

One Semiring to Rule Them All¹

Gianluca Giorgolo[†] & Ash Asudeh^{†‡}

[†]Faculty of Linguistics, Philology & Phonetics, University of Oxford

[‡]Institute of Cognitive Science, Carleton University

Abstract

In this paper we look at the well known phenomenon of *conjunction fallacies* (Tversky & Kahneman, 1983) from a linguistic perspective. We are particularly interested in understanding the impact of this phenomenon on our understanding of *compositionality*, a core assumption of contemporary linguistic theories about language meaning. We will argue that *contra* Hertwig, Benz, and Krauss (2008), conjunction fallacies do not arise because of the ambiguity of coordinating conjunctions such as *and*, but rather because there are two different strategies available for computing the composite likelihood of a collection of events. Crucially, in our analysis the two strategies are two variants of the same underlying general structure that simultaneously allows subjects 1. to reason in purely logical terms; 2. to follow the rules of probability and 3. to commit fallacies depending on the conditions under which they evaluate linguistic expressions relating uncertain events. We will explain the choice between the two strategies in terms of cognitive/computational economy and the consequences of overestimating the likelihood of an event.

Keywords: conjunction fallacy, natural language semantics, probability, category theory

Introduction

Conjunction fallacies have been an active area of research in cognitive science for more than three decades now. The phenomenon was first discussed by Tversky and Kahneman (1983). They noticed that in a task asking for ratings of the relative likelihoods of different events, the majority of the subjects consistently rated the likelihood of the conjunction of two events as higher than the likelihood of one of the conjoined events.

One of their examples is the well-known “Linda paradox”. Subjects were given the following statement and, as part of the experimental task, where asked to rank the probability that various statements were true of Linda; the resulting ranking for the relevant cases is given below the context.

- (1) 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.

Linda is active in the feminist movement. [F(eminist)]
Linda is a bank teller and is active in the feminist movement. [T&F]
Linda is a bank teller. [T(eller)]

The context is obviously designed to bias towards the label *Feminist* and it is unproblematic and unsurprising that the relevant proposition is ranked most likely, but the result that the joint probability T&F is ranked higher than T is interesting,

and constitutes an instance of *conjunction fallacy*: a conjunction of two propositions is reported by subjects to be more probably than the probability of one of the two propositions on its own.

This result disagrees with the rules of probability, as the probability of the conjunction of two events, being the intersection of the two events, cannot be higher than the likelihood of any of the two events, formally for any two events *A* and *B*:

$$P(A \text{ and } B) \leq P(A), P(B) \quad (2)$$

These results have been replicated by many studies that have investigated different ways in which this apparently fallacious response can be elicited (Yates & Carlson, 1986; Tentori, Bonini, & Osherson, 2004).

The original explanation of these results by Tversky and Kahneman (1983) was in terms of representativeness. The authors claimed that the observed responses are due to the fact that subjects do not operate in terms of probabilistic reasoning, but rather use a *representativeness heuristic*. According to Tversky and Kahneman (1983), subjects check the degree of correspondence between the events and a certain model of reality and select those events that are closer to what the model predicts as being the more likely events. Representativeness tends to covary with frequency but not necessarily. A crucial point of Tversky and Kahneman (1983)’s analysis is that this heuristic operates on the conjunction as a whole, or as they put it:

the judged probability (or representativeness) of a conjunction cannot be computed as a function (e.g., product, sum, minimum, weighted average) of the scale values of its constituents. (Tversky & Kahneman, 1983, p. 305)

This last property is rather problematic if we intend to integrate their observations with current linguistic theories of meaning composition. In fact, a core hypothesis of modern semantic theory is the *principle of compositionality* (Janssen, 2011, among many others), ultimately based on the philosophy of language of Frege (1891/1952). This principle states that the meaning of a composite expression is determined by the meaning of its constituents. This is in clear contradiction with the analysis of Tversky and Kahneman. However compositionality has proven a very important tool in linguistic research, in theoretical semantics and in the philosophy of language, representing the semantic counterpart of linguistic productivity. Moreover compositionality seems to be a pervasive, and indeed quite successful, strategy in human cognition. Therefore we should be careful before discarding it.

¹... and in the darkness ∗ them.

