Super Semantics:
a Unifying Framework for Meaning Phenomena in Nature[*]

Philippe Schlenker[abce***]; Salvador Mascarenhas[abc**]; Emmanuel Chemla[bcd*]

Preliminary draft, June 5, 2020

Abstract We argue that formal semantics, properly extended, can provide a unifying framework for diverse meaning phenomena beyond traditional linguistic objects. We summarize key achievements of semantics/pragmatics, pertaining to constraints on possible meanings, the role of variables, and the typology of linguistic inferences. We then show that semantics can help illuminate diverse non-standard objects, starting with primate calls, by: stating precise hypotheses about their lexical semantics; showing that no non-trivial rules of call composition are needed; and suggesting that rules of competition among calls might play an important pragmatic role. Turning to human language, new developments make it possible to ask how iconicity interacts with logical semantics in spoken language, in sign language and in gestures. Sign language plays a prominent role, for two reasons: it has the same type of logical resources as speech, but sometimes provides overt evidence for them (e.g. variables); and it also makes greater use of iconic enrichments. This motivates a systematic comparison between sign-with-iconicity and speech-with-gestures. Semantic results arguably differ across the two cases because iconic modulations in sign and gestural enrichments in speech are distributed differently within the inferential typology. The comparison also helps uncover a proto-grammar of gestures. Going beyond language, semantics has recently been extended to visual narratives and even to music, using a mix of new methods and standard tools (e.g. variables). Finally, semantics is increasingly used to illuminate constraints on logical and non-logical concepts in the 'language of thought', and to propose new accounts of reasoning.

Keywords: formal semantics, super semantics, primate semantics, sign language semantics, gesture semantics, picture semantics, visual narratives, music semantics, iconicity

---

[*] [Acknowledgments to be added later]
Funding: The research leading to these results received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC grant agreement N°324115–FRONTSEM (PI: Schlenker), under the European Union's Horizon 2020 Research and Innovation Programme (ERC grant agreement No 788077, Orisem, PI: Schlenker), and also from Agence Nationale de la Recherche, ANR-18-CE28-0008, LANG-REASON (PI: Mascarenhas). Research was conducted at Institut d'Etudes Cognitives, Ecole Normale Supérieure - PSL University. Institut d'Etudes Cognitives is supported by grant FrontCog ANR-17-EURE-0017.
[e] Department of Linguistics, New York University
[***] https://sites.google.com/site/philippeschlenkerhome/
[a] Institut Jean-Nicod (ENS - EHESS - CNRS)
[b] Département d'Etudes Cognitives, Ecole Normale Supérieure
[c] PSL Research University
[**] http://web-risc.ens.fr/~smascarenhas/
[d] LSCP (ENS - EHESS, CNRS)
[*] http://www.emmanuel.chemla.free.fr/

1 Introduction

Semiotics was initially conceived as a general theory of natural signs, be they human or animal, with a focus on their meaning (e.g. Morris 1938). But it is fair to say that initial hopes for a general science of signs have not been fulfilled: few, if any, formal and experimental sciences make sustained reference to its results. By contrast, there is a field, formal semantics, which studies meaning in human language with sophisticated formal tools from logic, and rich descriptive and experimental data from numerous languages (e.g. Maienborn et al. 2011, Gutzmann et al., to appear). But its goals have traditionally been restricted to human language, in fact to spoken languages, with only recent extensions to sign languages.

We argue that formal semantics can and should become a general theory of signs, and come to encompass aspects of human communication (such as gestures and facial expressions) that were traditionally left outside its purview. And it should be broader still, encompassing non-linguistic systems such as animal communication, music, and even reasoning and concept manipulation, where the form (the 'language of thought') must be inferred on indirect grounds.

