Communication and Computation

New Questions About Compositionality

Shane Noah Steinert-Threlkeld

Communication and Computation

New Questions About Compositionality

ILLC Dissertation Series DS-2017-05

Institute for Logic, Language and Computation

For further information about ILLC-publications, please contact

Institute for Logic, Language and Computation
Universiteit van Amsterdam
Science Park 107
1098 XG Amsterdam
phone: +31-20-525 6051
e-mail: illc@uva.nl
homepage: http://www.illc.uva.nl/

COMMUNICATION AND COMPUTATION: NEW QUESTIONS ABOUT COMPOSITIONALITY

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF PHILOSOPHY
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Shane Steinert-Threlkeld

June 2017

To my dad; you were taken too soon.

Abstract

The topic of this dissertation is the famous principle of compositionality, stating that the meanings of complex expressions are determined by the meanings of their parts and how the parts are put together. The title belies its intent: rather than advancing a central thesis, it instead asks and provides answers to new questions about compositionality. In particular, these questions arise by shifting from viewing compositionality as a property of symbolic systems – where what I call 'status questions' are naturally discussed – to viewing it from a procedural perspective, as operative in processes of production and interpretation.

The questions the dissertation asks arise at successively narrower levels of scale. At the level of our species, I ask: why are natural languages compositional in the first place? At the level of small conversations: what role does compositionality play in the broader theory of communication? And at the level of an individual language speaker: what is the algorithmic interpretation of compositionality and what demands on complexity does it impose? In the pursuit of answers to these questions, a wide variety of methods are employed, from simulations of signaling games, to logic and formal semantics, to theoretical computer science, as appropriately called for.

Acknowledgements

First and foremost, I must thank the members of my dissertation committee, each whose influence on me has been immense.

Johan van Benthem has been enormously gracious and supportive in my pursuits with time, discussion, and exposure to the wider intellectual world that he has helped develop over the previous decades. His abilities to never lose the forest for the trees, while also never losing the trees for the forest, and to find beautiful mathematical structures hiding in unexpected places represent a model of intellectual life to which I aspire. Being his student has been an honor that I will never forget.

Thomas Icard has been an extraordinary friend, collaborator, and advisor over the years. His well-honed knack for finding precise open questions and for formalizing intuitive ideas in myraid domains of inquiry has been both helpful and inspiring. It's hard to imagine having made it through this PhD program without his intellectual and social companionship.

Chris Potts first exposed me to the rigorous study of natural language semantics and pragmatics, through a two-quarter sequence of courses. It's fair to say that the tools developed there, as well as the central role played by compositionality in the courses, have exerted enormous influence on the trajectory of my research. The incredible detail that he puts into teaching materials, presentations, and comments on written work also represents an ideal that I hope to approximate in my own career.

Brian Skyrms first exposed me to the signaling games framework. Noticing the discrepancy between the beautiful explanations of the evolution of meaning it provides and the richer representations of meaning that I was learning about at the same time provided the impetus for much of my subsequent work. His writings are also a model,

having the highest density of information-per-sentence of any that I have ever read.

The Stanford Philosophy department provided a fertile intellectual community. I have benefited from conversations with many members, including Sam Asarnow, Rachael Briggs, Rahul Chaudhri, JT Chipman, Michael Cohen, Mark Crimmins, Lorenza D'Angelo, Tom Donaldson, Huw Duffy, Amos Espeland, Jonathan Ettel, the late Sol Feferman, Michael Friedman, Dave Gottlieb, Nathan Hathauler, Dustin King, Krista Lawlor, Anna-Sara Malmgren, Katy Meadows, Chris Mierzewski, the late Grisha Mints, Carlos Núñez, Adwait Parker, Grace Paterson, David Taylor, Ken Taylor, Paul Tulipana, John Turman, Yafeng Wang, Jessica Williams, Steve Woodworth, and Francesca Zaffora Blando.

The broader community at Stanford has also been ideal for pursuing interdisciplinary work like that in this dissertation. I am particularly grateful for conversations and exchanges with Cleo Condradvi, Lelia Glass, Noah Goodman, Dan Lassiter, Sven Lauer, Prerna Nadathur, Stanley Peters, Tania Rojas-Esponda, Ciyang Qing, and Ed Zalta.

Beyond Stanford, I must thank Peter Pagin and Dag Westerståhl for a very pleasant visit to Stockholm and for stimulating conversation on all matters compositional. During visits to Amsterdam, conversations with Maria Aloni, Alexandru Baltag, Michael Franke, Nina Gierasimczuk, and Sonja Smets have been very illuminating. During an extended stay at MIT, David Boylan, Kevin Dorst, Branden Fitelson, Samia Hesni, Justin Khoo, Josh Knobe, Matthias Jenny, Matt Mandelkern, and Ginger Schultheis provided many helpful comments and conversations.

As all of these thanks make evident, I consider the intellectual life best lived with other like minds. To that end, I must thank my collaborators. Jakub Szymanik and I started working on topics related to Chapter 4 upon meeting in the Summer of 2012 and have since had many fruitful projects, including co-teaching experiences. I am delighted to be joining him at the Institute for Logic, Language and Computation as a postdoc to continue our collaborations. Peter Hawke has been a tremendous friend, collaborator, and discussant on all topics. We first decided to collaborate in Amsterdam during an extended visit in December 2013. Our journey started with the problem of small versus large worlds in decision theory and wound up at epistemic

modals. Such is the curvy road that the intellect walks! As he also joins the ILLC, I look forward to continuing this intellectual friendship over the next few years where it began.

This dissertation was completed thanks to a – luckily, aptly named! – Mellon Foundation dissertation completion fellowship. The final draft was completed in Paris, while visiting the Institut Jean Nicod. I am very grateful to Philippe Schlenker for the impetus for and generosity in arranging this visit.

Of course, I would never have made it this far without the love and support of my family. My mom, Kayte, has been unwavering in her support of whatever life decisions I have made, trusting them to turn out well. Because of her upbringing, they always have. My brother, Zack, has been my best friend since (literally) before I can remember. Because he had been through many of the same things as me a few years earlier, my path has been better lit and therefore easier to follow. I lost my father, Tom, tragically in October 2013. He would be immensely proud were he able to see this finished product. In my own life, I can only hope to imitate his passion for his projects and the enormity of his laugh.

Last, because not least, thanks cannot do justice to the joy that coming to know and eventually marrying Meica Magnani has brought. Her comments on my writing and various talks, as well as late night conversations on all topics philosophical, have been enormously helpful. And without her support, I'm not sure I would have finished this dissertation. It's hard to remember life without you before we met; luckily, I do not have to imagine a future without you. Here's to bloobing around the world!

Erstaunlich ist es, was die Sprache leistet, indem sie mit wenigen Silben unübersehbar viele Gedanken ausdrückt, daß sie sogar für einen Gedanken, den nun zum ersten Male ein Erdbürger gefaßt hat, eine Einkleidung findet, in der ihn ein Anderer erkennen kann, dem er ganz neu ist.

It is astonishing what language can do. With a few syllables it can express an incalculable number of thoughts, so that even a thought grasped by a human being for the very first time can be put into a form of words which will be understood by someone to whom the thought is entirely new.

– Frege [1923], translated R.H. Stoothoff

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In the epigraph to this dissertation, Frege points to an awe-inspiring feature of human language: by producing a few sounds or marks on a sheet of paper, one can transmit a thought to another. Frege also mentions that an 'incalculable' number of thoughts can be so transmitted. The power of language largely comes from this expressivity. Human language allows us to do more than label a fixed, finite inventory of thoughts, as if communication were like choosing an item from a restaurant menu to order. Rather, language allows us to creatively express new thoughts. For example, if this dissertation is successful, I will have transmitted some thoughts to the reader that they have never before entertained. And I will do so by producing sentences that they have never read, heard, or rehearsed in inner speech before.

That the reader has the ability to recover the meaning of these never-before-seen sentences reflects the *productivity* of linguistic understanding, which has in turn been taken to motivate the famous principle of *compositionality* of meaning:[1]

(C) The meaning of a complex expression is determined by the meanings of the expression's parts and the manner in which they are combined.

The thought underlying the classic argument for (C) runs as follows: the best explanation for productivity – that we can understand new sentences when we encounter them for the first time – says that our process of understanding proceeds by relying

[1] See, for instance, Fodor [1987] (pp. 147-150) and Partee [1995].

on 'memorizing' the meanings of basic expressions and then building the meanings of larger units (phrases, sentences) on the basis of those and the structure of the larger units.

Compositionality occupies an absolutely central position in the field of natural language semantics. Witness the opening pages of any semantics textbook. §1.2 of Heim and Kratzer [1998] contains an extended version of the epigraph of this dissertation. §2.1 of Chierchia and McConnell-Ginet [2000] and §1.1.1 of Jacobson [2014] rehearse the productivity argument for compositionality. Despite this central role, debates at the border between philosophy of language, semantics, and cognitive science about the exact status of the principle have raged for decades.[2] These debates center on what I will call 'status questions', questions which focus on the status of (C).

This dissertation embodies a certain liberation from these status questions: while they are important, moving beyond the exact status of the principle of compositionality allows new and exciting questions about it to be asked and addressed, using methods that are not usually applied to questions about compositionality. With that being the main contention of this dissertation, I will not present an extended argument for a particular thesis, but rather *show* the fruitfulness of this liberated perspective over the subsequent chapters. In the remainder of this introduction, I will canvas some of the existing debates about compositionality, showing that they all view the principle as a property of abstract symbolic systems. From this purview, compositionality will be shown to have a privileged methodological status: whenever tensions arise between compositionality and the properties of a symbolic system, the latter must be changed. Because of this methodologically privileged position, we should move past status questions. Doing so, and thereby taking a broader perspective on compositionality – as an ability exercised in processes of interpretation and production – allows new questions to be asked. These questions, and my subsequent answers, will be summarized.

[2] See Janssen [1997], Szabó [2013], Pagin and Westerståhl [2010a,b] for excellent surveys.

1.1 Status Questions

1.1.1 Precisifications

The first status question is perhaps the most basic: what *is* the principle of compositionality? This question stems from the fact that (C), as stated, contains many ambiguities which remain to be made more precise. What are meanings? What are the part/whole structures of expressions? What modes of determination are permitted? Are the meanings of the parts of a complex expression to be taken distributively or collectively? Different answers to these questions lead to different precisifications of the principle of compositionality. See §1 of Szabó [2013] for an excellent overview of the choices here.

Both meanings and the part/whole syntactic structures of expressions are theoretical posits, introduced by the theorist to explain certain patterns of speaker judgments. In the case of syntax, these judgments primarily consist in judgments of the grammaticality of various strings. In the case of semantics, the judgments are a bit more variegated – involving truth-in-a-situation, entailment, ambiguity, et cetera – and are generally the product of literal meaning together with other pragmatic and cognitive factors.[3] Because of this variation and interaction, the semantic theorist has some flexibility in choosing what sort of things meanings are. Lewis [1970] (p. 22) captures the flexibility offered by this approach very nicely in his famous methodological dictum: "In order to say what a meaning *is*, we may first ask what a meaning *does*, and then find something that does that."

A bit more can be said about the notion of determination in question: at a minimum, the meaning of a complex expression must be *a function of* the meanings of the parts and their syntactic structure. Being a function captures the thought that specifying the meanings of the parts and their syntax suffices to determine, via the function, the meaning of the whole. In many theorists' eyes, there need not be a single composition function.[4] Rather, the semantic composition function can depend on the

[3] See §2.2 of Yalcin [2014] and §1.3 of Chierchia and McConnell-Ginet [2000] for useful overviews of the types of data to be explained by semantic theories.

[4] The leading account on which there is a single composition function follows Montague [1970, 1973] closely and uses *function application* as the only composition operation. Heim and Kratzer

syntactic operation that forms the complex expression. This leads to the following formulation, where μ is the function returning the meaning of expressions:[5,6]

(CH) For every syntactic operation o, there is a function f such that for all expressions $e_1, \ldots, e_n$, we have: $\mu\left(o\left(e_1, \ldots, e_n\right)\right) = f\left(\mu\left(e_1\right), \ldots, \mu\left(e_n\right)\right)$

More substantively than functional dependence, Szabó [2000] has argued that the dependence in question should be understood as metaphysical supervenience and Pagin [2012] that the function should have minimal computational complexity.

1.1.2 Counter-examples

Another status question asks about whether we should believe (C). That is: is the principle of compositionality *true*? In light of the last section, the question can be phrased more generally: on which disambiguations is the principle true? The literature on this question has centered around alleged counter-examples. Two sorts will be discussed here, both in the domain of quantification. The first sort takes the most common approach, in light of the functional interpretation (CH): if a complex expression changes meaning based on replacing a sub-expression with another expression that has the same meaning, then the meaning of the complex is not a function of the meanings of the parts and the syntax and so compositionality would fail.[7] The second sort shows that meanings of a certain type cannot be generated compositionally to deliver the right truth-conditions for sentences in a certain fragment.

The first sort of alleged counter-example is exemplified by the discussion of conditionals and 'unless' under quantifiers initiated by Higginbotham [1986]. To see the problem, consider the two sentences below:

(1) Every student will succeed unless they goof off.

[1998] attempt this form of type-driven interpretation, but end up introducing some other operations.

[5]See Pagin and Westerståhl [2010a]. Westerståhl [2012] discusses versions with context-dependence.

[6]The 'H' in '(CH)' stands for *homomorphism*, since the condition states that there is a homomorphism between syntax and semantics.

[7]Another example of this sort, from a domain besides quantification, comes from propositional attitude reports. See Partee [1982] and Pelletier [1994] for discussion.

(2) No student will succeed unless they work hard.

(1) has the following truth conditions: every student is such that either they will succeed or they will goof off, with the 'or' understood exclusively. A standard compositional semantics that treats 'unless' as meaning exclusive disjunction and gives 'every' its standard meaning will derive these truth conditions. (2) has the following truth conditions: no student is such that they both succeed and do not work hard. If, however, 'unless' contributes an exclusive disjunction, (2) would have the following truth conditions: every student both does not succeed and does not work hard. It thus looks like no uniform meaning can be given to 'unless' can be given that accounts for the contribution it makes to the sentences (1) and (2).

This appears to constitute a counter-example to compositionality because the syntactic structures of (1) and (2) appear to be the same and we have a firm grasp of the meaning of the quantifiers, but the contribution of 'unless' has to be sensitive to the context into which it is embedded. While this problem has spawned a vast literature,[8] I will here mention only the solution of Leslie [2009]. She takes 'unless' to do two things: to contribute an unpronounced necessity modal and to restrict the domain of the explicit quantifier.[9] Doing so delivers the right truth conditions for the target sentences while giving 'unless' a uniform meaning that does not depend on the linguistic context into which it is embedded. Higginbotham's argument, then, that (1) and (2) are counterexamples to compositionality required making assumptions about the range of possible meanings for 'unless', implicitly excluding the candidate that Leslie came up with.

For an example of the second sort, we will look at *branching* quantification. To see the problem, consider a simple quantified sentence.

(3) Every talk at the conference is great.

These are evaluated at variable assignment functions, so we can think of their semantic

[8] See, among others, von Fintel and Iatridou [2002] and Higginbotham [2003] and the references therein.

[9] This latter feature takes inspiration from the analysis of conditionals by Lewis [1975] and Kratzer [1986] as being domain restrictors.

values as being sets of such assignments. For example, we have that

$$[\![(3)]\!]^g = 1 \text{ iff } \{h : h \sim_x g \text{ and } [\![\text{talk}(x)]\!]^h = 1\} \subseteq \{h : h \sim_x g \text{ and } [\![\text{great}(x)]\!]^h = 1\}$$

where $h \sim_x g$ means that h and g are assignments that agree everywhere except possibly at the variable x. Now, consider the following examples.

(4) Some relative of each villager and some relative of each townsman hate each other.

[(37) from Hintikka [1974]]

(5) Most philosophers and most linguists agree with each other about branching quantification.

[(21) from Barwise [1979]]

(6) Each player of every baseball team has a fan, each actress in every musical has an admirer, and each aide of every senator has a friend, who are cousins.

[(50) from Hintikka [1974]]

The sentences (4) and (5) require the assignments of objects to variables of the two quantifiers to be done *independently* of each other. For example, in (4), there must be a relative depending on each villager and a relative depending on each townsman, but there are no dependency relations between the choices of those relatives.[10] This was thought to be a violation of compositionality because such truth-conditions couldn't be built 'from the bottom-up' in terms of the standard, first-order meanings of the individual quantifiers. So writes Barwise [1979], p. 79: "Linguistically, the discovery of branching quantification would force us to re-examine, and perhaps re-interpret, Frege's principle of compositionality according to which the meaning of a given expression is determined by the meanings of its constituent phrases."[11]

[10] Because of this, the logical syntax for branching quantification often stacks blocks of quantifiers since linear order creates scopal dependencies.

[11] Cameron and Hodges [2001] prove that no semantics for branching that evaluates formulas at single assignment functions can be given.

Nevertheless, Hodges [1997] gave a compositional semantics for a fragment containing these sentences with branching quantifiers. He did so by *changing the basic semantic value* to be sets of sets of variable assignments. That is, by replacing $[\![\cdot]\!]^g$ with a different interpretation function $[\![\cdot]\!]^G$, where G is a set of assignments, one can give a compositional semantics for branching quantification. For example, (4) will get the logical syntax $\forall x \exists y \forall v / \{x, y\} \exists w / \{x, y\} : (V(x) \wedge T(v)) \rightarrow R(x, y) \wedge R(v, w) \wedge H(y, w)$.[12] Then we have $[\![(4)]\!]^G = 1$ if and only if the set of assignments G has the following property: if any two assignments in G agree on what they assign to x, they also agree on what they assign to y; if any two assignments in G agree on what they assign to v, they also agree on what they assign to w; and every assignment in G satisfies, in the usual first-order sense, the quantifier-free matrix of the aforementioned logical form.[13]

Both of these arguments present a very common pattern. An alleged empirical counter-example to compositionality is proposed. It turns out only to be a counter-example once certain views about the syntax of the sentences and their semantic values are fixed. In the quantified conditionals case, the argument required assuming that 'unless' was truth-functional. In the branching quantification case, the argument required assuming that quantifiers denote sets of assignment functions. By enlarging the space of possible semantic values for expressions in the relevant fragment, compositionality can be restored. That the dialectic follows this trajectory reflects Lewis' methodological dictum mentioned in Section 1.1.1: the theorist has flexibility to choose meanings based on their job description and a central feature of this job description is that meanings must compose. So, upon finding that meanings of a certain kind do not compose, the theorist can change their view on what kinds of things meanings are. I take this general pattern to show that semanticists take compositionality to be, as Fodor [2001] writes, "non-negotiable".[14]

[12] Read $\exists y / \{x\} : \ldots$ as 'there is a y, not depending on x, such that ...'. Read V as 'is a villager', T as 'is a townsman', R as 'are related to each other', and H as 'hate each other'.

[13] The framework of *dependence logic* generates these truth conditions not by considering special logical forms for quantifiers, but by adding special atoms $= (x_1, \cdots x_n, y)$ saying that y functionally depends on the x_i. See Väänanen [2007], Galliani [2017].

[14] There's a subtlety in that Fodor takes only the language of thought, and not natural language, to be compositional. His arguments against the latter being compositional leave something to be

1.1.3 Vacuity

The freedom of the semanticist just discussed raises the third status question: does compositionality place any constraints on semantic theory or is it entirely *vacuous*? A number of authors have argued, often with mathematical proofs behind them, that compositionality is in fact vacuous. I will briefly survey these arguments and conclude that while compositionality does in fact require independent constraints from syntax and semantics to have empirical bite, compositionality is more methodologically entrenched than the sources of those constraints.

van Benthem [1986] (p. 200), succinctly summarizing Janssen [1983], notes that most syntactic algebras in the literature are freely generated from lexical items. In this situation, any assignment of meanings to the lexical items can be extended to a homomorphism between the syntactic algebra and a semantic one, even when a mapping between syntactic and semantic operations has already been specified. Because (CH) requires the existence of such a homomorphism and this result shows that they are easy to come by, van Benthem concludes that "by itself, *compositionality provides no significant constraint upon semantic theory*" (p. 200, emphasis in original). Two things should be noted about this result and the remark. First, the setting assumes that lexical meanings are given and need to be extended to a compositional semantics. This assumption is implausible and runs counter to the usual methodological direction in semantic theorizing, where one has a rough idea of sentential truth conditions and extracts lexical meanings that can generate them.[15] Second, even ignoring this criticism, van Benthem rightly hedges his claim with 'by itself', indicating that compositionality only constrains semantic theorizing in conjunction with other claims. While many theorists look to independent claims from syntax to provide additional

desired.

[15]This reflects a kind of methodological reading of Frege's context principle: "never to ask for the meaning of a word in isolation, but only in the context of a proposition" (Frege [1884], p. 10). This translation is J.L. Austin's; here is the German: "nach der Bedeutung der Wörter muss im Satzzusammenhange, nicht in ihrer Vereinzelung gefragt werden". While arguing for this lies beyond the present scope, I contend that the context principle is wholly compatible with the compositionality principle with the latter construed as a claim about interpretation and the former as a claim about methodology.

constraints, van Benthem suggests that providing an inference relation on semantic values that aligns with native speaker judgments can be a source of additional constraints.[16]

Zadrozny [1994] famously proved the following: given a semantics μ not satisfying (CH), a new function μ^* can be defined which is compositional and from which μ can be recovered.[17] To prove this theorem, however, Zadrozny has to define an entirely new meaning algebra for the function μ^*. While this in itself is not a huge problem – as we saw in Section 1.1.2, semanticists regularly change types of meaning, and so meaning algebras, in order to restore compositionality – Westerståhl [1998] notes that the meanings Zadrozny defines build syntactic facts into the meanings so that distinct expressions will always have distinct meanings. But such a function will always be compositional, since compositionality only has force when there are synonymous expressions.[18] Such a departure from ordinary semantic judgments seems too extreme.[19]

Both results show that, on its own, compositionality places little to no constraint on a semantic theory. While both results also make assumptions that make them inapplicable to actual semantic practice, the general moral of this result together with the case studies of counter-examples in Section 1.1.2 is this: compositionality only gains force in conjunction with additional constraints from syntax and semantics. But, when these additional constraints conflict with compositionality, theorists revise

[16]Looking wider for additional sources of constraints raises interesting questions about the shape of semantic theory and the role of compositionality therein. Traditionally, semantic values are used to explain all meaning-relevant facts: truth-in-a-context, valid inference, assertoric content, and other concepts are all defined in terms of semantic value. But it could be that separate 'semantic submodules' are to be used for explaining different linguistic phenomena. (The natural logic program, which develops separate representations to explain monotonicity inferences, can be seen as an example of this approach. See ch. 6 of van Benthem [1986], Sánchez Valencia [1991], MacCartney [2009], Icard III. and Moss [2014].) Would all of the submodules need to be compositional in a sense appropriate to the domain, or does compositionality only apply to 'denotational' representations? These questions, while fascinating, will be left for future pursuit. Thanks to Johan van Benthem for discussion.

[17]μ can be recovered in the sense that $\mu(e) = \mu^*(e)(e)$ for any expression e.

[18]See item (3) on p. 254 of Pagin and Westerståhl [2010a].

[19]Relatedly, Dever [1999] notes that the original meanings given by μ play no role in the composition process, which is entirely driven by what Dever calls 'occult meanings'. In this sense, the recoverability of μ is uninteresting, since it plays no role in composing meanings of complex expressions.

them instead of concluding that compositionality is false. This reflects a *methodological centrality* for the principle: because it reflects fundamental properties of our semantic competence, compositionality of a semantic theory will be preserved much more strongly than other of its properties.

1.2 New Questions

Considering that the long tradition of asking status questions about compositionality has resulted in recognition of its methodological centrality, we can be confident that some version of the principle will remain entrenched in semantic theory. Therefore, I believe that it is high time to ask *new questions* about compositionality. Whence the sub-title of this dissertation.[20]

The particular new questions that this dissertation addresses arise from a shift in perspective. In particular, the status questions just canvassed all view compositionality as a mathematical property of abstract symbol systems. While such a view can be useful, it operates at a level of abstraction that is highly removed from the actual speakers of natural languages. Because the original motivations for the principle of compositionality come from properties of human linguistic understanding, I propose to ask new questions about compositionality by viewing it from a procedural perspective, as something on display in processes of production and interpretation by actual speakers.

To this end, each chapter of the dissertation addresses one new question about compositionality, in order of decreasing levels of scale. At the level of our species: why and how might compositionality have arisen in our evolutionary history? At the level of small conversations: what role does compositionality play in our best theories of communication? At the level of the individual language user: what is the algorithmic interpretation of compositionality and what demands on complexity does it impose?

[20] Johan van Benthem, in personal communication, informs me that he and others joked in the early 1990s about placing a moratorium on discussions of compositionality because of a feeling that discussions of what I've called status questions were becoming stale. In a fictional review of Bolzano's *Wissenschaftslehre*, van Benthem [2013] (p. 302) refers to a 25-year moratorium beginning in 1990. This dissertation, then, appears at just the right time.

Because each chapter addresses a different question, they can be read independently of one another. In the remainder of this introduction, I outline each of them in turn, before summarizing and highlighting my previously published work in these areas.

1.2.1 Why are natural languages compositional?

The first question I ask stems from the observation that while various forms of communication exist widely in the animal kingdom, from insects to birds to monkeys[21] and apes, human language appears to be unique in its capacity for complex expressions and compositional interpretation thereof. One wonders, then, *why* humans have this ability. Why, that is, are natural languages compositional?

To address this question, I extend the signaling game framework – pioneered by Lewis [1969] and Skyrms [1996, 2004, 2010] – to model compositional language. This framework models the emergence of meaning through repeated interactions of the following kind: one agent, a Sender, has private information about the world which it tries to communicate to another agent, a Receiver. The Receiver can perform an action in the world and has common interests with the Sender. Starting from initially random transmission of signals, dynamics of learning and/or evolution can eventually create behavior which confers meaning on the signals. Because, however, all of the signals studied in the literature up to this point have lacked any syntactic structure, such studies offer at best explanations of the origins of lexical meaning.

To overcome this deficiency, Chapter 2 – *Compositional Signaling in a Complex World* – augments signaling games with syntactically complex signals which can be interpreted compositionally to capture what I call *minimal negation*. The chapter presents computer simulations of Senders and Receivers learning to signal via a particularly simple form of reinforcement learning. I compare agents who can only use simple signals with those who can use compositionally interpreted complex signals and find the following: simple signaling systems are easier to learn in simple situations while compositional ones are easier to learn in sufficiently complex situations,

[21] For a very useful overview of the literature on monkey communication and an application of linguistic methodology thereto, see Schlenker et al. [2016]. For a critical commentary on this program, see Steinert-Threlkeld [2016b].

where the complexity consists in how many things the agents need to communicate about.

This provides one possible answer to the 'why' question by identifying a purpose for which compositional language might have evolved: communicating about a multitude of diverse situations. Because many animals may only have selective pressure to communicate about a small number of topics – a few different predators, for instance – this chapter also helps explain the scarcity of compositional signaling in the animal kingdom. The chapter concludes by outlining future research directions.

1.2.2 What role does compositionality play in the theory of communication?

The second question I ask stems from the interaction between compositionality at the semantic level and the broader theory of communication in which semantics will be embedded. Traditionally, there has been a tight link between the two: the point of assertion is to transmit information by putting forth propositions. Moreover, traditional formal semantics has assigned propositions as the meanings of sentences, so a simple picture emerges whereby to assert that p, one can assertorically utter a sentence s that means that p. Famously, expressivists in metaethics have argued that moral language does not function this way but rather expresses approval or disapproval of actions. Classic proponents of expressivism take it to be a semantic thesis: moral terms and sentences containing them have different kinds of semantic value than others.

In Chapter 3 – *Pragmatic Expressivism and Non-Disjunctive Properties* – I begin by showing that distinguishing between compositional semantic value and assertoric content[22] allows a new form of expressivism, which I call *pragmatic expressivism*, to be formulated. On this view, expressivism is not a thesis first and foremost about the meanings of certain terms, but about what we *do* with them. By noting that the pragmatic expressivist can be seen as replacing the role played by propositional content in the standard theory of assertion with a role played by a property of an

[22]See Ninan [2010], Rabern [2012], Yalcin [2014].

agent's attitudes, I argue that pragmatic expressivists should endorse the following:

Non-Disjunctive Expressive Principle: The semantic value of a sentence, together with context, determines which *non-disjunctive* property of attitudes is expressed by that sentence in context.

The argument, in a nutshell, runs as follows: pragmatic expressivists identify a role in their theory of communication for a property of attitudes expressed by a sentence, but this role can only be played by non-disjunctive properties of attitudes.

After defending that argument, I show that the most well-developed version of pragmatic expressivism – that of Yalcin [2007, 2011, 2012] for epistemic modals – systematically violates the principle by allowing disjunctive properties to be expressed. A new pragmatic expressivism, based on an *assertability semantics* which evaluates sentences at information states (i.e. sets of worlds), is developed. I then state (and prove in Appendix A) a theorem showing that the new theory satisfies the principle in that *every* sentence expresses a non-disjunctive property of attitudes. As a welcome consequence of satisfying this demand, the theory also aptly handles other data concerning the interaction of epistemic modals and disjunction. The conclusion mentions future directions and preliminary experimental results – presented in Appendix B – corroborating an interesting prediction of the theory.

1.2.3 Does composition itself increase semantic complexity?

The third and final question I ask concerns the interaction of semantic composition with cognitive processes like sentence verification. In particular, thinking of such processes as algorithmic: how does compositionality work at an algorithmic level and does it increase the computational complexity of certain tasks? In Chapter 4 – *Composing Semantic Automata* – I address this question by looking at the task of quantifier sentence verification from the perspective of the semantic automata framework.

This framework shows how devices from theoretical computer science – in particular, automata – can be assigned as verification procedures for sentences containing quantified expressions like 'every student' and 'most apartments'. Certain quantifiers

– 'every', 'at least two', et cetera – can be captured by finite-state automata, the simplest kind of automaton. Others – 'most', 'less than two thirds of', et cetera – require a more complicated kind of automaton called a pushdown automaton which augments a finite-state one with a form of memory. After introducing all of these concepts and results, I summarize experimental work – both neuroimaging and behavioral – showing that this divide is psychologically real: when verifying the truth of sentences containing quantifiers requiring pushdown automata, people activate working memory in a way that they do not when the quantifiers require only a finite-state automaton.

After discussing the connection between semantic automata and semantic theory more generally, suggesting that automata provide 'level 1.5 explanations' and embody a distinction between semantic competence and performance, I present a new extension of the framework to handle sentences like 'most students take at least three classes', where quantified expressions appear as both subject and object of a transitive verb. The truth-conditions of such sentences are given as properties of binary relations built by a semantic composition operation called *iteration* of the two quantified expressions. Defining an analogous operation on automata allows us to address the question: does iteration increase the complexity of quantified sentence verification? In the Chapter, I prove that various classes of languages – star-free, regular, deterministic context-free – are *closed* under iteration: if two quantifiers have languages in that class, then the language for their iteration also belongs to that class. This shows that for a wide swath of natural language quantifiers, iteration does not increase the complexity of verification, when the complexity is measured by the Chomsky hierarchy of types of language.

I also use automata for iteration of quantifiers to address a question from the theory of generalized quantifiers, which are used to model the truth-conditions of the sentences in question. In particular, the Frege Boundary[23] divides the set of sentences which state properties of transitive verbs into those that can be expressed as iterations of quantifiers and those that cannot. While the Boundary itself has been

[23]As coined by van Benthem [1989b].

characterized,[24] it has been unknown whether the Boundary is *decidable*: is there an algorithm telling whether a given sentence of the right kind is an iteration or not? Using tools from automata and formal language theory, I answer this question in the affirmative, at least for the case where the sentence has a corresponding deterministic context-free language.

1.3 Summary and Previous Papers

This dissertation does not present an extended argument for one central thesis, as some do. Rather, it aims to show the fruitfulness of a shift in perspective on the nature of compositionality. By focusing on it as a procedural ability of speakers of natural languages, new questions naturally arise. As my previous summaries of the questions and their answers and the subsequent chapters show, this shift in perspective comes with a corresponding shift in methodology. While most of the status questions are addressed by using tools from algebra, this dissertation uses, among other tools, signaling games and computer simulations, logic and formal semantics, theoretical computer science, and experimental cognitive science. To my mind, this methodological diversity stands as a welcome consequence of the shift to a fresh perspective on compositionality. When new questions are asked, they should be addressed using the best tools available for addressing them. Should different questions require different tools, a researcher should try to expend their toolbox.

The dissertation does not stand on its own, but builds on earlier work of mine – much with wonderful co-authors – that has been previously published. I must here acknowledge these works and their publishers. The large majority of Chapter 2 appeared in *Journal of Logic, Language and Information* as Steinert-Threlkeld [2016c] and is reproduced here with Springer's permission. While the argumentative structure and final logical system of Chapter 3 are original, it has been heavily inspired by joint work with Peter Hawke, earlier versions of which have appeared as Hawke and Steinert-Threlkeld [2015, 2016a]. Once again, the permission of Springer is gratefully acknowledged. Chapter 4 contains a new and streamlined presentation of a mix of

[24] See Keenan [1992, 1996b], Dekker [2003].

original and joint (with Thomas Icard) work that has previously appeared as Steinert-Threlkeld and Icard III. [2013], Steinert-Threlkeld [2014, 2016a]. For a third and final time, Springer's permission is gratefully acknowledged.

Chapter 2

Compositional Signaling in a Complex World

> Denn der Mensch, als Thiergattung, ist ein singendes Geschöpf, aber Gedanken mit den Tönen verbindend.
>
> For man, as an animal species, is a singing creature, but associating thoughts with tones.
>
> von Humboldt [1836], p. 60

2.1 Introduction

In the epigraph, Humboldt points to an awe-inspiring feature of human language: that it allows us to use sensible devices to transmit thoughts across the gap between human minds. Perhaps even more remarkable than this feat is *how* we accomplish it. As Humboldt later puts it, through language we "make infinite use of finite means" (pp. 98-99). This has been taken to be an early statement of the productivity of our linguistic competence, which has in turn been used to motivate the famous principle of *compositionality*. Roughly:[1]

[1] The earliest statement in the Western tradition appears to be in Frege [1923]. Pagin and Westerståhl [2010a] find a similar statement in an Indian text dating from the fourth or fifth century.

(C) The meaning of a complex expression is determined by the meaning of its parts and how the parts are put together.

A sizable body of literature in philosophy and linguistics has debated the truth of (C).[2] Most arguments for (C) take the form of inferences to the best explanation: the truth of (C) gives us the best explanation of the learnability, systematicity, and productivity of our understanding of natural languages.[3] These last two properties refer, respectively, to the fact that anyone who understands a sentence like 'dogs like cats' automatically understands 'cats like dogs' and the fact that competent speakers can understand novel expressions that are built out of known simpler expressions and syntactic rules.

As discussed in Chapter 1, my aim in this dissertation is not to evaluate any of these arguments in particular or to answer the status question of whether (C) is true. Rather, I will simply suppose that (C) is true. In this chapter, then, I ask a new question: *why* is (C) true? In other words, what explains the fact that natural languages are compositional?

At first glance, it is hard to know what form an answer to this *why*-question would even take. The kind of answer I am interested in is broadly *evolutionary*. A crude first pass would transform one of the above arguments for the truth of (C) into an answer to the present question. Those arguments all take the form:

(7) Our linguistic abilities have property P.

(8) The best explanation for (7) is that (C).

(9) Therefore, (C).

By starting at the same point but taking a different fork in the argumentative road, one can turn this into an evolutionary argument along the following lines:

[2] A related literature also focuses on making the statement of (C) more precise so as to avoid worries that it is trivial and/or vacuous. See, for an overview, §2.1 of Pagin and Westerståhl [2010b]. In what follows, I will assume that (C) has been formulated in an appropriately more precise way.

[3] For learnability, see Davidson [1965]. For systematicity and productivity, see Fodor and Pylyshyn [1988], Fodor [1987], and Fodor [1998]. Some also take the principle to be a methodological one, guiding inquiry in semantics. See Szabó [2013], Pagin and Westerståhl [2010a,b], and Janssen [1997] for overviews of all of these proposals.

(7) Our linguistic abilities have property P.

(8)' That (7) would have been good for our ancestors.

(9)' Therefore, linguistic abilities with property P are likely to have evolved.

Unfortunately, this general form of argument is too unconstrained. Plenty of properties – say, the ability to run 50 miles per hour – would be good to have but have not evolved. In addition to constraints coming from the laws of physics and ecological pressures, the mechanisms of evolution – natural selection, in particular – operate locally in the sense that they can only select for genetic variations already present in a population of organisms.[4] Moreover, the mechanisms for introducing variation (drift, mutation, etc.) are unlikely to make every possible beneficial feature available. Haldane [1932] puts the point quite nicely:

> A selector of sufficient knowledge and power might perhaps obtain from the genes at present available in the human species a race combining an average intellect equal to that of a Shakespeare with the stature of Carnera. But he could not produce a race of angels. For the moral character or for the wings he would have to await or produce suitable mutations. (p. 110)

More generally, we can say that evolutionary answers to the *why*-question are constrained by a *how*-question: how, actually, did natural languages come to be compositional? Because the kind of evidence needed to answer this question – e.g. detailed archaeological records – is hard to come by and incomplete, the strongest constraint will be a different *how*-question: how, possibly, could natural languages come to be compositional? In this paper, I attempt to answer the *why*-question using models that are simple and widely-used enough to plausibly satisfy this how-possibly constraint. In particular, I will add a very rudimentary form of compositionality to Lewis-Skyrms signaling games and show that compositionality can help simple agents learn to signal in increasingly complex environments. This will exhibit a

[4]See chapter 3, section 5 of Bergstrom and Dugatkin [2012] for a discussion of constraints on natural selection. Futuyma [2010] considers more constraints than just a bottleneck in variation.

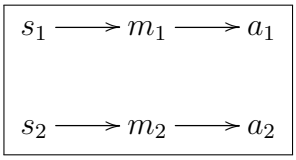
sense in which it would be good to have a compositional language. But the model of signaling and of learning used are simple enough that this learnability advantage could have conferred a fitness advantage in the prehistory of our species. Thus, while the results to be presented do not present a story about the gradual emergence of compositionality, they are likely to be compatible with any such story.