Tversky and Kahneman's explanation has been challenged by a number of researchers. In particular Hertwig (Hertwig & Gigerenzer, 1999; Mellers, Hertwig, & Kahneman, 2001; Hertwig et al., 2008) has proposed that conjunction fallacies are not real errors, but rather emerge because of the intrinsic ambiguity of linguistic operators such as the conjunction *and*. Another important contribution of this line of research has been the demonstration that in certain contexts conjunction fallacies do not arise that easily. This is particularly true of contexts in which subjects are in some way primed to reason in terms of frequencies; for instance, if subjects are presented with a scenario that explicitly introduces frequencies, or if subjects are required to express explicitly their judgements regarding the likelihood of different events in terms of numerical estimates rather than implicitly by ordering the events in terms of likelihood. A similar reduction in the number of fallacies measured can be obtained by "raising the stakes", for example by asking subjects to bet on their estimates.

In what follows we try to reconcile these different points of view on the basis of the data reported in the literature and our goal of combining the results with the notion of compositionality. Our model is similar in some way to the one proposed by Yates and Carlson (1986). They argue that conjunction fallacies arise because subjects use different strategies 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 \mathbb{R} line. 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:

$$\lambda(A \text{ and } B) = \lambda(A) + \lambda(B) \quad (3)$$

Our model starts from the same assumption that there are multiple strategies that are employed by subjects when evaluating the likelihood of conjoined uncertain events. But instead of assuming that unrelated computational processes underpin the different strategies, we will show 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, but that does so by using different but related representations of uncertainty, expressed as alternative algebraic structures. More specifically we will demonstrate how we can explain the results reported in the literature in terms of an algebraic structure known as a *semiring*. This mathematical object is at the heart of a specific instance of a mathematical structure known as a *monad*, which has independently been shown to be a good model for the composition of complex meanings in natural language semantics (Shan, 2001).

In our model, the observation made in the literature (Hertwig & Gigerenzer, 1999) 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. In fact the same computational

structure, the monad, can be used together with a number of different underlying semirings, one of them being the probability semiring. We predict that, in general, if the subjects are presented with a task where there are "no stakes" they will base their judgement on the basis of a reasoning modeled using a representation corresponding to a semiring defined over a relatively simple set with generally simple operations. Using this strategy will generally lead subjects to make overconfident estimations, which tend not to correspond to the reality of things. If, on the other hand, subjects are forced to evaluate the consequences of their judgements, such as in the context of a gaming scenario, or if explicitly primed to think in frequentist terms, then we will observe a switch to a more complex representation, with properties that better approximate those of probability theory.

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. In this sense, with respect to Hertwig et al. (2008)'s analysis, we switch the ambiguity to the context rather than assuming that a word like *and* has multiple meanings. This is in fact a problem with Hertwig et al. (2008)'s analysis, which effectively treats *and* as ambiguous, i.e. truly polysemous.

Our approach is similar in spirit to the standard Gricean approach (Grice, 1975), although it differs in some important ways. Similarly to the Gricean view we assume that logical words like *and* have a non-ambiguous core logical meaning. In the standard Gricean approach the core meaning is further elaborated by "implicatures", additional default meanings that are not explicitly stated by the speaker. Crucially, implicatures are determined by context and can be cancelled by explicitly stating that they do not hold. For instance, a numeral like *two* is usually understood to mean "exactly two", as in *Susy has two daughters*, but the implicature can be cancelled, as witnessed by examples like *Susy has two daughters, in fact she has three*.

Similarly, in our approach the core meaning of a word like *and* is its logical, non-probabilistic meaning. But instead of relying on implicatures defined by the context, we have a grammatical process (as part of the logic composition) that can lift this core meaning to the probabilistic level if other linguistic expressions introduce uncertainty in the interpretation process. For instance, if *and* is used to conjoin two uncertain events, then the conjunction itself will become uncertain. The role of context in our model is not to permit the derivation of related additional meanings, but instead it selects which specific mode of uncertainty is going to be used. Notice that this does not introduce a source of non-compositionality in our system: both modes of uncertainty — pure probability and the simpler approximation — function in a purely compositional fashion. The limited form of ambiguity that may be attributed to our system stems entirely from the choice of the form of ambiguity. In contrast, in Hertwig et al. (2008)'s analysis, *and* is treated as truly ambiguous, and its final interpretation requires context as an additional parameter.

