John Lee Hooker did not appear in The Blues Brothers.

     ⇏ John Lee Hooker was not from Tennessee.

     B: True, but actually John Lee Hooker was born in Mississippi.

Multidimensionality

- Potts introduced what has become an influential approach to conventional implicature: **multidimensional semantics**.
- He proposed that there are two dimensions of meaning:
  - The ‘at-issue dimension’, which represents the main assertion (what’s at issue, and hence open to negation, denial, etc.)
  - The ‘CI-dimension’, which collects the side comments. I’ll call this the ‘side-issue dimension’.
- For example, the interpretation of *Most fucking neighbourhood dogs crap on my damn lawn* would be:
  At issue:
  `Most neighbourhood dogs defecate on the speaker’s lawn`
  Side issue:
  `The speaker feels negatively towards neighbourhood dogs and/or their toilet habits on the speaker’s lawn`

Flow of information

- Potts contended that information flows from the at-issue dimension to the side-issue dimension, but not vice versa.
    - The conventional implicature is a comment on material in the main assertion, but the main assertion seems insensitive to the conventional implicature dimension.
    - This provides a natural explanation of some key properties of the phenomenon of conventional implicature; for example, that CIs do not interact with logical operators and are strongly commitments of the speaker — only contexts that somehow change the speaker, such as strongly quotative contexts, seem to allow change of CI attribution (*My father screamed that I would never marry that asshole*).
- In order to accomplish this, though, Potts developed a theory that had to replicate some aspects of the at-issue dimension in the side-issue dimension and had to introduce a mechanism for passing the conventional implicature information up to the top of the representation of the interpretation.

A challenge: Backflow

- In subsequent work, AnderBois, Brasoveanu & Henderson challenged the assumption about information flow, based on examples like the following.

3. Mary, a good **drummer**, is a good singer **too**.

4. Jake, who almost killed **a woman** with his car, visited **her** in the hospital.

5. Lucy, who doesn't **help her sister**, told Jane to ~~**help her sister**~~.

6. Melinda, who won three **games of tennis**, lost because Betty won six ~~**games of tennis**~~.

- ABH therefore propose that, contra Potts, there is a single dimension of meaning but two *modes of discourse update.*
    - At-issue content is *proposed*, whereas *side-issue* content is *imposed*.

Problems

- A potential problem for the ABH approach, quite apart from its potential dissolution of some of the original gains of multidimensionality, is that a unidimensional representation does not easily support analysis of certain other cases.

7. Luke Skywalker is so gullible that he believes that Jabba the Hutt, a notorious scammer, is a trustworthy business partner.

8. All **Cairo taxi drivers**, who by the way painted **their** taxis red in protest, are on strike.

9. * **Every Cairo taxi driver**, who by the way would threaten me with **his** gun, is on strike.

Our solution

- A function of monads in computer science is to model the notion of computation with a *side effect*: Some general computation is performed with other background effects.

- Shan had the intuition that we could use monads to similarly compute apparently non-compositional (or side-compositional, if you will) natural language semantics as a side effect of computing the main interpretation.

- We build on this intuition as follows:
  Expressions that contribute conventional implicatures are seen as computations that contribute possibly nothing* to the at-issue dimension, while also *logging* as a side-effect something to the side-issue dimension.

  [* Or possibly something, as in expressive verbs like *crap*, which can still also contribute a main predication about defecation, or slurs like *Snow Mexican,* which also denotes as a common noun — *Trudeau is just a Snow Mexican* — and is particularly interesting as it is a slur against one people by virtue of using a neutral term for another people derogatorily.]

Our solution

- The operation of logging can be modelled in theoretical computer science with the **`Writer`** monad.

- We also add an auxiliary monadic function **`Check`** that checks presuppositional or anaphoric conditions logged during the compositional process.

- In short, we treat the kinds of apparently exceptional phenomena noted by AnderBois et al. as involving a post-compositional check on discourse coherence, but this checking is lexically encoded in presupposition triggers, such as *also*.

- This allows us to maintain the benefits of the original multidimensional treatment in a more principled compositional setting, while also capturing the ABH phenomena lexically, without *ad hoc* operations in a unidimensional representation.

Conjunction Fallacies

Conjunction fallacies

- Conjunction fallacies have been an active area of research since the pioneering work of Tversky & Kahneman.
- T&K's experiments showed that in tasks asking for ratings of the relative likelihoods of different events, the majority of the participants consistently rated the likelihood of the conjunction of two events as higher than the likelihood of one of the conjoined events.
- The take-home for many cognitive scientists is that humans are bad at logical reasoning, about probabilities but also more generally, because they fail to obey the following axiom/theorem of probability theory:

$$P(A \text{ and } B) \leq P(A), P(B)$$

Linda, Linda, Linda

- One of T&K's famous examples is the so-called **Linda Paradox**.