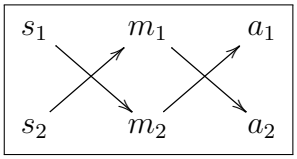
The chapter is structured as follows. In the next section, I introduce signaling games and the simple form of reinforcement learning that I will use later. Then, in section 3, I introduce the *negation game*, which enriches basic signaling games with non-atomic signals that are interpreted compositionally. Following that, section 4 presents simulation results of reinforcement learning in the negation game in increasingly large state spaces. We will see that compositional interpretation only makes learning easier in large state spaces and that how much compositionality helps learning strongly correlates with the number of states. From there, I drop the assumption that negation is built-in and ask whether the meaning of a function word can be learned. In a suitable set-up, I am able to prove that negation is learnable and offer simulation results that show that such learning happens very quickly. After that, I compare my work with other proposals in the literature and conclude with some future directions.

2.2 Signaling Games and Reinforcement Learning

Lewis [1969] introduced signaling games to understand how stable linguistic meaning could be conventional. In the simplest case, there are two agents – the Sender and the Receiver – who have common interests. The Sender can see which of two states of the world (call them s_1 and s_2) obtains and has two signals (call them m_1 and m_2) to send to the Receiver. The Receiver cannot see which state of the world obtains, but only which signal was sent by the Sender. Then, the Receiver can choose to perform one of two acts (call them a_1 and a_2). It is assumed that a_1 is appropriate in s_1 and a_2 in s_2 in that both the Sender and Receiver receive a benefit if a_i is performed in s_i, but not otherwise. Ideally, then, the Sender will send m_1 only in one state and m_2 only in the other and the Receiver will perform the appropriate act upon receiving m_i.

There are two ways of this happening, corresponding to the two so-called *signaling systems*. Schematically:





If we view the meaning of m_i as the state(s) in which it gets sent, this shows that the meaning depends on which signaling system the Sender and Receiver choose to adopt.[5] In this sense, it is conventional.

These ideas can be developed slightly more formally, in a way that will help for later. Writing $\Delta(X)$ for the set of probability distributions over a finite set X, we introduce a string of definitions.

Definition 2.1 (Signaling Game)**.** A signaling game is a tuple $\langle S, M, A, \sigma, \rho, u, P\rangle$ of states, messages, acts, a sender $\sigma : S \to \Delta(M)$, a receiver $\rho : M \to \Delta(A)$, a utility function $u : S \times A \to \mathbb{R}$, [6] and a distribution over states $P \in \Delta(S)$. σ and ρ have a common payoff, given by

$$\pi(\sigma, \rho) = \sum_{s \in S} P(s) \sum_{a \in A} u(s, a) \cdot \left(\sum_{m \in M} \sigma(s)(m) \cdot \rho(m)(a) \right) \tag{2.1}$$

We will also refer to $\pi(\sigma, \rho)$ as the *communicative success rate* of σ and ρ. If σ is not probabilistic but rather a deterministic function, we call it a *pure sender*. *Mutatis mutandis* for ρ and receiver.

[5] This corresponds to one of two types of informational content for a signal identified in chapter 3 of Skyrms [2010] and reflects the idea that a proposition is a set of possible worlds.

[6] In not having the utility function depend on the signal sent, we are assuming that no signal is more costly to send than any other.

The most well-studied class of signaling games are generalizations of the two-state game described above, where there is an equal number of states, signals, and acts, with exactly one act appropriate for each state. I will call these Atomic n-games.

Definition 2.2 (Atomic n-game)**.** The *Atomic n-game* is the signaling game where $|S| = |M| = |A| = n$, $u(s_i, a_j) = 1$ iff $i = j$, 0 otherwise, and $P(s) = 1/n$ for every $s \in S$.

Definition 2.3. The *Lewis game* is the Atomic 2-game.

Definition 2.4. A *signaling system* in a signaling game is a pair (σ, ρ) of a sender and receiver that maximizes $\pi(\rho, \sigma)$.

The reader can verify that the two signaling systems depicted at the beginning of this section are indeed the only ones in the Lewis game according to this definition. In fact, they are the only two strict Nash equilibria of the Lewis game. More generally, Trapa and Nowak [2000] show that the Atomic n-game always has strict Nash equilibria, all of which have the following form. By currying, we can view σ as an $n \times n$ stochastic matrix S, with $S_{ij} = \sigma(s_i)(m_j)$ and ρ as an $n \times n$ stochastic matrix R, with $R_{kl} = \rho(m_k)(a_l)$. Then: (σ, ρ) is a stict Nash equilibrium iff S is a permutation matrix[7] and $R = S^T$.[8] These strict equilibria have $\pi(\sigma, \rho) = 1$, which is the maximum value that $\pi(\sigma, \rho)$ can obtain. There are, however, other Nash equilibria in the Atomic n-game which have lower communicative success rates.

2.2.1 Reinforcement Learning in Signaling Games

Lewis' full analysis of conventional meaning in terms of signaling systems makes strong assumptions about common knowledge and rationality of the Sender and Receiver. Thus, while it provides a nice conceptual analysis, it cannot offer an explanation of how agents might *arrive at* or come to use a signaling system. In a long series of work, Skyrms [1996, 2004, 2010] and colleagues have weakened these assumptions

[7] A matrix with only 1s and 0s, where each row has a single 1 and distinct rows have 1s in distinct columns.

[8] Here, M^T denotes the transpose of M, i.e. $\left(M^T\right)_{ij} = M_{ji}$.

and explored various dynamic processes of learning and evolution in the context of signaling games. Here, I outline one exceedingly simple form of learning called Roth–Erev reinforcement learning.[9]

Now, instead of the signaling game being played once, we imagine a Sender and Receiver repeatedly playing the game. We also imagine that each is equipped with some urns: the Sender has an urn for each state with balls in it labeled by signals, while the Receiver has an urn for each signal with balls in it labeled by acts. When the Sender finds out what state obtains, it draws a ball from the corresponding urn and sends the signal written on the ball. When the Receiver hears the signal, it draws a ball from the corresponding urn and performs the act written on the ball. We assume initially that each urn has one ball of every appropriate kind. If the act was appropriate for the state, the Receiver adds a ball of that act to the signal's urn and the Sender adds a ball of that signal to the state's urn. Otherwise, nothing happens. The addition of balls of the same type after successful play of the game makes it more likely that similar choices will be made in the future, thus reinforcing the choices that led to successful signaling. In the context of the how-possibly constraint mentioned in the introduction, it must be noted that Schultz et al. [1997] show that dopamine neurons in certain areas of primate brains appear to implement a similar reinforcement learning procedure.[10]

Slightly more formally and generally,[11] we keep track of the *accumulated rewards* of the players' choices. That is, we have functions $ar_{\sigma,t} : S \times M \to \mathbb{R}$ and $ar_{\rho,t} : M \times A \to \mathbb{R}$. At $t = 0$, these are set to some initial values, usually 1. They are then incremented by

$$ar_{\sigma,t+1}(s_i, m_j) = ar_{\sigma,t}(s_i, m_j) + u(s_i, a_k)$$
$$ar_{\rho,t+1}(m_j, a_k) = ar_{\rho,t}(m_j, a_k) + u(s_i, a_k)$$

[9] See Roth and Erev [1995]. Sutton and Barto [1998] provides a detailed introduction to reinforcement learning.

[10] That seminal paper has spanwed a large body of research. See Schultz [2004] and Glimcher [2011] for overviews.

[11] For instance, the ball-in-urn metaphor essentially assumes that the utility function only has integer values.

where s_i, m_j, and a_k were the state, message, and act played at time t. How, though, do the Sender and Receiver choose their signals and acts? The simple idea captured by the urn metaphor is that they do so in accord with their accumulated rewards; that is, to the extent that those choices have been successful in the past. In other words:

$$\sigma_{t+1}(s_i)(m_j) \propto ar_{\sigma,t}(s_i, m_j)$$
$$\rho_{t+1}(m_j)(a_k) \propto ar_{\rho,t}(m_j, a_k)$$

The simplicity of this learning method does not prevent it from being effective. Consider, first, the Lewis game. In simulations with every urn containing one ball of each type and payoffs equaling one, after 300 iterations, the Sender and Receiver have $\pi(\sigma, \rho) \approx 0.9$ on average.[12] In fact, for this simplest case, one can prove that Roth-Erev learning converges to a signaling system.

Theorem 2.1 (Argiento et al. [2009])**.** *In the Lewis game, with probability 1,*

$$\lim_{t\to\infty} \pi(\sigma_t, \rho_t) = 1$$

Moreover, the two signaling systems are equally likely to occur: with probability 1/2,

$$0 = \lim_{t\to\infty} \frac{\sigma_t(s_1, m_1)}{\sigma_t(s_1, m_2)} = \lim_{t\to\infty} \frac{\sigma_t(s_2, m_2)}{\sigma_t(s_2, m_1)} = \lim_{t\to\infty} \frac{\rho_t(m_1, a_1)}{\rho_t(m_1, a_2)} = \lim_{t\to\infty} \frac{\rho_t(m_2, a_2)}{\rho_t(m_2, a_1)}$$

while with probability 1/2, the limits of the reciprocals are 0.

When considering the Atomic n-game for larger n, simulations have somewhat mixed results. Barrett [2006] finds that for $n = 3, 4$, after 10^6 iterations, agents have $\pi(\sigma, \rho) > 0.8$ at rates of .904 and .721, respectively. For $n = 8$, that number drops to .406. Nevertheless, the agents always perform significantly better than chance. I will present my own simulation results for this situation later and so do not dwell on them here. Analytically, Hu et al. [2011] show that there is a large class of non-strict Nash

[12] See Skyrms [2010], p. 94.

equilibria of the Atomic n-game (the so-called partial pooling equilibria) to which Roth-Erev learning converges with positive probability.[13]

2.3 The Negation Game

Having seen the basics of signaling games and reinforcement learning, we are in a position to use these tools to answer our original question: *why* would compositional languages arise? To address this question, we first need to extend the Lewis-Skyrms signaling games to handle rudimentary forms of compositional signaling. In particular, we will focus on having a signal corresponding to *negation*.

To get an intuition for how the model will work, consider vervet monkeys.[14] These monkeys have three predators: leopards, eagles, and snakes. It turns out that they have three acoustically distinct signals – a bark, a cough, and a chutter – that are typically sent only when a predator of a particular type is present. Moreoever, when a vervet receives one of the signals, it appears to respond in an adaptive way: if it hears a bark, it runs up a tree; if a cough, it looks up; if a chutter, it looks down. Schematically, their behavior looks like:

$$\text{leopard} \longrightarrow \text{bark} \longrightarrow \text{climb tree}$$

$$\text{eagle} \longrightarrow \text{cough} \longrightarrow \text{look up}$$

$$\text{snake} \longrightarrow \text{chutter} \longrightarrow \text{look down}$$

As the diagram above suggests, a signaling system in the Atomic 3-game provides a natural model for this signal/response behavior.

More speculatively, suppose that a vervet on the ground saw that a group-mate had run up a tree as if a leopard were present. The vervet on the ground, however,

[13] Similar results hold for the closely related replicator dynamics. See Huttegger [2007], Pawlowitsch [2008]. Hofbauer and Huttegger [2008, 2015] study the replicator-mutator dynamics for these games.

[14] See Seyfarth et al. [1980].

knows that no leopard is present. It would be useful if it could signal such a fact to the monkey in the tree so that it would come down and the group could get on with its business. To signal that no leopard was present, the vervet on the ground could attempt to use a completely new signal. Alternatively, it could send a signal along the lines of "no bark", relying on the fact that the other vervets already know how to respond to barks. This latter solution seems *prima facie* superior since it leverages the existing signaling behavior. Moreover, Zuberbühler [2002] finds a syntax of exactly this simple kind in Campbell monkeys: they have alarm calls for leopards and eagles, which can be prefixed with a low 'boom boom' sound to indicate that the threat is not immediate.[15] That some monkeys have such syntax makes it plausible that the following model will fit well with theories of the gradual emergence of compositionality.

To model signaling behavior of this sort, let us introduce *the negation game*. For a given n, there are $2n$ states and acts with the same utility function as in the Atomic $2n$-game. Now, however, instead of all messages being simple/atomic, the Sender can also send a *sequence* of two signals. The Sender has n "basic" signals $m_1, \ldots, m_n$ and can also send signals of the form $\boxminus m_i$ for some new signal $\boxminus$ and $1 \leqslant i \leqslant n$. Now, in the extended vervet case, and with negation in general, there are certain logical relations among the states. The model here will capture a few basic intuitions about the relations imposed by negation:

- Every state has a negation.
- The negation of a state is distinct from the state.
- Distinct states have distinct negations.

The relevant mathematical notion capturing these intuitions is that of a *derangement*: this is a bijection (a one-to-one and onto function) with no fixed points, i.e., no x such that $f(x) = x$.[16] So, writing $[n] = \{1, \ldots, n\}$, we also assume that we are equipped with a derangement $f : [2n] \to [2n]$.[17] We will consider f as a function on both the

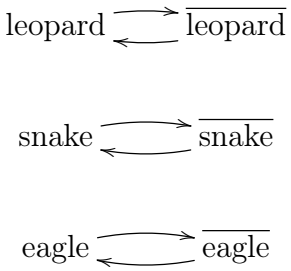
[15] See Schlenker et al. [2014] for a detailed semantic analysis of this form of signaling. They in fact find that the basic alarm calls have roots with a morphological suffix.

[16] See Hassani [2003] for the definition and applications.

[17] To fully model classical logical negation, we would also need to require that f is an *involution*,

states and acts by writing $f(s_i) := s_{f(i)}$ and $f(a_i) := a_{f(i)}$. Two examples of state spaces with derangements may be illuminating.

Example 2.1. There are six states: for each of a leopard, snake, and eagle, one state indicates the presence and another the absence thereof. The function sending presence states to the corresponding absence state and *vice versa* is a derangement, depicted here:



Example 2.2. Let W be a set of *worlds*. Then $\mathcal{P}(W)$, the set of subsets of W, is the set of propositions based on those worlds. *Set complement* is a derangement on $\mathcal{P}(W)$, often used to specify the meaning of negation: $[\![\neg\varphi]\!] = W \backslash [\![\varphi]\!]$.

These examples show how the notion of a derangement captures the minimal core of negation in terms of the three intuitions above. Because of this, I will use the term 'minimal negation' in this context. The question now becomes: how can the Sender and Receiver exploit this structure to communicate effectively?

First, consider the Sender. For simplicity, let us suppose that this is a pure sender, so that it always sends one signal in each state. Without loss of generality, we can even suppose that it sends m_i in s_i. Nature might inform the Sender, however, that state s_k obtains for some $k > n$. What signal, then, should it send? We can use the derangement f which captures the idea of minimal negation. For some i, $s_k = f(s_i)$. The Sender will therefore send $\boxminus m_i$. To put it more formally in order to foreshadow the generalization to a Sender using a mixed strategy, $\sigma(f(s_i)) = \boxminus\sigma(s_i)$.

---

i.e. that $f(f(i)) = i$. But since our syntax is so impoverished that sending a "double negation" signal is not possible, we can omit this requirement. It is also doubtful that natural language negation satisfies double negation elimination. This remark about involutions does, however, explain why we have $2n$ states and acts: a permutation that is an involution will only have cycles of length $\leqslant 2$. And being a derangement requires that there are no cycles of length 1. Together, this means that $[n]$ only has derangements which are involutions if n is even.

What about the Receiver? It will "interpret" $\boxminus$ as a minimal negation in the following way. When it receives a signal $\boxminus m_i$, it looks at the act it would take in response to m_i. Suppose that's a_i. Instead, however, of taking that act, it performs $f(a_i)$. Again, to put it formally: $\rho(\boxminus m_i) = f(\rho(m_i))$. Here already we see the rudiments of compositionality: a complex signal is 'interpreted' as a function of the interpretation of its part. Now, consider a Sender and Receiver playing in this way. If s_k obtains, the Sender sends $\boxminus m_i$. The Receiver then will play $f(\rho(m_i)) = f(a_i) = a_k$, which is the appropriate act.

This simple model shows how sending signals corresponding to minimal negations can enable the Sender and Receiver to reach a signaling system in a large and logically structured state space. As presented, however, we required that both agents played pure strategies and that they had an agreement on the meanings of the basic signals. We can generalize the above definitions to allow for mixed Sender and Receiver strategies. Once we have done that, we can use Roth-Erev reinforcement to try and *learn* those strategies.

Definition 2.5 (Negation n-game)**.** A *Negation n-game* is a tuple $\langle S, M, A, \sigma, \rho, u, P, f\rangle$ where $|S| = |A| = 2n$, $f : [2n] \to [2n]$ is a derangement, and $u(s_i, a_j) = 1$ iff $i = j$, 0 otherwise. $M = \{m_1, \ldots, m_n\} \cup \{\boxminus m_i : 1 \leqslant i \leqslant n\}$. Moreover, we require that:[18]

$$\sigma(f(s_j))(\boxminus m_i) \propto \sigma(s_j)(m_i) \tag{2.2}$$

$$\rho(\boxminus m_i)(f(a_j)) = \rho(m_i)(a_j) \tag{2.3}$$

Payoffs are given by (2.1) as before.

We can understand these constraints in terms of the urn model described above. Consider, first, the Sender. Now, the Sender's urns also contain balls labelled $\boxminus$. If one of these balls is drawn from an urn (say, s_j), then the Sender looks at the

[18] Or equivalently:

$$\sigma(s_j)(\boxminus m_i) \propto \sigma\left(f^{-1}(s_j)\right)(m_i)$$
$$\rho(\boxminus m_i)(a_j) = \rho(m_i)\left(f^{-1}(a_j)\right)$$

urn for $f^{-1}(s_j)$, draws a non-$\boxminus$ ball (say, m_i) and then sends $\boxminus m_i$. This Sender behavior exactly implements the constraint (2.2). The Receiver's behavior is even simpler: for signals of the form m_i, it simply draws a ball from the appropriate urn as before. When, however, encountering a signal $\boxminus m_i$, the Receiver draws a ball from the m_i urn (labelled, say, a_k) and then performs $f(a_k)$. This Receiver behavior exactly implements the constraint (2.3). In this way, the Sender uses and the Receiver interprets $\boxminus$ as minimal negation. Given this model of behavior, implementing Roth-Erev reinforcement learning for the Negation n-game is as simple as reinforcing all the choices made in a given iteration. That is, we perform the following reinforcements:

$$\begin{aligned} ar_{\sigma,t+1}(s_j, \boxminus) &= ar_{\sigma,t}(s_j, \boxminus) + u(s_j, f(a_k)) \\ ar_{\sigma,t+1}\left(f^{-1}(s_j), m_i\right) &= ar_{\sigma,t}\left(f^{-1}(s_j), m_i\right) + u(s_j, f(a_k)) \\ ar_{\rho,t+1}(m_i, a_k) &= ar_{\rho,t}(m_i, a_k) + u(s_j, f(a_k)) \end{aligned}$$

Learning in the Negation n-game provides a model to answer the following question: given the ability to use a signal to mean minimal negation, can the Sender and Receiver learn how to use the atomic words to communicate effectively?

2.4 Experiment

We are now in position to test whether compositional languages are in some sense better. In particular, we will ask whether such languages are easier to learn than languages with only atomic signals. Consider again Example 2.1, where the vervets want to communicate about both the presence and absence of leopards, snakes, and eagles. *Prima facie*, it seems like having minimal negation around would make learning easier: once signals for the three predators are known, the signals for their absence are also automatically known by prefixing with the negation signal. By contrast, with only atomic signals, three new unrelated signals would need to be introduced to capture the states corresponding to the lack of each predator. Of course, similar thoughts apply for more or fewer than three predators, so I will compare atomic versus compositional languages when there are different numbers of states.

2.4.1 Methods

To test the hypothesis that compositional languages will be easier to learn, I ran 100 trials of Roth-Erev learning for each of the Atomic $2n$-game and the Negation n-game for each $n \in \{2, 3, 4, 5, 6, 7, 8\}$. The initial value of accumulated rewards was 1 for every argument. Similarly, $u(s_i, a_j) = 1$ when $i = j$. Each trial consisted of 10,000 iterations. This relatively low value was chosen so that differences in speed of learning may also be apparent. I measured $\pi(\sigma, \rho)$ for each trial at the end of the iterations of learning. The code for running these simulations and performing the data analysis, in addition to the actual data, may be found at http://github.com/shanest/learning-evolution-compositionality. Let $\textsc{Atom}_{2n}$ (respectively, $\textsc{Neg}_n$) refer to the set of 100 such payoffs of the Atomic $2n$-game (respectively, Negation n-game). I will use $\overline{(\cdot)}$ to refer to the mean of a set of values.

2.4.2 Results

To measure the effectiveness of Roth-Erev learning, the mean payoffs of the trials of each game for each n were computed. Welch's t-test was used to test the hypothesis that those means are different. These results are summarized in Table 2.1, where $\textsc{Diff}_n = \overline{\textsc{Neg}_n} - \overline{\textsc{Atom}_{2n}}$.

n	2	3	4	5	6	7	8
$\overline{\textsc{Atom}_{2n}}$	0.914	0.863	0.819	0.758	0.699	0.617	0.506
$\overline{\textsc{Neg}_n}$	0.851	0.816	0.786	0.746	0.706	0.661	0.605
$\textsc{Diff}_n$	−0.064	−0.046	−0.033	−0.013	0.007	0.044	0.099
t	3.666	3.461	2.897	1.117	−0.724	−4.825	−10.88
p	0.0003	0.0007	0.0042	0.2654	0.4698	0.000003	0.6e-22

Table 2.1: Means and t-tests for Roth-Erev reinforcement learning.

To measure the effect, if any, of n on the effectiveness of Roth-Erev learning in the two types of game, a simple linear regression of $\textsc{Diff}_n$ on n was run. This yielded $r = 0.9658$ and $p = 0.00041$, indicating a very strong positive linear correlation. Figure 2.1 shows the line of best fit overlayed on the data.

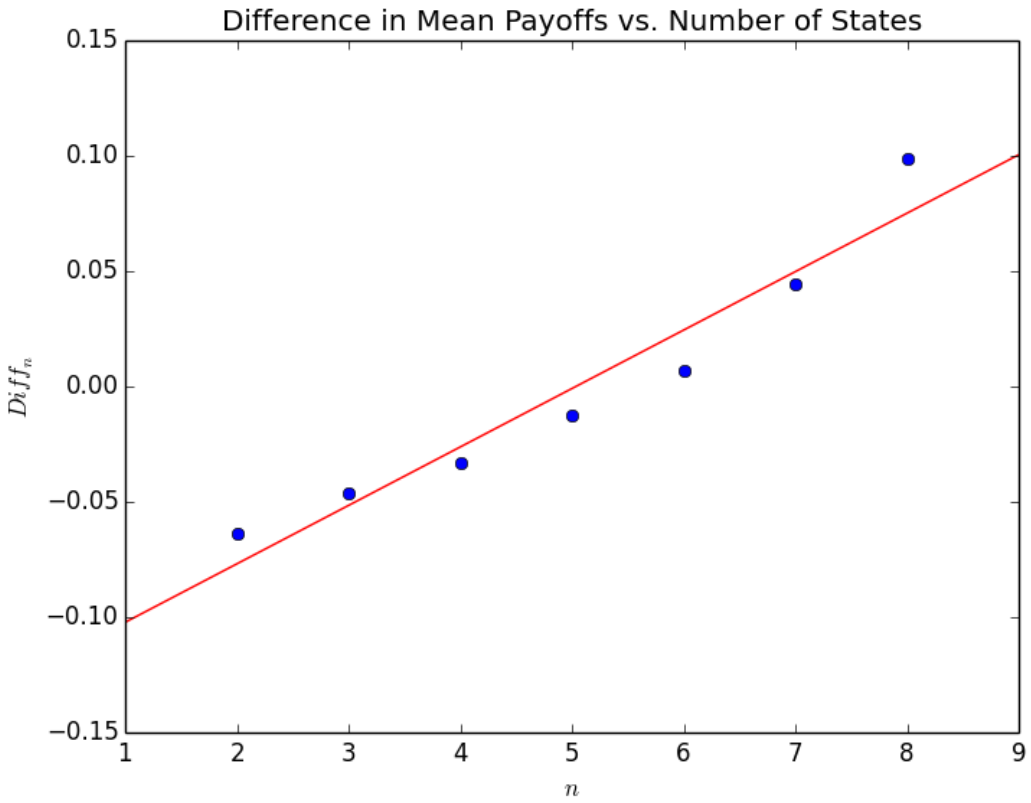


Figure 2.1: Fitting a line to $\textsc{Diff}_n$.

2.4.3 Discussion

There are two main take-aways from these results. First, having a compositional language does not always make learning easier. To see this, note that a negative value of $\textsc{Diff}_n$ means that agents learn to communicate more successfully with the atomic language than with the minimal negation language for that value of n. This occurs for $n = 2, 3, 4, 5$. Moreover, the p values from the Welch's t-test show that for $n = 2, 3, 4$, this difference in mean success is in fact statistically significant. Thus, when there are not many states (4, 6, and 8 respectively) for the agents to distinguish, an atomic language is easier to learn. By the time we reach $n = 7$ (14 states), however, the agents using a compositional language do perform statistically significantly better, as evidenced by the positive value for $\textsc{Diff}_7$ and the very low p value there. This

trend also continues for the $n = 8$ (16 state) case.

This leads to the second observation: the very strong positive correlation between n and $\textsc{Diff}_n$ ($r = 0.9658$) shows that it becomes increasingly more advantageous from a learning perspective to have a compositional language as the number of states increases. Thinking of the number of states as a proxy for the complexity of the world about which the agents must communicate, we can summarize these two results in a slogan: compositional signaling can help simple agents learn to communicate in a complex world.

Taken together, these results present a precise model identifying one ecological pressure that could explain the evolution of compositionality: the need to communicate about a large number of situations. Moreover, the surprising fact that compositional languages only confer a learning benefit in large state spaces shows another sense in which the crude evolutionary arguments from the introduction are too unconstrained. Recall that premise (2') claimed that, for example, productive understanding would have been 'good' for our ancestors. Perhaps one could even have identified the sense of goodness with learnability. These simulations, however, show that the situation is not so simple: the sense in which a compositional language is good depends very strongly on the size of the state space. One could not have figured that out without analysis of the kind presented here.

Note that this experiment examined only the most rudimentary form of compositional semantics. Nevertheless, the fact that this form only exhibited advantages in a suitably complex world suggests that even more sophisticated languages (with, for instance, binary operators and a partially recursive syntax) will only be advantageous when the world in which the agents are embedded also is substantially more complex. Moreover, that compositionality only confers an advantage in large state spaces may explain why it is rare in the animal kingdom: many species may not have pressure to talk about many different situations.

2.5 Learning Negation

While the above experiment shows how having a rudimentary form of compositional signaling can be beneficial, it leaves open the question: where did the compositional signaling come from? That is, can the agents *learn* to use a signal as negation?

To model the learning of negation itself, we need to relax the definition of the Negation n-game so that the Sender must learn how to use $\boxminus$ and the Receiver must choose how to interpret the signal $\boxminus$. Instead of putting one derangement in the model, we will put in a whole set of functions $\mathcal{F}$. The Receiver will have distributions over the acts for each basic signal and a distribution over $\mathcal{F}$ for $\boxminus$. Its choice on how to interpret $\boxminus m_i$ will directly refer to this latter distribution which reflects how it will interpret $\boxminus$ as a function.

Definition 2.6 (Functional n-game)**.** A *Functional n-game* is a tuple $\langle S, M, A, \sigma, \rho, u, \mathcal{F}\rangle$ where S, M, A, and u are as in the definition of a Negation n-game (see Definition 2.5). $\mathcal{F}$ is a set of functions $[2n] \to [2n]$ (not necessarily bijections). We require that exactly one $g \in \mathcal{F}$ is a derangement. For the Receiver, we require that

$$\rho\left(\boxminus m_i\right)\left(a_j\right) = \sum_{g \in \mathcal{F}} \rho\left(\boxminus\right)\left(g\right) \cdot \sum_{k:g(k)=j} \rho\left(m_i\right)\left(a_k\right) \tag{2.4}$$

Intuitively, this constraint captures the following behavior. The Sender no longer uses $\boxminus$ explicitly as minimal negation; its behavior is unspecified. Now, instead of the Receiver automatically interpreting $\boxminus$ as minimal negation, it first *chooses* a function by which to interpret it. In terms of the urn models, the Receiver now has another urn labeled $\boxminus$. Balls in this urn, however, are labeled with functions $g \in \mathcal{F}$. When the Receiver receives a signal $\boxminus m_i$, it draws a function g from the $\boxminus$ urn, an act a_i from the m_i urn and performs $g(a_i)$. Constraint (2.4) exactly captures that behavior. Note that when $\mathcal{F}$ is a singleton containing a lone derangement f, Constraint (2.4) reduces to Constraint (2.3) because $\rho\left(\boxminus\right)\left(f\right) = 1$ and there will be only one k such that $f(k) = j$ since f is a bijection.

To look at the emergence of minimal negation, we want to capture a natural idea about how agents would learn to use a function word: they are already capable of

communicating with atomic signals and then try to *introduce* the functional element. To model this kind of situation, we need a few more definitions. As before, let S be a set and f a derangement on S. Call a set $X \subset S$ *complement-free* iff $X \cap f[X] = \varnothing$. If X is a maximal complement-free subset (in the sense of not being contained in another complement-free set), we will call it a *complementizer* of S. As an example of a complementizer, note that both the sets $\{\text{leopard}, \text{eagle}, \text{snake}\}$ and its complement are complementizers with respect to the derangment mentioned in Example 2.1 above. Note that if $|S| = 2n$, then all complementizers of S have cardinality n. For a set S of size n and a subset $I \subseteq [n]$, write $S \restriction I = \{s_i : i \in I\}$.

The situation we are interested in is the following: the agents are playing a Functional n-game. For some complementizer X of $[2n]$ with respect to the derangement $f \in \mathcal{F}$, σ and ρ in addition constitute a signaling system on the subgame generated by $S \restriction X$ and $A \restriction X$ in which σ only sends basic signals from $\{m_1, \ldots, m_n\}$. In this setting, the Sender needs to choose when to send complex signals $\boxminus m_i$ and the Receiver needs to choose how to interpret $\boxminus$. We will suppose that the Receiver chooses between a few natural responses to hearing new signals with $\boxminus$. The Receiver might (a) ignore $\boxminus$, (b) treat $\boxminus$ as a new atomic word, or (c) treat $\boxminus$ as minimal negation. These options can be modeled by different functions: (a) corresponds to the identity function $id(j) = j$, (b) to a constant function $c_i(j) = j$ and (c) to a derangement f. Therefore, to model the Receiver choosing among these natural options for interpreting $\boxminus$, we assume that $\mathcal{F}$ contains exactly those three functions.

Call a Functional n-game with the above restrictions a *basic negation learning setup of size n*. Our question now is: does simple Roth-Erev reinforcement learning allow the Sender to start using $\boxminus$ like minimal negation and the Receiver to learn to interpret $\boxminus$ as minimal negation? We answer this question using both analytic and simulation results.

2.5.1 Analytic Result

It turns out that in this setup, we can actually *prove* that Roth-Erev learning works.

Theorem 2.2. *In a basic negation learning setup of size n, if $i \in X$ for the $c_i \in \mathcal{F}$,*

then:

- *With probability 1, for the derangement f, $\lim_{t\to\infty} \rho_t(\boxminus)(f) = 1$. In other words, the probability that $\boxminus$ is interpreted as minimal negation by the Receiver converges to 1.*
- *With probability 1, σ converges to a strategy where constraint (2.2) holds. In other words, with probability 1, the Sender learns to use $\boxminus$ as minimal negation.*

Proof. Theorem 4 of Beggs [2005] states that probabilities and empirical frequencies converge to 0 for strategies which do not survive iterated removal of strictly dominant strategies (when there are a finite number of players and actions) under Roth-Erev learning.

For the first part, this means that it suffices to show that choosing the derangement f is strictly dominant for the Receiver. To see this, note that σ only sends signals of the form $\boxminus m_i$ for states in $f[X]$. If the Receiver chooses c_i, then, the payoff will be 0 since $i \in X$. Similarly, since m_i is only sent in $s_i \in X$, if the Receiver chooses id, it will perform an act in $A \upharpoonright X$ and so receive a payoff of 0. Choosing the derangement is thus strictly dominant.

For the second part, recall that in a basic negation learning setup, σ and ρ are a signaling system on the subgame restricted to the complementizer X. This means that for $s_i \in X$, w.l.o.g., $\sigma(s_i)(m_i) = 1$. Now, consider $f(s_i)$. We must show that $\boxminus m_i$ is the only signal that survives iterated removal of dominant strategies. Suppose that σ sends a basic signal m_j in $f(s_i)$. By assumption, for some $k \in X$, $\rho(m_j) = a_k \neq f(a_i)$ since $i \in X$ and $X \cap f[X] = \varnothing$, so the payoff is 0. So no basic signal survives. Now, since we know that the Receiver will interpret $\boxminus$ as f, only $\boxminus m_i$ will lead to the Receiver performing $f(a_i)$ and to the Sender receiving a positive payoff. We thus have that $\sigma(f(s_i))(\boxminus m_i)$ converges to 1, which yields (2.2) above because $\sigma(s_i)(m_i) = 1$. □

This proof depends on the assumption that $i \in X$. If $i \in f[X]$, then choosing f is only *weakly* dominant over choosing c_i and so the result of Beggs [2005] no longer

applies. For this reason, and because the above convergence result does not tell us about *rate* of convergence, I also ran simulations of the basic negation learning setup.

2.5.2 Simulation Results

I ran 100 trials of only 1000 iterations of the basic negation learning setup for each $n \in \{2, 3, 4, 5, 6, 7, 8\}$ both with $i \in X$ and $i \in f[X]$. Table 2.2 shows the average payoffs and probability of interpreting $\boxminus$ as negation after the 100 trials.

	n	2	3	4	5	6	7	8
$i \in X$	$\pi(\sigma, \rho)$	0.995	0.989	0.978	0.962	0.937	0.899	0.855
	$\rho(\boxminus)(f)$	0.995	0.995	0.995	0.994	0.993	0.993	0.991
$i \in f[X]$	$\pi(\sigma, \rho)$	0.904	0.875	0.870	0.818	0.821	0.756	0.701
	$\rho(\boxminus)(f)$	0.816	0.844	0.807	0.666	0.922	0.760	0.830

Table 2.2: Statistics for Roth-Erev reinforcement in the basic negation learning setup.

These results show two things. First, learning happens *very fast* in this setup. This holds especially true for the Receiver, who has uniformly high values for $\rho(\boxminus)(f)$ in the $i \in X$ case. The Sender appears to need more time to fully learn its strategy in the large state spaces. Second, having $i \in f[X]$ does result in lower values for both the payoff and for $\rho(\boxminus)(f)$. Moreover, t-tests show that all of these differences are in fact statistically significant. This provides some initial reason to doubt that Theorem 2.2 generalizes to the case when $i \in f[X]$. Nevertheless, the Theorem and these simulations show that the task of learning the meaning of a function word when the meanings of atoms are known is rather easy even for very simple learners.

2.6 Comparison to Other Work

While the present paper does present a precise answer to the question of why natural languages are compositional, it is not the first to attempt to do so. Therefore, I will now discuss a few earlier proposals from the literature. In each case, I find that something distinctive about the nature of function words has been left out.

Nowak and Krakauer [1999] study the emergence of compositionality via a kind of natural selection of signal-object pairings. For our purposes, the most interesting case is their last, where they consider a state space consisting of pairs of two objects and two properties, for four total combinations. These combinations can be specified by four atomic words w_1, w_2, w_3, w_4 or with pairs $p_i o_j$. Nowak and Krakauer consider a strategy space where players use the atomic words with probability p and the 'grammatical' constructs with probability $1 - p$. They are able to show that the only two evolutionary stable strategies are when $p = 0$ and $p = 1$ and that their evolutionary dynamics evolves to use the grammatical rule with probability 1. In trying to use syntactic structure to mirror structure in the state space, the present approach does have something in common with Nowak and Krakauer's. There are, however, two reasons that their proposal does not go far enough. First, they only compare atomic and compositional signaling systems but do not analyze whether one or the other form of signaling makes it easier to arrive at a signaling system. One would like a dynamic that considers more than two possible signal-state mappings. Secondly, they are not explicitly interested in function words but only in subject-predicate structure. Both models are needed for a full story of the evolution of complex signaling: their emphasis is on the structure of minimally meaningful signals, while mine is on how function words can be used to modify the meanings of already meaningful signals.

In a series of papers, Barrett [2006, 2007, 2009] considers 'syntactic' signaling games with reinforcement learning where the number of states and acts exceeds the number of signals but where there is more than one sender. These are equivalent to games in which there is one sender who sends a sequence of signals for each state (the length of the sequence is fixed and the strategies are independent for each position in the sequence). Perfect signaling is achieved when sender-receiver strategies settle on a coding system with a sequence that elicits the ideal act in each state. For example, consider the 4-state case with 2 senders, each of which has 2 signals. An example of optimal signaling behavior has Sender 1 use m_1 in $\{s_1, s_2\}$ and m_2 in the other two states, while Sender 2 uses m_1 in $\{s_1, s_3\}$ and m_2 in the other two states. If the Receiver can take the intersection of the sets of states indicated by the

two messages received, it can learn the exact state and so perform the correct act. Barrett found that in simulations involving 4-state, 4-act, 2-signal, 2-sender signaling games, successful signaling[19] is achieved by reinforcement learning approximately 3/4 of the time and that in 8-state, 8-act, 2-signal, 3-sender games, successful signaling is achieved 1/3 of the time.

The difference, then, between this approach and our own is that there is no explicit signal that serves as a function word in Barrett's set-up; rather, signal concatenation is implicitly treated like conjunction. Of course, natural language contains many such function words, meaning that his model cannot provide a full account of the emergence of natural language compositionality. Moreover, as Franke [2014] observes, the Receiver in syntactic game appears to not actually interpret length-2 signals compositionally: "Although we can describe the situation as one where the meaning of a complex signal is a function of its parts, there is no justification for doing so. A simpler description is that the receiver has simply learned to respond to four signals in the right way" (p. 84). In other words, the Receiver's dispositions treat the complex signals as if they were atomic. Nevertheless, there is some similarity between the present paper and Barrett's work. The task in Experiment 1 can now be recast as the Sender having to learn a complementizer X on the state space. The simulation results show that this is actually a somewhat harder task than learning two complementary partitions.