As a final remark in this introduction, we want to stress that our approach is in its current form limited to the explanation of the conjunction of two events, as expressed by the conjunction of two propositions. The related phenomenon of “concept conjunction”, as discussed for instance by Hampton (1988), shares a number of similarities with conjunction fallacies, but at the same time displays some crucial differences, at least at the linguistic level. For example, in the case of noun-noun compounds, such as *school furniture*, the denotation of the expression is not constructed by simply intersecting the denotations of its component parts, but the function involved operates in non-trivial ways on these meanings.² This is not the case for the conjunction of events, where the result is not an entirely novel type of event, but it is indeed the Boolean meet of the two sub-events.

Monads and uncertainty

In what follows we will introduce two different mathematical structures, *semirings* and *monads*. For reason of space we will only sketch their definitions, trying to provide the reader with an intuitive understanding of their importance. The interested reader can find more details concerning semirings in any introductory text about algebra, and similarly monads are discussed in most recent textbooks about category theory.

A semiring is a set A with two distinguished elements 0 and 1 and equipped with two binary operations $+$ and \cdot , satisfying the following axioms, for all x, y and $z \in A$

$$(x + y) + z = x + (y + z) \quad (4)$$

$$x + y = y + x \quad (5)$$

$$x + 0 = 0 + x = x \quad (6)$$

$$(x \cdot y) \cdot z = x \cdot (y \cdot z) \quad (7)$$

$$x \cdot 1 = 1 \cdot x = x \quad (8)$$

$$x \cdot 0 = 0 \cdot x = 0 \quad (9)$$

$$x \cdot (y + z) = (x \cdot y) + (x \cdot z) \quad (10)$$

$$(x + y) \cdot z = (x \cdot z) + (y \cdot z) \quad (11)$$

In the case of the probability semiring, A is the real interval $[0, 1]$, with 0 and 1 representing the two units and $+$ and \cdot defined in the usual way.³

Monads are a mathematical structure that arises in category theory in the study of the algebra of the functors from a category to itself (Awodey, 2010). They have found successful application in the semantics of programming languages (Moggi, 1989; Wadler, 1992) and more recently they have been proposed in the linguistic literature as a method to model phenomena that challenge traditional formal models of meaning (Shan, 2001; Unger, 2011; Giorgolo & Asudeh, 2012a, 2012b). In this paper we use monads as a mathematical device to model uncertainty in natural language meaning, and

²As a matter of fact, the operation involved may not be a function at all; see, e.g., Partee (1995).

³These definitions allow the result of summing two probability measures to be higher than 1, but our system normalizes all probabilities such that the sum is less than or equal to 1.

how uncertainty is combined and propagated in composed expressions, such as the conjunction of two propositions or predicates.

Intuitively we can think of monads as ways to map between different types of objects, in particular as a way to map simple objects into more complex ones. Monads are special mappings because they represent the canonical way to map the simple space of objects to a more complex one in such a way that important properties that link the simple objects are preserved under the mapping.

The fact that probability distributions form a monad in the category of measurable spaces was an early discovery in category theory (Lawvere, 1962; Giry, 1982). Here we use a slightly different characterisation based on the use of monads in functional programming to model probabilistic calculi. In our case, the space of simple objects is represented by the collection of different semantic types of natural language expressions, e.g. expressions that refer to an individual, such as proper names, expressions that denote some truth about reality, such as propositions, and expressions that denote a collection of individuals, such as common nouns. 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 (when uttering *John sleeps* the verb *sleeps* maps the individual John to truth 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 (in our previous example 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).

Formally our monad is defined by the triple $\langle P, \eta, \star \rangle$. P is a functor that operates in the way described above: it maps our semantic types to probability distributions over the inhabitants of the type and lifts the mappings between the types so that they operate between probability distributions. η , pronounced “unit”, is an operation that creates a special type of probability distribution that corresponds to a categorical judgement, i.e. assigning the entire probability mass to a single element of a semantic type. Basically this is a way to integrate certainty into our model of uncertainty. Formally we define it as follows (where we represent probability distributions as functions from types to the interval $[0, 1]$, or in general to any base set of a semiring):

$$\eta(x) := y \mapsto \begin{cases} 1 & \text{if } y = x \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

where x represents an element of any semantic type (e.g. an individual, a truth value, a collection of individuals).