10. Linda is 31 years old, single*, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

    - Typical subject's ranking of probability that following statements about Linda are true:

      `Linda is active in the feminist movement. [Feminist]`

      `Linda is a bank teller and is active in the feminist movement. [Teller & Feminist]`

      `Linda is a bank teller. [Teller]`

[* Note the effortless, passive sexism/anti-feminism: "Don't practice feminism if you want to land a guy, girls!" 🤮]

Some Grice would be nice

- Most linguists or philosophers of language would probably look at this set-up and point out that nothing in the context is relevant to the question of whether Linda is a bank teller, which means that unless the `Feminist` judgement is involved, the task is basically asking the participant to assume that the "speaker" is acting uncooperatively, from a Gricean perspective.

- Nevertheless, many psychologists found this compelling and a cottage industry on human reasoning grew that took this result, and Tyversky & Kahneman's attendant *representativeness heuristic*, as more or less foundational.

- Our compositional monadic solution builds on a different intuition about humans.

  - We are lazy problem solvers, not necessarily bad reasoners —- we engage in *satisficing* (Simon), when possible.

Stakes

- Various researchers, most notably Hertwig and colleagues and Yates & Carlson, have shown that the conjunction fallacy disappears or is mitigated under certain circumstances:
  - When there are **stakes**, as in betting scenarios
  - When **frequencies** are explicitly introduced in the scenario
  - When subjects are asked for **numerical estimates** rather than likelihood rankings
- These results could be taken to show that there is not necessarily a fundamental problem of reasoning, but rather that subjects might pursue distinct cognitive strategies depending on the task.

An alternative model

- Yates & Carlson argue that conjunction fallacies arise in the absence of stakes because people use distinct strategies on such occasions to evaluate the combination of multiple uncertain events.

- They model the strategy that generates the fallacies with what they call a “signed summation” model.

- The idea is to substitute probabilities with a different likelihood measure, λ, that takes values in the entire range $\mathbb{R}$ of real number values (i.e., not just in the [0,1] interval).

- Likely events are assigned a positive number as their likelihood measure, whereas unlikely events are assigned a negative one. According to this model the joint likelihood of two events is the sum of their likelihoods:

$$\lambda(\texttt{A and B}) = \lambda(\texttt{A}) + \lambda(\texttt{B})$$

- Unfortunately the Y&C model is somewhat *ad hoc* and it is not clear how to relate their likelihood estimates in a principled way to general probability theory.

Our model: Intuitions

- Our model shares with Y&C the assumption that there are **multiple strategies** that are employed by people when evaluating the likelihood of joint uncertain events.
- But instead of assuming that **unrelated computational processes** underpin the different strategies, our model shows that it is possible to assume a single uniform process that computes the likelihood of the conjunction of two events from their two relative likelihoods.
    - The model does this by using different but related representations of uncertainty, expressed as alternative algebraic structures.
- More specifically, we can explain the results reported in the literature in terms of an algebraic structure known as a *semiring*.
- Monads can be understood in terms of semirings, so this is actually a generalization of the monadic approach we have already seen.

Conjunction fallacies as a consequence of cognitive/computational economy

- In our model the observation that conjunction fallacies arise only under specific conditions and can be cancelled if other conditions are imposed is explained in terms of **cognitive/computational economy**.

- The same computational structure, the monad, can be used together with different underlying semirings, one of them being the probability semiring.

Conjunction fallacies as a consequence of cognitive/computational economy

- Our model predicts that if people are presented with a task where there are “no stakes”, they will base their judgement on the basis of reasoning modelled using a representation corresponding to a semiring defined over a relatively simple set with generally simple operations.

  - This strategy is simpler, but will generally lead people to make overconfident likelihood estimations, which do not necessarily correspond to those of probability theory.

  - If, on the other hand, people are forced to evaluate the consequences of their judgements, as in the context of stakes, or if explicitly primed to think in terms of frequencies, then they will switch to a more complex representation, with properties that better approximate those of probability theory.

Keeping it nice, Keeping it Grice

- Crucially, in our model, logical operators such as *and* or *or* maintain their core logical meaning, while the probabilistic behaviour is determined by the context in which they operate.

- Unlike previous approaches, such as Hertwig et al., we place the ambiguity in the **context**, rather than assuming that a word like *and* has multiple meanings, which philosophers of language, notably Grice, have resisted strongly.

- We instead follow the standard Gricean approach: We treat *and* unambiguously as **logical conjunction** but allow it to derive additional *implicatures* depending on the context of use.