A final proposal comes from Franke [2014] where the explicit concern is with using reinforcement learning in signaling games to explain compositional meanings. Franke adds 'complex' signals of the form m_{AB} (as well as 'complex' states of the form s_{AB}). Formally, these are just new atomic signals. The sense in which they bear a relation to the basic signals is captured in a distance $s = d(m_{AB}, m_A) = d(m_{AB}, m_B)$ which intuitively represents a kind of similarity (and similarly for the set of complex states of the form s_{AB}). Franke modifies Roth-Erev learning by incorporating (i) spill-over and (ii) lateral inhibition. Spill-over refers to the fact that non-actualized message/state pairs (for the sender) are reinforced in proportion to their similarity to the played one (and similarly for the receiver). Lateral inhibition involves lowering the accumulated

[19]For Barrett, this means $\pi(\sigma, \rho) > 0.8$.

rewards for non-actualized pairs when the actualized pair was successful. Franke shows that with these two modifications, creative compositionality can arise in the following sense: if Sender/Receiver play only using 'basic signals', they can still learn to have $\sigma(m_{AB})(s_{AB}) > \sigma(m_j)(s_{AB})$ for all $m_j \neq m_{AB}$ and so on for the other complex states and signals. Since the 'complex' signal is the one most likely to be sent when the 'complex' state is seen for the first time, this purportedly captures creativity: using a new signal for a new state.[20]

There are two main differences between Franke's work and the present approach. First, in his model, signals do not have genuine syntactic structure which then gets compositionally interpreted. One does not find compelx signals with words treated like function words as one does in the Negation n-game. Secondly, his model claims to provide an explanation for the *emergence* of compositional signaling, whereas the present paper has simply identified an evolutionary advantage that compositional signaling would confer. One would like a model that incorporates both of these aspects: a story about the emergence of genuinely syntactically structured signals with compositional interpretation.

2.7 Conclusion and Future Directions

In this chapter, I have offered a potential story that answers the new question of *why* natural languages are compositional by enriching signaling games with the most rudimentary form of compositionality and exploring basic learning of such languages. The current answer is that compositional languages help simple agents learn to communicate more effectively in a complex world. This shows us, in a concrete and evolutionarily plausible scenario, one reason for which compositionality might have been selected. I also demonstrated that learning the meaning of a function word after knowing the meanings of atomic words is not difficult.

While these two strands present a promising start to answering the *why*-question, much more work remains to be done in the future. Firstly, there are areas of the

[20]Though it's worth noting that $\sigma(m_{AB})(s_{AB}) \leqslant 1/2$ so that m_{AB} never becomes strongly associated with s_{AB}.

parameter space of the current games that remain to be explored. For instance, in what ways are the simulation results robust against the form of the utility function? Is there an independently-motivated but wide class of such functions for which the results still hold? Secondly, one might wonder about other and more complex forms of compositionality. Can different languages be compared in order to find out whether different kinds of compositional language are better in different circumstances? Similarly, it is natural to continue extending the approach here to languages with multiple logical operators of varying arities (conjunction, for instance) and to start adding a more recursive syntax. Finally, considering such richer forms of compositional signaling might also require exploring more sophisticated learning algorithms than the very simple reinforcement learning considered here.

A particular approach that promises to yield insights consists in applying tools from machine learning to the problem, viewed from the following abstract perspective: the Sender has a piece of private information which – for whatever reason – it wants to transmit to the Receiver via communication. Allowing arbitrary sequences of basic signals to be sent and varying the kind of information – a possible worlds proposition, a credence function, e.g. – being transmitted will allow for a systematic survey of the emergence of different forms of structured language. A natural way to allow the Sender and Receiver to send arbitrary sequences of basic signals is to model them as simple kinds of *recurrent neural networks* (Hochreiter and Schmidhuber [1997]). Promising work applying these tools to the learning of communication protocols has started to appear (Foerster et al. [2016]). But the tools remain to be applied to the task of reconstructing a piece of information that the Sender wants to transmit. Even more generally, the framework could be applied to cases where what the Receiver needs to reconstruct from the Sender is not a piece of information, but some other object, say an image. In such a context, the question becomes: what kind of captions of an image best allow that image to be reconstructed? Because special problems arise in complex scenarios like this (what is a good similarity metric for images that extracts their high-level features?), this application will have to remain a promisory note. Such a scenario would take a fixed-dimensional representation,

produce a sequential representation, and try to reconstruct the original representation. As such, it represents a kind of 'inverse' of the famous sequence-to-sequence models (Sutskever et al. [2014]) that have gained widespread application in areas like machine translation. A thorough study of such 'fixed-to-fixed' models promises to be worthwhile.

Chapter 3

Pragmatic Expressivism and Non-Disjunctive Properties

This chapter asks the question of what role compositionality plays in the theory of communication and, in particular, the theory of assertion. Traditionally, it played a very direct role. According to the standard story,[1] the point of assertion is to transmit information. Similarly, the compositional semantic values of sentences have taken to be pieces of information. So to assert that p, one can utter in the appropriate way a sentence that means that p. However, observations that assertoric content does not compose in the same way that semantic values must has cast doubt on the direct role played by compositional semantics.

A second challenge to the standard story arises from domains of discourse that appear not to transmit information. Thus, expressivists have long been motivated by the idea that certain sentences do not state facts but serve some other function, such as expressing emotions or attitudes. Views of this type have been especially influential in metaethics because anti-realists about ethical properties must tell a non-standard story about how moral discourse works. But these views have very hard compositionality problems of their own: how do expressive sentences compose with each other, with factual sentences, and with logical operators? This question is

[1] See, among others, Stalnaker [1978].

a general form of the Frege-Geach Problem.[2]

The two challenges to the standard story, however, allow us to develop the relationship between compositionality and the theory of assertion in new and fruitful ways. In particular, the distinction between assertoric content and compositional semantic value opens the route to a new form of expressivism: *pragmatic* expressivism. On this family of views, expressivism is primarily a view not about the semantics of the sentences of the relevant domain of discourse but rather a pragmatic thesis about what we *do* with such sentences.

In this chapter, I contend that the pragmatic expressivist ought to be seen as replacing the notion of assertoric content in their theory of communication. Carefully identifying the concept that replaces assertoric content – a property of attitudes – and analyzing the similar role that it plays in communication allows us to formulate a principle, called the Non-Disjunctive Expressive Principle, that all pragmatic expressivists should adopt. This Principle states that semantic values of sentences in context should determine a *non-disjunctive* – in a sense to be made precise – property of attitudes. Against this backdrop, I will then argue that the most well-developed pragmatic expressivism – Seth Yalcin's expressivism about epistemic modals – violates this Principle in that it allows disjunctive properties of doxastic attitudes to be expressed. Such disjunctive properties prevent assertions from fulfilling the function that the expressivist claims they serve. Finally, I will present a new pragmatic expressivist theory which overcomes this difficulty, while simultaneously solving some other related linguistic problems. In this way, theorizing about the role that compositionality plays in communication leads to new constraints on the shape of compositional semantic theories in expressive domains.

This chapter is structured as follows. Section 3.1 presents and motivates the distinction between semantic value and assertoric content. Section 3.2 uses this distinction to develop pragmatic expressivism. Section 3.3 introduces the Non-Disjunctive Expressive Principle and argues that the pragmatic expressivist ought to accept it given what they identify as the function of assertion. Section 3.4 presents Yalcin's pragmatic expressivism about epistemic modals and shows how it runs afoul of the

[2]See Schroeder [2008].

Principle. Section 3.5 develops a new pragmatic expressivist theory based on assertability semantics and proves that this theory satisfies the Principle in that every sentence expresses a *non-disjunctive* – in a sense made precise in Section 3.3 – property of attitudes.

3.1 Semantic Value versus Content

To understand the distinction between semantic value and content, it will be helpful to recall a standard story about how communication works. Suppose Carlos and Meica are figuring out where to go to dinner and Meica says, "Mark's Deli is still open." On the standard story, Meica has a belief with a certain content and she wishes for Carlos to come to have the same belief (for the purposes of settling a currently live question, though that dimension will not presently concern us). To do so, she asserts a sentence that 'means', or 'expresses', that piece of content. Carlos, as a matter of his semantic competence, recognizes this fact and, if he accepts the assertion, comes to have a belief with that content. On the standard story, the sentence that Meica asserted means, or expresses, the content that it does as a matter of semantics. That is to say: the semantic value of a sentence (in context) is a content.

Recently, however, this last feature of the standard story about communication has come under fire.[3] The primary reason for drawing the distinction between semantic value and content is that doing so allows certain arguments against contextualism – understood as the thesis that a certain expression is context-sensitive – to be defused.[4] In particular, context-sensitivity has been thought to be hard to reconcile with the compositionality of linguistic meaning. The problems for contextualism arise when two sentences-in-context have the same content but embed differently. For a simple example, consider the pronoun 'she', which depends on context to determine its referent. In a context where the most salient woman is Susanne, assertions of the following sentences have the same content.

[3]See, for example, Ninan [2010], Rabern [2012], Yalcin [2014]. The line of reasoning goes back at least to Lewis [1980] and has a precursor in the distinction that Dummett [1973] makes between assertoric content and ingredient sense. See also Stanley [1997].

[4]See Rabern [2012] for some examples of this in action.

(10) Susanne is the strongest participant.

(11) She is the strongest participant.

But embedding (10) and (11) in the same linguistic environment can result in sentences with vastly different contents.

(12) Every participant thinks that Susanne is the strongest participant.

(13) Every participant thinks that she is the strongest participant.

In particular, the pronoun in (13) can be bound by 'Every participant', so that each participant thinks a different person – namely, herself – is the strongest. This difference in embedding behavior raises a problem because all parties agree that semantic values must be compositional: the semantic value of a complex expression is a function of the semantic values of its parts and its syntactic structure. The problem can be codified in the following argument.

(14) a) Semantic values are compositional: the semantic value of a complex expression (in context) is a function of the semantic values of its parts and its syntactic structure.

b) The semantic value of an expression in context is its assertoric content.

c) (10) and (11) have the same assertoric content.

d) So, (10) and (11) have the same semantic value.

e) (12) and (13) have the same syntactic structure, with parts that have the same semantic value.

f) By the compositionality principle (14a), (12) and (13) have the same semantic value.

g) But (12) and (13) do not have the same semantic value.

The argument results in a contradiction. The two critical assumptions are (14a) and (14b), so one of them must go. Because the identification of semantic value with assertoric content – premise (14b) – had often not been made explicit, such arguments were taken to be counter-examples to compositionality. But they seem not to be.

Instead, the proper reaction consists in rejecting (14b) and thus distinguishing between semantic value and assertoric content. We can observe one other feature of this argument to make a more general case for making this distinction. While (14e) is not true once (14b) is rejected, the following still holds: (12) and (13) have the same syntactic structure, with parts that have the same assertoric content. But the two sentences have different assertoric contents. This shows that content does not compose in the same way that semantic value does. This line of thought can be codified in what might be called Lewis' Master Argument against equating semantic value with content:[5]

(15) Lewis' Master Argument:

a) Semantic values are compositional.

b) Assertoric contents are not compositional.

c) So, semantic values are not assertoric contents.

Having distinguished the two notions, it does not follow that they are wholly unrelated. In particular, one still expects semantic theory to play a critical role in explaining linguistic communication, even if the standard story made that role too direct. Assertions can still transmit contents even if the semantic value of a sentence is not a content: it suffices that the semantic value of a sentence – whatever it may be – determines a content in context. Following Rabern [2012], it will be useful to isolate the principle underlying this revised story:

Determination Principle: The semantic value of an expression, together with the context, *determine* the assertoric content of that expression in context.

Following Kaplan [1989] and Lewis [1980], the standard way of implementing this goes as follows: semantic values are assigned relative to a context and an index, which will be a sequence of parameters that can be shifted by operators of the language in question. The content (in context) of an expression can be extracted from the semantic value by setting the index to be the index *of* the context, i.e. by setting all

[5] See p. 95 of Lewis [1980].

the parameters to the corresponding values of the context. For instance, to evaluate (11), we can take the index to just be a variable assignment g. We will then have that[6]

(16) $[\![\text{She}_i \text{ is the strongest participant.}]\!]^{c,g} = 1$ iff $g(i)$ is the strongest participant.[7]

Now, it is assumed that context determines an assignment function g_c; by the stipulations on salience, we assume that $g_c(i)$ is Susanne. Then the content of (11) will be that $g_c(i)$, aka Susanne, is the strongest participant. In this way, as long as semantic values (plus context) determine contents, the standard story need only be lightly revised.

3.2 Pragmatic Expressivism

Distinguishing between semantic value and assertoric content also opens the door to a new form of expressivism. To see how, consider a traditional statement of expressivism about a domain of discourse such as moral language: sentences in this domain do not state facts but rather express some kind of (non-doxastic) attitude. An early implementation would be the emotivism of Ayer [1936]. On this view, a typical utterance of

(17) Cheating is wrong.

does not state a fact (e.g. that the act-type of cheating has a certain property, namely being wrong) but rather expresses a negative emotional attitude towards cheating (e.g. "Boo, cheating!").

While the details of this view need not concern us presently, note that Ayer's emotivism and many subsequent brands of expressivism are *semantic* theses. On these views, the meaning of moral language (or whichever domain of discourse is

[6]Pronouns are assumed to be indexed in order to track co-reference relations.

[7]In a fully specified compositional semantics, the right-hand side would more precisely be generated as something like: $\text{strength}\,(g(i)) > \max\{\text{strength}(d) \mid d \in [\![\text{participant}]\!]^{c,g}\}$. But the present discussion only needs to highlight the role of the variable assignment in the interpretation of the pronoun, so I am neglecting compositional details.

currently being discussed) is of a different kind than standard, descriptive language: while sentences of the latter type express propositional contents and so state facts, those of the former kind do not. A common view has it that an expressivist must deliver a semantic theory which assigns attitudes, instead of contents or propositions, to sentences.[8] On such a view, sentences do not have a uniform type of meaning: the semantic values of sentences from the domain of discourse being discussed are *of a different kind* than those from the fact-stating domain.

The standard problem for semantic expressivism is the Frege-Geach problem. While the view as just sketched provides an account of basic moral statements, it does not yet explain the meanings of complex sentences like:

(18) If cheating is wrong, then hiring someone to cheat is wrong.

(19) Cheating is not wrong.

(20) Sam believes that cheating is wrong.

(21) If the system is corrupt, then cheating is not wrong.

In these sentences, (17) is embedded in a conditional, under negation, under an attitude verb, and in a conditional with a descriptive antecedent. All of these sentences are meaningful and, as importantly, they stand in logical relations to each other. While explaining the meanings of such complex sentences and their logical relations in the case of descriptive discourse is thought to be mostly unproblematic, the semantic expressivist incurs a special burden to offer new explanations since she takes moral language to have a different kind of meaning than descriptive language. At the most general level, then, the Frege-Geach problem can be stated: the expressivist needs to provide a compositional semantics for a language including the domain of discourse of interest. So doing will usually involve having to explain how terms from the domain of interest interact with those from the descriptive domain and with logical operators.[9]

[8]For example, Rosen [1998] writes that "The centerpiece of any quasi-realist 'account' is what I shall call a psychologistic semantics for the region: a mapping from statements in the area to the mental states they 'express' when uttered sincerely" (p. 387).

[9]See Schroeder [2008] for a useful summary of the problem and its history.

The distinction between semantic value and assertoric content opens the door to a new form of expressivism that I will call *pragmatic expressivism*, which has better hopes of solving the Frege-Geach problem. According to a pragmatic expressivist, the critical claim of expressivism is one about what we *do* when uttering moral sentences and not first and foremost about what such sentences mean. Understanding expressivism as a pragmatic thesis allows this new form of expressivist to assign the same type of semantic values to descriptive and moral sentences.[10] That is to say: the pragmatic expressivist can use tools from standard truth-conditional semantics to provide a compositional semantics for moral language and then appeal to these semantic values to explain inconsistencies, consequence relations, and other logical notions. Having a uniform type of semantic value for *all* sentences promises to dull the threat of the Frege-Geach problem. Once the move to a uniform type of semantic value has been made, the distinctions that semantic expressivists built into the semantic values of sentences will instead arise at the level of pragmatics, in how the semantic values that the pragmatic expressivist posits get used in communication.

As an example of a theory with this structure, consider the expressivism of Gibbard [1990]. Gibbard identifies a notion of *complete system of norms* that plays the same role for normative judgments that the notion of a possible world plays for factual judgments. Such a complete norm n makes a verdict for every course of action: forbids, requires, or permits but does not require it. So, the judgment that murder is wrong can be represented by a set of complete systems of norms that all forbid murdering. Using this notion, Gibbard can define semantic values as sets of pairs $\langle w, n\rangle$ consisting of a world w and a norm n and can treat negation as complement, conjunction as intersection, and so forth. He can then appeal to these semantic values to explain the logical relations among sentences, including those with normative terms. For instance, Gibbard can claim that "Murder is wrong" and "Murder is not wrong" are inconsistent because the semantic values of those two sentences have an empty intersection. Whether or not his theory proves ultimately satisfying does not presently concern us. Rather, the point to emphasize is that Gibbard provides a

[10] As Yalcin [2012] puts it (p. 140): "In particular, [pragmatic expressivism] is not a special kind of semantic theory."

uniform type of semantic value for factual, normative, and mixed sentences and uses these semantic values to explain the logical properties of sentences, thereby solving the Frege-Geach problem. A separate story about how these values are related to the attitudes expressed then needs to be told. Such a story will be pragmatic in nature.

While the pragmatic expressivist can make use of the distinction between semantic value and content in order to employ relatively standard methods from semantics, such an expressivist must tell a new story about the nature of communication. Such a story will not prominently feature content in the way that the standard story did. If not transmitting information by expressing propositional contents, then what is the purpose of communication? Here the pragmatic expressivist provides the paradigmatically expressivist line: an utterance in context expresses a property of one's attitudes, with the goal of engendering coordination on that property of attitudes amongst one's interlocutors. A selection of quotations from the literature shows that this conception is becoming standard:[11]

> "...in modeling the communicative impact of an epistemic possibility claim, we construe the objective as one of coordination on a certain global property of one's state of mind" [Yalcin, 2011, p. 310]
>
> "...an assertion of s is a suggestion that the conversational participants adopt some credal state in [the property expressed by s]" [Rothschild, 2012, p. 103]
>
> "**Coordination Pragmatics**: An utterance of φ by S to A is normally interpreted as a proposal by S that A psychologically approximate [the property of attitudes expressed by φ]" [Charlow, 2015, p. 34][12]
>
> "In asserting that φ, a speaker advises her addressees to conform their credences to the semantic value of 'φ'." [Swanson, 2016, p. 139]

[11] Many of these quotations come from papers exclusively about epistemic modals, which is the domain where pragmatic expressivism has taken the strongest hold. The view, however, can hold for any fragment of language. Yalcin [2012] and Charlow [2015] explore pragmatic expressivism for deontic language in particular.

[12] For reasons that need not concern us here – e.g. distinguishing assertions of deontic necessity modals from imperatives – Charlow enriches this conception of coordination.

> "The updated account: an assertion is like a proposal, not about a proposition that you should believe, but rather about a property that your credences should have." [Moss, 2015, p. 23][13]

3.3 The Non-Disjunctive Expressive Principle

The resulting picture of communication for the pragmatic expressivist replaces the notion of content with that of a property of attitudes and tells a new story about the role that this notion plays (e.g. that the goal is to engender coordination among the interlocutors on the possession of that property). Following the lead of those who distinguish between semantic value and content, the pragmatic expressivist need not and typically does not take semantic values to *be* such properties of attitudes. Nevertheless, such an expressivist must maintain a tight link between the two notions, along the lines of the Determination Principle discussed above. We can formulate such a link along expressivist lines as follows:[14]

Expressive Principle: The semantic value of a sentence, together with context, *determines* the property of attitudes expressed by that sentence in context.

In this section, I will argue that the pragmatic expressivist ought to adopt a more constrained form of the Expressive Principle. In particular, given the role that the property of attitudes expressed by a sentence plays in their theory of communication, they should require that the property of attitudes expressed be a *non-disjunctive* property. That is to say, they should embrace the following principle:

Non-Disjunctive Expressive Principle: The semantic value of a sentence, together with context, determines which *non-disjunctive* property of attitudes is expressed by that sentence in context.

[13] See also p. 5 of Moss [2013].

[14] The most explicit formulation of something like this principle comes from [Yalcin, 2012, p. 142]: "knowledge of the compositional semantic value of the sentence, together with any standing mutually known pragmatic norms and relevant facts of context, must be sufficient to determine this property". See also §5.3 of Charlow [2015] and the discussion of "bridging principles" in §IV of Rothschild [2012].

To argue for this stronger determination principle, I must first explain what it means for a property of attitudes to be non-disjunctive. After so doing, I will argue that disjunctive properties of attitudes are ill-suited to play the role in communication that pragmatic expressivists assign to them.

To get a handle on the notion of a non-disjunctive property of attitudes, it will help to introduce the notion of an attitudinal metalanguage. Such a language is motivated by the expressivist distinction between linguistic expressions that express propositional contents and those whose function is to express attitudes. In particular, an attitudinal metalanguage has expressions that denote *contents* and operators for the class of attitudes of interest. An example of such a metalanguage would be the language of propositional logic together with operators for belief and intention. In such a language, one can denote both contents and attitudes towards those contents.[15] It is in such a language that a pragmatic expressivist will want to identify the property of attitudes expressed by a sentence in their target object language. Crucially, an expressivist about an expression E will use an attitudinal metalanguage that does not contain E. We will call an expression in an attitudinal metalangauge an *attitudinal state description* since it describes a property of attitudinal states. But a more refined notion is needed. In an attitudinal metalanguage with attitudes $A_1, \ldots, A_n$, let an *attitudinal state type description* be a conjunction $\bigwedge_{i \in I} A_i p_i$ where each p_i is a content-denoting term of the metalanguage.[16] Such a description corresponds to a property of attitudinal states: the property of having all of the attitudes in the description. One can think of this property as the set of attitudinal states which are accurately described by the description. So, for example, $Bp \wedge Iq$ will be an attitudinal state type description that corresponds to the property of both believing that p and intending that q. A conjunction of attitudes like this is the notion that I will argue the pragmatic

[15] On the now-common Hintikkan [1962] conception of attitudes like belief, $B\varphi$ will also express a propositional content: the set of worlds w at which the agent's belief-worlds at w are all φ worlds. This will be useful in what follows, though the present discussion takes place at a level of abstraction where one need not commit to a particular conception of content or the Hintikkan analysis of attitudes.

[16] Here, I is some index set. I generally intend it to be finite, but certain cases involving quantification may warrant allowing infinite sets.

expressivist should use.[17]

Non-Disjunctive Property of Attitudes: A non-disjunctive property of attitudes is one which has an attitudinal state type description that corresponds to it.

To get a handle on the kinds of properties that are ruled out on this conception, consider the example attitudinal metalanguage from above. In that language, the expression $Bp \vee Bq$ would correspond to a disjunctive property of attitudes. There is no attitudinal state type that corresponds to the property of *either* believing p *or* believing q.[18]

A brief note before moving on is in order. Some expressivists – for example, those who take deontic language to express preferences[19] – may object that the properties of attitudes that they are interested in do not correspond to bearing attitudes towards contents. In response, one must note that the above story can be generalized by replacing 'contents' with 'objects of the relevant attitudes' throughout. In the case of preferences, for example, the object could be a pair of outcomes with the attitude then being strict preference for the first over the second outcome. In any event, non-disjunctive properties of attitudes – whatever you take the objects of attitudes to be – are those that can be expressed as conjunctions of attitudes.

My argument that pragmatic expressivists should endorse the Non-Disjunctive Expressive Principle will proceed in two steps. First, I will present an argument to the effect that endorsers of the standard story of assertion should maintain that assertions express a *determinate* content. Second, I will generalize this argument from the standard story to the pragmatic expressivist's story of assertion. This generalization will work because the argument fundamentally hinges on what type of thing is suited to carry out the characteristic function of assertion. Though expressivists disagree with the standard story on what this function is, the structure of the overall argument will still be analogous.

[17] I omit negations of attitudes from the conjunctions for the following reason: lacking an attitude toward a content can be re-conceived as its own attitude. So, for instance, an expressivist who wants to talk about intending and not intending can treat I and $\neg I$ as two attitudes in their metalanguage. I will do something similar in the epistemic case later.

[18] Note that, on this conception, the property of believing p-or-q will be determinate, described by the attitudinal state type description $B(p \vee q)$.

[19] See Silk [2015], Starr [2016a].

To deliver the first argument, I should first elaborate what I've been calling the standard story of assertion, largely due to Stalnaker [1978]. The standard story makes precise the intuitive idea that the *point* of assertions is to convey information. On this picture, conversation takes place against a *common ground*: a set of propositions that all discourse participants take for granted for the purposes of the conversation.[20] Taking propositions to be sets of possible worlds, the common ground determines a set of worlds: those worlds compatible with all of the propositions in the common ground. Call this set of worlds the *context set*. Against this backdrop, an assertion that p is taken to be a proposal to add p to the common ground and thereby shrink the context set. If all of the interlocutors accept the assertion so that p does get added, then the context set will be shrunk so that it only contains p worlds. In this sense, the assertion has conveyed the information that p.

With the standard story fleshed out in this way, we can look at an argument that to be an assertion, an utterance must express a determinate propositional content. Precursors of this argument can be found in Stalnaker [1978] when he defends his second principle governing assertions[21] and in §II of Glanzberg [2004].[22] Here is the argument.

(22) The Argument that Standard Assertions Require a Determinate Content:

a) The function of an assertion is to convey information.

i. Via an utterance expressing a proposition.

ii. And updating the context set by adding that proposition to the common ground.

b) If an assertion did not have a determinate content, it would be undetermined how to update the context set.

[20] In a more recent and refined version, a proposition is common ground if it is commonly believed to be accepted. See Stalnaker [2002] for motivations and details. This refinement does not concern the structure of the present argument.

[21] On page 325: "Any assertive utterance should express a proposition, relative to each possible world in the context set, and that proposition should have a truth value in each possible world in the context set."

[22] Note also that Stanley [2000] defends the more general claim that "... when a communicative act lacks a determinate content, it is not a linguistic speech act" (p. 408).

c) So, an assertion requires a determinate content to fulfill its function.

Some words of clarification are in order. First, consider the notion of a determinate content and its role in the premise (22b). While a determinate content generally means a complete possible worlds proposition,[23] the premise requires that the content determine, for every world in the context set, whether or not to include that world in the updated set. The thought is that an assertion makes a proposal to update the context set in a certain way. For that way to be recoverable by the audience, the content of the assertion must issue a verdict on every world in the context set. If it did not, the proposal would not be able to fulfill its function. Second, I intend the sense of requirement in the conclusion of the argument to be constitutive. That is: utterances which fail to express a determinate proposition are not faulty or defective assertions but rather fail to be assertions at all. Whether or not one endorses a stronger constitutive norm – knowledge[24] or reasonable belief,[25] for instance – of assertion, there are good reasons to accept this weak one.[26] The thought is that assertions are actions designed to carry out a certain function and so to be an assertion is to be capable of carrying out that function.

The argument for the Non-Disjunctive Expressive Principle will parallel the argument in (22). The general structure of that argument consists in identifying a function or purpose of assertion and then identifying what assertions must be like in order to carry out that function. While the pragmatic expressivist identifies a different function for assertion, that general structure can still be leveraged to make an analogous argument for the Non-Disjunctive Expressive Principle. The new argument runs as follows.

(23) The Argument that Expressive Assertions Require a Non-Disjunctive Property of Attitudes

a) The function of an assertion is to coordinate attitudes.

[23]Here, a complete possible worlds proposition is one that assigns a truth value to *every* possible world, leaving no 'gaps'.

[24]See Williamson [1996].

[25]See Lackey [2007].

[26]See Maitra [2011], which argues that stronger norms may capture what an asserter ought to be aiming at, but not what it is to make an assertion.

i. Via an utterance expressing a property of attitudes.

ii. And the interlocutors updating their attitudes to acquire that property.

b) If an assertion does not have a non-disjunctive property of attitudes, it would be undetermined how to update one's attitudes.

c) So, an assertion requires a non-disjunctive property of attitudes to fulfill its function.

While the argument in (23) does not conclude in the Non-Disjunctive Expressive Principle, we can get there quickly. All parties to this debate agree that semantic value plus context must determine the objects that their theory of assertion requires. So, if that object is a non-disjunctive property of attitudes – as I have just argued it should be for a pragmatic expressivist – then semantic value plus context must determine which such property a sentence expresses. But that just is the Principle.

The premise (23b) requires more justification than its counterpart (22b). In particular, its justification will not mirror that for (22b): it is not necessarily the case that there is a space of points corresponding to which attitudes to adopt and that a disjunctive property fails to tell the listener which points to rule out. To justify (23b), consider a prototypical disjunctive property of attitudes: say the property of either believing p or believing $\neg p$. If an utterer expresses that their attitudes have this property, what exactly are they asking their interlocutors to do? Answering this question proves difficult: perhaps the request is to flip a coin and choose on that basis which of the beliefs to have or perhaps the request is itself indeterminate. On either interpretation, however, such a request will fail to carry out the function of coordinating attitudes. The utterer could believe p and express the disjunctive property above,[27] while the listener 'coordinates' on the disjunctive property by coming to believe $\neg p$. Such an exchange appears to be a paradigm case of *anti*-coordination, at least on the question of whether p.

[27]There may be other reasons to criticize such a speaker, but I am supposing for the sake of argument that such an expression would still be an assertion.

It might help some readers to note a parallel between this argument and an argument against the existence of non-disjunctive properties in the metaphysics literature. Consider the property of being-red-or-hexagonal. A red cube and a black hexagon will both have this property, but appear to have nothing in common with respect to that property. Armstrong [1978] objects to there being such a property at all on these grounds: "disjunctive properties offend against the principle that a genuine property is identical in its different particulars" (p. 20). Similarly, Audi [2013] argues that genuine properties must guarantee 'similarity-in-a-respect' among their instances and that disjunctive properties fail to do so. While these authors take the possible lack of similarity among instances of a disjunctive property to tell against the existence of such properties, my argument takes the possible lack of similarity among agents' doxastic states sharing a disjunctive property of attitudes to tell against such properties playing a role in the theory of communication.

The general point can be put like this: a satisfying theory of assertion should provide a story about the dynamics of conversation that issues in deterministic updates. On the standard story, this requires determinate contents to be expressed, so that the assertion issues a deterministic update to the context set when accepted. On the pragmatic expressivist story, this requires non-disjunctive properties to be expressed, so that the assertion issues a deterministic update to the interlocutors' attitudes when accepted. In the terminology of Charlow [2015], assertions issue *cognitive directives* about how to update one's attitudes. But a disjunctive property of attitudes cannot constitute a genuine directive.[28]

3.4 Disjunctive Expression for Epistemic Modals

In this section, I want to look at the most well-developed version of pragmatic expressivism: the expressivism about epistemic modals due to Yalcin [2007, 2011, 2012]. In particular, while he does endorse the Expressive Principle, it will be shown that his

[28] Stalnaker [1978] makes a similar point in the case of the standard story when saying that an utterance that expresses different propositions in different worlds in the context set does not issue a determinate update to the context set because it "expresses an intention that is essentially ambiguous" (p. 327).

theory systematically violates the Non-Disjunctive Expressive Principle: disjunctions with wide-scope epistemic modal disjuncts will express disjunctive properties of attitudes. For present purposes, I will focus on a fragment containing logical connectives for negation, conjunction, and disjunction as well as the epistemic modal auxiliary 'might', symbolized by $\Diamond$. The problem that arises for this fragment will also arise for larger fragments containing probability operators such as 'likely' and 'probably'.

3.4.1 Yalcin's Pragmatic Expressivism

Yalcin's expressivism stems mostly from linguistic problems arising from the standard contextualist semantics due to Kratzer [1981, 1991].[29] Consider an utterance by Marco of:

(24) The keys might be on the table.

On the standard story of assertion, one must identify a content expressed. The typical answer will be that what is expressed is the proposition that it is compatible with some contextually-salient body of information – for instance, Marco's knowledge – that the keys are on the table. Numerous problems arise from an account like this. At an intuitive level, it gets the dynamics of an assertion of (24) wrong: a listener does not learn a fact about Marco's state of knowledge, but rather receives a suggestion to add the possibility that the keys are on the table to the range of possibilities that they consider to be open. Second, as Yalcin [2007] first emphasized, contextualism has a hard time accounting for the nearly uniform unacceptability of embedded epistemic contradictions[30,31] like

(25) # Marco believes the keys are not on the table and they might be.[32]

[29] Expressivism, of course, is not the only response to these kind of problems. They have also been used to argue for *relativism* by MacFarlane [2011, 2014]. My purpose here, however, is not to adjudicate the contextualism-expressivism-relativism debate but to raise a worry for the pragmatic expressivist. Dowell [2011] and von Fintel and Gillies [2011] present defenses of contextualism from the worries to be sketched below.

[30] Note that Yalcin's solution, presented below, draws heavy inspiration from the dynamic semantics tradition of Veltman [1996], Groenendijk et al. [1996]. See Muskens et al. [1997] for a survey.

[31] See Moss [2015] for some allegedly acceptable examples of embedded epistemic contradictions.

[32] A reader might think this sentence sounds acceptable. But that's likely only on a reading where

Third, the contextualist has a hard time explaining disagreement over claims like (24). Suppose Marco's interlocutor responds, "No, I already checked and they're not on the table." The problem is that if Marco has simply asserted that it is compatible with his knowledge that the keys are on the table, the interlocutor is not in a position to disagree with that fact. But if Marco's assertion did express a proposition with which the interlocutor could disagree – for instance, that it is compatible with the group's knowledge that the keys are on the table – then he would not be warranted in asserting it. In other words, it appears difficult to identify a single content that can be plugged in to the standard story of assertion that can account for assertability and disagreement facts.

To provide a more satisfying explanation of these features of the behavior of epistemic modals, Yalcin develops a pragmatic expressivism about this fragment of discourse. To develop this, he must assign compositional semantic values to the fragment and then explain what property of attitudes is expressed by a sentence in context. I will explain each of these dimensions of his theory in turn. For the compositional semantics, Yalcin takes as indices pairs of a world w and an information state $\mathbf{s}$, which is a set of worlds. The information state provides the domain of quantification for epistemic modals and, crucially, is not determined by the world parameter and something like an accessibility relation. The semantics is then relatively straightforward. Atoms are assigned truth-values relative to worlds as usual. The compositional clauses are:

(26) Yalcin's compositional semantics:

a) $[\![\neg\varphi]\!]^{w,\mathbf{s}} = 1$ iff $[\![\varphi]\!]^{w,\mathbf{s}} = 0$

b) $[\![\varphi \wedge \psi]\!]^{w,\mathbf{s}} = 1$ iff $[\![\varphi]\!]^{w,\mathbf{s}} = 1$ and $[\![\psi]\!]^{w,\mathbf{s}} = 1$

c) $[\![\varphi \vee \psi]\!]^{w,\mathbf{s}} = 1$ iff $[\![\varphi]\!]^{w,\mathbf{s}} = 1$ or $[\![\psi]\!]^{w,\mathbf{s}} = 1$

d) $[\![\Diamond\varphi]\!]^{w,\mathbf{s}} = 1$ iff $\exists w' \in \mathbf{s} : [\![\varphi]\!]^{w',\mathbf{s}} = 1$

To explain how Yalcin determines a property of attitudes from these semantic values, one needs to use his notion of acceptance: an information state $\mathbf{s}$ *accepts* φ

the conjunction takes scope over the belief operator. The present focus is on the reading where the belief operator scopes over the conjunction. Thanks to Kevin Dorst for discussion.

if and only if $[\![\varphi]\!]^{w,\mathbf{s}} = 1$ for every world $w \in \mathbf{s}$.[33] Now, the property determined can be extracted from his semantics for attitude reports. In particular, as is standard, Yalcin assumes that an agent A in a world w has a set of 'belief-worlds' $\mathbf{b}_{A,w}$: these are the set of worlds compatible with what the agent believes in w.[34] The semantics for belief reports then runs as follows:

(27) $[\![B_A\varphi]\!]^{w,\mathbf{s}} = 1$ iff $\mathbf{b}_{A,w}$ accepts φ

Note that the use of acceptance has two effects: it shifts the information state parameter to $\mathbf{b}_{A,w}$ and quantifies over the worlds therein. This semantics for attributions also provides the template for identifying the property of attitudes expressed by an assertion:

(28) Yalcin's Expressive Principle:

When an agent A utters a sentence φ, they express that $\mathbf{b}_{A,w}$ accepts φ.

This determination principle works quite nicely for pure factual and pure epistemic sentences. A sentence φ containing no epistemic vocabulary will not depend on $\mathbf{s}$ and so can be associated with a standard possible-worlds proposition. Then (28) just says that an assertion of φ expresses that all of the agent's belief-worlds are φ-worlds. A sentence $\Diamond\varphi$, where φ has no epistemic vocabulary, will express that φ is *compatible* with the agent's beliefs, i.e. that there is a φ-world among the agent's belief-worlds. Let us call this attitude *abelief*: an agent abelieves that p iff p is compatible with their beliefs, i.e. if they do not believe not-p.[35] In terms of the earlier discussion of attitudinal metalanguages, I can say that Yalcin will use such a metalanguage containing the language of propositional logic and two attitude operators: B for belief and A for abelief. Then an assertion of a purely factual sentence φ expresses the non-disjunctive property $B\varphi$ and an assertion of a basic epistemic sentence $\Diamond\varphi$ expresses the non-disjunctive property $A\varphi$.

[33] This becomes very close to the notion of 'support' in dynamic semantics à la Veltman [1996]. Yalcin's logic – defined as preservation of acceptance – thus resembles the 'test consequence' studied in van der Does et al. [1997]. See also van Benthem [1989a]. Schulz [2010] results the relating system in a very natural way to S5.