The second operation, \star , pronounced ‘bind’, is the way in which we combine uncertainty. Its definition is based on the definition of joint probability. It is important to notice that the joint probability we are discussing here is not the one we are trying to model in the case of a conjunction. The one discussed here is much more primitive and is at the core of the process that constructs the probability of an event from the linguistic elements that describe it. The joint probability that we will discuss in what follows arises as the byproduct of a large number of interactions described by the bind operation. There is also another difference between the way the bind operation is used and what we normally mean by joint probability. While the joint probability of two events gives us the likelihood of a third event (the occurrence of both atomic events), bind returns the probability distribution of what we can consider another atomic event. Bind is defined as in (13).

$$m \star k := y \mapsto \sum_{x \in A} m(x) \cdot k(x)(y) \quad (13)$$

where A is the set of elements measured by the probability distribution m . Bind takes two arguments, a probability distribution (m) and a function (k) from elements of the domain of the probability distribution m to probability distributions over elements of (possibly) another set. The resulting probability distribution is obtained by collecting all possible ways in which we can obtain the various results from this second set, and by scaling them (\cdot) using the likelihood that they emerge from the first distribution. The results are collected together using addition.

It is quite clear that despite the fact that we have been talking about a probability monad, all operations involved in its definition are those we have described for a semiring. This means that we can use the same general structure we have discussed here and replace the meanings of 0, 1, \cdot , and $+$ with constants and operations defined for other semirings. This is precisely what we are going to do in the next section in which we explain how we can reproduce the results reported in the literature using our uncertainty monad.

Conjunction fallacies, compositionally

We show how we can reproduce the results reported in (Yates & Carlson, 1986) using our monadic infrastructure. We use their data as it is the only example in the literature 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.

The first step is to define a suitable base for our semiring. Yates and Carlson (1986) employ a discrete scale based on the general prediction made by their summation model. Their model does not take into consideration the extreme cases, i.e. impossible and certain events. We want to include them in our model as they are necessary in order to model what we know about the logical entailment behaviour of the word *and*. Tentori et al. (2004) showed that subjects that commit conjunction errors correctly apply the rules of logic. Therefore

| $+$ | I | U | P | L | C | \cdot | I | U | P | L | C |
|-----|---|---|---|---|---|---------|---|---|---|---|---|
| I | I | U | P | L | C | I | I | I | I | I | I |
| U | U | U | P | L | C | U | I | U | U | U | U |
| P | P | P | P | L | C | P | I | U | P | P | P |
| L | L | L | L | L | C | L | I | U | P | L | L |
| C | C | C | C | C | C | C | I | U | P | L | C |

Table 2: The one semiring.

we will use a simple discrete set as the base for our semiring: $\{I(m)possible), U(un)likely), Possible), L(ikely), C(certain)\}$. I and C correspond to 0 and 1 respectively. The only additional condition we need to impose so that I and C behave as boolean values is that for all x in our set $x + C = C$. There are sixteen possible well behaved semirings that we can define for this set.

The next step is to select those semirings that reproduce the results reported in (Yates & Carlson, 1986) and here summarised in table 1. To do so we have to explain how we expect the uncertainty attached to the two atomic events to propagate to the conjunction. This is described formally by the function in (14).⁴

$$\langle p, q \rangle \mapsto p \star (x \mapsto q \star (y \mapsto \eta(x \wedge y))) \quad (14)$$

This function takes a pair of probability distributions over truth values (describing how likely an event is realised in the real world) and constructs a new probability distribution over the results of conjoining the truth value of the first event with the one of the second. In this particular case there are three ways in which the final result can be false, and only one in which it turns out true (when both events are realized).

Using this schema we can evaluate how the sixteen semirings match the simplified results in table 1. It turns out that only one of the sixteen semirings satisfies the conditions induced by the data. We show the one semiring in table 2. Notice that this semiring is not homomorphic to the probability semiring, meaning that we cannot reproduce its behaviour using probability theory.