We will sketch initial results of this broader field of 'Super Semantics': it makes it possible to investigate diverse meaning-bearing forms with increased precision; it offers an interesting typology of meaning operations in nature; and it draws surprising new connections among these domains, including between human words, animal concepts and general reasoning abilities. To get there, however, we must first introduce key achievements of linguistic semantics; each will serve to illuminate empirical domains that go beyond standard linguistics.

2 Achievements of formal semantics

Can one find rules that systematically explain how sentences of natural language convey information? This is the program of contemporary formal semantics, which despite initial quibbles has been closely associated with studies of formal syntax. In brief, the Chomskyan revolution of the 1950's and 1960's showed that human languages can be analyzed as formal languages. Following work by Montague and many others, the next step was to treat human language as a formal *interpreted* language, and thus to systematically associate syntactic rules with semantic (i.e. meaning) rules. But what is a semantic rule? The key insight, from the philosophy of language, was that *to know the meaning of a sentence is to know under what conditions it is true*. As a result, the key problem was to systematically associate truth conditions to sentences of arbitrary complexity. This reduced the problem to one that had been solved much earlier for logical languages: Tarski (1935, 1944) had shown how to define systematic truth conditions for mathematical logic. The challenge was thus to assess whether, and how, these logical methods could be extended to human languages.

This multi-generation enterprise has yielded systematic answers to three main questions, which we review in turn: (1) how is information about the world organized into words? (= semantics in the strict sense (2) how is this information connected to the beliefs of the speech act participants? (= pragmatics) (3) how are semantic and pragmatic operations acquired and processed in real time? (= psycholinguistics of meaning).

2.1 Semantics: what are the elementary operations by which language produces information?

The key idea of *compositionality* is that the meaning of a sentence is computed on the basis of the meaning of its elementary parts and the way they are put together (in other words, their syntax). Meaning is thus derived from two components, pertaining to the meaning of elementary expressions, and their mode of composition.

2.1.1 Expressive power and constraints on possible meanings

Elementary expressions may be closed-class or open-class, but in both cases there are systematic constraints on their semantic behavior. Formal semantics has particularly studied constraints on closed-class (= grammatical/logical) expressions, such as the determiner *some* in (1)a . Its contribution is partly similar to that of the quantifier *for some x* (written as $\exists x$) in logic. In this case, the truth conditions of the English sentence come close to those of the logical formula in (1)b, which belongs to Predicate Logic, the most common logic used to formalize mathematics and science in general.

(1) a. Some senator is honest.
b. ∃x(senator(x) and honest(x))

But as soon as further cases are considered, this analysis falls short. *Most* as in *most senators are honest* cannot in general be analyzed in terms of Predicate Logic: natural language is a more expressive system, one with 'generalized' quantifiers – not just *some* and *every* as in Predicate Logic, but also *most*, *no*, *exactly one*, etc. Natural language quantifiers differ from quantifiers of Predicate Logic in being more diverse, but also in being restricted by a Noun Phrase: *Most senators are honest* cannot be defined in terms of *Most things in the universe satisfying certain properties* (e.g. it won't do to treat it as: *most things are senators and are honest* – an obvious falsehood; other paraphrases fail as well in the general case). Work in formal syntax following the Chomskyan revolution sought to situate precisely the syntactic complexity of human language within a hierarchy of formal languages of increasing sophistication. The same enterprise has been conducted in logical semantics, but on the meaning side: semanticists can provide a reasonable approximation of the expressive power of human languages relative to all sorts of formal languages (e.g. Keenan and Westerståhl, 2011).