[34] See van Benthem [2011] for exposition of this model and various elaborations thereof.

[35] Note that this attitude is not to be confused with the very important notion of alief introduced by Gendler [2008].

3.4.2 Disjunctive Expression in Yalcin

While this story works exactly as a pragmatic expressivist would want for these basic cases, it violates the Non-Disjunctive Expressive Principle when more complicated sentences are considered. As an example, consider the following sentence, asserted in March 2016:

(29) Bernie Sanders might or might not win the Democratic nomination.

What property of their doxastic state did the asserter express? Intuitively, uncertainty: the speaker neither believes that he will nor that he will not win. Note that such a property is *non-disjunctive*: the speaker both abelieves that Bernie will win the nomination and abelieves that he will not win the nomination. The case that both attitudes are expressed can be bolstered by noting that an interlocutor can disagree in at least two ways:

(30) a) No, he has no chance of getting enough votes.

b) No, he will definitely win.

While I will not give a full analysis of the pragmatics of disagreement here, the rough idea would be: the assertion of (29) expressed both abeliefs and invited the interlocutors to update their doxastic states to also acquire those abeliefs. Someone who disagrees via (30a) firmly believes that Bernie has no chance and thus rejects the invitation to abelieve that he will win, i.e. refuses to make it compatible with their beliefs that Bernie will win. And *mutatis mutandis* for (30b). That a listener could disagree in either way suggests that both abeliefs were expressed by (29).

Unfortunately, Yalcin's Expressive Principle cannot deliver this result:

(31) According to Yalcin's Expressive Principle (28), an assertion of a sentence of the form $\Diamond p \vee \Diamond \neg p$ expresses that the agent *either* abelieves that p *or* abelieves that $\neg p$.[36,37]

[36]Proof: $\mathbf{b}_{A,w}$ accepts $\Diamond p \vee \Diamond \neg p$ iff for every $w' \in \mathbf{b}_{A,w}$: $[\![\Diamond p]\!]^{w',\mathbf{b}_{A,w}} = 1$ or $[\![\Diamond \neg p]\!]^{w',\mathbf{b}_{A,w}} = 1$. This occurs iff for every $w' \in \mathbf{b}_{A,w}$: for some $w'' \in \mathbf{b}_{A,w}$, $[\![p]\!]^{w'',\mathbf{b}_{A,w}} = 1$ or for some $w'' \in \mathbf{b}_{A,w}$, $[\![p]\!]^{w'',\mathbf{b}_{A,w}} = 0$. Now, there are two cases: if $\mathbf{b}_{A,w}$ is empty, the outer quantification is vacuous. If it is not empty, then a world will have to make either p or $\neg p$ true, so the whole disjunction will be true.

[37]In fact, the above proof shows that *every* information state accepts $\Diamond p \vee \Diamond \neg p$ for Yalcin.

In particular, the property expressed will be a disjunctive property of attitudes. In addition to being at odds with the intuitive analysis of (29), this represents a violation of the Non-Disjunctive Expressive Principle.

Moreover, the problem above has a deeper source. Consider the following sentence:

(32) The keys might be on the table or they might be in my car.

A number of theorists[38] have argued that this sentence seems to be equivalent in ordinary conversation to

(33) The keys might be on the table *and* they might be in my car.

A sense of this equivalence can be gleaned by using the same diagnostic as before. In particular, both of the following are felicitous and entirely normal continuations to an assertion of (32).

(34) a) No, they aren't on the table. (I already looked there. ...)

b) No, they aren't in the car. (I already looked there. ...)

This example shows that disjunctions of epistemic 'might' claims typically express a non-disjunctive property of attitudes: namely the property of abelieving the prejacent of both disjuncts. Unfortunately, Yalcin's Expressive Principle also yields a disjunctive property of attitudes for these cases.

(35) According to Yalcin's Expressive Principle (28), an assertion of a sentence of the form $\Diamond p \vee \Diamond q$ expresses that the agent *either* abelieves that p *or* abelieves that q.[39]

Thus, such epistemic disjunctions provide another source of violations of the Non-Disjunctive Expressive Principle on Yalcin's pragmatic expressivism.

[38] See, for instance, Zimmermann [2000], Simons [2005], Aloni [2007], Ciardelli et al. [2009], Roelofsen [2013].

[39] Proof: In general, $\mathbf{b}_{A,w}$ accepts $\Diamond p \vee \Diamond q$ iff it accepts $\Diamond p$ or it accepts $\Diamond q$. The right-to-left direction is easy. For the left-to-right direction: suppose that $\mathbf{b}_{A,w}$ accepts $\Diamond p \vee \Diamond q$. Unpacking as in footnote 36, we have that for every $w' \in \mathbf{b}_{A,w}$: for some $w'' \in \mathbf{b}_{A,w}$, $[\![p]\!]^{w'',\mathbf{b}_{A,w}} = 1$ or for some $w'' \in \mathbf{b}_{A,w}$, $[\![q]\!]^{w'',\mathbf{b}_{A,w}} = 1$. The outer quantifier is vacuous. If the left disjunct holds, then $\mathbf{b}_{A,w}$ accepts $\Diamond p$; if the right holds, then $\mathbf{b}_{A,w}$ accepts $\Diamond q$.

3.4.3 Objection: Going Gricean

Before moving along, I want to consider one possible rejoinder to this problem for Yalcin's pragmatic expressivism. In particular, it might be thought that broadly Gricean considerations can be used to generate the felt expression of the non-disjunctive properties of attitudes in examples (29) and (32). Such an explanation would proceed by assuming that a speaker is being cooperative and inferring what their doxastic state must be like in order to have made the utterance that they did. To see how this would go, first note that all of the examples are disjunctions. Moreover, consider factual disjunctions like

(36) The keys are either on the table or in the car.

All parties agree that this expresses a non-disjunctive property of attitudes: $B(p \vee q)$. Even so, an assertion of (36) can fail to engender coordination: the asserter could believe that the keys are on the table and assert (36), but the hearer could come to believe that they are in the car. These two facets suggest a strategy: resolve the problems about disjunctive expression and the resulting threat of anti-coordination for Yalcin using Gricean mechanisms that will be needed to solve anti-coordination problems even for factual disjunctions.

An implementation of this strategy would go as follows. Standard Gricean reasoning can be used to justify the following generalization:[40]

Disjunction Implicature: An assertion of $\varphi \vee \psi$ implicates that the speaker does not believe $\neg\varphi$ and does not believe $\neg\psi$.

In terms of the current brand of expressivism: someone who asserts $\varphi \vee \psi$ would also be prepared to assert $\Diamond\varphi \wedge \Diamond\varphi$. Why? If they believed $\neg\psi$, $\varphi \vee \psi$ would then be equivalent to φ, which would be briefer. And *mutatis mutandis* for $\neg\psi$. This resolves the threat of anti-coordination for factual disjunctions: a speaker who believes $\neg p \wedge q$ should not assert $p \vee q$; and a listener can rely on the above reasoning to infer that she should not update to come to believe $p \wedge \neg q$.

[40] See, for instance, (41) on p. 50 and pp. 59-61 of Gazdar [1979] as well as §4 of Aloni [2016].

Now, apply this reasoning to the epistemic disjunction case. An assertion of $\Diamond p \vee \Diamond q$ will typically communicate $(\Diamond p \vee \Diamond q) \wedge \Diamond\Diamond p \wedge \Diamond\Diamond q$. On Yalcin's semantics, this is equivalent to $\Diamond p \wedge \Diamond q$.[41] So the felt equivalence between an epistemic disjunction and its corresponding conjunction can be explained pragmatically. At the very least, an assertion of such a disjunction can be seen to express a non-disjunctive property of attitudes by relying on this Gricean reasoning.

In response to this objection, I will point out that the reasoning that justifies **Disjunction Implicature** in the factual case does not straightforwardly generalize to the epistemic case. This shows that a simple Gricean story cannot be used to rescue Yalcin. While I cannot rule out all possible pragmatic explanations, this shifts the burden back to the defender of Yalcin's Expressive Principle to offer a sophisticated pragmatic explanation of how such cases of indeterminate expression are to be avoided.

Consider the case of an assertion by a speaker S of $p \vee \neg p$. A hearer H can reason as follows. Exactly one of the following three exclusive and exhaustive attitudinal state type descriptions can accurately describe S's doxastic state: (i) Bp, (ii) $B\neg p$, (iii) $Ap \wedge A\neg p$. The hearer then reasons as follows: S cannot be in state (i) for they would have asserted p instead of the disjunction. And *mutatis mutandis* for state (ii) and $\neg p$. Therefore, S must be in state (iii).

This reasoning, however, does not generalize to an assertion of $\Diamond p \vee \Diamond \neg p$. Again, the speaker must be in exactly one of the states (i)-(iii). But in *each* of those states, the speaker could have said something more informative and either more or equally brief. In (i): p, in (ii): $\neg p$, and in (iii): $\Diamond p \wedge \Diamond \neg p$. From this perspective, then, the hearer will have a hard time making sense of why a speaker would ever have chosen to assert the epistemic disjunction. The conclusion would appear to be that such a speaker must not be rational and cooperative after all.

[41] Because iterated epistemic 'might's are redundant and disjunction introduction is valid.

3.5 A New Pragmatic Expressivism for Epistemic Modals

To recap: first, I argued that pragmatic expressivists, who replace the notion of content with that of a property of attitudes in their theory of assertion, should accept a strong bridge principle from semantic values to properties: sentences in context should express only *non-disjunctive* properties of attitudes. This is because disjunctive properties of attitudes will fail to fulfill the function of assertion identified by the pragmatic expressivist. I then looked at the most well-developed pragmatic expressivist theory – Yalcin's expressivism about epistemic modals – and found that it systematically allows for disjunctive properties to be expressed. The expression of such properties arises from the interaction of disjunction and epistemic modals. In this section, I will introduce a new pragmatic expressivist theory which overcomes these problems: I will prove that every sentence expresses a non-disjunctive property of attitudes.

3.5.1 Assertability Semantics

The semantics that I will present has three main innovations. First, sentences will be assigned truth-values relative to *only* an information state, not a pair of a world and an information state.[42] Note that Yalcin's concept of acceptance, in terms of which his Expressive Principle is couched, holds between an information state and a sentence. But it is defined derivatively by quantifying over the world parameter in his indices. One can look at the semantic values in this new theory, then, as capturing acceptance directly instead of indirectly. Second, a non-standard semantics for disjunction will be introduced.[43] Nevertheless, I will show that the new semantics in a sense reduces to classical logic for the fragment without epistemic modals. Finally, the semantics will be *bilateral* in the sense of simultaneously defining two relations –

[42]The move to evaluating at sets of worlds resembles the move Hodges [1997] made – discussed in Section 1.1.2 – to handle branching quantification by moving to sets of assignment functions.

[43]Precursors, in different theoretical contexts, for this disjunction can be found in Väänänen [2008], Ciardelli and Roelofsen [2015]. For further discussion of these two frameworks, see footnote 49.

assertability and deniability – between information states and sentences. This helps capture the fundamentally trivalent notion of the semantics: some sentences will be neither assertable nor deniable at some information states (e.g. an atomic proposition p at an information state with both p and not-p worlds). The primary motivation for making this move will become clear in Section 3.5.3.[44]

In defining semantic values relative to information states, one should read $\mathbf{s} \Vdash \varphi$ as "φ is assertable relative to the information $\mathbf{s}$" and $\mathbf{s} \dashv\Vdash \varphi$ as "φ is deniable relative to the information $\mathbf{s}$".[45] We will assume that sets of worlds are assigned to atomic letters via a valuation function V. The whole semantics can then be given by:

(37) φ is assertable/deniable relative to information state $\mathbf{s}$:

a) $\mathbf{s} \Vdash p$ iff $\mathbf{s} \subseteq V(p)$
$\mathbf{s} \dashv\Vdash p$ iff $\mathbf{s} \cap V(p) = \varnothing$

b) $\mathbf{s} \Vdash \neg\varphi$ iff $\mathbf{s} \dashv\Vdash \varphi$
$\mathbf{s} \dashv\Vdash \neg\varphi$ iff $\mathbf{s} \Vdash \varphi$

c) $\mathbf{s} \Vdash \varphi \wedge \psi$ iff $\mathbf{s} \Vdash \varphi$ and $\mathbf{s} \Vdash \psi$
$\mathbf{s} \dashv\Vdash \varphi \wedge \psi$ iff $\mathbf{s}_1 \dashv\Vdash \varphi$ and $\mathbf{s}_2 \dashv\Vdash \psi$ for some $\mathbf{s}_1, \mathbf{s}_2$ such that $\mathbf{s} = \mathbf{s}_1 \cup \mathbf{s}_2$

d) $\mathbf{s} \Vdash \varphi \vee \psi$ iff $\mathbf{s}_1 \Vdash \varphi$ and $\mathbf{s}_2 \Vdash \psi$ for some $\mathbf{s}_1, \mathbf{s}_2$ such that $\mathbf{s} = \mathbf{s}_1 \cup \mathbf{s}_2$
$\mathbf{s} \dashv\Vdash \varphi \vee \psi$ iff $\mathbf{s} \dashv\Vdash \varphi$ and $\mathbf{s} \dashv\Vdash \psi$

e) $\mathbf{s} \Vdash \Diamond\varphi$ iff $\{w\} \Vdash \varphi$ for some $w \in \mathbf{s}$
$\mathbf{s} \dashv\Vdash \Diamond\varphi$ iff $\mathbf{s} \dashv\Vdash \varphi$

f) $\mathbf{s} \Vdash B_A\varphi$ iff $\mathbf{b}_{A,w} \Vdash \varphi$ for every $w \in \mathbf{s}$
$\mathbf{s} \dashv\Vdash B_A\varphi$ iff $\mathbf{b}_{A,w} \Vdash \Diamond\neg\varphi$ for every $w \in \mathbf{s}$

That φ is assertable relative to $\mathbf{s}$ encodes the following fact: if an agent has $\mathbf{s}$ as their belief-worlds, then it is epistemically appropriate for them to assert φ. That φ is deniable relative to $\mathbf{s}$ encodes the following fact: if an agent has $\mathbf{s}$ as their belief-worlds, then it is epistemically appropriate for them to deny φ. One can motivate

[44] For earlier bilateral systems, see Veltman [1985], Groenendijk and Roelofsen [2010], Fine [2014].
[45] If you prefer, you can also read $\mathbf{s} \dashv\Vdash \varphi$ as "φ is rejectable given the information $\mathbf{s}$".

the clauses, then, by reflecting on inquiry. The information state represents the worlds that have not been ruled out as candidates for the actual world. A sentence is assertable relative to an information state if it is guaranteed to hold come what may in the rest of inquiry. A sentence is deniable relative to an information state if it is guaranteed to not hold come what may in the rest of inquiry. So, for instance, an atom is assertable if it is true *throughout* the information state and deniable if it is false throughout the information state. Now, consider the clause for disjunction: a disjunction $\varphi \vee \psi$ will be assertable relative to an information state if that information state is *covered* by a φ zone and a ψ zone. So the information can guarantee that φ or ψ holds even though it might not yet guarantee that either one of them holds. For a disjunction to be deniable, however, one must be able to deny each disjunct. Because of this epistemic dimension of these semantic clauses, one can also read assertability as "$\mathbf{s}$ supports φ"[46] or "φ must be accepted given $\mathbf{s}$".[47]

Before defining a new expressive principle and showing that no disjunctive properties are thereby expressed, I note a few facts about this semantics. First, I can recover a notion of propositional content by considering assertability at singletons. Think of this via the following slogan: what is true is what would be assertable in the absence of uncertainty, i.e. at the end of inquiry.[48] To make this precise, define the *truth set of* φ as follows:

(38) $[\varphi] = \{w : \{w\} \Vdash \varphi\}$

One can prove that, under the semantics in (37), the truth sets behave classically for the fragment without epistemic modals.

(39) Classicality. If φ and ψ do not contain $\Diamond$, then:

[46] Groenendijk et al. [1996], Portner [2009].

[47] Veltman [1996], Yalcin [2007]. There is a sense, elucidated in van Benthem [1989a], in which Yalcin's semantics is a "static" version of Veltman's update semantics. For some differences between a closely related logic of assertability and Veltman's, consult Hawke and Steinert-Threlkeld [2016a].

[48] There are hints of a Peircean account of truth here: "The opinion which is fated to be ultimately agreed to by all who investigate, is what we mean by the truth, ..." (Peirce [1878], p. 273). See Misak [2007] for thorough exposition and elaboration of a Peircean conception of truth in terms of the end of inquiry. While the present connection between truth and assertability echoes this pragmatist thought, my present concern is with the theoretical concept of truth-at-a-world as used in semantics. The connection between this concept and the ordinary concept of truth lies beyond the present scope. Thanks to Matthias Jenny here.

a) $[p] = V(p)$

b) $[\neg\varphi] = W \backslash [\varphi]$

c) $[\varphi \wedge \psi] = [\varphi] \cap [\psi]$

d) $[\varphi \vee \psi] = [\varphi] \cup [\psi]$

e) $[B_A\varphi] = \{w : \mathbf{b}_{A,w} \subseteq [\varphi]\}$

We can thus think of the fact that $w \in [\varphi]$ as the fact that φ is true at world w.

Second, while the main intuition for assertability is that φ is assertable at $\mathbf{s}$ iff it is true throughout $\mathbf{s}$, the clause for epistemic 'might' undercuts this thought. The modal reflects a global property of information states, one that does not distribute to properties of the individual worlds therein.[49] Luckily, however, it turns out that 'might' is the sole source of such failures:

(40) Let φ be a $\Diamond$-free sentence. Then:

$\mathbf{s} \Vdash \varphi$ iff $\{w\} \Vdash \varphi$ for every $w \in \mathbf{s}$

$\mathbf{s} \dashv\Vdash \varphi$ iff $\{w\} \dashv\Vdash \varphi$ for every $w \in \mathbf{s}$

These equivalences also allow us to explain the felt infelicity of epistemic contradictions – sentences of the form $\varphi \wedge \Diamond\neg\varphi$ and $\neg\varphi \wedge \Diamond\varphi$ – by observing that they are never assertable when φ is $\Diamond$-free.[50]

[49] More precisely, sentences $\Diamond p$ are not *persistent* (i.e. preserved under sub-states of information states): there are $\mathbf{s}$ and $\mathbf{t} \subsetneq \mathbf{s}$ such that $\mathbf{s} \Vdash \Diamond p$ but $\mathbf{t} \nVdash \Diamond p$. For an example, let $\mathbf{s}$ contain one p world and one $\neg p$ world, with $\mathbf{t}$ being the singleton containing the $\neg p$ world. Hawke and Steinert-Threlkeld [2016b] suggest that persistence demarcates fact-stating from non-factual domains of discourse. Both of the systems of modal dependence logic of Väänänen [2008] and inquisitive dynamic epistemic logic of Ciardelli and Roelofsen [2015] evaluate sentences at information states and use a clause like ours for disjunction. But in both systems, truth at an information state (which Ciardelli and Roelofsen [2015] call 'support') is in fact persistent. This shows that our 'might' cannot be defined in their systems. Extensions of those frameworks and deeper exploration of the connections between all of them are interesting pursuits for future work.

[50] All of the examples from the literature are cases where φ contains no 'might's. Note, however, that on our semantics, $\Diamond p \wedge \Diamond\neg\Diamond p$ will be assertable at any $\mathbf{s}$ which has both a p and a $\neg p$ world. Such an information state will not accept, in Yalcin's sense, this complex epistemic contradiction. Because judgments about complex sentences with embedded epistemic modals are not firm, it is unclear whether this is a problem for the present view.

Finally, I note two facts that are critical in proving that this theory satisfies the Non-Disjunctive Expressive Principle and that are of independent interest. To do so, I must introduce a notion of entailment and equivalence. Say that a set Γ assertorically entails ψ just in case if $\mathbf{s} \Vdash \gamma$ for every $\gamma \in \Gamma$, then $\mathbf{s} \Vdash \psi$ for every information state $\mathbf{s}$. Say that φ and ψ are assertorically equivalent if (the singleton of) each sentence entails the other one. The following then hold:

(41) $B_A\Diamond\varphi$ is assertorically equivalent to $\neg B_A\neg\varphi$[51]

(42) $\Diamond\varphi \vee \Diamond\psi$ is assertorically equivalent to $\Diamond\varphi \wedge \Diamond\psi$[52]

This fact also contains the core of the explanation of the sentences (29) and (32) which we saw were problematic in the last section.

3.5.2 Non-Disjunctive Expression

We can now move towards the proof that every assertion expresses a non-disjunctive property of attitudes by providing a bridge principle from the new semantic values to such properties. Recall that we are employing an attitudinal metalanguage with two attitudes: B and A. Our bridge principle links first and foremost to attitudinal state descriptions which describe properties of attitudes.

(43) New Expressive Principle:

> φ expresses attitudinal state description δ just in case: φ is assertable relative to an agent's doxastic state (i.e. $\mathbf{b}_{A,w} \Vdash \varphi$) if and only if δ is a true description of the agent's state (i.e. $w \in [\delta]$)

According to this definition, factual assertions express beliefs, and widest-scope epistemic 'mights' express abeliefs, just as before. But now I can do more. Recall that an attitudinal state *type* description is a *conjunction* of attitudes from the attitudinal metalanguage; in this case, a conjunction of beliefs and abeliefs. Such descriptions

[51]Proof: $\mathbf{s} \Vdash B_A\Diamond\varphi$ iff $\mathbf{b}_{A,w} \Vdash \Diamond\varphi$ for every $w \in \mathbf{s}$, iff $\mathbf{b}_{A,w} \Vdash \Diamond\neg\neg\varphi$, iff $\mathbf{s} \dashv\Vdash B_A\neg\varphi$ iff $\mathbf{s} \Vdash \neg B_A\neg\varphi$.

[52]Proof: $\mathbf{s} \Vdash \Diamond\varphi \vee \Diamond\psi$ iff $\mathbf{s}_1 \Vdash \Diamond\varphi$ and $\mathbf{s}_2 \Vdash \Diamond\psi$ for some $\mathbf{s}_1, \mathbf{s}_2$ such that $\mathbf{s} = \mathbf{s}_1 \cup \mathbf{s}_2$. This holds iff $\{w_1\} \Vdash \varphi$ and $\{w_2\} \Vdash \psi$ for some $w_1 \in \mathbf{s}_1, w_2 \in \mathbf{s}_2$. In turn, this holds iff $\{w_1\} \Vdash \varphi$ and $\{w_2\} \Vdash \psi$ for some $w_1, w_2 \in \mathbf{s}$, which is equivalent to $\mathbf{s} \Vdash \Diamond\varphi \wedge \Diamond\psi$.

correspond to non-disjunctive properties of attitudes. On this conception, I can now prove that every sentence determines a non-disjunctive property of attitudes by showing that it expresses an attitudinal state type description.

Theorem 3.1 (Non-Disjunctive Property Expression)**.** *For every sentence φ in the language, there is an attitudinal state type description δ_φ such that φ expresses δ_φ in the sense of (43).*

Proof. See Appendix A. □

The resulting package constitutes a pragmatic expressivist treatment of epistemic 'might' that also makes good on the Non-Disjunctive Expressive Principle. Semantic values are given as functions from information states to truth values. To determine a property of doxastic attitudes, one can set the information state parameter to the asserter's belief-worlds. While this is a reasonable default in assertoric conversational contexts, the current framework also allows properties of other information states to be expressed. The theorem above also generalizes.[53] In the doxastic setting, however, the theorem guarantees that an arbitrary assertion will express a *non-disjunctive* property of attitudes. For example, recall the sentence (29) which expresses a disjunctive property for Yalcin. According to the Theorem, such a sentence would express that attitudinal state type description of abelieving that Bernie will win and abelieving that Bernie will not win. In general, $\Diamond p \vee \Diamond \neg p$ will express the description $Ap \wedge A\neg p$, as desired.

3.5.3 'Must'

While the language used thus far only has an epistemic possibility modal, it is natural to extend the language with 'must' as the syntactical dual of 'might': $\Box\varphi := \neg\Diamond\neg\varphi$. This definition gives the following assertability and deniability conditions:

(44) a) $\mathbf{s} \Vdash \Box\varphi$ iff $\mathbf{s} \Vdash \varphi$[54]

b) $\mathbf{s} \dashv\Vdash \Box\varphi$ iff $\mathbf{s} \Vdash \Diamond\neg\varphi$

[53] See §8.1 of Hawke and Steinert-Threlkeld [2016b].

[54] Proof: $\mathbf{s} \Vdash \neg\Diamond\neg\varphi$ iff $\mathbf{s} \dashv\Vdash \Diamond\neg\varphi$ iff $\mathbf{s} \dashv\Vdash \neg\varphi$ iff $\mathbf{s} \Vdash \varphi$.

This has the interesting consequence that $\Box\varphi$ and φ are assertorically equivalent, but their negations are not. Such a situation can only arise in the bilateral setting.

This is as it should be. From the present epistemic perspective, wherein we do not model features like indirect evidence, one is in a position to assert $\Box\varphi$ just when one is in a position to assert φ. But $\neg\Box\varphi$ and $\neg\varphi$ still behave differently, especially under embeddings. Consider the following context.[55]

> **Traveling.** Michael's partner Shari and her friend Molly are traveling to see their mutual friend Emily. Shari has already bought her ticket. Molly told both Shari and Emily that she will be there, but is also notoriously absent-minded, often needing to be reminded to take care of practical matters.

With the information in this context, then, the following judgments seem natural:

(45) Shari believes that it's not the case that Molly must have her ticket.

(46) # Shari believes that Molly does not have her ticket.

The present definition of 'must' can offer a natural explanation of these. If Shari is genuinely uncertain about whether Molly has bought her ticket, Shari's belief worlds will contain some worlds where Molly does have her ticket and some where she does not. When that property of her belief worlds is known, (45) will be assertable, but (46) will not be.

3.5.4 Further Phenomena Involving Disjunction

Before concluding, I observe that the interaction between modals and disjunction in the assertability setting can also explain other phenomena from the literature. In each case, Yalcin's theory cannot straightforwardly explain the data, while the present theory can. Note that I do not intend to hereby undertake a thorough examination of the rich topic of the interaction between epistemic modals and disjunctions. Rather,

[55] Thanks to Matt Mandelkern for making us think about scenarios like this one and to Meica Magnani for the particular example. Thinking about these cases was the main impetus for moving to a bilateral semantics.

because such interaction was the source of disjunctive expression, I consider it a virtue of the present solution to the problem of disjunctive expression that it can also handle other related phenomena with aplomb.

First, consider some examples due to **?**.[56]

(47) Either Jack went to Alaska or he must have gone to Hawaii.

(48) The dice is less than four or probably even.

[Adapted from (38) of Moss [2015]]

Such disjunctions have the following distinguishing feature: they can be asserted even when neither disjunct can. For example, someone who only knows that Jack went to Alaska or Hawaii can assert (47) but is not in a position to assert either disjunct. Yalcin cannot straightforwardly accommodate such facts. On his semantics, sentences with widest-scope epistemic modals are *world-invariant*.[57] Then, one can observe a fact generalizing the results from the discussion of (29) and (32): if ψ is world-invariant, then if $\mathbf{s}$ accepts $\varphi \vee \psi$, it either accepts φ or accepts ψ.[58] Now, note that (47) has the logical form $p \vee \Box q$ and that $\Box q$ is world-invariant for Yalcin. Therefore, if $\mathbf{s}$ accepts (47), it either accepts p or accepts $\Box q$. Now, if one assumes that assertability amounts to the asserter being in a certain attitudinal state – knowledge, belief, or something weaker – and follows Yalcin in cashing out the attitudes in terms of acceptance by a body of information, one cannot explain the assertability properties of (47): because it is accepted if and only if one of the disjuncts is accepted, it will also be assertable if and only if one of the disjuncts is.

The present framework can handle this fact in the following way. Recall that by the definition of 'must' in Section 3.5.3, $\Box p$ and p are assertorically equivalent. This has the result that (47) can be assertable even when neither disjunct is for the same reason that factual disjunctions can be so assertable: when the information

[56] For discussion of similar cases, see Rothschild [2012], Moss [2015], Cariani [2016], Swanson [2016], Yu [2016].

[57] Definition: φ is world-invariant iff for every $\mathbf{s}$, $[\![\varphi]\!]^{w,\mathbf{s}} = [\![\varphi]\!]^{w',\mathbf{s}}$ for every w, w'.

[58] Proof: let ψ be world-invariant and suppose $\mathbf{s}$ accepts $\varphi \vee \psi$. Case 1: for some $w' \in \mathbf{s}$, $[\![\psi]\!]^{w',\mathbf{s}} = 1$. Because ψ is world-invariant, this will hold for every $w'' \in \mathbf{s}$, so $\mathbf{s}$ accepts ψ. Case 2: for all $w' \in \mathbf{s}$, $[\![\psi]\!]^{w',\mathbf{s}} = 0$. Then, $[\![\varphi]\!]^{w',\mathbf{s}} = 1$ for every $w' \in \mathbf{s}$, i.e. $\mathbf{s}$ accepts φ.

state is undecided about which disjunct holds, but contains only Jack-in-Alaska and Jack-in-Hawaii worlds, it will render (47) assertable.

Second, Dorr and Hawthorne [2013] observe that epistemic modals are often interpreted in a constrained manner: instead of ranging over all epistemically possible worlds, they range over a restricted subset thereof. Moreover, the constraint on worlds can be *inherited* from surrounding linguistic context. As an example, consider:

(49) Jennifer is feeling sick and might stay home.

Intuitively, (49) can only be asserted when some epistemically possible world is such that Jennifer stays at home *and* is sick, but not when she stays at home simply to play hooky. Yalcin can handle this fact: if $\mathbf{s}$ accepts $p \wedge \Diamond q$, it accepts $\Diamond (p \wedge q)$ as well.[59] Problems arise, however, from Dorr and Hawthorne's observation that the inherited constraints phenomenon survives many embeddings. For instance, embed (49) under a disjunction:[60]

(50) Either Jennifer is feeling sick and might stay home, or she'll be at the office.

Such a sentence cannot be asserted in a case where the only worlds in which Jennifer stays home are ones where she plays hooky but is not sick or in a case where there are no worlds in which she stays home. Because Yalcin's semantics allows arbitrary disjunction introductions, it cannot explain this phenomenon: (50) will be assertable in a case where the asserter believes that Jennifer will be at the office and so does not countenance any staying-home possibilities.

On the other hand, the present assertability semantics handles this phenomenon very elegantly:

(51) $(p \wedge \Diamond q) \vee r$ is assertorically equivalent to $(p \vee r) \wedge \Diamond (p \wedge q)$[61]

Thus, in an information state where all of the Jennifer-at-home worlds are ones where she's not sick but rather is playing hooky, (50) will not be assertable, as desired.

[59] Proof: every world in $\mathbf{s}$ is a p-world and there must also be a q-world. The latter will be a $p \wedge q$-world.

[60] Compare also example (56) on page 153 of Klinedinst and Rothschild [2012].

[61] See clause (3) of Proposition A.1.

Finally, one last piece of data concerning the interaction of epistemic modals and disjunction comes from Mandelkern [2017], who observes that embedding two epistemic contradictions under a disjunction sounds just as bad as a bare epistemic contradiction or embedding one under a conditional or an attitude verb. Here are two examples:

(52) # Jo isn't tall but she might be, or Jim isn't tall but he might be.

[from Mandelkern [2017]]

(53) # Johan is in Amsterdam but he might not be, or Thomas is in California but might not be.

I take the infelicity of (52) and (53) to show that sentences of the forms $(\neg p \wedge \Diamond p) \vee (\neg q \wedge \Diamond q)$ and $(p \wedge \Diamond \neg p) \vee (q \wedge \Diamond \neg q)$ are never assertable.[62]

First, observe that Yalcin cannot explain this generalization. Consider an information state $\mathbf{s}$ with two worlds: one – w_p – at which p is true but q is not and another – w_q – at which q is true but p is not. According to his semantics and the definition of acceptance in §3.4.1, $\mathbf{s}$ accepts $(\neg p \wedge \Diamond p) \vee (\neg q \wedge \Diamond q)$.[63] Mandelkern [2017] explains the infelicity of these sentences by appealing to a conception of bounded modality on which the interpretation of modals is constrained by their *local context*.[64]

While evaluating that proposal and comparing it to the present one lies beyond the present scope, observe that the present assertability semantics also explains the uniform infelicity of (52) and (53): there are no information states at which either sentence is assertable. That is to say: there is no information state $\mathbf{s}$ such that $\mathbf{s} \Vdash (\neg p \wedge \Diamond p) \vee (\neg q \wedge \Diamond q)$. Such a state $\mathbf{s}$ would have to be composed of two substates $\mathbf{s}_1$ and $\mathbf{s}_2$ at which each disjoined epistemic contradiction is assertable. Because epistemic contradictions are never assertable (when the conjunct contains no 'might', as in these examples), this will never happen.

[62] As Mandelkern [2017] observes, reversing the order of the conjuncts in each disjunct does not effect this judgment.

[63] A disjunction is accepted in $\mathbf{s}$ iff every $w \in \mathbf{s}$ is such that one of the disjuncts is true at w and $\mathbf{s}$. Then, by Yalcin's semantics, $[\![\neg p \wedge \Diamond p]\!]^{w_q,\mathbf{s}} = 1$ since w_q is a $\neg p$-world but there is a p world in $\mathbf{s}$. Similarly, $[\![\neg q \wedge \Diamond q]\!]^{w_p,\mathbf{s}} = 1$. So the whole disjunction is accepted at $\mathbf{s}$.

[64] See Schlenker [2009].

3.6 Conclusion

Pragmatic expressivism promises to help expressivism solve many of its problems by drawing on the distinction between semantic value and object of assertion. In this chapter, I have done two things. First, I argued by drawing analogies between the standard story and the pragmatic expressivists' story of assertion that the pragmatic expressivist ought to accept the Non-Disjunctive Expressive Principle. Failing to do so countenances assertions which would fail to fulfill the function of assertion identified by the expressivist. Against this backdrop, I showed that Yalcin's expressivism about epistemic modals runs afoul of this Principle because of how epistemic modals interact with disjunction. I then presented a new expressivist theory which one can prove satisfies this Principle.

Returning to the question that motivated this chapter – what role does compositionality play in the theory of communication? – a more general lesson can be drawn. By theorizing about the pragmatic expressivist's theory of communication, I was able to formulate a Principle that provides a bridge between compositional semantics and assertion. Moreover, the discussion of Yalcin and subsequent development of a new theory shows that this Principle has real bite: paying careful attention to the role that compositional semantics must play in the theory of communication can help adjudicate between rival semantic theories and foster the development of new ones.

More work remains to be done. In particular, to show that the Non-Disjunctive Expressive Principle can be satisfied in a general way, the theory developed here should be extended to a wider fragment of natural language. For conditionals and quantifiers, see §8 of Hawke and Steinert-Threlkeld [2016b]. Extending it to handle probability operators such as 'likely' and 'probably' will require enriching the parameter of evaluation from an information state to something with more structure. It will also be important to extend this theory to handle deontic language. The theories of Charlow [2015] and Starr [2016a,b] appear to be especially friendly companions. As discussed above, appropriately generalizing the notion of a non-disjunctive property of attitudes to the deontic setting will take some care. But such care promises to pay dividends: the result will be a pragmatic expressivist theory of a large and important

region of natural language which avoids a pitfall faced by many current theories.

An important component of the extension to the deontic case will be a thorough comparison of the behavior of epistemic and deontic modals. The assertability framework in this chapter predicts *wide-scope free choice* as an entailment (i.e. $\Diamond p \vee \Diamond q \Vdash \Diamond p \wedge \Diamond q$). Most of the literature on free choice has focused on the narrow-scope case (i.e. the inference from $\Diamond(p \vee q)$ to $\Diamond p \wedge \Diamond q$) and for deontic modals in particular. In Appendix B, I present preliminary experimental results showing that the predictions of the present theory for epistemic modals are good: people draw free choice inferences more readily with epistemic modals when the disjunction takes wide scope. Subsequent work will conduct similar experiments for deontic modals. If a different pattern in free choice emerges, an adequate theory of natural language modals will have to account for this difference.

On the logical side, future investigation will map out possible meanings for 'might' and disjunction that can deliver a result analogous to Theorem 3.1. Because of the implicit partiality of our assertability logic (it can be that neither p nor $\neg p$ are assertable relative to $\mathbf{s}$), a natural more general setting for pursuing this work comes in the form of the *data semantics* of Veltman [1985] (§III) (see also van Benthem [1988], chapters 7 and 8 and van Benthem [1991], chapter 15).

Chapter 4

Composing Semantic Automata

4.1 Introduction

This chapter pursues a third new question about compositionality, at the level of the individual language user: what is the algorithmic interpretation of compositionality and what demands on complexity does it impose on linguistic tasks? Pursuing questions of this sort requires having an algorithmic model of meaning in the context of particular tasks. A natural place to look for such a model comes from quantifier sentence verification. Verification of a sentence against a context is one strong manifestation of lingusitic competence. Moreover, the meanings of quantifiers are especially likely to be amenable to algorithmic treatment, since they appear not to introduce new content of their own. In fact, such a treatment has been given, as we will soon see. The task, then, will be to show what composition means in the algorithmic context and to assess whether composition increases complexity in a relevant sense.

To make all of this more precise, consider the following three sentences:

(54) Every student attends classes.

(55) At least three people will attend.

(56) Most people enjoyed the show.

Each one begins with a determiner – 'every', 'at least three', 'most' – followed by a noun and a verb phrase. Analyzing sentences like these has been one of the primary focuses, and one of the strongest successes, of natural language semantics in the last four decades. The determiners are analyzed as generalized quantifiers, relations between the sets of entities denoted by their complement noun and the verb phrase. So, for instance 'every' denotes the subset relation, yielding for (54) the truth-conditions that the set of students is a subset of the set of class attendees.