Keeping it nice, Keeping it Grice

- We thus recuperate Grice's fundamental intuition that, despite the fact that the same words are used with different meanings, speakers are not necessarily confused about their semantic meanings, as linguistic expressions are always evaluated with respect to a context.

- The strategy is to try to explain distinct meanings as contextual modification of a core meaning.

  Grice's Razor ("Modified Occam's Razor")
  Senses are not to be multiplied beyond necessity.

- This is, in effect, one of the foundational assumptions of modern pragmatics.

The probability semiring

- Our monadic solution depends on the `Probability` monad and the notion of a semiring

- A semiring is a **set A** with **two distinguished elements**, 0 and 1, associated with **two operations**, + and ×, such that 0 is the unit for + and 1 is the unit for ×.

    - There is also a set of axioms that define the semiring based on these operations and their associated unit elements, but I omit these here.

- We are particularly interested in the probability semiring, where A is the real interval [0,1], with 0 and 1 representing the two units and + and × are defined in the usual way.

The probability monad

- The fact that probability distributions form a monad in the category of measurable spaces was an early discovery in category theory (Lawvere).

- We use a slightly different characterization based on the use of monads in functional programming to model probabilistic calculi.

- The space of simple objects is represented by the collection of different semantic types of natural language expressions, such as entities/individuals, propositions, and collections of entities.

The probability monad

- The mappings between these objects are expressed by similar expressions that “bring” us from one type of expression to another: For example we take a predicate expressed by a verb as a way to map the individual denoted by the subject expression to a truth value.
    - For example, for *John sleeps* the interpretation of the verb *sleeps* maps the individual John to true if John is indeed asleep and to false otherwise.
- The **`Probability`** monad lifts these simple objects to probability distributions over inhabitants of the types, and so the mappings are transformed so that they relate different probability distributions
    - *John sleeps*: The predicate would map from the probability distribution corresponding to the denotation of the subject, one that possibly assigns the entire probability mass to the individual John, to the probability distribution over the truth values, effectively giving us an estimate of the likelihood of the event that John is asleep, given the model.

The monad and the semiring

- All operations involved in the definition of the **`Probability`** monad, details of which I leave aside here, are those of a semiring.

- This means that we can use the same general structure of the semiring but replace the meanings of 0, 1, ×, and + with constants and operations defined for other semirings.

- This is what allows us to reproduce the reasoning results reported in the literature using the **`Probability`** monad.

Conjunction fallacies, compositionally

- Yates & Carlson's data is the only instance in the literature, at least that we are aware of, where the relative likelihood of the atomic events in conjunctions has been (at least partially) controlled for.
- This gives us the possibility of deriving the overall likelihood of the conjoined event in a compositional fashion starting from the atomic events.
- Our monadic infrastructure can reproduce the results reported by Y&C.

Yates & Carlson's results (simplified)

Likelihood of event A	Likelihood of event B	Observed rating
U	U	$P(A \text{ and } B) \leq P(A), P(B)$
U	L	$P(A) \leq P(A \text{ and } B) \leq P(B)$
L	L	$P(A), P(B) \leq P(A \text{ and } B)$

Put a semiring on it

- The first step is to define a suitable base for our semiring.
- Y&C use a discrete scale based on the general prediction made by their summation model.
- Their model does not take into account the limiting cases, i.e. impossible and certain events.
- These are nevertheless included in our approach, as they are necessary in order to model what we know about the logical entailment behaviour of the word *and*.
  - Tentori, Bonini & Osherson showed that participants who make conjunction errors can nevertheless correctly apply the rules of logic.

Put a semiring on it

- Therefore we will use a simple discrete set as the base for our semiring:
  { I(mpossible), U(unlikely), P(ossible), L(ikely), C(ertain) }

- I and C correspond to the distinguished elements 0 and 1 respectively.

- The only additional condition that is imposed, so that I and C behave as boolean values, is that for all **x** in our set
  x+C = C.

- There are sixteen possible semirings that can then be defined for this set, given the additional condition.

Picking the right semiring

- We now evaluate how the sixteen semirings match the simplified Y&C results, repeated here.

Likelihood of event A	Likelihood of event B	Observed rating
U	U	$P(A \text{ and } B) \leq P(A), P(B)$
U	L	$P(A) \leq P(A \text{ and } B) \leq P(B)$
L	L	$P(A), P(B) \leq P(A \text{ and } B)$

- It turns out that only one of the sixteen semirings yields this pattern of results.

- This semiring is not homomorphic to the probability semiring: We cannot reproduce its behaviour using probability theory.

- We call this the **one semiring**.