A panoply of constraints on natural language semantics have been discussed. One is that **elementary grammatical words are either positive or negative** (Barwise and Cooper 1981). Positivity and negativity are defined in terms of entailments: from *Sam is in Paris,* we obtain a valid entailment by replacing *in Paris* with the less restrictive property *in France*: this is the hallmark of a positive expression (here: *Sam is* __). By contrast, *Sam is not* ___ licenses an entailment in the opposite direction: from *Sam is not in France*, one can infer that *Sam is not in Paris*, as seen in (2):

(2)	Positive operators:	**Sam is** in Paris	=>	**Sam is** in France
	Negative operators:	**Sam is not** in France	=>	**Sam is not** in Paris

Similarly, *some* as in *some senator* __ is positive, *no* as in *no senator*__ is negative. By contrast, *exactly one* __ is neither positive nor negative: *exactly one archbishop is in France* doesn't entail that *exactly one archbishop is in Paris* (e.g. there might be zero in Paris and one in Lyons); and conversely, *Exactly one archbishop is in Paris* doesn't entail that *Exactly one archbishop is in France* (e.g. there might be one in Paris and another one in Lyons). This conforms with the constraint: *some* and *no* are elementary expressions but *exactly one* (two words) isn't. This illustrates one of the simplest cases of a semantic universal, which puts constraints on the possible meanings of natural language elementary expressions.

Further constraints on the meaning of open-class words like *chair* have been investigated as well. **Gärdenfors (2004, 2014) proposes that one constraint on lexical expressions is connectedness**. Formally, the denotation of a word is connected if for any two elements in it, any elements that are between those two are also in the denotation. Take the word *animal*, with the assumption that the duck and the badger are instances of animals. If one has reason to think that the platypus is intermediate between ducks and badgers, one will infer by connectedness that the platypus is an animal. Interestingly, this notion has recently been unified with a constraint on quantifiers: it can be seen as a generalization of monotonicity, i.e. the property of being positive or negative in the above sense.[1] Strikingly, we will see that these constraints have precursors in non-human animals .

2.1.2 Variables

Relative to its logical machinery, there are further respects in which natural language both resembles and differs from Predicate Logic. Variables, often notated as *x, y, z* in mathematics, are thought to be real yet unexpressed in spoken language.[2] To illustrate, *Sarkozy told Obama that he would be elected* is ambiguous between a reading on which *he* refers to Sarkozy and one on which it refers to Obama. Many linguists have argued on indirect grounds that disambiguation is effected by way of invisible variables, as illustrated in (3)a: *he* may refer to *Sarkozy*$_x$ or *Obama*$_y$, and disambiguation is effected because *he* is mentally processed with the variable *x* or *y*. (3)b displays the same type of disambiguation, but the variables are now dependent on ('bound by') quantifiers.

[1] Specifically: a quantifier *Q* is monotonic if Q and its negation *not Q* are connected (Chemla et al. 2019a), and conversely a quantifier Q is connected if it is a conjunction of monotonic quantifiers (Thijsse 1983).
[2] Note that there exist alternative, 'variable-free' analyses, for both spoken and sign language (e.g. Barker and Jacobson 2007; Kuhn 2016).

(3) a. **Sarkozy**$_x$ told **Obama**$_y$ that he$_x$ / he$_y$ would be elected.
b. **[A representative]**$_x$ told **[a senator]**$_y$ that he$_x$ / he$_y$ would be re-elected.
c. **[Some senator]**$_x$ is honest. She$_x$ opposed the proposal.

Once a framework with variables was in place, semanticists uncovered some differences between natural language variables and those of Predicate Logic. In logic, a quantifier such as $\exists x$ in $\exists xF$ can only control variables x that appear in the immediately following formula F. By contrast, in (3)c the quantifier *[some senator]*$_x$ controls a variable x (on *she*$_x$) that appears in another sentence altogether. To allow for such dependencies, non-standard logics called 'dynamic logics' have been developed for natural language.

Positing abstract variables was initially motivated indirectly by their explanatory force. Remarkably, variables can arguably be overt in sign languages, namely as positions in signing space, called 'loci'. In (4), *past senator* co-occurs with a pointing sign (or 'index') *IX-a* towards position *a,* on the right, and *current senator* co-occurs with pointing sign *IX-b* towards position *b*, on the left. The two pronouns in the second sentence can then be fully disambiguated: *IX-a* refers to the former senator, *IX-b* refers to the current senator. The positions *a* and *b* play the role of the variables *x* and *y* in (3), disambiguating the meaning of pronouns and making overt the relation between a variable and an existential quantifier in a previous clause; the latter point also strengthens the case of dynamic logics (Schlenker 2011).