The interpretation of natural language determiner phrases as generalized quantifiers has led to deep and subtle insights into linguistic quantification. While the original goal of interpreting determiner phrases uniformly as higher-order properties is now seen as perhaps too simplistic,[1] the very idea that determiners can be assigned meanings which correctly predict their contribution to the meanings of sentences is of fundamental importance in semantics. Generalized quantifier theory, and arguably model-theoretic semantics in general, has largely developed independently of detailed questions about language processing. If one's aim is to understand how language can express truths, abstracting away from language users, then this orientation is arguably justified.[2] However, if the aim is to understand the role quantification plays in human cognition, model-theoretic interpretation by itself is too abstract. Patrick Suppes [1980] aptly summarized the point more than three decades ago:

> "It is a surprising and important fact that so much of language . . . can be analyzed at a nearly satisfactory formal level by set-theoretical semantics, but the psychology of users is barely touched by this analysis." (p. 27)

Consider, for instance, the basic question of how a quantified sentence is verified as true or false. Generalized quantifier theory by itself has nothing to say about this

[1] See Szabolcsi [2010] for an overview of some recent developments in quantifier theory. As she notes (p.5), "these days one reads more about what [generalized quantifiers] cannot do than about what they can."

[2] A classic statement of this approach to semantics can be found in Lewis [1970] (p.170): "I distinguish two topics: first, the description of possible languages or grammars as abstract semantic systems whereby symbols are associated with aspects of the world; and second, the description of the psychological and sociological facts whereby one of these abstract semantic systems is the one used by a person or population. Only confusion comes of mixing these two aspects." Lewis, Montague, and others clearly took the first as their object of study.

question. One may worry that the psychological details would be too complex or unsystematic to admit useful and elegant theorizing of the sort familiar in formal semantics. However, in the particular case of verification, we believe the analysis of quantifier phrases by *semantic automata* provides a promising intermediate level of study between the abstract, ideal level of model theory and the mosaic, low-level details of processing.

Semantic automata, originally pioneered by Johan van Benthem [1986], offer an algorithmic, or procedural, perspective on the traditional meanings of quantifier phrases as studied in generalized quantifier theory. They are thus ideally suited to modeling verification-related tasks. A semantic automaton represents the *control structure* involved in assessing whether a quantified sentence is true or false. While there has been relatively little theoretical work in this area since van Benthem [1986] (though see Mostowski [1991, 1998]), a series of recent imaging and behavioral experiments has drawn on semantic automata to make concrete predictions about quantifier comprehension (McMillan et al. [2005, 2006], Szymanik and Zajenkowski [2010a,b]). These experiments establish (among other results, to be discussed further below) that working memory is recruited in the processing of sentences involving certain quantifiers, which corresponds to an analogous memory requirement on automata. Such studies provide impetus to revisit the semantic automata framework from a theoretical perspective.

This chapter undertakes the extension of the semantic automata framework to a wider swath of natural language. In particular, determiners can occur not just as subjects of a sentence but also as the object of transitive verbs:

(57) Every student takes at least three classes.

(58) Most professors teach two classes.

The truth-conditions of these sentences are given by performing an operation called *iteration* on the two quantified phrases: this operation takes the two determiners and generates a relation between two subsets (the students and the classes in (57)) and the binary relation that is the denotation of the transitive verb (the taking relation in

(57)). This raises a first question: can automata for iterations be built from automata for the given determiners?

Thinking of the truth-conditions of (57) and (58) as properties of their transitive verbs, this raises a second question: do all properties of transitive verbs denoted by natural language sentences arise via the operation of iteration? Because Frege can be read as advocating a 'yes' answer to this question, the line dividing the properties of binary relations that arise from iteration from those that do not has been dubbed *the Frege Boundary* by van Benthem [1989b]. Keenan [1992, 1996b] does find examples of natural language sentences whose truth-conditions lie on the other side of the Frege boundary, such as:

(59) Different students answered different questions.

(60) A majority of the students read those two books.

(61) The two professors graded a total of fifty exams.

The sentence (59) is to be understood as having the following truth-conditions: any two distinct students answered distinct sets of questions.

Keenan [1992, 1996b] also provides an exact characterization of the Frege Boundary.[3] This characterization, however, has a highly complex formulation, making it hard to apply in practice. One would like to be able to turn the characterization into an algorithm for deciding whether a given property of binary relations is an iteration of quantifiers or not. This would allow one to determine whether a given sentence with truth-conditions of the appropriate kind is equivalent to one with iterated quantifiers. But it has remained unknown whether his characterization is *effective*. In other words: is the Frege Boundary deciable? We will also address this question using the framework of semantic automata.

This chapter is structured as follows. Section 4.2 gives a quick review of generalized quantifiers. It also contains more precise definitions of the operation of iteration and of the Frege Boundary. Section 4.3 introduces the semantic automata framework, including all of the relevant background in formal language and automata theory.

[3] Dekker [2003] generalizes these results to handle more than one iteration.

Section 4.4 provides a more detailed discussion of how semantic automata might fit into semantic theory more broadly, at a level in between model-theoretic semantics and language processing. Section 4.5 shows how to define formal languages for the iteration of two quantifiers. In Section 4.6, we study the case when quantifiers with deterministic finite-state automata are iterated. We show that the iteration also has a finite-state automaton and that the Frege Boundary is in fact decidable in this special case. Section 4.7 moves beyond to the case where pushdown automata are required. While the two results above do not carry over in full generality, we prove that they do hold for languages accepted by a *deterministic* pushdown automaton. Finally, we conclude with general remarks and future directions in Section 4.8.

4.2 Generalized Quantifiers

Definition 4.1 (Mostowski [1957], Lindström [1966])**.** A *generalized quantifier* Q of type $\langle n_1, \ldots, n_k \rangle$ is a class of models $\mathcal{M} = \langle M, R_1, \ldots, R_k \rangle$ closed under isomorphism, where each $R_i \subseteq M^{n_i}$.[4] A generalized quantifier is *monadic* if $n_i = 1$ for all i, and *polyadic* otherwise.

We write $Q_M \ R_1 \ \ldots \ R_k$ as shorthand for $\langle M, R_1, \ldots, R_k \rangle \in Q$. Usually the subscripted M is omitted for readability. Thus, e.g., for a type $\langle 1, 1 \rangle$ quantifier Q, we write $Q \ A \ B$, where A and B are subsets of the domain. This connects with the more familiar definition given in linguistic semantics. For example, the determiners in the sentences (54)-(56) have the following denotations:

$$\begin{aligned} \textsc{All} &= \{\langle M, A, B \rangle \mid A \subseteq B\} \\ \textsc{At least three} &= \{\langle M, A, B \rangle \mid |A \cap B| \geqslant 3\} \\ \textsc{Most} &= \{\langle M, A, B \rangle : |A \cap B| > |A \backslash B|\} \end{aligned}$$

The isomorphism closure condition (partially) captures the intuition that quantifiers are sensitive only to the size of the relevant subsets of M and not the identity of any

[4] For more complete introductions to the theory of generalized quantifiers, see Barwise and Cooper [1981], van Benthem [1986], Westerståhl [1989], Keenan [1996a], Keenan and Westerståhl [1997].

particular elements or the order in which they are presented.

A useful classification of generalized quantifiers is given by the standard logical hierarchy. For the time being, we restrict attention to type $\langle 1,1\rangle$, and we distinguish only between first-order and higher-order definability.

Definition 4.2. A generalized quantifier Q of type $\langle 1,1\rangle$ is *first-order definable* if and only if there is a first-order language $\mathcal{L}$ and an $\mathcal{L}$-sentence φ whose non-logical vocabulary contains only two unary predicate symbols A and B such that for any model $\mathcal{M} = \langle M, A, B\rangle$,

$$Q_M AB \quad \Leftrightarrow \quad \langle M, A, B\rangle \models \varphi$$

The generalization to higher-order (non-first order) definability is obvious. As examples, All, Some, and At least three are first-order definable:

$$\begin{aligned}
\textsc{All}_M AB &\Leftrightarrow \langle M, A, B\rangle \models \forall x\,(Ax \rightarrow Bx)\,;\\
\textsc{Some}_M AB &\Leftrightarrow \langle M, A, B\rangle \models \exists x\,(Ax \wedge Bx)\,;\\
\textsc{At least three}_M AB &\Leftrightarrow \langle M, A, B\rangle \models \exists x, y, z\ \varphi(x, y, z),
\end{aligned}$$

where $\varphi(x, y, z)$ is the formula

$$x \neq y \wedge y \neq z \wedge x \neq z \wedge Ax \wedge Bx \wedge Ay \wedge By \wedge Az \wedge Bz.$$

On the other hand, Most, An even number of, and An odd number of are (only) higher-order definable. For Most, see, e.g., Appendix C of Barwise and Cooper [1981].

Because the space of type $\langle 1,1\rangle$ quantifiers places few constraints on possible determiner meanings, several properties have been offered as potential semantic universals, narrowing down the class of possible meanings. These properties seem to hold of (at least a majority of) quantifiers found in natural languages. Two of these will play a pivotal role in the development of semantic automata, due to their role in

Theorem 4.1 below:

$$\textbf{Cons} \quad Q_M AB \text{ iff } Q_M A(A \cap B).$$
$$\textbf{Ext} \quad Q_M AB \text{ iff } Q_{M'} AB \text{ for every } M \subseteq M'.$$

The property **Cons** – short for Conservativity – states that the truth of a quantified sentence depends only on the entities in the scope of the quantifier. For example, the equivalence of the two sentences in (62) below show that *every* is conservative.

(62) a) Every dog makes its owner happy.

b) Every dog is a dog that makes its owner happy.

The property **Ext** – short for Extensionality – states that the truth-value of a quantified sentence in a model cannot be flipped simply by expanding the domain of the model.

Lemma 4.1. A quantifier Q satisfies **Cons** + **Ext** iff, for all $\mathcal{M} = \langle M, A, B \rangle$:

$$Q_M AB \Leftrightarrow Q_A A(A \cap B).$$

Theorem 4.1. *A quantifier Q satisfies* **Cons** *and* **Ext** *if and only if for every M, M' and $A, B \subseteq M$, $A', B' \subseteq M'$, if $|A - B| = |A' - B'|$ and $|A \cap B| = |A' \cap B'|$, then $Q_M AB \Leftrightarrow Q_{M'} A'B'$.*

Proof. Suppose Q satisfies **Cons** and **Ext**. If $|A - B| = |A' - B'|$ and $|A \cap B| = |A' \cap B'|$, then we have bijections between the set differences and intersections which can be combined to give a bijection from A to A'. Thus $Q_A A(A \cap B) \Leftrightarrow Q_{A'} A'(A' \cap B')$ by isomorphism closure. By two applications of Lemma 4.1, $Q_M AB \Leftrightarrow Q_{M'} A'B'$.

In the other direction, for any given $\langle M, A, B \rangle$, let $M' = A' = A$ and $B' = A \cap B$. The assumption yields $Q_M AB \Leftrightarrow Q_{M'} A'B' \Leftrightarrow Q_A A(A \cap B)$, which by Lemma 4.1 implies **Cons** + **Ext**. □

In other words, quantifiers that satisfy **Cons** and **Ext** can be summarized succinctly as binary relations on natural numbers. Given Q we define the relation $Q^c \subseteq \mathbb{N} \times \mathbb{N}$ as follows:

$$Q^c_M xy \quad \Leftrightarrow \quad \exists\, A, B \subseteq M \text{ s.t. } Q_M AB \text{ and } |A - B| = x, |A \cap B| = y.$$

Standard generalized quantifiers can thus be seen as particular simple cases.

$$\textsc{All}^c_M\ xy \Leftrightarrow x = 0$$
$$\textsc{Some}^c_M\ xy \Leftrightarrow y > 0$$
$$\textsc{At least three}^c_M\ xy \Leftrightarrow y \geqslant 3$$
$$\textsc{Most}^c_M\ xy \Leftrightarrow y > x$$
$$\textsc{An even number of}^c_M\ xy \Leftrightarrow y = 2n \text{ for some } n \in \mathbb{N}$$

Theorem 4.1 guarantees that the relation Q^c is always well defined.

4.2.1 Iterating Monadic Quantifiers

To handle sentences such as

(57) Every student takes at least three classes.

(58) Most professors teach two classes.

in which quantified phrases appear both in object position and subject position, we need to look at so-called polyadic lifts of monadic quantifiers. Intuitively, these sentences express complex properties of the respective transitive verbs. Since these verbs take two arguments, it will be impossible to give truth-conditions using monadic predicates.

In particular, we will need iterations of type $\langle 1, 1\rangle$ quantifiers. For notation, if R is a binary relation, we write

$$R_x = \{y \mid Rxy\}$$

If Q_1 and Q_2 are type $\langle 1,1\rangle$, then $\mathrm{It}(Q_1, Q_2)$ will be of type $\langle 1,1,2\rangle$, defined:

$$\mathrm{It}(Q_1, Q_2)\ A\ B\ R \quad \Leftrightarrow \quad Q_1\ A\ \{x \mid Q_2\ B\ R_x\}$$

We will sometimes use the alternative notation $Q_1 \cdot Q_2$ for $\mathrm{It}(Q_1, Q_2)$.[5]

Sentences with embedded quantifiers can be formalized as iterations. For instance, the sentences (57) and (58) would typically be assigned truth conditions (63) and (64), respectively:[6]

(63) It(All, At least three) ⟦student⟧ ⟦classes⟧ ⟦take⟧

(64) It(Most, Two) ⟦professors⟧ ⟦classes⟧ ⟦teach⟧

For example, (63) – the truth-conditions of (57) – holds iff

$$[\![\text{student}]\!] \subseteq \{x : |[\![\text{class}]\!] \cap [\![\text{take}_x]\!]| \geqslant 3\}$$

which says that each student is such that the number of classes taken by that student is at least three.

4.2.2 The Frege Boundary

As we just saw, iteration generates a quantifier of type $\langle 1,1,2\rangle$, i.e. a relation between two subsets and a binary relation over a domain. Because one of the members of such a relation has a higher arity than 1, such a quantifier will be *polyadic*. In the same way that researchers attempted to discover universals like **Cons** and **Ext** limiting the range of monadic quantifiers expressed in natural language, one can seek for plausible restrictions on the space of all polyadic quantifiers. A particularly strong universal would be:

Frege's Thesis All polyadic quantification in natural language is iterated monadic quantification.

[5] This is a special case of a general definition for iterating quantifiers. For details, see Chapter 10 of Peters and Westerståhl [2006].

[6] I am using the interpretation brackets $[\![\cdot]\!]$ to denote the extension in a model (ommitting the relativity to a model for ease of exposition) of an expression.

Some results allow us to examine Frege's Thesis in a precise way. Call a type $\langle 2 \rangle$ quantifier *Fregean* if it is an iteration of monadic quantifiers (or a boolean combination thereof). Say a quantifier 'lies beyond the Frege Boundary' if it is not Fregean.

Definition 4.3. A type $\langle 2 \rangle$ Q is *right-oriented* iff for every M, $R, S \subseteq M^2$ and $a \in M$:
if $|R_a| = |S_a|$ then $(M, R) \in Q \iff (M, S) \in Q$

Theorem 4.2 (van van Benthem [1989b])**.** *A type $\langle 2 \rangle$ GQ is Fregean iff it is permutation-invariant and right-oriented.*

Another characterization says when two Fregean, i.e. iterated, quantifiers are equal.[7]

Theorem 4.3 (Fregean Equivalence; Keenan [1992])**.** *If Q_1, Q_2 are Fregean type $\langle 2 \rangle$ quantifiers, then $Q_1 = Q_2$ iff: for all (finite) M, $P, R \subseteq M$, $\langle M, P \times R \rangle \in Q_1$ iff $\langle M, P \times R \rangle \in Q_2$.*

This provides a method for showing that a quantifier is not Fregean: show that it is equivalent to a particular Fregean quantifier on all products ($P \times R$), but that it is not equivalent in general.

For a purported example of a quantifier that lies beyond the Frege Boundary, consider again (from Keenan [1992]):

(59) Different students answered different questions.

The intended truth-conditions for (59) are that any distinct students did not answer all of the same questions:

$$\forall a \neq b \in \text{students} : \ \text{answered}_a \neq \text{answered}_b$$

We will consider these truth-conditions as a type $\langle 2 \rangle$ quantifier, i.e. as a property of binary relations, called DIFF. We can use Fregean Equivalence to show that DIFF lies beyond the Frege Boundary.

[7] See Keenan [1996a], Dekker [2003] for generalizations.

Consider the type $\langle 1 \rangle$ quantifier $\bot$ which is uniformly false. Its iteration $\mathrm{It}(\bot, \bot)$ will be uniformly false of all binary relations.[8] We will show that DIFF agrees with $\mathrm{It}(\bot, \bot)$ on products, but not in general. Without loss of generality, fix a domain M with at least three students and let $P \times R$ be a product. There are two cases. (i) There are two students not in P. Then, they bear $P \times R$ to the same set of questions: the empty set. So DIFF is false at $\langle M, P \times R \rangle$. (ii) There are at least two students in P. Then, they bear $P \times R$ to the same set of questions: the questions that are in R. Again, DIFF is false. So it agrees with $\mathrm{It}(\bot, \bot)$ on products. But DIFF is true for the binary relation relating s_1 to q_1 and s_2 to q_2, where the s_i are distinct students and the q_i distinct questions.

As can be seen from this example, telling whether a quantifier is Fregean or not is not always easy. And fully general characterizations of the boundary are even more complex. This raises the interesting question: is the Frege Boundary decidable? That is, does there exist an algorithm which, given a quantifier of type $\langle 2 \rangle$, decides whether or not it is an iteration of quantifiers? We return to this question in Sections 4.6.3 and 4.7.4, using tools from semantic automata theory that will be introduced in the intervening sections.

4.3 Semantic Automata

Having completed a brief introduction to generalized quantifiers, we proceed in this section to begin the study of semantic automata, in which tools from formal language and automata theory are applied to generalized quantifiers. Subsection 4.3.1 contains the necessary preliminary definitions from language and automata theory. Subsection 4.3.2 then shows how to assign languages and their corresponding automata to type $\langle 1, 1 \rangle$ quantifiers and elucidates connections between definability and types of language.

[8] While I treated iteration in the previous section as applying to type $\langle 1, 1 \rangle$ quantifiers, it is easy to generalize the definition. For type $\langle 1 \rangle$, $\mathrm{It}(Q_1, Q_2)\ R \Leftrightarrow Q_1 \{x \mid Q_2 R_x\}$.

4.3.1 Preliminaries

Though some familiarity with the theory of regular languages and deterministic finite-state automata will be assumed, we here introduce basic concepts and notation.

An alphabet Σ is a finite set, whose elements are called characters. A language L is a subset of Σ^*, where Σ^* is the set of all finite sequences in Σ. We will use L, L_1, L_2 to denote languages. Elements $w \in \Sigma^*$ are called *words*. ε denotes the empty sequence. For $w \in \Sigma^*$ for finite alphabet Σ, $w_i \in \Sigma$ denotes the character at the ith position of w. We use $|\cdot|$ as both a set cardinality function and a word-length function. The functions $\#_a : \Sigma^* \to \mathbb{N}$ for each $a \in \Sigma$ are recursively defined in such a way to return the number of as in a word $w \in \Sigma^*$.[9]

A *deterministic finite-state automaton* (DFA) is a tuple $\langle \Sigma, Q, \delta, F, q_0 \rangle$ where Σ is an alphabet (a finite set of letters), Q is a set of states, $q_0 \in Q$ is the starting state, $F \subseteq Q$ is the set of final (or accepting) states, and $\delta : Q \times \Sigma \to Q$ is the transition function. We will use $\mathsf{M}, \mathsf{M}_1, \mathsf{M}_2$ to denote automata. The components of an automaton are denoted by $Q(\mathsf{M})$, $\delta(\mathsf{M})$, et cetera. We often write $|\mathsf{M}|$ instead of $|Q(\mathsf{M})|$.

A DFA accepts a language in the following sense. Define $\delta^* : Q \times \Sigma^* \to Q$ by induction on Σ^*:

$$\delta^*(q, \varepsilon) = q$$
$$\delta^*(q, aw) = \delta(\delta^*(q, w), a)$$

In other words, $\delta^*(q, w)$ is the state arrived at by starting at q, transitioning according to w_1, then according to w_2, and so on through w_n, where n is the length of w. We can now define the language of an automaton M by

$$L(\mathsf{M}) = \{w \in \Sigma^* : \delta^*(q_0, w) \in F\}$$

The *regular languages* are the languages L such that $L = L(\mathsf{M})$ for some DFA M.[10] If

[9] Here's the definition: $\#_a(\varepsilon) = 0$ and $\#_a(cw) = \#_a(w) + 1$ if $c = a$; $\#_a(w)$ otherwise.

[10] Many alternative characterizations exist: acceptance by a non-deterministic finite-state automaton and generation by a regular expression, for example. See Hopcroft and Ullman [1979] for details.

L is a regular language, we denote the minimal automaton accepting L by $\min(L)$.[11]

Given a machine M, we define

$$\operatorname{sgn}(q, \mathsf{M}) = \chi_{F(\mathsf{M})}(q)$$

and $\operatorname{sgn}(\mathsf{M})$ as an abbreviation for $\operatorname{sgn}(q_0(\mathsf{M}), \mathsf{M})$ where χ_S is the characteristic function of set S. In other words, $\operatorname{sgn}(q, \mathsf{M})$ is 1 if q is an accepting state of M, 0 otherwise. We omit the second argument when context permits.

For an example, consider the language $L_\forall = \left\{w \in \{0,1\}^* : \#_0(w) = 0\right\}$ of words in $\{0,1\}^*$ containing all and only 1s. This language is regular; its minimal automaton is depicted in Figure 4.1. The choice of name for this language will become clear in the next section. In this diagram and others like it later in the paper, states are denoted by circles. q_0 is the state with an arrow leading in to it. States in F have two circles. An arrow from state q to q' labeled by a means that $\delta(q, a) = q'$.

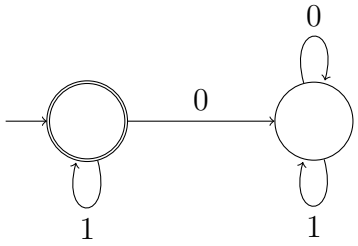


Figure 4.1: The automaton $\min(L_\forall)$.

The *star-free languages* in Σ is the smallest set of languages which contains Σ^*, $\{a\}$ for each $a \in \Sigma$ and which is closed under finite union, concatenation, and complementation. These languages are accepted by the *acyclic* or counter-free DFAs; see McNaughton and Papert [1971]. Thus, every star-free language is regular. The converse, however, is not true: a paradigmatic regular language which is not star-free is $L_{\text{even}} = \left\{w \in \{0,1\}^* : \#_1(w) \text{ is even}\right\}$. The minimal automaton accepting this language is depicted in Figure 4.2. A similar construction shows that $L_{/m} =$

[11] Again, see Hopcroft and Ullman [1979] for a proof that every regular language has a minimal automaton, i.e. one with as few states as any other automaton accepting the same language.

$\left\{w \in \{0,1\}^* : \#_1(w) \text{ is divisible by } m\right\}$ is also regular and that its minimal automaton has m states.

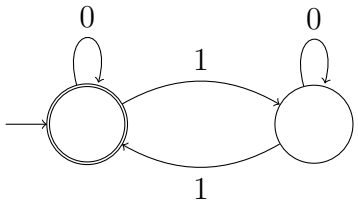


Figure 4.2: The automaton $\min(L_{\text{even}})$.

We will use another characterization of the star-free languages in terms of first-order definability.[12] Given an alphabet Σ, consider a standard first-order language with unary predicate symbols P_a for each $a \in \Sigma$, a binary relation symbol $<$, and a countably infinite set of variables denoted VAR. We write $\text{Form}(\Sigma)$ and $\text{Sent}(\Sigma)$ for, respectively, the set of first-order formulas and sentences in this signature. We interpret formulas in this language in words $w \in \Sigma^*$ with respect to variable assignments $g : \text{VAR} \to \{1, 2, \ldots, |w|\}$. The relevant semantic clauses are given by

$$w, g \models x_i < x_j \text{ iff } g(x_i) < g(x_j)$$
$$w, g \models P_a(x_i) \text{ iff } w_{g(x_i)} = a$$

A first-order sentence φ defines the language

$$L_\varphi = \{w \in \Sigma^* \mid w \models \varphi\}$$

where the variable assignment g is omitted since a sentence is satisfied with respect to some g iff it is with respect to all g. The theorem alluded to above can now be stated.

Theorem 4.4 (McNaughton and Papert [1971])**.** *A language $L \subset \Sigma^*$ is star-free iff it is first-order definable in the following sense: there is a sentence $\varphi \in \text{Sent}(\Sigma)$ such that $L = L_\varphi$.*

[12] See Diekert and Gastin [2007] for a self-contained presentation of this equivalence and others.

While the star-free languages form a natural and interesting sub-class of the regular languages, we will also be interested in natural extensions of the regular languages. In particular, we will study the *context-free* languages (and the detemrinistic subset thereof), which again will be defined in terms of the machines accepting them. Essentially, these are automata which extend DFAs with a form of memory called a stack. A stack stores symbols. The top symbol can be read and removed (referred to as popping). Moreover, a string of symbols can be pushed on to the top of the stack.

A *pushdown automaton* (PDA) is a tuple $\langle Q, \Sigma, \Gamma, \delta, q_0, Z_0, F\rangle$ where Q, Σ, q_0, and F are as in a DFA. Γ is another alphabet called the stack alphabet; Z_0 is a special symbol designating an empty stack; $\delta : Q \times (\Sigma \cup \{\varepsilon\}) \times \Gamma \to \mathcal{P}(Q \times \Gamma^*)$ is the transition function. Intuitively, $(q', \gamma) \in \delta(q, a, b)$ means that if the PDA reads letter a while in state q_1 with letter $b \in \Gamma$ on top of the stack, it can transition into q_2 and replace b with γ on top of the stack.[13] The case when $\gamma = \varepsilon$ captures popping the top of the stack. In transition diagrams, we will draw an arrow from q to q' labelled by $a, b/\gamma$.

To define the language accepted by a PDA, we need a bit more notation. A triple $\langle q, w, \gamma\rangle$ of a state, word in Σ, and word in Γ will be called an *instantaneous description* of a PDA. We write $\langle q_1, aw, b\gamma\rangle \vdash_{\mathsf{M}} \langle q_2, w, \gamma'\gamma\rangle$ whenever $\langle q_2, \gamma'\rangle \in \delta(q_1, a, b)$. We omit the subscript when context allows and write $\vdash^*$ for the reflexive, transitive closure of $\vdash$. The language accepted by a PDA M is:

$$L(\mathsf{M}) = \{w \in \Sigma^* : \langle q_0, w, Z_0\rangle \vdash^* \langle q, \varepsilon, \gamma\rangle \text{ for some } q \in F \text{ and } \gamma \in \Gamma^*\}$$

A language L is *context-free* iff $L = L(\mathsf{M})$ for some PDA M.[14]

4.3.2 Automata Corresponding to Quantifiers

We will now see how these tools can be used to study quantification in natural language. Recall that in formal semantics of natural language, the denotations of the

[13] The 'can' here reflects that the present definition defines nondeterministic PDAs which, unlike in the finite-state case, are strictly more powerful than their deterministic counterparts, to be discussed below.

[14] These languages can equally be characterized as those generated by a context-free grammar.

determiners 'every', 'at least three', and 'most' in examples (54)-(56) above have been given as type $\langle 1,1\rangle$ generalized quantifiers. A type $\langle 1,1\rangle$ generalized quantifier is a class of finite models of the form $\langle M, A, B\rangle$ with $A, B \subseteq M$. As we saw, the determiners above have the following denotations, where I am introducing new abbreviations for the former two:[15]

$$\begin{aligned}\forall &= \{\langle M, A, B\rangle : A \subseteq B\}\\ \geqslant_3 &= \{\langle M, A, B\rangle : |A \cap B| \geqslant 3\}\\ \textsc{Most} &= \{\langle M, A, B\rangle : |A \cap B| > |A \backslash B|\}\end{aligned}$$

Semantic automata theory shows how such denotations can be given corresponding formal languages and automata accepting these languages. This works as follows. Under standard assumptions about generalized quantifiers – **Cons** and **Ext** – it follows that $\langle M, A, B\rangle \in Q$ iff $\langle A, A, A \cap B\rangle \in Q$ (recall Lemma 4.1). Now, given an enumeration $\vec{a}$ of A, a word in alphabet $\{0,1\}$ corresponding to such a model can be defined by

$$\left(\tau\left(\vec{a}, B\right)\right)_i = \begin{cases} 0 & a_i \in A \backslash B \\ 1 & a_i \in A \cap B \end{cases}$$

The *language of* Q – denoted L_Q – is then defined as the image of Q under τ; that is, as the set of all strings that can be generated from models in Q (together with an enumeration of A) by τ. One can verify that $L_\forall$ from the previous section is the language of $\forall$ in precisely this sense because $A \subseteq B$ iff $|A \backslash B| = 0$ iff $\#_0\left(\tau\left(\vec{a}, B\right)\right) = 0$ for any enumeration $\vec{a}$. Similarly, we have

$$\begin{aligned}L_{\geqslant_3} &= \left\{w \in \{0,1\}^* : \#_1(w) \geqslant 3\right\}\\ L_{\textsc{Most}} &= \left\{w \in \{0,1\}^* : \#_1(w) > \#_0(w)\right\}\end{aligned}$$

Note that the quantifier $\geqslant_3$ and its corresponding language are instances of a schema

[15] Note that I am using the symbol $\forall$ to denote a type $\langle 1,1\rangle$ generalized quantifier whereas it is standardly used to denote the type $\langle 1\rangle$ quantifier $\{\langle M, A\rangle : A = M\}$. They are, however, intimately related: the $\langle 1,1\rangle$ one here is the *relativization* of the normal type $\langle 1\rangle$ quantifier. See §4.4 of Peters and Westerståhl [2006].

$\geqslant_n$ for any n. In particular, $\geqslant_1$ is the standard denotation for 'some' and so will be denoted $\exists$; the corresponding language is $L_\exists = \left\{w \in \{0,1\}^* : \#_1(w) \geqslant 1\right\}$.

Example 4.1. Consider a model given as follows: $M = B = \{a, b, c, d\}$, $A = \{a, b, c\}$ with $[\![\text{student}]\!] = A$ and $[\![\text{attended}]\!] = B$. Clearly,

(54) Every student attends classes.

will be true in this model with this interpretation, i.e. $M \in \forall$. For any ordering $\vec{a}$, we have that $\tau(\vec{a}, B) = 111$, which is in $L_\forall$. The automaton in Figure 4.1 can thus be seen as a verifier of the truth of sentences of the form 'Every A is a B'.

Example 4.2. Figure 4.3 shows an automaton accepting $L_{\geqslant_3}$.

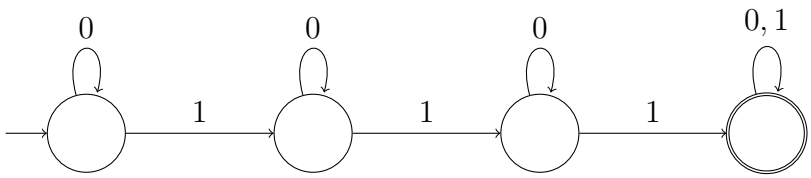


Figure 4.3: The automaton $\min(L_{\geqslant_3})$.

Not all quantifiers, however, have regular languages. In other words, some quantifiers have corresponding languages that require more powerful machines – pushdown automata, for example – to be accepted. The paradigm example is MOST.

Fact 4.1. L_{MOST} is not regular.

L_{MOST} is, however, context-free. Figure 4.4 depicts a pushdown automaton accepting that language. The states are labeled for future reference. Intuitively, this automaton implements a 'pair-matching' verification strategy. If it reads a 1 while a 1 is on the stack, it pushes the new 1 on to the stack as well. If, however, a 0 is on the stack, that 0 is popped off the stack, matching a 1 and a 0. And *mutatis mutandis* for reading a 0. If only 1s are left after this process of matching 1s and 0s, that means that there were more 1s in the string.

We have now observed that MOST is special in two ways: it is not definable in first-order logic and has a non-regular language. As it turns out, these two special features

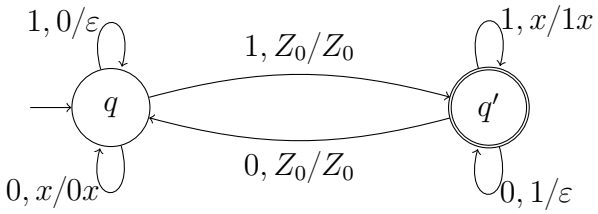


Figure 4.4: PDA accepting L_{most}.

are related. Definability of a generalized quantifier in certain logics corresponds to the type of automaton recognizing its language. The following result characterizes the first-order definable quantifiers in terms of their corresponding languages.

Theorem 4.5 (van Benthem [1986])**.** *Q is definable in first-order logic iff L_Q is star-free.*

Because the star-free languages are a strict subset of the regular languages, one would like to know which quantifiers have corresponding regular languages. The following result exactly characterizes them.

Theorem 4.6 (Mostowski [1998])**.** *L_Q is regular iff Q is definable in first-order logic augmented with all divisibility quantifiers $D_n = \{\langle M, A\rangle : |A| \text{ is divisible by } n\}$*

More characterization results exist. van Benthem [1986] characterizes the quantifiers with corresponding context-free languages in terms of their binary relation Q^c: L_Q is context-free if and only if Q^c is *semi-linear* (see the reference for details). Similarly, Kanazawa [2013] characterizes the quantifiers whose languages are *deterministic* context-free. See §4.3 of Szymanik [2016] for a more detailed exposition of these results. In the next section, we discuss the significance of these results.

4.4 Automata and Processing

How exactly does this work on automata relate to questions about processing? On one hand, the machine representations of quantifiers discussed in the previous section are inspired directly by the standard model-theoretic meanings assumed in classical

quantifier theory. On the other hand, the fine structure of these representations promises a potential bridge to lower level questions about language processing. In this section we discuss two important dimensions of this connection:

(i) Semantic automata suggest a relatively clean theoretical separation between semantic competence and performance errors.

(ii) Recent imaging and behavioral experiments have revealed that structural differences in automata may be predictive of neuroanatomical demands in quantifier processing.

4.4.1 Explaining Performance Errors

The logical theory of generalized quantifiers is occasionally dismissed as being irrelevant to the psychological facts about how people produce and assess quantified sentences, typically by citing data that suggests human reasoning with quantifiers does not match the patterns assumed in standard logical analysis. Indeed, experiments reveal people diverge from standard logical prescriptions, not only in reasoning tasks such as the classical syllogism (see, e.g. Chater and Oaksford [1999]), but even in simple verification tasks (see, e.g. McMillan et al. [2005] with a similar pattern in Szymanik and Zajenkowski [2010a]). One could take this to show, or at least to reinforce the idea, that logical/truth-conditional semantics and the psychology of language are best kept separate, with the former studying abstract normative aspects of meaning and the latter studying the actual mechanisms involved in processing.[16] Yet the normative aspects of meaning, and of quantifier meanings in particular, are clearly relevant to questions about how people use quantifiers, and *vice versa*. While a complete discussion of this issue is beyond the scope of this paper, we would like to point out that semantic automata have the potential to serve as a natural bridge. In principle, they allow for a separation between abstract *control structure* involved in quantifier verification and innumerable other variables that the framework leaves underspecified: order of evaluation, predication judgments, domain restriction, and

[16] Recall the quotation from Lewis [1970] in Footnote 2 above.

any contextual or practical factors that might affect these and other variables. This could be viewed as a modest distinction between *competence* and *performance* for quantifier expressions.[17]

We might hypothesize that competence with a particular quantifier involves, at least in part, internalizing the right abstract computational mechanism for verification, in particular that given by the appropriate automaton. How precisely verification is implemented on a given occasion will depend on many factors quite independent from the meanings of quantifiers: prototype effects, saliency effects, time constraints, and so on. Consider, for instance, how one might verify or falsify a sentence like (65):

(65) All U.S. presidents have served at least one year in office.

Supposing one does not already know whether this is true or false, but that an answer must be computed using information stored in memory, one might check famous presidents like George Washington or Abraham Lincoln first and only then move to less salient presidents. Alternatively, if one actually knew James Garfield or William Harrison had served less than a year, such information might be retrieved quickly without first checking more salient presidents. The subtleties of such strategies are fascinating, but arguably go beyond the meanings of quantifiers. Indeed, they arise in tasks and phenomena having nothing to do with quantifiers.

The same can be said for cases where a search is terminated too soon. In example (65), a person might think about a few salient examples and conclude that the sentence is true. For a particular task it may take too long to think about all forty-four presidents, or it may not be worth the effort, and a quick guess is sufficient. However, even in such cases, upon being shown a counterexample, a subject will not insist that the sentence is actually true just because they were unable to identify the counterexample. It is in that way that people are reasonably attuned to the proper, "normative" meanings. Ordinary speakers have a good sense for what needs to be

[17] Our suggestion is compatible with many interpretations of what this distinction comes to. For instance, the relatively non-committal interpretation of Smolensky [1988] says competence of a system or agent is "described by hard constraints" which are violable and hold only in the ideal limit, with unbounded time, resources, and other enabling conditions. The actual implementational details are to be given by "soft constraints" at a lower level, which have their own characteristic effects on performance.

checked to verify a quantified sentence, even if in practice going through the necessary steps is difficult or infeasible. Semantic automata allow separating out control structure from the specific algorithm used to implement the procedure.

In terms of the Marr's famous *levels of explanation* (Marr [1982]), the standard model-theoretic semantics of quantification could be seen as a potential computational, or level 1, theory. Semantic automata offer more detail about processing, but, as we have just argued, less than one would expect by a full algorithmic story about processing (level 2), which would include details about order, time, salience, et cetera. Thus, we might see the semantic automata framework as aimed at level 1.5 explanation,[18] in between levels 1 and 2, providing a potential bridge between abstract model theory and concrete processing details.