One semiring to rule them all

+	I	U	P	L	C
I	I	U	P	L	C
U	U	U	P	L	C
P	P	P	P	L	C
L	L	L	L	L	C
C	C	C	C	C	C

×	I	U	P	L	C
I	I	I	I	I	I
U	I	U	U	U	U
P	I	U	P	P	P
L	I	U	P	L	L
C	I	U	P	L	C

−	I	U	P	L	C
	C	L	P	U	I

One process, two representations

- If there are no real stakes, and therefore there is no incentive to use a probabilistically accurate but costlier representation, people will employ a form of shortcut, represented in our model by the simpler **one semiring**.

- If, on the other hand, people are pushed to think about the consequences of their judgements, the cognitively costlier probability semiring is used as the representation.

- The **one semiring** is undoubtedly simpler than the standard probability semiring, because the **one semiring** is based on a simpler base set.

One process, two representations

- The **one semiring** also has another important property: It is in fact possible to reconstruct the entire semiring on the basis of only one of the two operations (+ and ×) and a simpler complement operation, −.

−	I	U	P	L	C
	C	L	P	U	I

- If we define the complement as above, which seems to be the most intuitive way to define it — since we just reverse the order of the relative likelihoods — then we can observe that for all **y** and **z** in our base set we have that $\mathbf{y} \times \mathbf{z} = -(-\mathbf{y} + -\mathbf{z})$ or alternatively $\mathbf{y} + \mathbf{z} = -(-\mathbf{y} \times -\mathbf{z})$.
  - These laws also demonstrate the logical naturalness of the one semiring, given the standard relationship between × and ∧ (logical conjunction) and between + and ∨ (logical disjunction):
    - The laws correspond to **De Morgan's Laws** in the semiring of truth values.
- In sum, this particular encoding of uncertainty has much lower representational costs than other competing possibilities.
- And its symmetry makes it a particular simple and efficient computational object.

One process, two representations

- From a processing perspective this means that we can posit a single compositional process that computes the likelihood of two conjoined events in a way that is completely blind to the specific details of how uncertainty is encoded.
- We just require that the encoding satisfies the axioms of a semiring.
- The two encodings (the one semiring and the probability semiring) are instead prompted by the context in which the description of the conjoined event is evaluated.
    - If there are no real risks and judgements have no real consequences, then people select a computationally and representationally cheap encoding (the **one semiring**).
    - Otherwise they will use the probability semiring and apply the rules of probability, which require a much higher degree of computation and are representationally more costly.

Are you satisficed?

- In other words, we do assume that conjunction fallacies are true errors, in the sense that they lead to overestimation of the likelihood of events.

- However, they are errors due to a standard and independently motivated heuristic of cognitive/computational economy, Simon's *satisficing heuristic.*

- They are not due to Tversky & Kahneman's *representativeness heuristic*.

- Note that we are not claiming that the *representativeness heuristic* could not be deployed in other cognitive tasks, just that it is not, in fact, the right explanation of conjunction fallacies.

The monadic approach: Further advantages

- The advantage of this approach over other previous models is that it bridges an important hypothesis in the study of natural language semantics, *compositionality*, also in light of standard Gricean pragmatics, and a pervasive apparent reasoning problem, the conjunction fallacy.

- The model also explains why different strategies for evaluating uncertain events are selected, based on a simple computational criterion, and moreover the notion of cognitive/computational economy that we formally capture can be understood in light of the one of the foundational principles of cognitive science: satisficing.

The monadic approach: Further predictions

- The model also makes predictions that could be tested in future work.

- The purely compositional nature of our model means that we can apply it, as is, to other cases that involve the combination of different events via logical operators, such as disjunction and implication.

- The model predicts that similar effects should also be observable in cases where the conjunction of events is implicit, such as in the case of universally quantified sentences, which can be understood as an iterated conjunction over the domain of quantification.

    - If the equivalent of conjunction fallacies are also observed in these cases, this would provide important evidence for a compositional model.

The monadic approach: Further predictions

- The model also suggests new ways in which we can prime participants to not commit reasoning fallacies related to uncertainty.

- If we are correct in assuming that people select a specific representation for evaluating uncertain events on the basis of the possible repercussions of their choices, then we predict that we could force participants to select an accurate, but cognitively more expensive, probabilistic reasoning strategy by simply introducing such consequences.

- Betting is of course one example, but other possibilities that suggest themselves are situations with emotional or social costs.

Substitution Puzzles

Frege's puzzle

- An important problem in the philosophy of language and the linguistic semantics and pragmatics concerns co-referential terms and substitutability in different contexts.

- This problem is commonly associated with Frege, and is often called *Frege's puzzle*.

- The puzzle can be presented in various ways, but its essence can be captured as follows: given two co-referential linguistic expressions, why is it that in certain linguistic contexts substitution of one expression for the other is truth-preserving, while in others it is not?