(4) IX-1 KNOW **[PAST SENATOR PERSON] IX-a** IX-1 KNOW **[NOW SENATOR PERSON] IX-b**.
IX-b SMART BUT **IX-a** NOT SMART.
'I know a former senator and I know a current senator. He [= the current senator] is smart but he [= the former senator] is not smart.' (ASL; 4, 179; Schlenker 2014)

Strikingly, we will see that variables (in their linguistic version) have been argued to play a semantic role in pictorial narratives and possibly even in music.

2.1.3 Compositionality

How can elementary meanings be combined? This question has often received a relatively uniform semantic answer: a combination of two expressions is interpreted by treating one expression as a function and the other as an argument. For instance, in the sentence *[some senator][is honest]*, *[is honest]* has as its meaning the set of people who are honest. And *[some senator]* has as its meaning a function f from sets x to truth values, with f(x) = true just in case x contains at least one senator.

This is a versatile framework that extends to other quantifiers, including ones that are not found in Predicate Logic. To illustrate, the meaning of *most senators* can be taken to be a function g for which g(x) = true just in case x contains more than half the senators. What matters for our purposes is that this function-argument mechanism is very different from the mere addition of meanings by way of concatenation (= juxtaposition), of the sort we could get for: *It's raining. It's cold.* Here we can treat each sentence as making an independent contribution: the two meanings can be conjoined without further ado. This strategy won't normally work when function-argument combinations are involved (e.g. it won't do to treat *some senator is honest as: There is some senator. There is honesty*). Importantly, we will see that concatenation appears to suffice to combine monkey calls.

2.2 Pragmatics: the inferential typology of language

Contemporary semantics has discovered that language doesn't just convey information but simultaneously comments on its status relative to the beliefs of the speech-act participants, whence a surprisingly rich typology of inferences and a new question: **how is information divided among the slots of this inferential typology**?

2.2.1 Implicatures

To illustrate, let us go back to *Some senator is honest*. It makes two contributions: its literal meaning is that at least *one senator is honest*. But in addition, the sentence strongly suggests that *not all are*. This cannot be due to the literal meaning of *some* because it is no contradiction to say: *Some senator is honest – in fact all are*. The inference is defeasible, and sometimes absent (e.g. an answer *no* to *Is some senator* *honest?* denies that *any* is, not that *some but not all are*). A standard analysis (following Grice 1975) involves competition between *some* and *every*: the statement *every senator is honest* is more informative, hence if the speaker didn't use it, chances are that this was because she took it to be false. This inference, called an implicature, is crucially based on an Informativity Principle:

(5) If S and S' are competing utterances, and if S' is more informative than S, prefer S' over S unless S' is false.

As we will see, implicatures are found throughout human communication (including gestures). More surprisingly, versions of them have been argued to illuminate the meaning of several monkey calls.

2.2.2 Presuppositions

The typology of linguistic inferences doesn't end there. *Mary stopped/continued/regretted smoking* triggers the inference that Mary smoked before. This is called a presupposition because it is presented as already established in the discourse. Unlike the at issue entailment from *Mary is in Paris* to *Mary is in France*, it is preserved in questions and in negative statements, as shown in (6):

(6) a. Did Mary stop/continue/regret smoking? => Mary smoked before
b. Mary did not stop/continue/regret smoking. => Mary smoked before

Presuppositions are ubiquitous. If I ask about someone in the distance: *Is she approaching?*, I am asking whether the person is approaching but presupposing that this person is female (unlike in the question: *Is this a woman approaching?*). Thus the gender specifications of pronouns generate presuppositions. While presuppositions are often thought to be hard-wired in the meaning of words, we will see that new gestures and even visual animations can generate presuppositions. This suggests that a productive rule lies at their source: presuppositions need not be encoded in words.