One process, two representations

As we have already explained, we assume that conjunction fallacies are true errors, in the sense that they lead to overestimation of the likelihood of events, but at the same time they are correct applications of a different strategy for computing the likelihood of composed events. Our explanation is given in terms of cognitive/computational economy. If there are no real stakes on the table, and therefore there is no incentive in using a safer but costlier computational system, subjects will employ a form of shortcut, represented in our model by a simpler semiring. If on the other hand, subjects are pushed to

⁴This function is actually generated in a completely compositional fashion by the grammatical system. For details on how such a system may work see Benton, Bierman, and de Paiva (1998) for a logical perspective and Giorgolo and Asudeh (2012a) for a more linguistic one.

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

Table 1: Results reported by (Yates & Carlson, 1986)

| | | | | | |
|---|---|---|---|---|---|
| — | I | U | P | L | C |
| | C | L | P | U | I |

Table 3: Complement operation for the one semiring.

think about the consequences of their judgements, the more expensive solution is selected.

The semiring described in table 2 is undoubtedly simpler than the standard probability semiring. First of all it is based on a simpler base set. But it 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 a simpler complement operation, $-$. If we define the complement as in table 3 (which seems to be most intuitive way to define it, as we just reverse the order of the relative likelihoods), then we can observe that for all x and y in our base set we have that $x \cdot y = -(-x + -y)$ or alternatively $x + y = -(-x \cdot -y)$.⁵ This means that this particular encoding of uncertainty has much lower representational costs than other competing possibilities. Its symmetry makes it a particularly simple and efficient computational object.

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. Moreover we observe the same general process, here formalized in terms of monads, at play in other areas of natural language meaning (Shan, 2001; Unger, 2011; Giorgolo & Asudeh, 2012a, 2012b). The two encodings we have discussed are instead selected by the context under which the description of the conjoined event is evaluated. If there are no real risks and judgements have no real consequences, then we predict that subject will select a computationally and representationally cheap encoding (the one semiring), otherwise they will apply the rules of probability, which require a much higher degree of computation and are representationally more costly. In terms of psychological heuristics, our proposal is that conjunction fallacies involve the heuristic of *satisficing* (Simon, 1956), either in addition to or instead of the heuristic of *representativeness* (Tversky & Kahneman, 1983).

Conclusion

We have presented a uniform model that provides a uniform model for the interpretation of uncertain conjoined events. In

particular, we have shown how it is possible to account for the presence of conjunction fallacies in judgements related to the likelihood of conjoined events, together with the correct application of the rules of probability theory. Our model starts from the idea that linguistically described events are interpreted starting from their linguistic description through a compositional process. This process is blind to the specific semantic content of the linguistic elements, in particular to the representation of the measure of uncertainty. We take that the context of evaluation has the effect of selecting one of two possible representations. One is computationally less expensive, but has the side effect of generating over-confident estimates for conjoined events. The other is computationally more expensive, but leads to judgements that are closer to the expected odds. We assume that the first representation is selected if there are no real consequences in case of an erroneous estimate, while the second is preferred if a mistake may have consequences.

The advantage of this approach over previous models is that it bridges an important hypothesis in the study of natural language semantics, compositionality, and a pervasive cognitive illusion such as the conjunction fallacy. Our 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 *satisficing heuristic*.

Our model also makes novel predictions that we plan to study in future work. First of all, 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 the cases of disjunction (Carlson & Yates, 1989) and implication. But our 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. One simple interpretation of such examples is in fact in terms of an iterated conjunction over the domain of quantification. If we also observe conjunction fallacies in these cases, this should provide important evidence for a compositional model. But our model also suggests novel ways in which we can prime subjects to not commit reasoning fallacies related to uncertainty. If we are correct in assuming that subjects select the specific strategy for evaluating uncertain events on the basis of the possible repercussions of their choices, then we predict that we can force subjects to select a frequency based style of reasoning by simply introducing such consequences. Betting is of course one example, but other possibilities include some kind of emotional or social

⁵It is interesting to note that these laws correspond to De Morgan's Laws in the semiring of truth values.

feedback to the judgements of subjects.