A standard sort of demonstration

- For example, given that (11) is true, since *Hesperus* and *Phosphorus* are different names for the planet Venus, how can it be that (12) can be true while (13) is false?

11. Hesperus is Phosphorus.

12. Kim believes that Hesperus is a planet.

13. Kim believes that Phosphorus is a planet.

- Alternatively, we could characterize the puzzle by observing that a sentence like the following can be true without entailing that Kim does not believe a tautology:

14. Kim doesn’t believe that Hesperus is Phosphorus.

Frege & Quine

- Frege's own solution was that in addition to a *reference*, nominals have a *sense*, or 'mode of presentation', and that in certain contexts, such as those involving propositional attitudes, it is these distinct senses that block substitutability.
- Frege's puzzle is thus clearly related to the *problem of referential opacity* in the study of propositional attitudes, as discussed by Quine and many others.

Expanding the scope

- We take it for granted that there is an empirical phenomenon to be explained here — differing truth value judgements despite substitution of co-referential terms — and offer a formal mechanism for capturing and explaining it semantically.

- We follow Saul in observing that problems of substitutability also arise in 'simple sentences', which lack embedding under propositional attitudes.

- There are also cases of differential interpretation of the *same* expression, as in Kripke's famous Paderewski puzzle.

- Moreover, our approach yields insight into cases that concern differential interpretation of expressions other that referential expressions, as discussed by Carnap, Mates, and Kripke, among others.

- Our proposal is not only formally well-founded on the notion of monads, it also gives us the beginnings of a *general* semantics of what we might informally call **perspective**.

Shortcomings of the standard view

- The substitutability puzzle is standardly characterized as involving two factors:

  1. Embedding under a modal or propositional attitude expression, such as *believe*; and

  2. Co-referential but distinct terms, such as *Hesperus* and *Phosphorus*

- It has been shown in the literature that neither of these factors is necessary for the substitutability puzzle or related puzzles to arise, which yields a typology of four major classes of substitution puzzles.

The actual problem space

	Simple	Embedded
Same term	#Dr. Octopus punched Spider-Man but he didn't punch Spider-Man.	Kim doesn't believe Sandy is Sandy.
Distinct term	Mary Jane loves Peter Parker but she doesn't love Spider-Man. #Dr. Octopus killed Spider-Man but he didn't kill Peter Parker.	Kim doesn't believe Hesperus is Phosphorus.

Saul: Simple sentences

- Saul points out that lack of substitutability can hold even in ‘simple sentences’ that ‘contain no attitude, modal or quotational constructions’.

  - Assuming it is common knowledge that Clark Kent is Superman’s secret identity, she notes that if (15) is true, substitution of Clark Kent for Superman seems to render (16) false.

15. Clark Kent went into the phone booth, and Superman came out.

16. Clark Kent went into the phone booth, and Clark Kent came out.

Saul: Simple sentences

- With respect to this pair, an obvious out presents itself: Why would someone *say* (16) if they *meant* (15)?
  - That is, it seems that we could say these sentences *are* in fact semantically synonymous (so (16) is not false when (15) is true), but pragmatically distinct.
  - This is in fact basically what Saul argues: What is mistaken is our *intuition* that (15) is true while (16) is false.

A simple sentence with less wiggle room

- Let's focus instead on a different simple sentence that makes the same point, but seems to not offer as much pragmatic wiggle room:*

  17. Mary Jane loves Peter Parker, but she doesn't love Spider-Man.

- Let's assume that the time of evaluation is a point in the stories before Mary Jane knows that Peter Parker is Spider-Man.

- There is a non-contradictory reading of this sentence.

- Saul would seem to have to treat this sentence as simply false, but this seems to entirely set aside Mary Jane's say in the matter, which strikes us as problematic.

  [* We prefer to use Spider-Man in our examples, because Superman is frankly kind of boring, but also because the Peter Parker/Spider-Man case involves a different (yet still familiar) set-up: it is not as clear which is whose secret identity, since Peter Parker is as much the "main character" in those stories as Spider-Man is. This avoids the problem of Pitt's concept of *primum egos*. It also avoids the problem that both Superman and Clark Kent are in fact secret identities of a third identity, Kal-El, a Kryptonian **refugee**.]

A simple sentence with less wiggle room

- It would be very strange to insist that if Mary Jane loves Peter Parker, then she really does love Spider-Man.
    - *She* certainly wouldn't agree to that.
- In short, (17) shows a lack of substitutability in a grammatically very simple sentence and this lack of substitutability does not yield as readily to a pragmatic explanation as do Saul's examples.
- Instead, (17) seems to crucially involve Mary Jane's **perspective**.
- This is the intuition that our approach builds on.

Guises, guys?

- We thus take seriously Castañeda's insight that "All references are made and conceived from a point of view."