2.2.3 Supplements

A fourth inferential type pertains to supplements, which are typically triggered by non-restrictive relative clauses. Unlike presuppositions, they are presented as informative. But unlike at issue entailments, they trigger the same inferences as independent clauses even when they are embedded under logical operators. They thus contrast with embedded conjunctions, hence (7)a,b but not (7)c trigger the inference that *if Ann lifts weights, this will adversely affect her health.*

(7) a. If Ann lifts weights, which will adversely affect her health, we should talk to her.
b. If Ann lifts weights, we should talk to her. This will adversely affect her health.
c. If Ann lifts weights and this adversely affects her health, we should talk to her.

We will see that supplements too can be generated with new gestures or visual animations, ones that follow the words they specify.

2.3 Psycholinguistics of meaning: processing and acquiring meaning

The last question is **how semantic operations are put to actual use, and how they are acquired by children**. In its early days, formal semantics repudiated questions of mental reality (Thomason 1974). Later developments disavowed this position. Here the case of implicatures offers an enlightening case study. Their formal description suggests that they should be an add-on to the literal meaning of words, obtained by applying the Informativity Principle. This could be expected to take time, and processing studies have shown that this is indeed the case. In addition, their derivation requires a comparison between an utterance and its competitors, and finding the competitors might be expected to be difficult. This arguably accounts for an important finding: young children often fail to compute implicatures where adults do, although when given an explicit choice among alternatives they avoid the under-informative one (Chierchia et al. 2001). The key lesson is that semantic theories make predictions about language acquisition and processing, highlighting the fact that semantics is part of cognitive science.

In sum, contemporary semantics offers a precise description of the formal properties of elementary words, of composition rules, of the enrichment and division of meaning among inferential types, and of its real time processing and acquisition. One key question within the field is which aspects of semantic information are encoded in words, and which are due to productive rules; investigating new, sometimes invented meaning-bearing forms (such as gestures) will help address the productivity question. More broadly, we would want to determine how these semantic operations compare to those found in other semantic systems in nature, and what we can learn from the comparison.

3 Semantics before language: primate communication

3.1 Primate semantics and pragmatics: the import of formal models

Far simpler semantic systems are found in non-human primates. Contemporary primatology has established that some calls and gestures are naturally produced in some situations but not in others (e.g. Zuberbühler 2009, Byrne et al. 2017). Furthermore, field experiments have shown that the monkeys themselves know this correlation and thus derive information from the calls they hear. Under any name ("appropriate" vs. "inappropriate" or "true" vs. "false"), this means that these signals have a semantics – which is unsurprising since they clearly convey information.

Applying the methods of formal semantics to them has made two contributions: (i) it has clarified the space of possible theories by stating fully explicit analyses about (a) the meaning of elementary calls and (b) the ways they are combined; (ii) it has generated new potential explanations, especially about the division of labor between semantics and pragmatics, leading to the hypothesis that there are 'animal implicatures', cases in which the meaning of a call is enriched by the assumption that it was the most informative call that could be appropriately used. The main results are as follows:

(8) a. Elementary calls can often be assigned simple and explanatory meanings.
b. Call combination can usually be analyzed without structure, by treating each call as a separate utterance reflecting the caller's information as it is produced: concatenation appears to suffice. There is one plausible case of compositional composition at the level of call suffixes.
c. Call meaning might be enriched by an animal version of the Informativity Principle, although this hypothesis needs to be tested further.