4.4.2 Experimental Results

While this separation between competence and performance remains somewhat speculative, recent experimental work has shown that certain structural features of semantic automata are concretely reflected in the neuroanatomical demands on quantifier verification. Recall that certain quantifiers can be computed by memoryless finite state automata whereas others (such as *most*) require a pushdown automaton which has a form of memory in its stack data structure. McMillan et al. [2005] used fMRI to test "the hypothesis that all quantifiers recruit inferior parietal cortex associated with numerosity, while only higher-order quantifiers recruit prefrontal cortex associated with executive resources like working memory." In their study, twelve native English speakers were presented with 120 grammatically simple sentences using a quantifier to ask about a color feature of a visual array. The 120 sentences included 20 instances of 6 different quantifiers: three first-order (*at least 3*, *all*, *some*) and three higher-order (*less than half*, *an odd number of*, *an even number of*). Each subject

[18] Peacocke [1986] coined the term "level 1.5", though in a slightly different context. Soames [1984] independently identified three levels of (psycho)linguistic investigation which appear to (inversely) correlate with Marr's three levels. Pietroski et al. [2009] also articulate the idea that verification procedures for quantifiers (for *most*, specifically) provide a level 1.5 explanation. Independently, the idea of level 1.5 explanation has recently gained currency in Bayesian psychology in relation to *rational process models* (Griffiths et al. [2015]).

was first shown the sentence alone on a screen for 10s, then the sentence with a visual scene for 2500ms, then a blank screen for 7500ms during which they were to assess whether the sentence accurately portrayed the visual scene.

While behavioral results showed a statistically significant difference in accuracy between verifying sentences with first-order quantifiers and those with higher-order quantifiers, more interesting for our present purposes was the fact that activation in brain regions (dorsolateral prefrontal and inferior frontal cortices) typically linked with executive functioning such as working memory was found only in processing higher-order quantifiers. They concluded that the formal difference between the machines required to compute quantifier languages seems to reflect a physical difference in neuro-anatomical demands during quantifier comprehension.

This first piece of evidence does not tell the whole story, however. The first-order vs. higher-order distinction does not map directly on to the distinction between DFAs and PDAs because of parity quantifiers such as *an even number of*, which are computable by cyclic DFAs. Szymanik [2007] observed that McMillan et al. did not make this distinction and began investigating the demands placed on memory by parity quantifiers. In a subsequent study by Szymanik and Zajenkowski [2010b], reaction times were found to be lowest for Aristotelian quantifiers (*all* , *some*), higher for parity (*an even number of*), yet higher for cardinals of high rank (*at least 8*), and highest for proportionality quantifiers (*most*). This provides some evidence that the complexity of the minimal automaton, and not simply the kind of automaton, may be relevant to questions about processing.[19] A subsequent study by Szymanik and Zajenkowski [2011] showed that proportionality quantifiers place stronger demands on working memory than parity quantifiers. This result is consistent with the semantic automata picture, since a PDA computing the relevant parity quantifiers never needs to push more than one symbol on to the stack.

Finally, it is also relevant that McMillan et al. [2005] found no significant difference in activation between judgments involving *at least 3* where the relevant class had cardinality near or distant to three. This suggests subject are invoking a precise

[19]Indeed, a general notion of complexity for automata in the context of language processing would be useful in this context.

number sense in such cases. This contrasts with recent studies by Pietroski et al. [2009] and Lidz et al. [2011], which attempt to distinguish which of several verification procedures are actually used when processing sentences with *most*. Although the present paper shares much in spirit with this project, their protocols show a scene for only 150 or 200ms, which effectively forces the use of the *approximate number system*.[20] The finding is that in such cases people do not seem to be using a pair-matching procedure such as that defined above. We conjecture that with more time the pair-matching algorithm would be used, at least approximately.

The experimental work described in this section has been based solely on automata defined for monadic quantifiers. In order to carry this project further, and to understand its potential and its shortcomings as a high-level processing model, we need to define machines appropriate for more complex quantificational types. In the next section we take an important first step in this direction, defining languages for iterations of quantifiers. After defining these languages, we then ask whether certain classes of languages are closed under iteration: if the languages Q_1 and Q_2 are, e.g., regular, then is the language of $\mathrm{It}(Q_1, Q_2)$ also regular?[21]

4.5 Iterated Languages

While the framework as described has proven fruitful both logically and empirically, the above methods can only handle sentences of the form '$Q\ A\ B$'. A natural larger fragment of natural language comes from considering examples like (57) and (58) in which a quantified noun phrase appears as both subject and object of a transitive verb. As we saw in Section 4.2.1, these sentences can be given truth-conditions using an operation called *iteration* of quantifiers, which takes two type $\langle 1,1\rangle$ quantifiers and generates a new type $\langle 1,1,2\rangle$ quantifier.

To extend the semantic automata framework to handle iterated quantifiers, one

[20] See Dehaene [1997] for the distinction between precise and approximate number systems.

[21] Szymanik [2010] investigated the computational complexity of polyadic lifts of monadic quantifiers. This approach, however, deals only with Turing machines. Our development can be seen as investigating the fine structure of machines for computing quantifier meanings.

must show how to define formal languages corresponding to iterations. Steinert-Threlkeld and Icard III. [2013] take an approach modeled on the previous case: they define an encoding $\tau\left(\vec{a},\vec{b},R\right)$ for finite models of the form $\langle M,A,B,R\rangle$ (with respect to enumerations of A and B) as strings in an alaphabet $\{0,1,\boxdot\}$. A slightly different approach is pursued here: because the quantifier $\mathrm{It}(Q_1,Q_2)$ is 'built from' Q_1 and Q_2, we provide a definition of a language for iterations in terms of two given languages. In particular, given a language L in some alphabet Σ, define the function $\sigma_L:\{0,1\}\to\mathcal{P}\left((\Sigma\cup\{\boxdot\})^*\right)$ by

$$1\mapsto L\cdot\{\boxdot\}\qquad 0\mapsto L^c\cdot\{\boxdot\}$$

where $L^c=\Sigma^*\backslash L$ and $\boxdot$ is a fresh symbol not in Σ. Here, $\cdot$ is concatenation of languages. For example, $L\cdot\{\boxdot\}:=\{w\boxdot:w\in L\}$. Given $L_1,L_2\subseteq\{0,1\}^*$, we are interested in languages of the form

$$L=\sigma_{L_2}\left[L_1\right]$$

where σ_{L_2} is extended to a substitution on the whole language L_1 in the usual way.[22] We call this language *the iteration of* L_1 *and* L_2 and denote it by $L_1\bullet L_2$. This language is aptly named: if L_1 and L_2 are the languages of type $\langle 1,1\rangle$ quantifiers Q_1 and Q_2, then $L_1\bullet L_2$ will be the language of the iteration of Q_1 and Q_2.[23,24]

Example 4.3. We can apply this definition to our example sentence (57):

(57) Every student takes at least three classes.

Observe that

$$L_\forall\bullet L_{\geqslant 3}=\left\{w\in(w_i\boxdot)^*:\left|\{w_i:\#_1(w_i)<3\}\right|=0\right\}$$

where w_i ranges over maximal subwords of w containing only 0s and 1s. Consider

[22] That is: $\sigma_{L_2}(\varepsilon)=\varepsilon$, $\sigma_{L_2}(aw)=\sigma_{L_2}(a)\,\sigma_{L_2}(w)$ and $\sigma_{L_2}[L_1]=\bigcup_{w\in L_1}\sigma_{L_2}(w)$.

[23] See Steinert-Threlkeld and Icard III. [2013] and references therein. The present definition of iteration can be seen as a more concise representation of their Definition 8.

[24] Szymanik et al. [2013] contains a preliminary experiment looking at processing consequences of semantic automata for iterated quantifiers.

the following model and interpretation (where the s_i are the students, the c_i are the classes, and a ✓ means that $\langle s_i, c_j \rangle \in [\![\text{take}]\!]$):

	c_1	c_2	c_3	c_4
s_1	✓	✓	✓	✓
s_2	✓		✓	✓
s_3	✓	✓		✓

The encoding from Steinert-Threlkeld and Icard III. [2013] will generate the string $1111 \boxdot 1011 \boxdot 1101 \boxdot$ which is in $L_\forall \bullet L_{\geqslant_3}$. This is at it should be: every student does take at least three classes.

Now that we have shown how to define formal languages for iterations of quantifiers, we can now turn to the study of automata accepting these languages. As in the case with simple semantic automata, a natural questions concerns stratification of these languages: which iterations have regular languages and which require the jump to context-free? In the context of iterations, this can be transformed into a closure question: if L_1 and L_2 are both regular (resp. context-free), is $L_1 \bullet L_2$ also regular (resp. context-free)?

In addition to asking this closure question, we can use iterated languages and automata to address a question from pure generalized quantifier theory raised in Section 4.2.2: is the Frege Boundary deciable? In the context of iterated languages, this question becomes: given a language $L \subseteq \{0, 1, \boxdot\}^*$, is it decidable whether there are two languages $L_1, L_2 \subseteq \{0, 1\}^*$ such that $L = L_1 \bullet L_2$?

We first address all of these questions in the case of regular languages, in Section 4.6. We will prove the closure of the regular and star-free and languages under iteration and define automata for iterated languages in order to establish the state-complexity of iteration. After that, we answer the decidability question for regular languages in the affirmative. Then, in Section 4.7, we look at semantic automata for and iteration of context-free and deterministic context-free languages. These languages are needed for sentences like (56) and (58) with proportional quantifiers.

4.6 Iterated Regular Languages

4.6.1 Closure

Fact 4.2. The regular languages are closed under iteration. In other words, if L_1 and L_2 are regular, then $L_1 \bullet L_2$ is regular.

Proof. This follows immediately from the closure of the regular languages under complement and substitution. □

The proof for star-free langauges will have to be more complex because this class of languages is not closed under substitution in general. As an example, let $\Sigma = \{1\}$ and consider $L_1 = \{1\}^*$, $L_2 = \{11\}$ and $\sigma(1) = L_2$. Then $\sigma[L_1] = \{w \mid \#_1(w) \text{ is even}\}$, which we have seen is quintessentially not star-free.

Proposition 4.1. *The star-free languages are closed under iteration.*

Proof. Let $\varphi_1(P_0, P_1)$ and $\varphi_2(P_0, P_1)$ be first-order sentences defining L_1 and L_2 (see Theorem 4.4). We will use the ⧈ symbol, with its corresponding predicate $P_{⧈}$ to convert predications in φ_1 into quantifications over words in $\{0,1\}^*$. We need the following string of defined symbols:

$$\begin{aligned}
\text{prev}(x_j) = x_k &:= (x_k < x_j \wedge \forall x_i (x_i < x_j \rightarrow x_i \leqslant x_k)) \\
&\quad \vee (x_k = x_j \wedge \neg\exists x_i (x_i < x_j)) \\
\text{next}(x_j) = x_k &:= (x_j < x_k \wedge \forall x_i (x_j < x_i \rightarrow x_k \leqslant x_i)) \\
&\quad \vee (x_k = x_j \wedge \neg\exists x_i (x_j < x_i)) \\
\text{Start}(x_j) &:= \text{prev}(x_j) = x_j \vee P_{⧈}(\text{prev}(x_j)) \\
\text{End}(x_j) &:= \text{next}(x_j) = x_j \vee P_{⧈}(\text{next}(x_j)) \\
\text{Word}(x_j, x_k) &:= x_j \leqslant x_k \wedge \text{Start}(x_j) \wedge \text{End}(x_k) \wedge \\
&\quad \forall x_i (x_j < x_i < x_k \rightarrow \neg P_{⧈}(x_i))
\end{aligned}$$

Inspection of these definitions shows that $w, g \models \text{Word}(x_j, x_k)$ iff the subword $w_{g(x_j)} \cdots w_{g(x_k)}$ is a maximal sub-word of w containing only 0s and 1s.

Given a formula φ, we denote by $\varphi^{[x_i,x_j]}$ the formula obtained by gaurding all quantifiers in φ over x with $x_i \leqslant x \leqslant x_j$. We will now define a translation

$$\tau : \mathrm{Form}(\{0,1\}) \times \mathrm{Form}(\{0,1\}) \to \mathrm{Form}(\{0,1,\boxdot\})$$

In the definition below, we stipulate that if $i \neq j$, then $i_k \neq j_k$ for $k \in \{1,2\}$.

$$\begin{aligned}
\tau(x_i = x_j, \varphi_2) &= x_{i_1} = x_{j_1} \wedge x_{i_2} = x_{j_2} \\
\tau(x_i < x_j, \varphi_2) &= x_{i_2} < x_{j_1} \\
\tau(P_1(x_i), \varphi_2) &= \varphi_2^{\left[x_{i_1}, x_{i_2}\right]} \\
\tau(P_0(x_i), \varphi_2) &= \neg\varphi_2^{\left[x_{i_1}, x_{i_2}\right]} \\
\tau(\neg\varphi_1, \varphi_2) &= \neg\tau(\varphi_1, \varphi_2) \\
\tau(\varphi_1 \wedge \psi_1, \varphi_2) &= \tau(\varphi_1, \varphi_2) \wedge \tau(\psi_1, \varphi_2) \\
\tau(\exists x_i \varphi_1, \varphi_2) &= \exists x_{i_1} \exists x_{i_2} (\mathrm{Word}(x_{i_1}, x_{i_2}) \wedge \tau(\varphi_1, \varphi_2))
\end{aligned}$$

It's easy to see that if φ_1 and φ_2 are sentences, then so too is $\tau(\varphi_1, \varphi_2)$.

Claim 4.1. Let $\varphi_1, \varphi_2 \in \mathit{Sent}(\{0,1\})$. Then

$$L_{\tau(\varphi_1,\varphi_2)} = L_{\varphi_1} \bullet L_{\varphi_2}$$

Proof. Put φ_1 in prenex normal form and generate $\tau(\varphi_1, \varphi_2)$ 'from the outside in'. Quantifiers over positions in words in L_1 are converted into quantifiers over maximal subwords in $\{0,1\}^*$. The translation of $<$ ensures that the order of the subwords reflects the order of the characters. The translation of P_0 and P_1 ensure that 0s are replaced by words in $L^c_{\varphi_2}$ and 1s by words in L_{φ_2}. But that is just the definition of $L_{\varphi_1} \bullet L_{\varphi_2}$. □

By Theorem 4.4, this shows that $L_1 \bullet L_2$ is star-free. □

Example 4.4. $\varphi_\forall := \forall x P_1(x)$ defines $L_\forall$ and $\varphi_\exists := \exists x P_1(x)$ defines $L_\exists$. We have

that

$$\begin{aligned}\tau(\varphi_\forall, \varphi_\exists) &= \forall x_1 \forall x_2 \left(\text{Word}(x_1, x_2) \rightarrow \tau(P_1(x), \varphi_\exists)\right) \\ &= \forall x_1 \forall x_2 \left(\text{Word}(x_1, x_2) \rightarrow \varphi_\exists^{[x_1, x_2]}\right) \\ &= \forall x_1 \forall x_2 \left(\text{Word}(x_1, x_2) \rightarrow \exists x \left(x_1 \leqslant x \leqslant x_2 \wedge P_1(x)\right)\right)\end{aligned}$$

which clearly defines $L_\forall \bullet L_\exists$.

4.6.2 State Complexity

The state complexity of a (binary) operation O on regular languages is the number of states sufficient and necessary in the worst case for a DFA M to accept $O(L_1, L_2)$ given DFAs M_1 and M_2 for the languages L_1 and L_2; see Yu et al. [1994], Cui et al. [2012]. In our case, this requires knowing how to build an automaton for an iterated language out of automata for the two given languages. To rule out certain edge cases (see Example 4.5 below), we need one helper definition.

Definition 4.4. We will define the *one-step unraveling of* M, denoted M^+. If $\delta(q_0, c) \neq q_0$ for all $c \in \Sigma$, then $\mathsf{M}^+ = \mathsf{M}$. Otherwise:

- $Q^+ = \{*\} \cup Q(\mathsf{M})$
- $q_0^+ = *$
- $F^+ = F(\mathsf{M}) \cup \begin{cases} \{*\} & \text{if } \operatorname{sgn}(\mathsf{M}) = 1 \\ \varnothing & \text{otherwise} \end{cases}$
- $\delta^+ = \delta(\mathsf{M}) \cup \{\langle *, c, q\rangle \mid \langle q_0, c, q\rangle \in \delta(\mathsf{M})\}$

Essentially, if the start state has any loops, we unwind those loops by one step. It's easy to verify that the above definition does not change the language accepted.

Lemma 4.2. $L(\mathsf{M}^+) = L(\mathsf{M})$

We are now in a position to define an automaton for the iteration of two languages. First, we intuitively describe how it works. Let M_1 and M_2 be DFAs for the two languages. By the definition of iteration, we want to replace 1 transitions in M_1 by accepting runs of M_2. We do this as follows: at each state of M_1, we attach a copy of M_2^+ and remove all 0 and 1 transitions from M_1. Now, any time M_1 would have made a 1 transition from q to q', we add a $\boxdot$ transition from all of the accepting states of the q copy of M_2^+ to the start state of the q' copy of M_2^+. This works because reading a $\boxdot$ while in an accepting state of M_2^+ means the previous subword is done and has been accepted, corresponding to a 1 in M_1; the machine then goes to the start state of the q' copy of M_2^+ to begin processing the next subword. Similarly, 0 transitions in M_1 are replaced by $\boxdot$ transitions from rejecting states of the relevant copy of M_2^+. These intuitions are made precise in the following definition, after which some examples are given.

Definition 4.5 (Iteration Automaton)**.** Let M_1 and M_2 be DFAs in alphabet $\{0,1\}$. We define *the iteration of* M_1 *and* M_2, denoted $\mathsf{It}(\mathsf{M}_1, \mathsf{M}_2)$ as follows:

- $\Sigma = \{0, 1, \boxdot\}$
- $Q = Q(\mathsf{M}_1) \times Q\left(\mathsf{M}_2^+\right)$
- $q_0 = \langle q_0(\mathsf{M}_1), q_0\left(\mathsf{M}_2^+\right)\rangle$
- $F = F(\mathsf{M}_1) \times \left\{q_0\left(\mathsf{M}_2^+\right)\right\}$
- Transition function:[25]

$$\delta = \left\{\left\langle\langle q, q_1\rangle, c, \langle q, q_2\rangle\right\rangle \mid q \in Q(\mathsf{M}_1) \text{ and } \langle q_1, c, q_2\rangle \in \delta\left(\mathsf{M}_2^+\right)\right\} \cup \tag{1}$$

$$\left\{\left\langle\langle q_1, q\rangle, \boxdot, \langle q_2, q_0\left(\mathsf{M}_2^+\right)\rangle\right\rangle \mid \langle q_1, \mathrm{sgn}(q, \mathsf{M}_2^+), q_2\rangle \in \delta(\mathsf{M}_1)\right\} \tag{2}$$

[25]This definition can be re-written in more traditional function notation:

$$\delta(\langle q_1, q_2\rangle, c) = \begin{cases} \left\langle q_1, \delta\left(\mathsf{M}_2^+\right)(q_2, c)\right\rangle & c \in \{0,1\} \\ \left\langle \delta(\mathsf{M}_1)\left(q_1, \mathrm{sgn}\left(q, \mathsf{M}_2^+\right)\right), q_0\left(\mathsf{M}_2^+\right)\right\rangle & c = \boxdot \end{cases}$$

Some readers may find this definition easier to comprehend.

Example 4.5. Figure 4.5 shows the minimal automata for $L_\forall$, $L_\exists$, Figure 4.6 shows $\mathsf{It}(\mathsf{M}_\exists, \mathsf{M}_\forall)$, and Figure 4.7 shows $\mathsf{It}(\mathsf{M}_\forall, \mathsf{M}_\exists)$. The states are labelled to enhance readability; the pair (a, b) is abbreviated ab.

Both iteration machines illustrate the need to use one-step unraveling. In the machine $\mathsf{It}(\mathsf{M}_\exists, \mathsf{M}_\forall)$ (Figure 4.6), if $\mathsf{M}_\forall$ were not unraveled, the word $111\boxdot 11$, which does not end in $\boxdot$, would be accepted. Similarly, if $\mathsf{M}_\exists$ were not unraveled in $\mathsf{It}(\mathsf{M}_\forall, \mathsf{M}_\exists)$, then $0^* \subset L(\mathsf{It}(\mathsf{M}_\forall, \mathsf{M}_\exists))$, which we want to prevent.

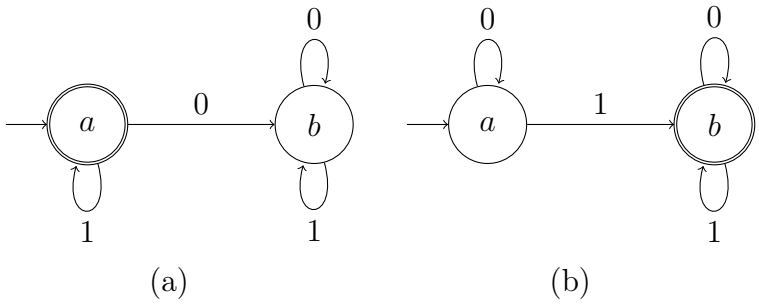


Figure 4.5: The minimal automaton for (a) $L_\forall$, (b) $L_\exists$.

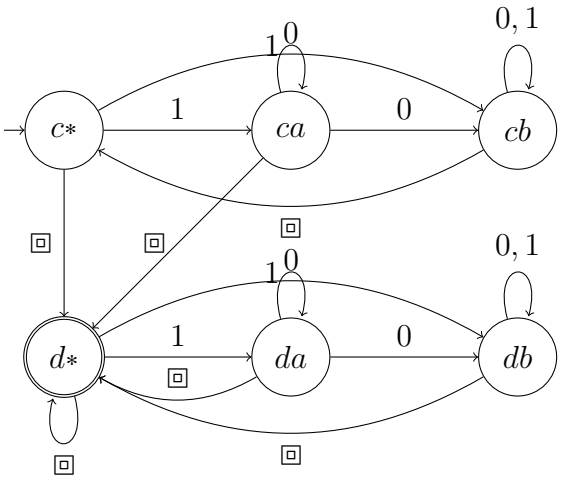


Figure 4.6: The automaton $\mathsf{It}(\min(L_\exists), \min(L_\forall))$.

Fact 4.3. $|\mathsf{It}(\mathsf{M}_1, \mathsf{M}_2)| = |\mathsf{M}_1| \cdot |\mathsf{M}_2^+| \leqslant |\mathsf{M}_1| \cdot |\mathsf{M}_2 + 1|$

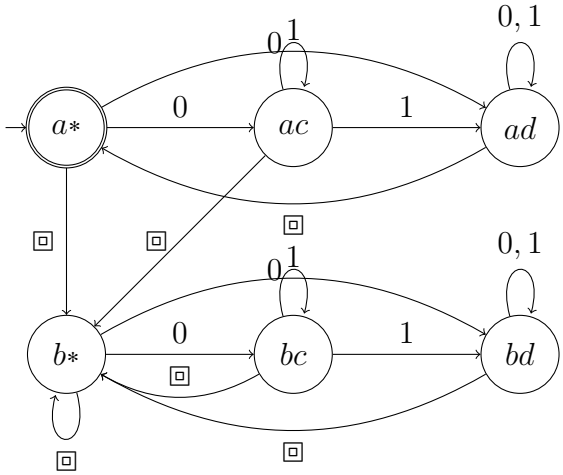


Figure 4.7: The automaton $\mathsf{It}\left(\min\left(L_\forall\right), \min\left(L_\exists\right)\right)$.

Proposition 4.2. $L\left(\mathsf{It}\left(\mathsf{M}_1, \mathsf{M}_2\right)\right) = L\left(\mathsf{M}_1\right) \bullet L\left(\mathsf{M}_2\right)$

Proof. For simplicity, in this proof we write L_1 and L_2 for $L\left(\mathsf{M}_1\right)$ and $L\left(\mathsf{M}_2\right)$ and It for $\mathsf{It}\left(\mathsf{M}_1, \mathsf{M}_2\right)$.

$\supseteq$: Write $L_1^n = \{w \in \{0,1\}^n \mid w \in L_1\}$. We show by induction on n that $L_1^n \bullet L_2 \subseteq L\left(\mathsf{It}\right)$ for all n.

The base case is $n = 0$. We have that L_1^0 is either $\varnothing$ or $\{\varepsilon\}$. The former case is trivial. For the latter case, we have that $L_1^0 = \{\varepsilon\}$ iff $\mathrm{sgn}\left(\mathsf{M}_1\right) = 1$ iff $q_0\left(\mathsf{M}_1\right) \in F\left(\mathsf{M}_1\right)$. By the definition of the iteration automaton, this holds iff $q_0\left(\mathsf{It}\right) \in F\left(\mathsf{It}\right)$ iff $\varepsilon \in L\left(\mathsf{It}\right)$. Since $L_1^0 \bullet L_2 = \{\varepsilon\}$ iff $L_1^0 = \{\varepsilon\}$, this completes the base case.

Now, suppose that $L_1^n \bullet L_2 \subseteq L\left(\mathsf{It}\right)$ and let $w \in L_1^{n+1} \bullet L_2$. This means that there is a word $w' = a_1 \ldots a_{n+1} \in L_1$ and $w_i \in \{0,1\}^*$ such that $w = w_1 \boxdot \ldots w_{n+1} \boxdot$ with $w_i \in L_2$ iff $a_i = 1$.

By clause (2) of the definition of $\delta\left(\mathsf{It}\right)$, the run on $w_1 \boxdot \ldots w_n \boxdot$ ends in a state $\langle q, q_0\left(\mathsf{M}_2^+\right)\rangle$. By our inductive hypothesis, this state is in $F\left(\mathsf{It}\right)$ iff $a_1 \ldots a_n \in L_1$. Now, we have that $w_{n+1} \in L_2$ iff $a_{n+1} = 1$. Moreover, by clause (1) of the definition of $\delta\left(\mathsf{It}\right)$ and Lemma 4.2, reading w_{n+1} will lead to a state $\langle q, q_2\rangle$ such that $\mathrm{sgn}\left(q_2, \mathsf{M}_2^+\right) = 1$ iff $w_{n+1} \in L_2$. In other words, $\mathrm{sgn}\left(q_2, \mathsf{M}_2^+\right) = a_{n+1}$. Since, by assumption, $a_1 \ldots a_{n+1} \in L_1$, we know that $\langle q, a_{n+1}, q_1\rangle \in \delta\left(\mathsf{M}_1\right)$ for some $q_1 \in F\left(\mathsf{M}_1\right)$. Then clause (2) of the

definition of $\delta(\mathsf{It})$ yields a transition $\langle\langle q, q_2\rangle, \boxdot, \langle q_1, q_0\left(\mathsf{M}_2^+\right)\rangle\rangle$. This latter state, by definition, is in $F(\mathsf{It})$, so $w \in L(\mathsf{It})$, as desired.

This completes the induction. Now, $w \in L_1 \bullet L_2$ iff $w \in L_1^n \bullet L_2$ for some n, whence it follows that $L_1 \bullet L_2 \subseteq L(\mathsf{It})$.

$\subseteq$: The definitions of iteration automaton and one-step unraveling ensure that It accepts only words of the form $(w_i\boxdot)^*$ where $w_i \in \{0,1\}^*$. By reasoning similar to the previous direction, we have that

$$\langle\langle q_1, q_0\left(\mathsf{M}_2^+\right)\rangle, w_i\boxdot, \langle q_2, q_0\left(\mathsf{M}_2^+\right)\rangle\rangle \in \delta^*(\mathsf{It}) \Leftrightarrow \langle q_1, \chi_{L_2}(w_i), q_2\rangle \in \delta(\mathsf{M}_1)$$

where δ^* is as defined in Section 4.3.1. Combined with the definition of $F(\mathsf{It})$, this shows that $w \in L(\mathsf{It}) \Rightarrow w \in L_1 \bullet L_2$. □

Inspection of the definition above shows that if M_1 and M_2 are counter-free (and thus accept star-free languages), then so too is the iterated machine. This provides an alternative proof of Proposition 4.1.

The above proposition and fact show that given DFAs with m and n states, there is a DFA with $m \cdot (n+1)$ states which accepts the iteration of the languages of the given DFAs. Moreover, this number of states is necessary in the worst case. The following fact can be verified using standard techniques from automata theory (e.g. by counting the number of equivalence classes of the Myhill-Nerode equivalence relation). It provides an exact characterization of the state complexity of iteration.

Proposition 4.3. *The minimal automaton accepting $L_{/m} \bullet L_{\geqslant n-1}$ has $m \cdot (n+1)$ states, where the minimal automata accepting $L_{/m}$ and $L_{\geqslant n-1}$ have m and n states respectively.*

Theorem 4.7. *Iteration has state complexity $m \cdot (n+1)$.*

4.6.3 Decidability of the Frege Boundary

We now pursue the language-theoretic version of the question whether the Frege Boundary is decidable: given a language $L \subseteq \{0, 1, \boxdot\}^*$, is it decidable whether or

not there are languages L_1 and L_2 such that $L = L_1 \bullet L_2$? We start with the case where the languages in question are regular.

Theorem 4.8 (Decidability)**.** *Let $L \subseteq \{0, 1, \boxdot\}^*$ be a regular language. Then it is decidable whether there are regular languages $L_1, L_2 \subseteq \{0, 1\}^*$ such that $L = L_1 \bullet L_2$.*

The proof works as follows: first, a case distinction is made between L having a certain 'uniformity' property or not. It is shown to be decidable which case obtains. Then, in each case, languages L_1 and L_2 are extracted such that L is an iteration if and only if it is the iteration of L_1 and L_2. All of the methods used for extraction are effective and language equality is decidable, so this suffices to prove the main result.[26]

Proof. Let $L \subseteq \{0, 1, \boxdot\}^*$ be a regular language. Write $S_n := \left(\{0, 1\}^* \boxdot\right)^n$ and $S_* := \left(\{0, 1\}^* \boxdot\right)^*$. There are two cases:

$$\forall n \geqslant 1, w, w' \in S_n,\ w \in L \Leftrightarrow w' \in L \tag{4.1}$$

$$\exists n \geqslant 1, w, w' \in S_n,\ w \in L \text{ and } w' \notin L \tag{4.2}$$

First, we can decide which case holds. Let $h : \{0, 1, \boxdot\} \to \{0, 1, \boxdot\}^*$ be the string homomorphism given by

$$h(0) = h(1) = \varepsilon$$
$$h(\boxdot) = 0\boxdot$$

It is clear that (1) holds iff $L \cap S_* = h^{-1}(L) \cap S_*$. Because automata for intersection and inverse homormosphism can be effectively constructed and language equality is decidable, we can check the right-hand side of this equivalence.

If (1) holds, we proceed as follows. Define the homomorphism $g : \{0, 1\} \to \{0, 1, \boxdot\}^*$ by

$$g(0) = g(1) = 0\boxdot$$

[26] I am thankful to Makoto Kanazawa for suggesting this particular proof.

and let

$$L_1 = g^{-1}(L)$$
$$L_2 = \varnothing$$

It is clear that L_1 and L_2 are regular and that $L \cap S_* = L_1 \bullet L_2$. Thus, L is an iteration of two regular languages iff $L = L \cap S_*$. This can easily be decided.

If (2) holds, there are words $w_1, \ldots, w_n, w'_1, \ldots, w'_n \in \{0,1\}^*$ such that

$$w = w_1 \boxdot \cdots w_n \boxdot$$
$$w' = w'_1 \boxdot \cdots w'_n \boxdot$$

with $w \in L$ and $w' \notin L$. Then there must be an $i \in \{1, \ldots, n\}$ s.t.

$$\begin{aligned} w_1 \boxdot \cdots w_{i-1} \boxdot w_i \boxdot w'_{i+1} \boxdot \cdots w'_n \boxdot \in L \\ w_1 \boxdot \cdots w_{i-1} \boxdot w'_i \boxdot w'_{i+1} \boxdot \cdots w'_n \boxdot \notin L \end{aligned} \tag{4.3}$$

Let $f : \{0,1\} \to \{0,1,\boxdot\}^*$ be the string homomorphism

$$f(0) = w'_i \boxdot$$
$$f(1) = w_i \boxdot$$

and

$$L_1 = f^{-1}(L)$$
$$L_2 = \left\{ v \in \{0,1\}^* \mid w_1 \boxdot \cdots w_{i-1} \boxdot v \boxdot w'_{i+1} \boxdot \cdots w'_n \boxdot \in L \right\}$$

It is clear that L_1 and L_2 are regular.

Now, L is an iteration $L'_1 \bullet L'_2$ iff $L = L_1 \bullet L_2$. To see this, suppose that $L = L'_1 \bullet L'_2$. The condition (4.3) implies that $\chi_{L'_2}(w_i) \neq \chi_{L'_2}(w'_i)$ and that

$$w_1 \boxdot \cdots w_{i-1} \boxdot v \boxdot w'_{i+1} \boxdot \cdots w'_n \boxdot \in L \text{ iff } \chi_{L'_2}(v) = \chi_{L'_2}(w_i)$$

Here again there are two cases. (i) $\chi_{L_2'}(w_i) = 1$ and $\chi_{L_2'}(w_i') = 0$. Then $L_2' = L_2$ and $L_1' = L_1$. (ii) $\chi_{L_2'}(w_i) = 0$ and $\chi_{L_2'}(w_i') = 1$. Then $L_2' = \{0,1\}^* \setminus L_2$ and $L_1' = k(L_1)$ where k is the homomorphism $k(0) = 1$ and $k(1) = 0$. It's easy to see that $L_1' \bullet L_2' = L_1 \bullet L_2$.

All that remains is to show that automata for L_1 and L_2 can be found effectively. There must be $w_1, \ldots, w_n, w_1', \ldots, w_n' \in \{0,1\}^*$ satisfying (4.3). We can find these by brute-force search (since membership in L is decidable). With these strings in hand, automata for L_1 and L_2 can be obtained by the usual constructions. □

4.7 Beyond Regular

We now look at the closure and decidability questions for classes of languages beyond the regular languages. First, we look at the context-free languages. Because these lack many of the nice properties of the regular languages, most of the results do not carry over. Therefore, we then turn to a natural and more well-behaved sub-class: the *deterministic* context-free languages. We will prove that this class of languages is closed under iteration and that it is decidable whether a given such language is an iteration.

4.7.1 Context-Free Languages

The closure under iteration result does not immediately extend to context-free languages because these languages are not closed under complement. But certain sub-classes are. For instance, the permutation-closed languages on a two-letter alphabet are:

Theorem 4.9 (van Benthem [1986])**.** *The permutation-closed context-free languages on a two-letter alphabet are closed under complement.*

We cannot conclude from this, however, that these languages are closed under iteration since, in general, an iterated language will not be closed under permutations. Nevertheless, we get a 'semi-closure' result:[27]

[27] Theorem 6 of Steinert-Threlkeld and Icard III. [2013] states this result as a full closure result

Proposition 4.4. *If L_1 and L_2 are permutation-closed context-free languages in a two-letter alphabet, then $L_1 \bullet L_2$ is context-free.*

Proof. Immediate from Theorem 4.9 and the closure of context-free languages under substitution. □

The main obstacle to using the strategy in Section 4.6.3 to try and prove the decidability of iteration for context-free languages is that language equality is undecidable. Several steps in the proof of Theorem 4.8 depended on the fact that language equality is decidable for regular languages. This leads us to conjecture:

Conjecture 4.1. It is undecidable whether a given context-free language $L \subseteq \{0, 1, \boxdot\}^*$ is an iteration of two context-free languages in alphabet $\{0, 1\}$.

4.7.2 Deterministic Context-Free Languages

Some hope, however, comes from finding a large subclass of the context-free languages for which equality is decidable. One such class is the *deterministic* context-free languages; these are the languages accepted by deterministic pushdown automata, which are those having at most one choice at every machine configuration in a sense to be made precise below. It is famously known that language equality is decidable for deterministic context-free langauges (Sénizergues [1997, 2001, 2002]). In this section, we will introduce these languages and then extend both the closure and decidability results to them.

A *deterministic pushdown automaton* (DPDA) is a PDA such that:

- for every $q \in Q$ and $b \in \Gamma$, if $\delta(q, \varepsilon, b) \neq \varnothing$, then $\delta(q, a, b) = \varnothing$ for every $a \in \Sigma$
- for every $q \in Q$, $a \in \Sigma \cup \{\varepsilon\}$, and $b \in \Gamma$, $|\delta(q, a, b)| \leqslant 1$.

for context-free languages. This is because in that paper, the authors assumed – in addition to conservativity and extension – the isomorphism closure of quantifiers, which has the result that all of the languages discussed were permutation-closed. While this assumption is often made, natural language determiners like 'the first five' and 'every other' seem to be counterexamples.

These two constraints have the effect that at most one transition can occur at any instantaneous description of the DPDA.[28] A language L is deterministic context-free iff $L = L(\mathsf{M})$ for some DPDA M. Note that the PDA in Figure 4.4 is in fact a DPDA, so L_{MOST} is in fact deterministic context-free. Another example is given below. For a complete characterization of the generalized quantifiers with languages accepted by DPDAs, see Kanazawa [2013].

Example 4.6. The language $L_{\geqslant 1/3} = \left\{w \in \{0,1\}^* : \#_1(w) \geqslant \frac{1}{3}(\#_1(w) + \#_0(w))\right\}$, which can be used to verify the truth of sentences like (66), is deterministic context-free.

(66) At least one third of the students are happy.

A DPDA accepting the language is given in Figure 4.8. The states are labeled for future reference.

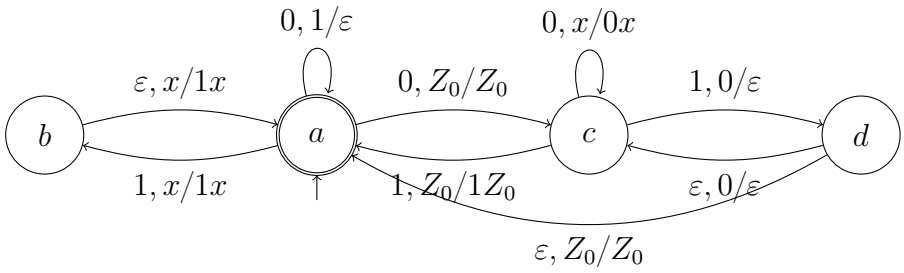


Figure 4.8: A DPDA accepting $L_{\geqslant 1/3}$.

4.7.3 Closure for DCFLs

Importantly, the deterministic context-free languages are in fact closed under complement; see, e.g., Theorem 10.1 on p. 238 of Hopcroft and Ullman [1979]. Unfortunately, however, closure under iteration does not follow immediately since these languages

[28] As presently defined, a DPDA may not read the entire input. However, for every DPDA M, there is another DPDA M' that does read the entire input such that $L(\mathsf{M}) = L(\mathsf{M}')$. See Lemma 10.3 (p. 236) of Hopcroft and Ullman [1979]. Because of this, in what follows we will assume that all DPDAs read the entire input.

are not closed under substitution. However, as in the star-free case, we can use the separator ⧈ to prove closure under iteration. We first offer a proof of closure under complement since some intermediate steps will be useful in defining a DPDA for iteration.