Acknowledgements

We would like to thank our anonymous reviewers for their comments. We would also like to thank John Baez and members of the Glue Group at Oxford University for comments and discussion. Any remaining errors are our own. This research is supported by a Marie Curie Intra-European Fellowship from the European Commission under contract number 327811 (Giorgolo) and an Early Researcher Award from the Ontario Ministry of Research and Innovation and NSERC Discovery Grant #371969 (Asudeh).

References

Awodey, S. (2010). *Category theory* (2nd ed.). Oxford University Press.

Benton, N., Bierman, G. M., & de Paiva, V. (1998). Computational types from a logical perspective. *Journal of Functional Programming*, 8(2), 177–193.

Carlson, B. W., & Yates, J. F. (1989). Disjunction errors in qualitative likelihood judgment. *Organizational Behavior and Human Decision Processes*, 44(3), 368–379.

Frege, G. (1891/1952). Function and concept. In P. T. Geach & M. Black (Eds.), *Translations from the philosophical writings of Gottlob Frege* (pp. 22–41). Oxford: Blackwell. (Translation of *Funktion und Begriff*, in *Der Jenaischen Gesellschaft für Medizin und Naturwissenschaft*.)

Giorgolo, G., & Asudeh, A. (2012a). $\langle M, \eta, \star \rangle$ Monads for conventional implicatures. In A. Aguilar Guevara, A. Chernilovskaya, & R. Nouwen (Eds.), *Proceedings of sinn und bedeutung 16* (Vol. 1, pp. 265–278). MIT Working Papers in Linguistics.

Giorgolo, G., & Asudeh, A. (2012b). Missing resources in a resource-sensitive semantics. In M. Butt & T. H. King (Eds.), *Proceedings of the LFG12 conference* (pp. 219–239). Stanford, CA: CSLI Publications.

Giry, M. (1982). A categorical approach to probability theory. In *Categorical aspects of topology and analysis, Proc. int. Conf., Ottawa 1981, Lect. Notes Math. 915*, 68–85 (1982).

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Speech acts* (Vol. 3, pp. 41–58). San Diego, CA: Academic Press.

Hampton, J. A. (1988). Overextension of conjunctive concepts: Evidence for a unitary model of concept typicality and class inclusion. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14(1), 12–32.

Hertwig, R., Benz, B., & Krauss, S. (2008). The conjunction fallacy and the many meanings of *and*. *Cognition*, 108, 740–753.

Hertwig, R., & Gigerenzer, G. (1999). The “conjunction fallacy” revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, 12, 275–305.

Janssen, T. M. V. (2011). Compositionality. In J. van Benthem & A. ter Meulen (Eds.), *Handbook of logic and language* (2nd ed., pp. 495–553). London: Elsevier.

Lawvere, F. W. (1962). *The category of probabilistic mappings*. (Seminar handout)

Mellers, B., Hertwig, R., & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? an exercise in adversarial collaboration. *Psychological Science*, 12, 269–275.

Moggi, E. (1989). Computational lambda-calculus and monads. In *LICS* (p. 14–23). IEEE Computer Society Press.

Partee, B. H. (1995). Lexical semantics and compositionality. In L. R. Gleitman & M. Liberman (Eds.), *An invitation to cognitive science: Language* (2nd ed., Vol. 1, pp. 311–360). Cambridge, MA: MIT Press.

Shan, C. (2001). Monads for natural language semantics. In K. Striegnitz (Ed.), *Proceedings of the ESSLLI-2001 student session* (pp. 285–298).

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63(2), 129–138.

Tentori, K., Bonini, N., & Osherson, D. (2004). The conjunction fallacy: a misunderstanding about conjunction? *Cognitive Science: A Multidisciplinary Journal*, 28(3), 467–477.

Tversky, A., & Kahneman, D. (1983). Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychological Review*, 90(4), 293–315.

Unger, C. (2011). Dynamic semantics as monadic computation. In M. Okumura, D. Bekki, & K. Satoh (Eds.), *New frontiers in artificial intelligence — JSAI-isAI 2011* (p. 68–81).

Wadler, P. (1992). Comprehending monads. In *Mathematical structures in computer science* (pp. 61–78).

Yates, J., & Carlson, B. (1986). Conjunction errors: Evidence for multiple judgment procedures, including “signed summation”. *Organizational Behavior and Human Decision Processes*, 37, 230–253.