3.2 Call meaning vs. pragmatic enrichment: Campbell's monkeys[3]

A rich example pertaining to the meaning of individual calls is afforded by Campbell's monkeys of the Tai Forest (= (9)). Male adults have a non-predation-related call, *boom*. They use a call *krak* to raise leopard alerts, and *hok* for raptor alerts. Remarkably, they also have suffixed calls: *krak-oo* is used for unspecific alerts, and *hok-oo* for non-ground disturbances. The challenge is to assign meanings to *boom, krak, hok*, and *-oo*. Further complexity is added by Campbell's call use on Tiwai Island, where leopards haven't been seen for decades: the Tai calls are used, but *krak* raises unspecific alerts (as does *krak-oo*), rather than leopard alerts. Should we conclude that call meaning is subject to a kind of dialectal variation – as it is for *pants* in American English (meaning "trousers") vs. British English (meaning "underpants")?

(9) **Male Campbell's monkey calls**

a. Description

CAMPBELLS MONKEYS

Call	Typical situations
boom boom	non-predation alert
hok	presence of an eagle
krak	*Tai:* presence of a leopard *Tiwai:* unspecific alert
hok-oo	alert from above
krak-oo	unspecfic alert

b. Analysis

Literal meanings

boom boom	non-predation alert
hok	non-ground alert
krak	alert
R-oo	weak *R*-alert

Informativity Principle
"Prefer more informative expressions!"

c. Results of call competition

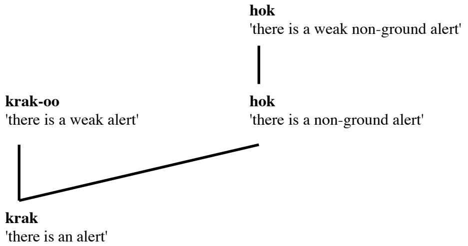
Calls	Competitors	Enriched meanings
hok	hok-oo	serious non-ground alert
krak	krak-oo, hok	*Tai:* alert, serious, ground *Tiwai:* useless enrichment, hence literal meaning only

Schlenker et al. 2014 propose an analysis without dialectal variation, but with a crucial use of the Informativity Principle (see (5)). First, they take *krak* to trigger general alerts, and *hok* to trigger non-ground alerts. Second, in order to analyze the meaning of the suffix *-oo*, they assume that if *R* is *krak* or *hok, R–oo* indicates a weak alert of the R-type. Thus *hok-oo* indicates a weak *(-oo)* non-ground (*hok*) alert – which is more informative than *hok*.

Finally, they use the Informativity Principle, with the informativity relations in (10):

[3] Sections 3.2 and 3.3 borrow from Schlenker et al. 2016c.

(10) **Informativity relations among Campbell's calls (for two calls that are linked, 'higher' means 'more informative')**



Hok competes with other calls, and because *hok-oo* (pertaining to *weak* non-ground alerts rather than to any non-ground alert) is more specific, the meaning of *hok* is enriched to *hok but not hok-oo*: it only applies to aerial (*hok*) non-weak *(not hok-oo)* alerts – hence the raptor uses. By the same logic, the unspecific alert *krak* competes with *krak-oo,* but also with *hok*. Due to this competition with two more informative calls, in the end *krak* can only be used for *serious (not krak-oo) ground (not hok) disturbances*. This comes very close to the leopard uses observed in Tai. Importantly, while in human language the Informativity Principle is often taken to derive from more general rules of communicative rationality, it might conceivably be hard-wired and automatic in non-human primates.

3.3 Call combination vs. context change: Titi monkeys

The calls of the Titi monkeys of South America make a different point: not everything that looks like a complex syntax/semantics interaction requires complex composition rules.

The initial puzzle is that with two calls (*A* and *B*) re-arranged in various ways, Titi monkeys can provide information about both predator type (cat, raptor) and predator location (on the ground, in the canopy). Simplifying somewhat, and writing X^+ for a series of iterations of call X, B^+ is used for non-predation alerts and for situations involving a cat on the ground, while $A\ B^+$ is used in situations in involving a cat in the canopy. A raptor on the ground gives rise to an A^+B^+ sequence, while a raptor in the canopy triggers an A^+ sequence. The main patterns are summarized in (11). Should we conclude that these sequences have a complex syntax/semantics interface, as hinted by Cäsar et al. 2013? Or should they be treated as very long idioms, with no internal semantics?