As a first step, we will define a DPDA that keeps track of whether or not a given DPDA has entered a final state since reading the last input symbol. This helps in distinguishing acceptance from rejection because a DPDA may make ε-transitions through both accepting and rejecting states after reading an input symbol. The states of the new DPDA will have a second component: 0, 1, or 2. A 0 will mean that the machine has not entered an accepting state since the last input, while a 1 means that it has entered an accepting state.

Definition 4.6. Given a DPDA M, define the *tracking extension of* M – denoted M^T – as follows:

- $Q = Q(\mathsf{M}) \times \{0, 1, 2\}$
- $\Sigma = \Sigma(\mathsf{M})$
- $\Gamma = \Gamma(\mathsf{M})$
- Transition function:

$$\delta = \left\{ \begin{array}{c} \langle\langle q, 1\rangle, \varepsilon, b, \langle q', 1\rangle, \gamma\rangle \\ \langle\langle q, 0\rangle, \varepsilon, b, \langle q', \operatorname{sgn}(q', \mathsf{M})\rangle, \gamma\rangle \end{array} : \langle q, \varepsilon, b, q', \gamma\rangle \in \delta(\mathsf{M}) \right\} \cup$$
$$\left\{ \begin{array}{l} \langle\langle q, 0\rangle, \varepsilon, b, \langle q, 2\rangle, b\rangle \\ \langle\langle q, 1\rangle, a, b, \langle q', \operatorname{sgn}(q', \mathsf{M})\rangle, \gamma\rangle : \langle q, a, b, q', \gamma\rangle \in \delta(\mathsf{M}) \text{ for } a \in \Sigma \\ \langle\langle q, 2\rangle, a, b, \langle q', \operatorname{sgn}(q', \mathsf{M})\rangle, \gamma\rangle \end{array} \right\}$$

- $q_0 = \langle q_0, \operatorname{sgn}(\mathsf{M})\rangle$
- $F = \{\langle q, 1\rangle : q \in Q(\mathsf{M})\}$

Lemma 4.3. $L\left(\mathsf{M}^T\right) = L(\mathsf{M})$

Theorem 4.10. *The deterministic context-free languages are closed under complement.*

Proof. Given L, let M be a DPDA such that $L = L(\mathsf{M})$. Consider the automaton M^c constructed as follows: M^c is just like M^T except that

$$F(\mathsf{M}^c) = \{\langle q, 2\rangle : q \in Q(\mathsf{M})\}$$

Clearly, M^c is a DPDA. The accepting states of M^c are those ending in a 2. Note that these are only entered from states ending in a 0 by an ε-transition. But a state ending in a 0 encodes that the machine has not entered an accepting state since reading the last input. The reader can use these ideas to show that $L(\mathsf{M}^c) = L^c$ or consult p. 239 of Hopcroft and Ullman [1979] for details. □

This proof shows that membership in L versus L^c is fully encoded by the second component of states of M^T. This allows us to prove the following proposition – a light generalization of Lemma 6.1.1 of McWhirter [2014] – which will be crucial in proving closure under iteration.

Proposition 4.5. *Given a DPDA* M*, there is another DPDA* $M^{\boxdot}$ *with* $\Sigma(\mathsf{M}^{\boxdot}) = \Sigma(\mathsf{M}) \cup \{\boxdot\}$ *and unique states* q_{acc} *and* q_{rej} *such that for any* γ *with* $\gamma_1 \notin \Gamma(\mathsf{M})$*:*

- $\langle q_0, w\boxdot, \gamma\rangle \vdash^* \langle q_{acc}, \varepsilon, \gamma\rangle$ *iff* $w \in L(\mathsf{M})$
- $\langle q_0, w\boxdot, \gamma\rangle \vdash^* \langle q_{rej}, \varepsilon, \gamma\rangle$ *iff* $w \notin L(\mathsf{M})$

Proof. $\mathsf{M}^{\boxdot}$ will be just like M^T but with two new states q_{acc} and q_{rej} and the following modification to the transition function:

$$\begin{aligned}\delta\left(\mathsf{M}^{\boxdot}\right) =& \delta\left(\mathsf{M}^T\right) \\ &\cup \left\{\langle q, \boxdot, b, q_{\text{acc}}, b\rangle : q \in F\left(\mathsf{M}^T\right) \text{ and } b \in \Gamma(\mathsf{M})\right\} \\ &\cup \{\langle q, \boxdot, b, q_{\text{rej}}, b\rangle : q \in F(\mathsf{M}^c) \text{ and } b \in \Gamma(\mathsf{M})\} \\ &\cup \{\langle q_i, \varepsilon, b, q_i, \varepsilon\rangle : i \in \{\text{acc}, \text{rej}\} \text{ and } b \in \Gamma(\mathsf{M})\}\end{aligned}$$

In other words: upon reading ⧈, the machine transitions to q_{acc} or q_{rej} according to whether the machine has entered an accepting state since the last input or not and then empties the stack of all contents from $\Gamma(\mathsf{M})$. This clearly implements the desired behavior. □

We are now in position to define a DPDA for the iteration of two others. As before, we want to replace 1 transitions from M_1 with accepting runs from M_2 (and *mutatis mutandis* for 0 transitions and rejecting runs). By Proposition 4.5, we know that an accepting run of M_2 corresponds to a run ending in q_{acc} without altering the stack in $\mathsf{M}_2^{⧈}$. So we do the following: we have a copy both of M_1 and of $\mathsf{M}_2^{⧈}$. All ε-transitions in M_1 are left intact. When M_1 is about to read a 1 in state q, it does the following: it takes an ε-transition to the start state of M_2 while pushing a q on to the stack. This has two effects: the q both 'insulates' M_1's use of the stack from $\mathsf{M}_2^{⧈}$'s and also tells the latter machine which state M_1 was previously in. Now, when $\mathsf{M}_2^{⧈}$ gets to q_{acc}, it sees the q on the stack and transitions to a new state named a_q while popping the q. This has the effect of 'lifting the lid' off of M_1's stack. The state a_q then makes an ε-transition to the state that M_1 would have transitioned on reading a 1 in q while also making the appropriate stack manipulations. The 0 transitions are treated similarly, but with q_{rej} and other new states r_q. The following definition and lemma make this all precise. An example demonstrating the construction is given at the end of the section.

Definition 4.7 (Iteration of DPDAs)**.** Let M_1 and M_2 be DPDAs in alphabet $\{0,1\}$. We define *the iteration of* M_1 *and* M_2, denoted $\mathsf{It}(\mathsf{M}_1, \mathsf{M}_2)$ as follows:

- $Q = Q(\mathsf{M}_1) \cup Q\left(\mathsf{M}_2^{⧈}\right) \cup \{a_q, r_q : q \in Q(\mathsf{M}_1)\}$
- $\Sigma = \{0, 1, ⧈\}$
- $\Gamma = \Gamma(\mathsf{M}_1) \cup \Gamma(\mathsf{M}_2) \cup Q(\mathsf{M}_1)$

- Transition function:

$$\begin{aligned}\delta = {} & \delta\left(\mathsf{M}_2\right) \cup \left\{\left\langle q, \varepsilon, b, q', \gamma\right\rangle : \left\langle q, \varepsilon, b, q', \gamma\right\rangle \in \delta\left(\mathsf{M}_1\right)\right\} \\ & \cup \left\{\left\langle q, \varepsilon, b, q_0\left(\mathsf{M}_2\right), qb\right\rangle : \left\langle q, a, b, q', \gamma\right\rangle \in \delta\left(\mathsf{M}_1\right) \text{ for some } a \in \{0,1\}\right\} \\ & \cup \left\{ \begin{matrix} \left\langle q_{\text{acc}}, \varepsilon, q, a_q, \varepsilon\right\rangle \\ \left\langle a_q, \varepsilon, b, q', \gamma\right\rangle \end{matrix} : \left\langle q, 1, b, q', \gamma\right\rangle \in \delta\left(\mathsf{M}_1\right) \right\} \\ & \cup \left\{ \begin{matrix} \left\langle q_{\text{rej}}, \varepsilon, q, r_q, \varepsilon\right\rangle \\ \left\langle r_q, \varepsilon, b, q', \gamma\right\rangle \end{matrix} : \left\langle q, 0, b, q', \gamma\right\rangle \in \delta\left(\mathsf{M}_1\right) \right\}\end{aligned}$$

- $q_0 = q_0\left(\mathsf{M}_1\right)$
- $F = F\left(\mathsf{M}_1\right)$

When context allows, $\mathsf{It}\left(\mathsf{M}_1, \mathsf{M}_2\right)$ will be abbreviated as It.

Lemma 4.4. Let M_1 and M_2 be as in Definition 4.7.

(1) If $\left\langle q, 1w, x\beta\right\rangle \vdash_{\mathsf{M}_1} \left\langle q', w, \gamma\beta\right\rangle$, then

$$\left\langle q, w \boxdot w', x\beta\right\rangle \vdash_{\mathsf{It}}^{*} \left\langle q', w', \gamma\beta\right\rangle \text{ iff } w \in L\left(\mathsf{M}_2\right)$$

(2) If $\left\langle q, 0w, x\beta\right\rangle \vdash_{\mathsf{M}_1} \left\langle q', w, \gamma\beta\right\rangle$, then

$$\left\langle q, w \boxdot w', x\beta\right\rangle \vdash_{\mathsf{It}}^{*} \left\langle q', w', \gamma\beta\right\rangle \text{ iff } w \notin L\left(\mathsf{M}_2\right)$$

Proof. We only show (1), since (2) is completely analogous. Suppose $\left\langle q, 1w, x\beta\right\rangle \vdash_{\mathsf{M}_1} \left\langle q', w, \gamma\beta\right\rangle$, i.e. $\left\langle q, a, x, q', \gamma\right\rangle \in \delta\left(\mathsf{M}_1\right)$. First, in a configuration $\left\langle q, w \boxdot w', x\beta\right\rangle$, It will take an ε-transition to $q_0\left(\mathsf{M}_2^{\boxdot}\right)$ while pushing a q on to the stack. By Proposition 4.5,

$$\left\langle q_0\left(\mathsf{M}_2^{\boxdot}\right), w \boxdot w', qx\beta\right\rangle \vdash_{\mathsf{It}}^{*} \left\langle q_{\text{acc}}, w', qx\beta\right\rangle$$

iff $w \in L\left(\mathsf{M}_2\right)$. Then, from q_{acc}, It will make an ε-transition to a_q while popping the q off the top of the stack. It will then make an ε-transition to q' while replacing x by γ on top of the stack. □

Theorem 4.11. $\mathsf{It}(\mathsf{M}_1, \mathsf{M}_2)$ *is a DPDA with* $L(\mathsf{It}(\mathsf{M}_1, \mathsf{M}_2)) = L(\mathsf{M}_1) \bullet L(\mathsf{M}_2)$. *Thus, the deterministic context-free languages are closed under iteration.*[29]

Proof. That It is a DPDA can be seen from inspection of Definition 4.7. ε-transitions in M_1 are left in-tact, while 0 and 1 transitions are replaced by ε-transitions to $q_0(\mathsf{M}_2)$, which preserves determinism. Transitions into and out of the new states a_q and r_q are also deterministic.

That $L(\mathsf{It}) = L(\mathsf{M}_1) \bullet L(\mathsf{M}_2)$ follows from repeated application of Lemma 4.4 and the observation that the only paths to accepting states which are not $q_0(\mathsf{It})$ must process an input of the form $w\boxdot$ for $w \in \{0,1\}^*$. □

Example 4.7. Figure 4.9 depicts the automaton $\mathsf{It}\left(\mathsf{M}_{\text{most}}, \mathsf{M}_{\geqslant 1/3}\right)$. Note that the DPDA could be pruned: the states $d1$, $b0$, $a0$, $b2$, $a2$, and $d2$ are not reachable. They have been included to illustrate Definition 4.7 exactly.

4.7.4 Decidability for DCFLs

Theorem 4.12. *It is decidable whether a given deterministic context-free language* $L \subseteq \{0, 1, \boxdot\}^*$ *is an iteration.*

Proof. As noted earlier, language equality for DCFLs is decidable. Moreover, the DCFLs are effectively closed under inverse homomorphism and intersection with a regular language. Because S_*, as defined in Theorem 4.8, is regular, the proof of Theorem 4.8 carries over to the DCFL case unchanged. □

4.8 Conclusion

This chapter has covered a lot of ground, which I will briefly summarize. First, after presenting an introduction to and overview of generalized quantifier theory and the semantic automata framework, it suggested that semantic automata can be seen as providing "level 1.5" explanations, capturing features of semantic competence related

[29] McWhirter [2014] has obtained this result independently. My exposition of this result has improved greatly from reading her work.

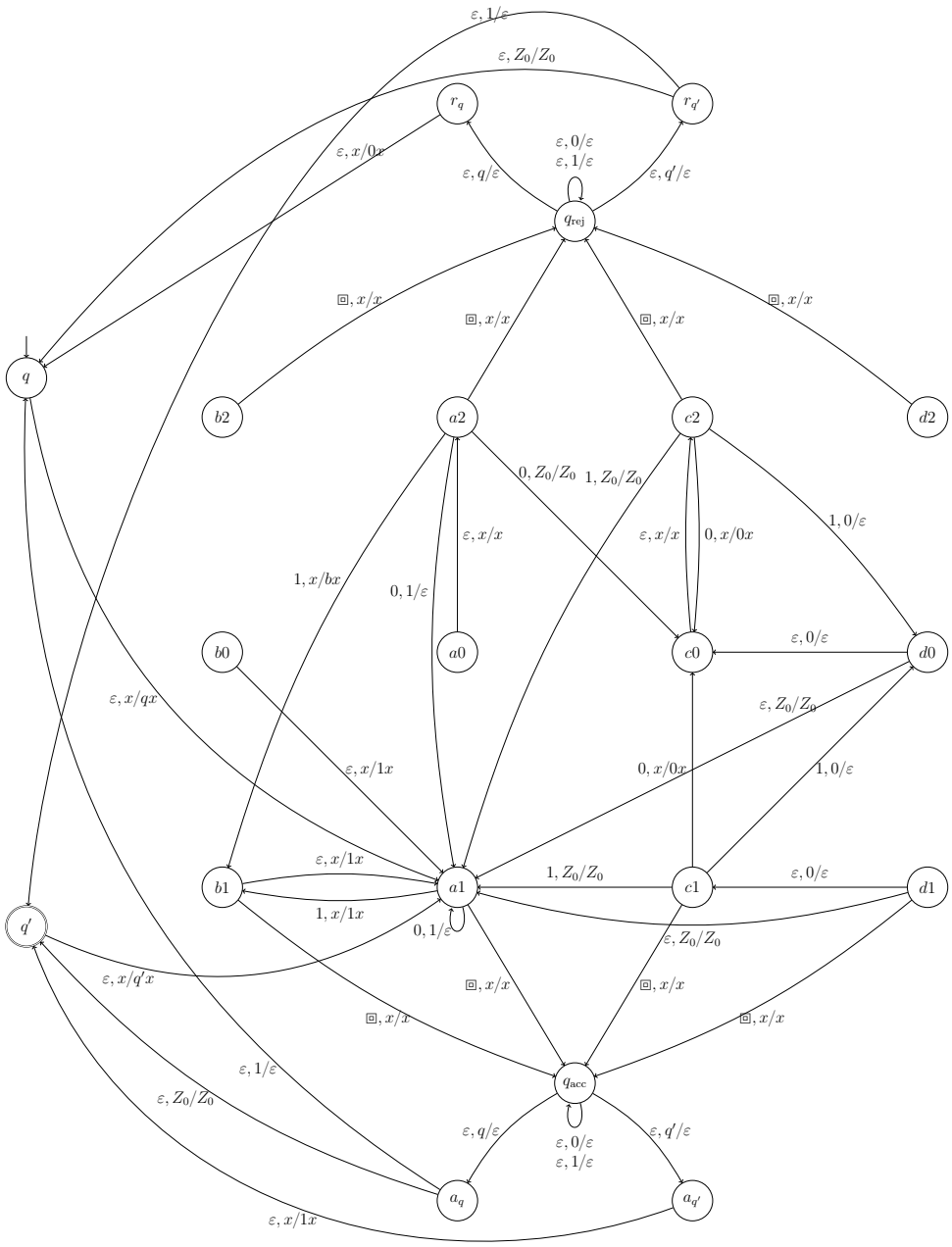


Figure 4.9: The DPDA $\mathsf{It}\left(\mathsf{M}_{\text{most}}, \mathsf{M}_{\geqslant 1/3}\right)$.

to control structure and resource demands. Next, the first extension of the theoretical framework was made. Namely, formal languages for *iterations* of quantifiers were defined, expanding the domain of applicability of the framework. This answers one version of the question of what compositionality amounts to from an algorithmic perspective.

We then showed that the star-free languages, the regular languages, and the deterministic context-free languages are closed under this operation, showing that iteration – a form of composition – alone does not introduce a fundamentally new degree of complexity in this framework, at least when complexity is measured by levels in the Chomsky hierarchy. Thus, we have partially answered the question of whether composition itself increases the computational complexity of certain linguistic tasks by demonstrating a case in which it does not.

We also used the semantic automata framework to address a question in pure generalized quantifier theory: is the Frege Boundary decidable? For two important special cases – when the property of binary relations has a corresponding regular or deterministic context-free language – we answered in the affirmative. This shows that considering compositionality from an algorithmic perspective also has the power to shed light on more 'traditional' questions that arise in the course of compositional semantic theorizing.

Many avenues for future research remain open, both on the theoretical and the experimental side. On the theoretical side, a first goal would be to tackle Conjecture 4.1: that the Frege Boundary is undecidable for context-free languages. One would also like to provide a characterization of the Frege Boundary in purely language- or automata-theoretic terms. A more natural setting for that pursuit may be the so-called 'picture languages', which are binary matrices (instead of words) of alphabet symbols.[30] Another avenue concerns extending the procedural perspective on meaning embodied in the automata framework to other domains of natural language besides quantifiers. As an example, the superlative morpheme *-est* has a natural procedural interpretation: search the relevant set of entitites, holding in memory the tallest yet

[30]See Giammarresi and Restivo [1997] for details.

seen, until all have been scanned.[31]

On the experimental side, one would like to generate concrete empirical predictions from semantic automata for iterated quantifiers. Szymanik et al. [2013] present a preliminary study in which it's shown that It (Some, Every) uses more working memory and cognitive control than It (Every, Some). Stronger bridge principles from iterated automata to performance in sentence verification need to be made in order to generate this prediction. Most generally, we hope to pursue the development of a uniform measure of complexity that cuts across the finite-state/pushdown automata distinction but that correlates with memory and other cognitive demands in sentence verification.

A final avenue that bridges both of these sides concerns the encoding function $\tau(\cdot)$ which is used to generate formal languages from quantifier meanings (classes of models). Following the literature, we have studied a particular – admittedly natural – choice for this encoding. From the psychological perspective, however, this represents how a visual scene is encoded by a language user. Moreover, on which side of the regular/context-free language divide a quantifier falls can depend on this encoding. Steinert-Threlkeld et al. [2015] develop this thought more thoroughly and provide initial experimental support for the dependence of memory demands in verification on the method of encoding.[32] A full exploration of the theoretical and experimental consequences of different encodings promises to be fruitful.

[31] See van Benthem [1986] and Suppes [1980, 1982] for some suggestive remarks. Note that van Benthem [1987], §7, endorses the pursuit of procedural semantics primarily for *functional* vocabulary like quantifiers and the superlative morpheme, while Suppes [1980] wants to extend the approach to include content words like adjectives and nouns. Although an argument lies beyond the present scope, I am inclined to endorse van Benthem's restriction because I believe it avoids objections of an anti-verificationist style.

[32] See §8.2 of van Benthem [1987] for other motivation.

Chapter 5

Conclusion

This dissertation has not advanced one thesis, but rather has attempted to show the fruitfulness of a change in perspective on the topic of compositionality. In particular, in the Introduction I suggested that moving from viewing compositionality as a property of symbolic systems, which naturally leads to asking status questions, to viewing it as a procedural ability at play in concrete linguistic episodes allows new and exciting questions to be asked. Moreover, providing answers to these questions has required a wide variety of different methodologies. Having a larger toolbox with which to address compositionality can only be a positive development.

The new questions asked occur at successively more local levels of scale. At the level of our species, Chapter 2 asked why human languages, but not other animal communication systems, are compositional. This question was addressed by enriching the signaling game framework – which had already been fruitfully applied to the evolution of lexical meaning – with simple forms of compositional signaling. In particular, I compared the ease of learning non-compositional and compositional signaling systems using a form of reinforcement learning known to be implemented by dopamine pathways in primate brains. The results showed that compositional languages are easier to learn, but only when the agents have to talk about many different things. This simulation work thus identifies one possible evolutionary pressure which may have caused language to become compositional: the need to talk about many different topics.

At the level of conversational dynamics, Chapter 3 asked what role compositionality plays in the theory of communication. In particular, the fact that assertoric contents do not compose like semantic values do and the desire to treat certain bits of language as (merely) expressive have caused a gap to grow between compositional semantics and the theory of assertion. In the chapter, I identified a new form of expressivism – pragmatic expressivism – and argued for a new principle – the Non-Disjunctive Expressive Principle – which shows how compositional semantics should fit with an expressive theory of communication. Moreover, I showed that this Principle has real consequences for compositional semantics. The most well-developed pragmatic expressivism – Seth Yalcin's about epistemic modals – systematically violates the Principle. I then developed a new theory based on assertability and deniability which provably satisfies it. This dialectic thus shows that constraints from how compositional semantics must fit into the theory of assertion can have real consequences for semantic theory itself.

Finally, at the level of an individual language user, Chapter 4 asked about the algorithmic meaning of compositionality and whether it increases the computational complexity of certain linguistic tasks. These questions were asked in the context of quantifier sentence verification, where the semantic automata model had previously been developed and shown to capture important properties of the cognitive demands of the task. Providing definitions for languages and automata for sentences with iterated quantifiers (where quantifiers are both subject and object of a transitive verb) provides one algorithmic implementation of a semantic composition operation. To address the question about whether composition itself increases complexity, it was shown that several important classes of languages (star-free, regular, deterministic context-free) are in fact closed under the operation of iteration. This demonstrates that – using position in the Chomsky hierarchy of languages as a measure of complexity – composition itself need not increase the complexity of the verification task. Moreover, we were also able to use automata for iteration of quantifiers to answer the question of whether the Frege Boundary – which demarcates iterated from non-iterated quantifiers – is in fact decidable. This represents an application of the algorithmic perspective on compositionality to a question arising at the purely truth-conditional

level.

In some ways, then, the proof has been in the pudding: the change in perspective about compositionality has allowed new questions to be asked and many lessons to be drawn from the attempts to answer them. Moreover, along the way, I identified much more work to tackle on each of the questions. And yet more new questions about compositionality are sure to arise. To that end, this dissertation can also stand as a rallying call, urging others to start pursuing questions of their own about compositionality from a procedural perspective.

Appendix A

Proof of Non-Disjunctive Property Expression Theorem

We prove Theorem 3.1, restated below.

Theorem. *For every sentence φ in the language, there is an attitudinal state type description δ_φ – a conjunction of beliefs and abeliefs – such that $\boldsymbol{b}_{A,w} \Vdash \varphi$ if and only if $w \in [\delta_\varphi]$.*

We start with some important equivalence facts. Recall that two sentences φ and ψ are assertorically equivalent just in case they are assertable relative to the exact same information states.

Proposition A.1. *(1) $\Diamond\big(\psi \wedge \Diamond\varphi_1 \wedge \ldots \wedge \Diamond\varphi_n\big)$ is assertorically equivalent to $\Diamond\big(\psi \wedge \varphi_1 \wedge \ldots \wedge \varphi_n\big)$*

(2) $\neg\,(\psi \wedge \Diamond\varphi_1 \wedge \ldots \wedge \Diamond\varphi_n)$ is assertorically equivalent to $\neg\,(\psi \wedge \varphi_1 \wedge \ldots \wedge \varphi_n)$

(3) If ψ^1 and ψ^2 are both $\Diamond$-free, then:

$$\big(\psi^1 \wedge \Diamond\varphi^1_1 \wedge \ldots \wedge \Diamond\varphi^1_m\big) \vee \big(\psi^2 \wedge \Diamond\varphi^2_1 \wedge \ldots \wedge \Diamond\varphi^2_n\big)$$

is assertorically equivalent to

$$\big(\psi^1 \vee \psi^2\big) \wedge \Diamond\left(\psi^1 \wedge \varphi^1_1\right) \wedge \ldots \wedge \Diamond\left(\psi^1 \wedge \varphi^1_m\right) \wedge \Diamond\left(\psi^2 \wedge \varphi^2_1\right) \wedge \ldots \wedge \Diamond\left(\psi^2 \wedge \varphi^2_n\right)$$

Proof. In each case, let $\mathbf{s}$ be an arbitrary information state.

(1) $\mathbf{s} \Vdash \Diamond(\psi \wedge \Diamond\varphi_1 \wedge \cdots \wedge \Diamond\psi_n)$ iff $\{w\} \Vdash \psi \wedge \Diamond\varphi_1 \wedge \cdots \wedge \Diamond\psi_n$ for some $w \in \mathbf{s}$. Because $\{w\} \Vdash \Diamond\varphi$ iff $\{w\} \Vdash \varphi$ for any w, it follows that $\{w\} \Vdash \psi \wedge \varphi_1 \wedge \cdots \wedge \varphi_n$, as desired.

(2) $\mathbf{s} \Vdash \neg(\psi \wedge \Diamond\varphi_1 \wedge \cdots \wedge \Diamond\varphi_n)$ iff the conjunction is deniable at $\mathbf{s}$, which holds iff there are $s_0, \cdots, s_n$ such that $\mathbf{s} = \bigcup_{0 \leqslant i \leqslant n} \mathbf{s}_i$ such that $\mathbf{s}_0 \dashv\Vdash \psi$ and $\mathbf{s}_i \dashv\Vdash \Diamond\varphi_i$. This last conjunct holds iff $\mathbf{s}_i \dashv\Vdash \varphi_i$, from whence it then follows that $\mathbf{s} \dashv\Vdash \psi \wedge \varphi_1 \wedge \cdots \wedge \varphi_n$.

(3) Let ψ^1, ψ^2 be $\Diamond$-free. $\mathbf{s} \Vdash (\psi^1 \wedge \Diamond\varphi^1_1 \wedge \ldots \wedge \Diamond\varphi^1_m) \vee (\psi^2 \wedge \Diamond\varphi^2_1 \wedge \ldots \wedge \Diamond\varphi^2_n)$ iff there are $\mathbf{s}_1, \mathbf{s}_2$ such that $\mathbf{s} = \mathbf{s}_1 \cup \mathbf{s}_2$, $\mathbf{s}_1 \Vdash \psi^1$, $\mathbf{s}_1 \Vdash \Diamond\varphi^1_i$, $\mathbf{s}_2 \Vdash \psi^2$, and $\mathbf{s}_2 \Vdash \Diamond\varphi^2_j$. Using the fact (40) that assertability for $\Diamond$-free formulas amounts to assertability at all singletons, we have that $\mathbf{s}_1 \Vdash \Diamond(\psi^1 \wedge \varphi^1_i)$ and $\mathbf{s}_2 \Vdash \Diamond\left(\psi^2 \wedge \varphi^2_j\right)$. These, in turn hold iff the corresponding formula are assertable at $\mathbf{s}$ itself. We thus have that $\mathbf{s} \Vdash \psi^1 \vee \psi^2$ and that $\mathbf{s} \Vdash \Diamond(\psi^1 \wedge \varphi^1_i)$ and $\mathbf{s} \Vdash \Diamond\left(\psi^2 \wedge \varphi^2_j\right)$, as needed. □

In the following, we will make use of an equivalence fact from the main chapter, reproduced here:

(41) $B_A\Diamond\varphi$ is assertorically equivalent to $\neg B_A\neg\varphi$

Lemma A.1. Let φ be a sentence in the assertability language. Then there exist sentences β, α_1, …, α_n (for some $n \geqslant 0$) such that:

- β, α_1, …, α_n contain no occurrences of $\Diamond$,
- φ is assertorically equivalent to $\beta \wedge \Diamond\alpha_1 \wedge \ldots \wedge \Diamond\alpha_n$

Proof. By induction on the complexity of formulae. The non-trivial cases (taking the assumption that φ is assertorically equivalent to $\beta \wedge \Diamond\alpha_1 \wedge \ldots \wedge \Diamond\alpha_n$ as the induction hypothesis):

- $\neg\varphi$: using Fact (2) of Proposition A.1, we conclude that $\neg\varphi$ is assertorically equivalent to $\neg(\beta \wedge \alpha_1 \wedge \ldots \wedge \alpha_n)$.

- $\varphi_1 \vee \varphi_2$: assume that φ_1 is assertorically equivalent to $\beta^1 \wedge \Diamond\alpha^1_1 \wedge \ldots \wedge \Diamond\alpha^1_m$, and that φ_2 is assertorically equivalent to $\beta^2 \wedge \Diamond\alpha^2_1 \wedge \ldots \wedge \Diamond\alpha^2_n$. Now use Fact (3) of Proposition A.1.

- $\Diamond\varphi$: Fact (1) of Proposition A.1 shows that this is assertorically equivalent to $\Diamond(\beta \wedge \alpha_1 \wedge \ldots \wedge \alpha_n)$.

- $B_A\varphi$: first, note the following: $\mathbf{s} \Vdash B_A\varphi$ iff $\mathbf{b}_{A,w} \Vdash \varphi$ for every $w \in \mathbf{s}$, iff (by the inductive hypothesis) $\mathbf{b}_{A,w} \Vdash \beta \wedge \Diamond\alpha_1 \wedge \cdots \wedge \Diamond\alpha_n$ for every $\mathbf{w} \in \mathbf{s}$ iff $\mathbf{s} \Vdash B(\beta \wedge \Diamond\alpha_1 \wedge \cdots \wedge \Diamond\alpha_n)$.

  Second, note that it is easy to show that assertability is preserved when B is distributed over a conjunction. Using fact (41), this yields that $\mathbf{s} \Vdash B(\beta \wedge \Diamond\alpha_1 \wedge \cdots \wedge \Diamond\alpha_n)$ just in case $\mathbf{s} \Vdash B\beta \wedge \neg B\neg\alpha_1 \wedge \cdots \wedge \neg B\neg\alpha_n$. □

The following definition of a sentence expressing an attitudinal state description was given.

(43) φ expresses attitudinal state description δ just in case: φ is assertable relative to an agent's doxastic state (i.e. $\mathbf{b}_{A,w} \Vdash \varphi$) if and only if δ is a true description of the agent's state (i.e. $w \in [\delta]$)

Recall that an attitudinal state *type* description is a conjunction of beliefs ($B\varphi$) and abeliefs ($\neg B\neg\varphi$), where the sentences in the complement of the two attitudes do not contain $\Diamond$. We can now prove Theorem 3.1.

Theorem. *For every sentence φ in the language, there is an attitudinal state type description δ_φ such that φ expresses δ_φ in the sense of (43).*

Proof. By Lemma A.1, there are $\Diamond$-free sentences β, α_1, …, α_n such that $\mathbf{b}_{A,w} \Vdash \varphi$ iff $\mathbf{b}_{A,w} \Vdash \beta \wedge \Diamond\alpha_1 \wedge \cdots \wedge \Diamond\alpha_n$. Note that fact (41) has the special case that $\{w\} \Vdash B_A\Diamond\varphi$ iff $\{w\} \Vdash \neg B_A\neg\varphi$, with the former being equivalent to $\mathbf{b}_{A,w} \Vdash \Diamond\varphi$. Then, unpacking the definition of conjunction and applying this observation, we have that the above assertability holds iff $w \in [B\beta]$, $w \in [\neg B\neg\alpha_1]$, …, and $w \in [\neg B\neg\alpha_n]$. This in turn holds just in case $w \in [B\beta \wedge \neg B\neg\alpha_1 \wedge \cdots \wedge \neg B\neg\alpha_n]$, so the latter is the attitudinal state type description δ_φ that φ expresses. □

Appendix B

Experimenting with Epistemic Free Choice

So-called free choice inferences arise when a disjunction interacting with a possibility modal entails a certain conjunction. The paradigm examples come from the deontic domain:[1]

(66) a) You may have ice cream or cake. $\rightsquigarrow$ $[\Diamond(p \vee q)]$

b) You may have ice cream *and* you may have cake $[\Diamond p \wedge \Diamond q]$

Many authors[2] have noted that wide-scope disjunctions tend to give rise to the same phenomenon:[3]

(67) a) You may have ice cream or you may have cake. $\rightsquigarrow$ $[\Diamond p \vee \Diamond q]$

b) You may have ice cream *and* you may have cake $[\Diamond p \wedge \Diamond q]$

Because of this, many authors have adopted what we might call *the free choice uniformity thesis*: free choice effects arise equally in cases where the disjunction takes

[1] At this level of generality, I'm leaving it open whether the entailment is semantic or pragmatic and use '$\rightsquigarrow$' to denote the relevant sense of felt entailment.

[2] Kamp [1978] and Zimmermann [2000] are prominent examples.

[3] There are differences. For example, the free choice inference in (67) seems easier to cancel than in the narrow case (66). These details do not concern the present experiment and so will be ommitted.

wide scope and where it takes narrow scope with respect to the possibility modal. There are roughly three ways of endorsing the uniformity thesis. *Wide reductionists* take free choice in wide scope disjunctions to be primary and reduce the narrow scope case to it. Zimmermann [2000] takes this approach.[4] *Narrow reductionists* take free choice in narrow scope disjunctions to be primary and reduce the wide scope case to it. Simons [2005] takes this approach. *Non-reductionists* provide a treatment of free choice which generates the inferences for both wide and narrow scope disjunctions that does not reduce one to the other. Starr [2016b] takes this approach.

Now, observe that a critical component to proving that the assertability semantics in Chapter 3 meets the desideratum of only allowing the expression of non-disjunctive properties of attitudes is treating wide scope free choice in the epistemic domain as an assertability entailment.[5] In particular, we have the following entailment:

(68) a) The keys might be on the table or they might be in the drawer. $\Vdash$

b) The keys might be on the table *and* they might be in the drawer.

Interestingly, however, the assertability semantics does not have narrow scope free choice as an entailment: $\Diamond(p \vee q) \nVdash \Diamond p \wedge \Diamond q$.[6]

This situation suggests what we reject the free choice uniformity thesis for epistemic modals. In particular, we predict the following: in the case of epistemic 'might', speakers should be more willing to accept free choice inferences when the disjunction takes wide scope than when it takes narrow scope. In what follows, I present an experiment that provides support for this prediction.

B.1 Methods

To test whether people drew free choice inferences, participants were presented with a dialogue between two people who were considering what could be the case. One person would assert the sentence containing a disjunction and an epistemic possibility

[4] So, too, does Geurts [2005].

[5] Recall: Γ assertorically entails φ – written $\Gamma \Vdash \varphi$ – iff for every information state $\mathbf{s}$: if $[\![\gamma]\!]^{\mathbf{s}} = 1$ for every $\gamma \in \Gamma$, then $[\![\varphi]\!]^{\mathbf{s}} = 1$. See (37) in Section 3.5.1 for the semantic clauses.

[6] Counter-example: an information state $\mathbf{s}$ containing a single $p \wedge \neg q$ world.

– either with wide scope or narrow scope disjunction – and the other would respond by agreeing with the assertion but then denying one of the possibilities. If the free choice inference is drawn, that should be an incoherent action. Here is a concrete example of a vignette used, in the wide scope condition:

> Two siblings, Lorna and Charlie, love dessert. They both know that there are three desserts in the kitchen – pie, ice cream, and cookies – and that their mom is eating dessert right now.
>
> Lorna and Charlie have the following conversation.
>
> Charlie: do you know what mom's having?
> Lorna: it might be ice cream or it might be pie.
> Charlie: That's true, but it can't be pie.
>
> Does Charlie's statement make sense to you?

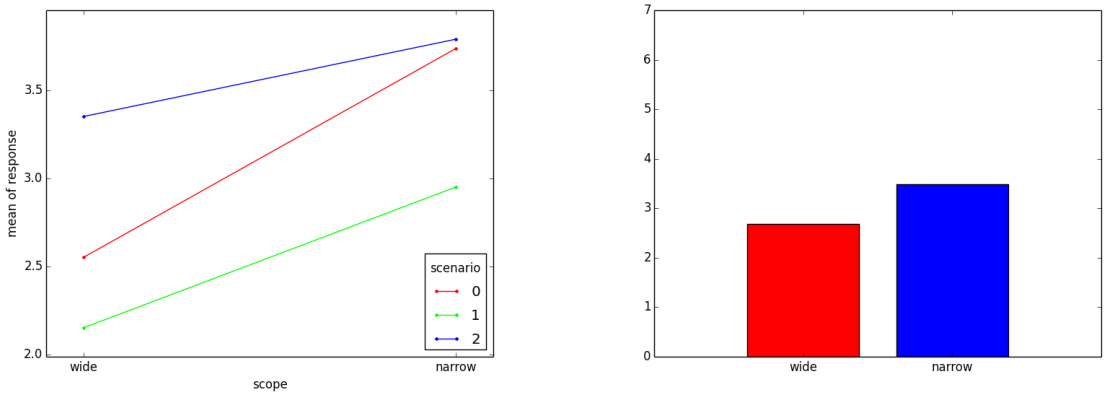
For the study, 118 participants were recruited on Amazon's Mechanical Turk. Each was randomly assigned to the wide scope or narrow scope condition. Furthermore, each was randomly assigned to see one of three vignettes: one about ice cream (above), one about the location of a colleague, and one about what card game was being played by a group. This factor should have no effect and was included to ensure that any results do not depend on the particular content of a vignette. After signing a consent form, participants were shown the vignette, plus one reading comprehension question, and the target question ("Does Charlie's statement make sense to you?"). They responded on a seven point Likert scale.[7]

B.2 Results

To analyze the responses, we performed both a parametric test and a nonparametric test. First, a two-way ANOVA revealed a significant main effect of scope ($p < 0.05$), no main effect of vignette, and no interaction effect between scope and vignette.