- Despite this, the perspectival approach actually distinguishes our approach from one such as Castañeda's own, where a sentence like *Mary Jane loves Peter Parker, but she doesn't love Spider-Man* is interpreted as simply saying that Mary Jane loves only one *guise* of the entity that corresponds to Peter Parker but not another.

- Guises are also explored in important work by Heim.

- If it is indeed the case that different co-referring expressions simply pick out different guises of the same individual, then a sentence like (18) should have a non-contradictory reading, but this does not seem to be the case (assuming it is indeed Peter Parker who is Spider-Man at the time):

18. # Dr. Octopus killed Spider-Man but he didn't kill Peter Parker.

Love, Punching, Killing, and Murder

- Moreover, the theory must capture the difference between, for example, *kill* and *murder*, since murder does involve intention and the minimal pair of (18) with *murder* substituted for *kill* is not contradictory (in the same circumstance):

    19. Dr. Octopus murdered Spider-Man but he didn't murder Peter Parker.

- Although *murder* involves intention, it is not a propositional attitude verb and there is no obvious evidence of embedding.

- Lastly, whatever analysis we give must not lose sight of the fact that there are genuine cases of contradiction that must still be derivable, such as the following:

    20. # Dr. Octopus punched Spider-Man but he didn't punch Spider-Man.

- What unites *murder* and *love* versus *kill* and *punch* is the fact that for the former, the subject/agent's perspective can be part of the interpretation.

Non-distinct terms but distinct beliefs: The Paderewski Puzzle

- Kripke presents a puzzle that is closely related to the substitutability puzzle, but which relates to the second factor mentioned above: whether the terms involved must be distinct.

- He considers the case of ‘phonetically identical tokens of a single name’.

- He provides the example of an individual, Peter, who has learned that Paderewski was the name of an accomplished Polish pianist. The following then seems true:

21. Peter believes that Paderewski had musical talent.

Non-distinct terms but distinct beliefs: The Paderewski Puzzle

- Peter then hears of a Polish politician named Paderewski, and concludes that this is a different person, since he has no reason to believe that politicians make good musicians. Given that in fact the same Paderewski was in fact both a politician and a pianist, is the following true or not?

  22. Peter believes that Paderewski had no musical talent.

- Kripke argues that this is a true paradox and we can neither conclude that (22) is true nor false, given the situation.

Disentangling the Paderewski Puzzle

- Kripke's conclusion that we are dealing with a paradox seems to us motivated by the interplay between the perspectival dimension introduced by the verb *believe* together with the ambiguous nature of the name *Paderewski* in the context of Peter's lexicon.

- In this case we not only have different perspectives regarding the interpretation of a term (the speaker's and Peter's), but the two interpretations also have different cardinalities.

- Given that Peter can use the name *Paderewski* to refer to two different (from Peter's perspective) entities, in an example like (22) it is not possible to resolve whether we are talking about Peter's belief with regard to the pianist entity or the politician one.

Disentangling the Paderewski Puzzle

- Therefore (22) seems to lack a determinate truth value: It is true with respect to Paderewski the politician, but false with respect to Paderewski the musician.

- There are competing interpretations, but each one is fully interpretable and can be assigned a truth value.

- Of course, this move itself only makes sense if the two instances of the name Paderewski in fact do not refer to one and the same entity for Peter, which is not possible for Kripke, but is possible on our monadic perspectival approach.

Identity Statements: Delusions

- We demonstrate the generality of substitution puzzles in an even simpler way, by starting with basic identity statements involving two homophonous tokens of the same name.

- Statements such as the following are normally taken to be uninformative tautologies semantically (although they may have contentful pragmatic implicatures).

23. Sandy is Sandy.

- If this is true, then a statement like the following should mean that Kim does not believe a tautology, an *unsatisfiable* reading:

24. Kim doesn't believe Sandy is Sandy.

Identity Statements: Delusions

- However, sentences like *Kim doesn't believe Sandy is Sandy* also have *satisfiable* readings in the right context:
    - Kim suffers from Capgras Syndrome, also known as the Capgras Delusion, a condition "in which a person holds a delusion that a friend, spouse, parent, or other close family member has been replaced by an identical-looking impostor."
- In this context, it is clear that one instance of Sandy is interpreted from the speaker's perspective, call this Sandy Sandy, and the other from Kim's, call this Impostor Sandy.
- The speaker is then asserting that Kim does not believe that Impostor Sandy = Sandy Sandy.
- This is a simple, limiting case of the puzzles we have been looking at.