(11) **Titi monkey calls**
a. Description b. Analysis c. Results of call competition

TITI MONKEYS

Call	Typical situations	
B^+	non-predation alert	
B^+	cat	on the ground
A B^+	cat	in the canopy
A^+ B^+	raptor	on the ground
A^+	raptor	in the canopy

Literal meanings

A	serious non-ground alert
B	alert

Informativity Principle
"Prefer more informative expressions!"

B^+ correspond to ground or weak alerts (or alerts that *have become* weak alerts)

(Raptors in the canopy remain serious threats even after having been signaled by As)

Due to their length and slow time course, it is unlikely that these sequences are interpreted non-compositionally as idioms because hearers would need to wait for too long for the meaning of the message to be effective. A simpler analysis has been explored, in which each call is in effect an independent utterance and thus contributes its meaning independently from the others (Schlenker et al. 2016d). Since the B-call is used in predatory and non-predatory situations alike, one may take it to trigger an unspecific alert. In field experiments, the A-call triggers a 'looking up' behavior, and thus one can posit that it is indicative of *serious non-ground alerts*. These assumptions explain why one finds B^+-sequences (= series of B-calls) in 'cat on the ground' situations, and A^+-sequences in 'raptor in the canopy' situations. Importantly, the Informativity Principle is important to explain why the B-call *fails* to be used at the beginning of sequences in the latter case (= it would be under-informative).

But why does one find A^+B^+ in 'raptor on the ground' situations? A remark about hunting techniques proved suggestive: raptors on the ground usually attack by *flying*, hence the serious non-ground alerts A^+. Still, being on the ground isn't a typical hunting position, and after a while the alert stops being serious, which only leaves B as a possibility. In 'cat in the canopy' situations, one finds AB^+

sequences, possibly because a serious non-ground danger is indicated, which then transitions to a weaker danger because a cat becomes less dangerous after detection (Zuberbühler et al. 1999).

On this view, which is still the subject of active debates,[4] the apparent complexity of Titi sequences might reflect the interaction between simple meanings and the evolution of the contextual environment as the sequence unfolds, rather than a complex syntax/semantics interface or very long idioms.

In sum, the Campbell's and Titi studies illuminate several points: (i) they clarify the possible meanings of individual calls; (ii) in the case of Campbell's -oo, they make it plausible that there is a tiny composition rule at the level of a suffix; (iii) in both species, there seem to be general alert calls, but their use is constrained by the Informativity Principle; (iv) in all cases, a call can be treated as a separate utterance, and no non-trivial composition rules are needed at this point – and in particular no call-external application of functional application, as in human language.

3.4 New questions about the evolution of animal meanings

The investigation of primate signals has another benefit: in some cases, it yields insights into the evolution of animal meanings. As an example, *booms* are non-predation-related calls present, not just in Campbell's monkeys, but also in many cousin subspecies of the family cercopithecines. Inspection of their distribution is strongly indicative of their presence in the most recent common ancestor of entire subgroups: *booms* probably existed several million years ago (Schlenker et al. 2016a).

Turning to ape gestures, Byrne et al. 2017 argue that 36 gestures are shared among the great apes, and likely part of an innate repertoire. On the form side, Hobaiter and Byrne 2017 argue on that the similarity of the repertoires is unlikely to be due to limitations of the articulatory possibilities for gestures. On the meaning side, Byrne et al. 2017 argue that partial similarities in gesture use among bonobos and chimpanzees are also unlikely to be due to chance.