[7]The '1' response was labeled "makes no sense" and the '7' response was labeled "makes perfect sense".

Secondly, a Mann-Whitney U test comparing all wide scope to all narrow scope responses showed that participants found the free-choice-contradicting response less acceptable in the wide scope condition than the narrow scope condition ($p < 0.01$). Figure B.1 shows an interaction plot and a bar plot of these results.



(a) (b)

Figure B.1: (a) Interaction plot showing difference in mean responses between wide and narrow scope in each vignette. (b) Bar plot showing mean responses between all responses in wide and narrow scope conditions.

B.3 Discussion

These results confirm the prediction of our assertability semantics: the free choice uniformity thesis should be rejected, at least in the epistemic domain. Participants are more likely to draw free choice inferences when disjunction takes wide scope over epistemic modals than when it takes narrow scope. These results leave open the exact mechanism that links the semantics to speaker judgments. But the fact that wide scope free choice is an assertability entailment should make it more difficult for listeners to make sense of Charlie's statement in the wide scope case than in the narrow case, where free choice is not an entailment.

It also remains open whether the free choice uniformity thesis holds for deontic (or, more generally, root) modals. If free choice inferences are drawn uniformly for both

scopes in the deontic domain, this would point to an asymmetry in the free choice potential of epistemics and deontics. Such a result could begin to call into question the standard assumption that the two types of modals are very closely related. Minimally, in order to preserve the Kratzerian thought that epistemics and deontics are flavors of the same lexical meaning, one would have to trace the difference in free choice to something about the differences in the flavors.

Bibliography

Maria Aloni. Free choice, modals, and imperatives. *Natural Language Semantics*, 15 (1):65–94, 2007. doi: 10.1007/s11050-007-9010-2.

Maria Aloni. Disjunction. In Edward N Zalta, editor, *Stanford Encyclopedia of Philosophy*. Summer edition, 2016. URL http://plato.stanford.edu/archives/sum2016/entries/disjunction/.

Raffaele Argiento, Robin Pemantle, Brian Skyrms, and Stanislav Volkov. Learning to signal: Analysis of a micro-level reinforcement model. *Stochastic Processes and their Applications*, 119(2):373–390, 2009. doi: 10.1016/j.spa.2008.02.014.

David M Armstrong. *A Theory of Universals, volume 2: Universals and Scientific Realism*. Cambridge University Press, 1978.

Paul Audi. How to Rule Out Disjunctive Properties. *Noûs*, 47(4):748–766, 2013. doi: 10.1111/nous.12016.

Alfred Jules Ayer. *Language, Truth, and Logic*. Victor Gollancz LTD, London, 1936.

Jeffrey A Barrett. Numerical Simulations of the Lewis Signaling Game: Learning Strategies, Pooling Equilibria, and the Evolution of Grammar. Technical report, Institute for Mathematical Behavioral Sciences, 2006. URL http://escholarship.org/uc/item/5xr0b0vp.

Jeffrey A Barrett. Dynamic Partitioning and the Conventionality of Kinds. *Philosophy of Science*, 74:527–546, 2007.

Jeffrey A Barrett. The Evolution of Coding in Signaling Games. *Theory and Decision*, 67(2):223–237, 2009. doi: 10.1007/s11238-007-9064-0.

Jon Barwise. On Branching Quantifiers in English. *Journal of Philosophical Logic*, 8 (1):47–80, 1979. doi: 10.1007/BF00258419.

Jon Barwise and Robin Cooper. Generalized Quantifiers and Natural Language. *Linguistics and Philosophy*, 4(2):159–219, 1981.

Alan W Beggs. On the convergence of reinforcement learning. *Journal of Economic Theory*, 122(1):1–36, 2005. doi: 10.1016/j.jet.2004.03.008.

Carl T Bergstrom and Lee Alan Dugatkin. *Evolution*. W. W. Norton, New York, 2012.

Peter Cameron and Wilfred Hodges. Some Combinatorics of Imperfect Information. *Journal of Symbolic Logic*, 66(2):673–684, 2001.

Fabrizio Cariani. Choice Points for a Modal Theory of Disjunction. *Topoi*, 2016. doi: 10.1007/s11245-015-9362-z.

Nate Charlow. Prospects for an Expressivist Theory of Meaning. *Philosophers' Imprint*, 15(23):1–43, 2015.

Nick Chater and Mike Oaksford. The Probability Heuristics Model of Syllogistic Reasoning. *Cognitive Psychology*, 38(2):191–258, mar 1999. doi: 10.1006/cogp.1998.0696.

Gennaro Chierchia and Sally McConnell-Ginet. *Meaning and Grammar*. The MIT Press, 2000.

Ivano Ciardelli, Jeroen Groenendijk, and Floris Roelofsen. Attention! Might in Inquisitive Semantics. In Ed Cormany, Satoshi Ito, and David Lutz, editors, *Proceedings of Semantics and Linguistic Theory (SALT) 19*, pages 91–108, 2009.

Ivano A Ciardelli and Floris Roelofsen. Inquisitive dynamic epistemic logic. *Synthese*, 192(6):1643–1687, 2015. doi: 10.1007/s11229-014-0404-7.

Bo Cui, Yuan Gao, Lila Kari, and Sheng Yu. State complexity of combined operations with two basic operations. *Theoretical Computer Science*, 437:82–102, 2012. doi: 10.1016/j.tcs.2012.02.030.

Donald Davidson. Theories of Meaning and Learnable Languages. In Yehoshua Bar-Hillel, editor, *Proceedings of the International Congress for Logic, Methodology, and Philosophy of Science*, pages 3–17. 1965.

Stanislas Dehaene. *The Number Sense: How the Mind Creates Mathematics*. Oxford University Press, Oxford, 1997.

Paul Dekker. Meanwhile, Within the Frege Boundary. *Linguistics and Philosophy*, 26(5):547–556, 2003.

Josh Dever. Compositionality as Methodology. *Linguistics and Philosophy*, 22(3): 311–326, 1999.

Volker Diekert and Paul Gastin. First-order definable languages. In Jörg Flum, Erich Grädel, and Thomas Wilke, editors, *Logic and Automata: History and Perspectives*, Texts in Logic and Games, pages 261–306. Amsterdam University Press, Amsterdam, 2007.

Cian Dorr and John Hawthorne. Embedding Epistemic Modals. *Mind*, 122(488): 867–913, 2013. doi: 10.1093/mind/fzt091.

Janice L Dowell. A Flexible Contextualist Account of Epistemic Modals. *Philosophers' Imprint*, 11(14):1–25, 2011.

Michael Dummett. *Frege: Philosophy of Language*. Harper & Row, New York, 1973.

Kit Fine. Truth-Maker Semantics for Intuitionistic Logic. *Journal of Philosophical Logic*, 43(2-3):549–577, 2014. doi: 10.1007/s10992-013-9281-7.

Jerry Fodor. *Psychosemantics*. The MIT Press, Cambridge, MA, 1987.

Jerry Fodor. *In Critical Condition: Polemical Essays on Cognitive Science and the Philosophy of Mind*. The MIT Press, Cambridge, MA, 1998.

Jerry Fodor and Zenon W. Pylyshyn. Connectionism and Cognitive Architecture. *Cognition*, 28(1-2):3–71, 1988. doi: 10.1016/0010-0277(88)90031-5.

Jerry A Fodor. Language, Thought and Compositionality. *Mind & Language*, 16(1): 1–15, 2001. doi: 10.1111/1468-0017.00153.

Jakob N Foerster, Yannis M Assael, Nando de Freitas, and Shimon Whiteson. Learning to Communicate with Deep Multi-Agent Reinforcement Learning. In Daniel D Lee, Masashi Sugiyama, Ulrike V Luxburg, Isabelle Guyon, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 29 (NIPS 2016)*, 2016.

Michael Franke. Creative Compositionality From Reinforcement. In Erica A Cartmill, Seán Roberts, Heidi Lyn, and Hannah Cornish, editors, *The Evolution of Langauge (Proceedings of EvoLang 10)*, pages 82–89, Singapore, 2014. World Scientific.

Gottlob Frege. *Die Grundlagen der Arithmetik*. Verlag von Wilhelm Koebner, 1884.

Gottlob Frege. Logische Untersuchungen. Dritter Teil: Gedankengefüge ("Compound Thoughts"). *Beiträge zur Philosophie des deutschen Idealismus*, III:36–51, 1923. doi: 10.1093/mind/LI.202.200.

Douglas J Futuyma. Evolutionary Constraint and Ecological Consequences. *Evolution*, 64(7):1865–1884, 2010. doi: 10.1111/j.1558-5646.2010.00960.x.

Pietro Galliani. Dependence Logic. *Stanford Encyclopedia of Philosophy*, (Spring), 2017. URL https://plato.stanford.edu/archives/spr2017/entries/logic-dependence/.

Gerald Gazdar. *Pragmatics: Implicature, Presupposition and Logical Form*. Academic Press, New York, 1979.

Tamar Szabó Gendler. Alief and Belief. *The Journal of Philosophy*, 105(10):634–663, 2008.

Bart Geurts. Entertaining Alternatives: Disjunctions as Modals. *Natural Language Semantics*, 13:383–410, 2005. doi: 10.1007/s11050-005-2052-7.

Dom Giammarresi and Antonio Restivo. Two-Dimensional Languages. In Grzegorz Rozenberg and Arto Salomaa, editors, *Handbook of Formal Languages (vol. 3): Beyond Words*, volume 3, pages 215–268. Springer, Berlin, 1997.

Allan Gibbard. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Harvard University Press, 1990.

Michael Glanzberg. Against Truth-Value Gaps. In JC Beall, editor, *Liar and Heaps: New Essays on Paradox*, pages 151–194. Oxford University Press, 2004.

Paul W Glimcher. Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences*, 108(42):15647–15654, 2011. doi: 10.1073/pnas.1014269108.

Thomas L. Griffiths, Falk Lieder, and Noah D. Goodman. Rational Use of Cognitive Resources: Levels of Analysis Between the Computational and the Algorithmic. *Topics in Cognitive Science*, 7(2):217–229, 2015. doi: 10.1111/tops.12142.

Jeroen Groenendijk and Floris Roelofsen. Radical inquisitive semantics. In *Sixth International Symposium on Logic, Cognition, and Communication*, 2010.

Jeroen Groenendijk, Martin Stokhof, and Frank Veltman. Coreference and Modality. *The Handbook of Contemporary Semantic Theory*, pages 179–213, 1996. doi: 10.1111/b.9780631207498.1997.00010.x.

J. B. S. Haldane. *The Causes of Evolution*. Longmans, Green, and co., London, 1932.

Mehdi Hassani. Derangements and Applications. *Journal of Integer Sequences*, 6, 2003.

Peter Hawke and Shane Steinert-Threlkeld. Informational Dynamics of 'Might' Assertions. In Wiebe van der Hoek, Wesley H Holliday, and Wen-fang Wang, editors, *Proceedings of Logic, Rationality, and Interaction (LORI-V)*, volume 9394 of *Lecture Notes in Computer Science*, pages 143–155, 2015. doi: 10.1007/978-3-662-48561-3_12.

Peter Hawke and Shane Steinert-Threlkeld. Informational dynamics of epistemic possibility modals. *Synthese*, 2016a. doi: 10.1007/s11229-016-1216-8.

Peter Hawke and Shane Steinert-Threlkeld. The Pragmatics of Epistemic Modality. 2016b.

Irene Heim and Angelika Kratzer. *Semantics in Generative Grammar*. Blackwell Textbooks in Linguistics. Wiley-Blackwell, 1998. ISBN 0631197133.

James Higginbotham. Linguistic Theory and Davidson's Program in Semantics. In Ernest Lepore, editor, *Truth and Interpretation*, pages 29–48. Basil Blackwell, Oxford, 1986.

James Higginbotham. Conditionals and Compositionality. *Philosophical Perspectives*, 17:181–194, 2003.

Jaakko Hintikka. *Knowledge and Belief*. Cornell University Press, Ithaca, 1962.

Jaakko Hintikka. Quantifiers vs. Quantification Theory. *Linguistic Inquiry*, 5(2): 153–177, 1974.

Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9:1735–1780, 1997.

Wilfrid Hodges. Compositional Semantics for a Language of Imperfect Information. *Logic Journal of the IGPL*, 5(4):539–563, 1997.

Josef Hofbauer and Simon M Huttegger. Feasibility of communication in binary signaling games. *Journal of Theoretical Biology*, 254(4):843–849, 2008. doi: 10.1016/j.jtbi.2008.07.010.

Josef Hofbauer and Simon M Huttegger. Selection-Mutation Dynamics of Signaling Games. *Games*, 6(1):2–31, 2015. doi: 10.3390/g6010002.

John E Hopcroft and Jeffrey D Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, 1979.

Yilei Hu, Brian Skyrms, and Pierre Tarrès. Reinforcement learning in signaling game. 2011. URL http://arxiv.org/abs/1103.5818.

Simon Huttegger. Evolution and the Explanation of Meaning. *Philosophy of Science*, 74(1):1–27, 2007. doi: 10.1086/519477.

Thomas F. Icard III. and Lawrence S. Moss. Recent Progress on Monotonicity. *Linguistic Issues in Language Technology*, 9:167–194, 2014.

Pauline Jacobson. *Compositional Semantics: An Introduction to the Syntax/Semantics Interface*. Oxford Textbooks in Linguistics. Oxford Univeristy Press, 2014.

Theo M V Janssen. *The Foundations and Applications of Montague Grammar*. PhD thesis, Universiteit van Amsterdam, 1983.

Theo M V Janssen. Compositionality. In Johan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language*, chapter 7, pages 417–473. Elsevier Science, 1997. doi: 10.1016/B978-044481714-3/50011-4.

Hans Kamp. Semantics Versus Pragmatics. In Jack Kulas, James H Fetzer, and Terry L Rankin, editors, *Philosophy, Language, Artificial Intelligence: Resources for Language Processing*, volume 2 of *Studies in Cognitive Systems*, pages 349–380. Springer, 1978.

Makoto Kanazawa. Monadic Quantifiers Recognized by Deterministic Pushdown Automata. In Maria Aloni, Michael Franke, and Floris Roelofsen, editors, *Proceedings of the 19th Amsterdam Colloquium*, pages 139–146, 2013.

David Kaplan. Demonstratives. In Joseph Almog, John Perry, and Howard Wettstein, editors, *Themes from Kaplan*, pages 481–563. Oxford University Press, New York, 1989.

Edward L Keenan. Beyond the Frege Boundary. *Linguistics and Philosophy*, 15(2): 199–221, 1992.

Edward L Keenan. The Semantics of Determiners. In Shalom Lappin, editor, *The Handbook of Contermporary Semantic Theory*, number 1989, chapter 2, pages 41–63. Blackwell, Oxford, 1996a.

Edward L Keenan. Further Beyond the Frege Boundary. In Jaap van der Does and Jan van Eijck, editors, *Quantifiers, Logic, and Language*, volume 54 of *CSLI Lecture Notes*, pages 179–201. CSLI Publications, Stanford, 1996b.

Edward L Keenan and Dag Westerståhl. Generalized Quantifiers in Linguistics and Logic. In Johan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language*, pages 837–893. Elsevier, 1997.

Nathan Klinedinst and Daniel Rothschild. Connectives without truth tables. *Natural Language Semantics*, 20(2):137–175, 2012. doi: 10.1007/s11050-011-9079-5.

Angelika Kratzer. The Notional Category of Modality. In Hans-Jürgen Eikmeyer and Hannes Rieser, editors, *Words, Worlds, and Context*, pages 38–74. Walter de Gruyter, 1981.

Angelika Kratzer. Conditionals. In *Proceedings of the Chicago Linguistics Society*, volume 22, pages 1–15, 1986.

Angelika Kratzer. Modality. In Arnim von Stechow and Dieter Wunderlich, editors, *Semantik*, chapter 29, pages 639–650. Mouton de Gruyter, Berlin and New York, 1991.

Jennifer Lackey. Norms of Assertion. *Noûs*, 41(4):594–626, 2007. doi: 10.1111/j.1468-0068.2007.00664.x.

Sarah-Jane Leslie. "If", "Unless", and Quantification. In Robert J. Stainton and Christopher Viger, editors, *Compositionality, Context and Semantic Values*, number 1986 in Studies in Linguistics and Philosophy, pages 3–30. Springer Netherlands, Dordrecht, 2009. doi: 10.1007/978-1-4020-8310-5.

David Lewis. *Convention.* Blackwell, 1969.

David Lewis. General semantics. *Synthese*, 22:18–67, 1970.

David Lewis. Adverbs of Quantification. In Edward L Keenan, editor, *Formal Semantics of Natural Language*, pages 178–188. Cambridge University Press, 1975.

David Lewis. Index, Context, and Content. *Philosophy and Grammar*, (2):79–100, 1980.

Jeffrey Lidz, Paul Pietroski, Justin Halberda, and Tim Hunter. Interface transparency and the psychosemantics of most. *Natural Language Semantics*, 19(3):227–256, 2011. doi: 10.1007/s11050-010-9062-6.

Per Lindström. First order predicate logic with generalized quantifiers. *Theoria*, 32 (3):186–195, 1966.

Bill MacCartney. *Natural Language Inference*. PhD thesis, Stanford University, 2009.

John MacFarlane. Epistemic Modals are Assessment-Sensitive. In Andy Egan and Brian Weatherson, editors, *Epistemic Modality*, pages 144–179. Oxford University Press, 2011.

John MacFarlane. *Assessment Sensitivity*. Oxford University Press, 2014. doi: 10.1093/acprof:oso/9780199682751.001.0001.

Ishani Maitra. Assertion, Norms, and Games. In Jessica Brown and Herman Cappelen, editors, *Assertion: New Philosophical Essays*, pages 277–296. Oxford University Press, jan 2011. doi: 10.1093/acprof:oso/9780199573004.003.0012.

Matthew Mandelkern. Bounded Modality. In *Coordination in Conversation*, chapter 1. 2017.

David Marr. *Vision*. Freeman, San Francisco, 1982.

Corey T McMillan, Robin Clark, Peachie Moore, Christian Devita, and Murray Grossman. Neural basis for generalized quantifier comprehension. *Neuropsychologia*, 43 (12):1729–1737, 2005. doi: 10.1016/j.neuropsychologia.2005.02.012.

Corey T McMillan, Robin Clark, Peachie Moore, and Murray Grossman. Quantifier comprehension in corticobasal degeneration. *Brain and Cognition*, 62(3):250–260, 2006. doi: 10.1016/j.bandc.2006.06.005.

Robert McNaughton and Seymour A Papert. *Counter-free Automata*, volume 65 of *MIT Research Monographs*. The MIT Press, 1971.

Sarah McWhirter. *An Automata-Theoretic Perspective on Polyadic Quantification in Natural Language*. Masters of logic thesis, Universiteit van Amsterdam, 2014.

Cheryl Misak. *Truth and the End of Inquiry*. Oxford University Press, 2007.

Richard Montague. Universal Grammar. *Theoria*, 36(3):373–398, 1970. doi: 10.1111/j.1755-2567.1970.tb00434.x.

Richard Montague. The Proper Treatment of Quantification in Ordinary English. In Jaako Hintikka, Julius Moravcsik, and Patrick Suppes, editors, *Approaches to Natural Language*, pages 221–242. Reidel, Dordrecht, 1973.

Sarah Moss. Epistemology Formalized. *The Philosophical Review*, 122(1):1–43, 2013. doi: 10.1215/00318108-1728705.

Sarah Moss. On the semantics and pragmatics of epistemic vocabulary. *Semantics and Pragmatics*, 8(5):1–81, 2015. doi: 10.3765/sp.8.5.

Andrzej Mostowski. On a generalization of quantifiers. *Fundamenta Mathematicae*, 44(2):12–36, 1957.

Marcin Mostowski. Divisibility Quantifiers. *Bulletin of the Section of Logic*, 20(2):67–70, 1991.

Marcin Mostowski. Computational semantics for monadic quantifiers. *Journal of Applied Non-Classical Logics*, 8(1-2):107–121, 1998.

Reinhard Muskens, van Johan Benthem, and Albert Visser. Dynamics. In Johan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language*, pages 567–648. Elsevier Science, 1997.

Dilip Ninan. Semantics and the objects of assertion. *Linguistics and Philosophy*, 33 (5):355–380, 2010. doi: 10.1007/s10988-011-9084-7.

Martin A Nowak and David C Krakauer. The evolution of language. *Proceedings of the National Academy of Sciences of the United States of America*, 96:8028–8033, 1999.

Peter Pagin. Truth Theories, Competence, and Semantic Computation. In Gerhard Preyer, editor, *Davidson on Truth, Meaning, and the Mental*, pages 49–75. Oxford University Press, Oxford, 2012.

Peter Pagin and Dag Westerståhl. Compositionality I: Definitions and Variants. *Philosophy Compass*, 5(3):250–264, 2010a. doi: 10.1111/j.1747-9991.2009.00228.x.

Peter Pagin and Dag Westerståhl. Compositionality II: Arguments and Problems. *Philosophy Compass*, 5(3):265–282, 2010b. doi: 10.1111/j.1747-9991.2009.00229.x.

Barbara Hall Partee. Belief-Sentences and the Limits of Semantics. In Stanley Peters and Esa Saarinen, editors, *Processes, Beliefs, and Questions*, volume 16 of *Texts and Studies in Linguistics and Philosophy*, pages 87–106. Springer, 1982.

Barbara Hall Partee. Lexical Semantics and Compositionality. In Lila Gleitman and Mark Liberman, editors, *Invitation to Cognitive Science, Part 1: Language*, chapter 11, pages 311–360. MIT Press, Cambridge, 1995.

Christina Pawlowitsch. Why evolution does not always lead to an optimal signaling system. *Games and Economic Behavior*, 63(1):203–226, 2008. doi: 10.1016/j.geb.2007.08.009.

Christopher Peacocke. Explanation in Computational Psychology: Language, Perception and Level 1.5. *Mind and Language*, 1(2):101–123, 1986.

Charles Sanders Peirce. How to Make Our Ideas Clear. In Christian J.W. Kloesel, editor, *Writings of Charles S. Peirce: A Chronological Edition*, volume 3, pages 257–276. Indiana University Press, 1878.

Francis Jeffry Pelletier. The Principle of Semantic Compositionality. *Topoi*, 13(1): 11–24, 1994. doi: 10.1007/BF00763644.

Stanley Peters and Dag Westerståhl. *Quantifiers in Language and Logic*. Clarendon Press, Oxford, 2006.

Paul Pietroski, Jeffrey Lidz, Tim Hunter, and Justin Halberda. The Meaning of 'Most': Semantics, Numerosity and Psychology. *Mind and Language*, 24(5):554–585, 2009.

Paul Portner. *Modality*. Oxford University Press, Oxford, 2009.

Brian Rabern. Against the identification of assertoric content with compositional value. *Synthese*, 189(1):75–96, 2012. doi: 10.1007/s11229-012-0096-9.

Floris Roelofsen. A bare bone attentive semantics for might. In Maria Aloni, Michael Franke, and Floris Roelofsen, editors, *The dynamic, inquisitive, and visionary life of ϕ, $?\phi$ and $\Diamond\phi$*. 2013.

Gideon Rosen. Essays in Quasi-Realism by Simon Blackburn. *Noûs*, 32(3):386–405, 1998.

Alvin E. Roth and Ido Erev. Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term. *Games and Economic Behavior*, 8:164–212, 1995.

Daniel Rothschild. Expressing Credences. *Proceedings of the Aristotelean Society*, 112(1):99–114, 2012. doi: 10.1111/j.1467-9264.2012.00327.x.

Víctor Sánchez Valencia. *Studies on Natural Logic and Categorial Grammar*. PhD thesis, Universiteit van Amsterdam, 1991.

Philippe Schlenker. Local Contexts. *Semantics and Pragmatics*, 2(4):1–78, 2009. doi: 10.3765/sp.2.3.

Philippe Schlenker, Emmanuel Chemla, Kate Arnold, Alban Lemasson, Karim Ouattara, Sumir Keenan, Claudia Stephan, Robin Ryder, and Klaus Zuberbühler. Monkey semantics: two 'dialects' of Campbell's monkey alarm calls. *Linguistics and Philosophy*, 37:439–501, 2014. doi: 10.1007/s10988-014-9155-7.

Philippe Schlenker, Emmanuel Chemla, Anne M Schel, James Fuller, Jean-Pierre Gautier, Jeremy Kuhn, Dunja Veselinović, Kate Arnold, Cristiane Cäsar, Sumir Keenan, Alban Lemasson, Karim Ouattara, Robin Ryder, and Klaus Zuberbühler. Formal monkey linguistics. *Theoretical Linguistics*, 42(1-2):1–90, 2016. doi: 10.1515/tl-2016-0001.

Mark Schroeder. What is the Frege-Geach Problem? *Philosophy Compass*, 4:703–720, 2008.

Wolfram Schultz. Neural coding of basic reward terms of animal learning theory, game theory, microeconomics and behavioural ecology. *Current Opinion in Neurobiology*, 14(2):139–47, 2004. doi: 10.1016/j.conb.2004.03.017.

Wolfram Schultz, Peter Dayan, and P Read Montague. A Neural Substrate of Prediction and Reward. *Science*, 275:1593–1599, 1997.

Moritz Schulz. Wondering what might be. *Philosophical Studies*, 149(3):367–386, jul 2010. doi: 10.1007/s11098-009-9361-2.

Géraud Sénizergues. The Equivalence Problem for Deterministic Pushdown Automata is Decidable. In Pierpaolo Degano, Roberto Gorrieri, and Alberto Marchetti-Spaccamela, editors, *Automata, Languages and Programming*, volume 1256 of *Lecture Notes in Computer Science*, pages 671–681. Springer Berlin Heidelberg, Berlin, 1997.

Géraud Sénizergues. L(A) = L(B)? decidability results from complete formal systems. *Theoretical Computer Science*, 251(1-2):1–166, 2001.

Géraud Sénizergues. L(A) = L(B)? A simplified decidability proof. *Theoretical Computer Science*, 281(1-2):555–608, 2002.

Robert M Seyfarth, Dorothy L Cheney, and Peter Marler. Vervet Monkey Alarm Calls: Semantic Communication in a Free-ranging Primate. *Animal Behavior*, 28: 1070–1094, 1980.

Alex Silk. How to Be an Ethical Expressivist. *Philosophy and Phenomenological Research*, 91(1):47–81, 2015. doi: 10.1111/phpr.12138.

Mandy Simons. Dividing things up: The semantics of *or* and the modal/*or* interaction. *Natural Language Semantics*, 13(3):271–316, 2005. doi: 10.1007/s11050-004-2900-7.

Brian Skyrms. *Evolution of the Social Contract*. Cambridge University Press, 1996.

Brian Skyrms. *The Stag Hunt and the Evolution of Social Structure*. Cambridge University Press, 2004.

Brian Skyrms. *Signals: Evolution, Learning, and Information*. Oxford University Press, 2010.

Paul Smolensky. On the proper treament of connectionism. *Behavioral and Brain Sciences*, 11:1–23, 1988.

Scott Soames. Linguistics and Psychology. *Linguistics and Philosophy*, 7(2):155–179, 1984.

Robert Stalnaker. Assertion. In P Cole, editor, *Syntax and Semantics 9: Pragmatics*, pages 315–332. Academic Press, New York, 1978.

Robert Stalnaker. Common Ground. *Linguistics and Philosophy*, 25:701–721, 2002. doi: 10.1023/A:1020867916902.

Jason Stanley. Names and Rigid Designation. In Bob Hale and Crispin Wright, editors, *A Companion to the Philosophy of Language*, pages 555–585. Blackwell Publishers, Oxford, 1997.

Jason Stanley. Context and logical form. *Linguistics and Philosophy*, 23:391–434, 2000.

William B Starr. Dynamic Expressivism about Deontic Modality. In Nate Charlow and Matthew Chrisman, editors, *Deontic Modality*, pages 355–394. Oxford University Press, 2016a.

William B Starr. Expressing permission. In *Proceedings of Semantics and Linguistic Theory 26*, pages 325–349, 2016b.

Shane Steinert-Threlkeld. On the Decidability of Iterated Languages. In Oleg Prosorov, editor, *Proceedings of Philosophy, Mathematics, Linguistics: Aspects of Interaction (PhML2014)*, pages 215–224, 2014.

Shane Steinert-Threlkeld. Some Properties of Iterated Languages. *Journal of Logic, Language and Information*, 25(2):191–213, 2016a. doi: 10.1007/s10849-016-9239-6.

Shane Steinert-Threlkeld. Compositionality and competition in monkey alert calls. *Theoretical Linguistics*, 42(1-2):159–171, 2016b. doi: 10.1515/tl-2016-0009.

Shane Steinert-Threlkeld. Compositional Signaling in a Complex World. *Journal of Logic, Language and Information*, 25(3):379–397, 2016c. doi: 10.1007/s10849-016-9236-9.

Shane Steinert-Threlkeld and Thomas F. Icard III. Iterating semantic automata. *Linguistics and Philosophy*, 36(2):151–173, 2013. doi: 10.1007/s10988-013-9132-6.

Shane Steinert-Threlkeld, Gert-Jan Munneke, and Jakub Szymanik. Alternative Representations in Formal Semantics: A case study of quantifiers. In Thomas Brochhagen, Floris Roelofsen, and Nadine Thelier, editors, *Proceedings of the 20th Amsterdam Colloquium*, pages 368–378, 2015.

Patrick Suppes. Procedural Semantics. In R Haller and W Grassl, editors, *Language, Logic, and Philosophy: Proceedings of the 4th International Wittgenstein Symposium*, pages 27–35. Hölder-Pichler-Tempsy, Vienna, 1980.

Patrick Suppes. Variable-Free Semantics with Remarks on Procedural Extensions. In TW Simon and RJ Scholes, editors, *Language, Mind, and Brain*, pages 21–34. Erlbaum, Hillsdale, NJ, 1982.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to Sequence Learning with Neural Networks. In Zoubin Ghahramani, Max Welling, Corinna Cortes, Neil D Lawrence, and Kilian Q Weinberger, editors, *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, 2014.

Richard S Sutton and Andrew G Barto. *Reinforcement Learning: an Introduction*. Bradford, 1998.

Eric Swanson. The Application of Constraint Semantics to the Language of Subjective Uncertainty. *Journal of Philosophical Logic*, 45(2):121–146, 2016. doi: 10.1007/s10992-015-9367-5.

Zoltán Gendler Szabó. Compositionality as Supervenience. *Linguistics and Philosophy*, 23(5):475–505, 2000.

Zoltán Gendler Szabó. Compositionality. *The Stanford Encyclopedia of Philosophy*, Fall, 2013. URL http://plato.stanford.edu/archives/fall2013/entries/compositionality/.

Anna Szabolcsi. *Quantification*. Research Surveys in Linguistics. Cambridge University Press, Cambridge, 2010.

Jakub Szymanik. A comment on a neuroimaging study of natural language quantifier comprehension. *Neuropsychologia*, 45(9):2158–2160, may 2007. ISSN 0028-3932. doi: 10.1016/j.neuropsychologia.2007.01.016. URL http://www.ncbi.nlm.nih.gov/pubmed/17336347.

Jakub Szymanik. Computational complexity of polyadic lifts of generalized quantifiers in natural language. *Linguistics and Philosophy*, 33(3):215–250, nov 2010. ISSN 0165-0157. doi: 10.1007/s10988-010-9076-z. URL http://www.springerlink.com/index/10.1007/s10988-010-9076-z.

Jakub Szymanik. *Quantifiers and Cognition: Logical and Computational Perspectives*, volume 96 of *Studies in Linguistics and Philosophy*. Springer, 2016.

Jakub Szymanik and Marcin Zajenkowski. Comprehension of simple quantifiers: empirical evaluation of a computational model. *Cognitive Science*, 34(3):521–532, 2010a. doi: 10.1111/j.1551-6709.2009.01078.x.

Jakub Szymanik and Marcin Zajenkowski. Quantifiers and Working Memory. *Lecture Notes in Artificial Intelligence*, 6042:456–464, 2010b.

Jakub Szymanik and Marcin Zajenkowski. Contribution of working memory in parity and proportional judgments. *Belgian Journal of Linguistics*, 25(1):176–194, 2011. doi: 10.1075/bjl.25.08szy.

Jakub Szymanik, Shane Steinert-Threlkeld, Marcin Zajenkowski, and Thomas F. Icard III. Automata and Complexity in Multiple-Quantifier Sentence Verification. In *Proceedings of the 12th International Conference on Cognitive Modeling*, 2013.

Peter E Trapa and Martin A Nowak. Nash equilibria for an evolutionary language game. *Journal of Mathematical Biology*, 41(2):172–188, 2000. doi: 10.1007/s002850070004.

Jouko Väänanen. *Dependence Logic: A New Approach to Independence Friendly Logic*, volume 29 of *London Mathematical Society Student Texts*. Cambridge University Press, 2007.

Jouko Väänänen. Modal Dependence Logic. In Krzysztof R Apt and Robert van Rooij, editors, *New Perspectives on Games and Interaction*, volume 4 of *Texts in Logic and Games*, pages 237–254. Amsterdam University Press, Amsterdam, 2008.

Johan van Benthem. *Essays in Logical Semantics*. D. Reidel Publishing Company, Dordrecht, 1986.

Johan van Benthem. Toward a Computational Semantics. In Peter Gardenfors, editor, *Generalized Quantifiers: Linguistic and Logical Approaches*, pages 31–71. Kluwer Academic Publishers, 1987.

Johan van Benthem. *A Manual of Intensional Logic*. CSLI Publications, second edition, 1988.

Johan van Benthem. Semantic Parallels in Natural Language and Computation. In HD Ebbinghaus, J Fernandez-Prida, M Garrido, D Lascar, and M Rodriguez Artalejo, editors, *Logic Colloquium '87*, volume 129 of *Studies in Logic and the Foundations of Mathematics*, pages 331–375. 1989a. doi: 10.1016/S0049-237X(08)70133-2.

Johan van Benthem. Polyadic quantifiers. *Linguistics and Philosophy*, 12(4):437–464, 1989b. doi: 10.1007/BF00632472.

Johan van Benthem. *Language in Action*, volume 130 of *Studies in Logic and the Foundations of Mathematics*. Elsevier, Amsterdam, 1991.

Johan van Benthem. *Logical Dynamics of Information and Interaction*. Cambridge University Press, Cambridge, 2011.

Johan van Benthem. Bernard Bolzano's *Wissenschaftslehre*. *Topoi*, 32(2):301–303, 2013. doi: 10.1007/s11245-013-9160-4.

Jaap van der Does, Willem Groeneveld, and Frank Veltman. An Update on "Might". *Journal of Logic, Language, and Information*, 6:361–381, 1997.

Frank Veltman. *Logics for Conditionals*. PhD thesis, Universiteit van Amsterdam, 1985.

Frank Veltman. Defaults in Update Semantics. *Journal of Philosophical Logic*, 25(3):221–261, 1996. doi: 10.1007/BF00248150.

Kai von Fintel and Anthony S Gillies. 'Might' Made Right. In Andy Egan and Brian Weatherson, editors, *Epistemic Modality*, pages 108–130. Oxford University Press, Oxford, 2011.

Kai von Fintel and Sabine Iatridou. If and When *If*-Clauses Can Restrict Quantifiers. In *Workshop in Philosophy and Linguistics*, number June, 2002.

Wilhelm von Humboldt. *On Language: On the Diversity of Human Language Construction and its Influence on the Mental Development of the Human Species*. Cambridge Texts in the History of Philosophy. Cambridge University Press, 1836.

Dag Westerståhl. Quantifiers in Formal and Natural Languages. In D Gabbay and F Guenthner, editors, *Handbook of Philosophical Logic (vol. 4)*, pages 1–132. D. Reidel Publishing Company, Dordrecht, 1989.

Dag Westerståhl. On Mathematical Proofs of the Vacuity of Compositionality. *Linguistics and Philosophy*, 21(6):635–643, 1998. doi: 10.1023/A:1005401829598.

Dag Westerståhl. Compositionality in Kaplan Style Semantics. In Wolfram Hinzen, Edouard Machery, and Markus Werning, editors, *The Oxford Handbook of Compositionality*, pages 192–219. Oxford Univeristy Press, 2012. ISBN 9780199541072. doi: 10.1093/oxfordhb/9780199541072.013.0009.

Timothy Williamson. Knowing and Asserting. *The Philosophical Review*, 105(4): 489–523, 1996.

Seth Yalcin. Epistemic Modals. *Mind*, 116(464):983–1026, 2007. doi: 10.1093/mind/fzm983.

Seth Yalcin. Nonfactualism About Epistemic Modality. In Andy Egan and Brian Weatherson, editors, *Epistemic Modality*, pages 295–332. Oxford University Press, Oxford, 2011.

Seth Yalcin. Bayesian Expressivism. *Proceedings of the Aristotelian Society*, 112(2): 123–160, 2012. doi: 10.1111/j.1467-9264.2012.00329.x.

Seth Yalcin. Semantics and Metasemantics in the Context of Generative Grammar. In Alexis Burgess and Brett Sherman, editors, *Metasemantics: New Essays on the Foundations of Meaning*, pages 17–54. Oxford University Press, Oxford, 2014.

Andy Demfree Yu. Epistemic Modals and Sensitivity to Contextually-Salient Partitions. *Thought: A Journal of Philosophy*, 2016. doi: 10.1002/tht3.203.

Sheng Yu, Qingyu Zhuang, and Kai Salomaa. The state complexities of some basic operations on regular languages. *Theoretical Computer Science*, 125(2):315–328, 1994. doi: 10.1016/0304-3975(92)00011-F.

Wlodek Zadrozny. From Compositional to Systematic Semantics. *Linguistics and Philosophy*, 17(4):329–342, 1994.

Thomas Ede Zimmermann. Free Choice Disjunction and Epistemic Possibility. *Natural Language Semantics*, 8:255–290, 2000.

Klaus Zuberbühler. A syntactic rule in forest monkey communication. *Animal Behaviour*, 63(2):293–299, 2002. doi: 10.1006/anbe.2001.1914.