Identity Statements: Hucksters

- These kinds of expressions are not restricted to pathological cases.
- We can even construct similar examples involving mathematical terms, a domain that we would not expect to be open to interpretation in the same way, because mathematical terms are paradigmatically assumed to be strongly rigid in reference.
- Consider the following piece of American history:
  - In 1897 Dr. Edwin J. Goodwin presented a bill to the Indiana General Assembly for “[. . .] introducing a new mathematical truth and offered as a contribution to education to be used only by the State of Indiana free of cost”. He had copyrighted that $\pi = 3.2$ and offered this “new mathematical truth” for free use to the State of Indiana (but others would have to pay to use it).

Identity Statements: Hucksters

- At the appropriate historical juncture, it is clear that the following sentence had a satisfiable reading:

  25. Dr. Goodwin doesn't believe that π is π.

- Dr. Goodwin was a huckster (someone trying to profit no matter what), but given the context, it seems that (25) accurately reported his beliefs.

- I won't go into details here, but in the *Perspectives* paper, we argue that a standard Montegovian *de re/de dicto* distinction is too weak to capture Goodwin's beliefs.

The expanse

- Intuitively, what the Capgras and Indiana Pi Bill cases share is a mix of the speaker's perspective with some other perspective: that of the subject of the sentence.
- Thus, it seems to us that the key to these puzzles, as mentioned above, is a notion of perspective, which can also potentially explain the lack of substitutability in simple sentences involving verbs like *love* and *murder*.
- If we cross the factors of same/distinct terms with simple/embedded context, we obtain the space of possibilities that we saw earlier, with cells filled by examples from the previous sections.

	Simple	**Embedded**
Same term	#Dr. Octopus punched Spider-Man but he didn't punch Spider-Man.	Kim doesn't believe Sandy is Sandy.
Distinct term	Mary Jane loves Peter Parker but she doesn't love Spider-Man. #Dr. Octopus killed Spider-Man but he didn't kill Peter Parker.	Kim doesn't believe Hesperus is Phosphorus.

Our solution

- We formalize the notion of perspective by using the monad that describes values that are made dependent on some external factor, commonly known as the `Reader` monad.

- Linguistic expressions that can be assigned potentially different interpretations are represented as functions from **perspective indices** to values, in the enriched monadic space.

- Effectively, this allows us to construct a kind of lexicon that not only represents the linguistic knowledge of a single speaker but also her (possibly partial) knowledge of the language of other speakers.

- In other words, following Chomsky, Jackendoff, and others, we construe lexicons to be aspects of the **knowledge of language of individuals**, and take standard circumlocutions like the “lexicon of English” to be atheoretical folk talk at best, or simply incoherent at worst.

A sample speaker's lexicon

Word	Denotation	Type
Reza	$\mathbf{r}_\sigma$	e
Kim	$\mathbf{k}_\sigma$	e
Dr. Octopus	$\mathbf{o}_\sigma$	e
Mary Jane	$\mathbf{mj}_\sigma$	e
Peter Parker	$\mathbf{pp}_\sigma$	e
not	$\lambda p.\neg p$	$t \to t$
but	$\lambda p.\lambda q.p \wedge q$	$t \to t \to t$
is	$\lambda x.\lambda y.x = y$	$e \to e \to t$
punch	$\lambda o.\lambda s.\mathbf{punch}(s, o)$	$e \to e \to t$
believe	$\lambda c.\lambda s.\mathbf{B}(s, c(\kappa(s)))$	$\Diamond t \to e \to t$
love	$\lambda o.\lambda s.\mathbf{love}(s, o(\kappa(s)))$	$\Diamond e \to e \to t$

Word	Denotation	Type
Hesperus	$\lambda i.\begin{cases}\mathbf{es}_{\mathtt{k}} & \text{if } i = \mathtt{k},\\ \mathbf{v}_\sigma & \text{if } i = \sigma\end{cases}$	$\Diamond e$
Phosphorus	$\lambda i.\begin{cases}\mathbf{ms}_{\mathtt{k}} & \text{if } i = \mathtt{k},\\ \mathbf{v}_\sigma & \text{if } i = \sigma\end{cases}$	$\Diamond e$
Spider-Man	$\lambda i.\begin{cases}\mathbf{sm}_i & \text{if } i = \mathtt{o} \text{ or } i = \mathtt{mj},\\ \mathbf{pp}_\sigma & \text{if } i = \sigma\end{cases}$	$\Diamond e$
Jesus	$\lambda i.\begin{cases}\mathbf{j}_{\mathtt{r}} & \text{if } i = \mathtt{r},\\ \mathbf{j}_\sigma & \text{if } i = \sigma\end{cases}$	$\Diamond e$
Sandy	$\lambda i.\begin{cases}\mathbf{imp}_{\mathtt{k}} & \text{if } i = \mathtt{k},\\ \mathbf{s}_\sigma & \text{if } i = \sigma\end{cases}$	$\Diamond e$

Result: Distinct terms, Embedding

26. Kim doesn't believe Hesperus is Phosphorus.

- $\neg$`B(k)(es`$_k$ `= ms`$_k$`)`
- $\neg$`B(k)(v`$_\sigma$ `= ms`$_k$`)`
- $\neg$`B(k)(v`$_\sigma$ `= es`$_k$`)`
- $\neg$`B(k)(v`$_\sigma$ `= v`$_\sigma$`)`

Result:
Same Terms, No Embedding
Contradictory

27. # Dr. Octopus punched Spider-Man but he didn't punch Spider-Man.

- $\texttt{punch}(\texttt{o}_\sigma)(\texttt{pp}_\sigma) \wedge \neg\texttt{punch}(\texttt{o}_\sigma)(\texttt{pp}_\sigma)$

Result:
Distinct Terms, No Embedding
Not Necessarily Contradictory

28. Mary Jane loves Peter Parker but she doesn't love Spider-Man.

- Non-contradictory reading, using MJ's perspective
  - $\mathtt{love}(\mathtt{mj}_\sigma)(\mathtt{pp}_\sigma) \wedge \neg\mathtt{love}(\mathtt{mj}_\sigma)(\mathtt{sm}_{\mathtt{mj}})$
- Contradictory reading, using speaker's perspective
  - $\mathtt{love}(\mathtt{mj}_\sigma)(\mathtt{pp}_\sigma) \wedge \neg\mathtt{love}(\mathtt{mj}_\sigma)(\mathtt{pp}_\sigma)$

Result:
Same Term, Embedding

29. Kim doesn't believe Sandy is Sandy.

- $\neg \mathtt{B(k)(s_{\sigma} = imp_{k})}$

Generalizing the approach

30. Elena loves dolphins, but she doesn’t love marine mammals.

- Suppose Elena thinks that Flipper is a dolphin and Hoover is a seal, but she thinks only Hoover is a marine mammal; i.e., she thinks seals are marine mammals, but dolphins are not.

- Suppose also that the speaker and Elena are in agreement about which entities the names *Flipper* and *Hoover* refer to, so the names are not controversial.

- The table on the next slide sketches (the relevant part of) the lexicon for the speaker of (30).

- We do not mean to imply that this extension of our approach is trivial, since matters of compositionality of, e.g., *marine mammal*, have not been addressed here, but the extension is at least a natural candidate for further exploration.

Generalizing the approach

Word	Denotation	Type
dolphin	$\{\mathbf{flipper}_\sigma\}$	$e \to t$
seal	$\{\mathbf{hoover}_\sigma\}$	$e \to t$
marine mammal	$\lambda i. \begin{cases} \{\mathbf{hoover}_\sigma\} & \text{if } i = \mathbf{mj}, \\ \{\mathbf{flipper}_\sigma, \mathbf{hoover}_\sigma\} & \text{if } i = \sigma \end{cases}$	$\Diamond(e \to t)$

Conclusion

- Monads are a key way to understand **computation** in **computer science**.
- We have investigated the following phenomena:
    - **Conventional Implicature**, a phenomenon important to linguistic semantics and pragmatics and philosophy of language
      [`Writer` monad]
    - **Conjunction Fallacies**, a phenomenon important to cognitive psychology, logic, and philosophy of language
      [`Probability` monad]
    - **Substitution Puzzles**, a phenomenon important to semantics, philosophy of language, and philosophy of mind
      [`Reader` monad]
- In our work, we see an idea from **computer science** helping to explain problems in **linguistics**, **psychology**, and **philosophy**, all from a **linguistic and logical perspective**.

Related Work by Other Researchers

- Our work is part of a growing tradition that applies category theory and monads to problems in natural language semantics.
  - Christina Unger: Anaphora, dynamic semantics
  - Christina Unger, Jan van Eijck: Computational semantics textbook, based on functional programming with Haskell
  - Simon Charlow: Indefinites, anaphora, scope, dynamic semantics
  - Dylan Bumford: Split-scope definites, relative superlatives
  - Chris Barker, Chung-chieh (Ken) Shan: Continuations, related to monads

Thank you!

Collaborators

Dr. Gianluca Giorgolo

Students

Luke Burke (UCL)

Support

Natural Sciences and Engineering Research Council of Canada

Ministry of Research and Innovation (Ontario)

References

Please see the following papers on my website:

$\langle \mathbf{M}, \boldsymbol{\eta}, \star \rangle$ *Monads for Conventional Implicatures*

One Semiring to Rule Them All

Perspectives

Web
users.ox.ac.uk/~cpgl0036

“CPGL0036 reporting for duty, Sir!”

Logic, Language & Information Lab
Carleton University