Strikingly, ape gestures are an area in which a genuine connection with human communication might be non-speculative, as argued by Kersken et al. 2018. They uncover a human infant gestural repertoire with 50 gestures (96%) that are shared with apes A further step will be to connect these infant gestures with adult human gestures, and/or with signs. Yet another question will be to assess the extent of iconicity in ape gestures, discussed for pointing and pantomime by several groups (e.g. Genty and Zuberbühler 2014, Douglas and Moscovice 2015).

Call evolution raises questions of its own, in particular the following: under what conditions can the simultaneous presence of specific and general calls be stable? Steinert-Threlkeld et al. 2019 show that under broad modelling conditions, general calls are unstable: a Titi-style general B-call competing with a specific A-call meaning *serious non-ground alert* should lead in later generations to a division of lexical labor, with the B-call meaning not-A, i.e. *non-[serious non-ground] alert*. If so, the Informativity Principle won't have anything to operate on, since call meanings will never be in a specific-to-general relation. To block Steinert-Threlkeld et al's results, however, we might surmise that *non-[serious non-ground] alert* is not a possible concept because it does not correspond to a natural class of objects. But showing this requires an independent study of constraints on animal concepts, a point we revisit in Section 7.

4 Integrating logical and iconic semantics

Iconicity is an important way of producing meaning in spoken and especially signed languages. A key challenge is to understand, both empirically and formally, how iconic and compositional semantics interact; this will have benefits for the analysis of meaning in speech, signs and gestures.

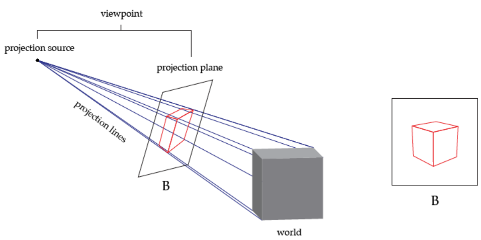
4.1 Pictorial semantics

Pioneering work by Greenberg (2013, 2019) on the formal semantics of pictures clarifies the workings of iconicity. The semantic content of a picture is obtained by asking which situations could give rise to

[4] Berthet et al. 2019a,b fail to replicate the AB^+ sequences in 'cat in the canopy' situations, finding instead B^+, with A's interspersed; they also develop a somewhat different theory.

the markings found on the picture given the projection method used. One among several is perspective projection, illustrated in (12) (acoustic iconicity would have to be handled by different principles).

(12) Perspective projection (Greenberg 2019)



This naturally gives rise to a definition of truth for pictures: a picture P is true of those situations that can project onto P relative to the viewpoint and the system of projection used.

(13) A picture P is true in <u>situation</u> w relative to viewpoint v along the system of projection S iff w projects to P from viewpoint v along S, abbreviated as: $\mathrm{proj}_S(w, v) = P$

Abusch 2013, 2015 uses this framework to account for the semantics of silent comics, a point to which we return in Section 6.1. But iconic semantics also interacts in interesting ways with compositional semantics in sign languages, to which we now turn.

4.2 Logical visibility and iconicity in sign language

Sign languages have been argued to display the same general grammatical and semantic properties as spoken languages (e.g. Sandler and Lillo-Martin 2006), and sometimes to express aspects of Logical Forms in a more transparent fashion, as we saw in Section 2. But they also make greater use of iconicity than spoken languages to modulate the form of conventional words. A simple example is the verb *GROW* in ASL, which is (conventionally) realized as shown by the pictures in (14)a. But it can also be modulated in iconic ways by making the endpoints more or less broad, and by realizing the sign more or less quickly, with unmistakable semantic effects: the broader the end points of the sign, the larger the final size of the group; and the more rapid the movement, the quicker the growth process.

(14) POSS-1 GROUP GROW_ 
'My group has been growing.' (ASL, 8, 263; Schlenker et al. 2013)

This phenomenon is not unique to sign language, however: the adjective *long* can be modulated by lengthening the vowel to evoke a greater duration (and doing the same thing with *short* is… odd). Here the rule of iconic modulation seems to be that *the longer the vowel, the greater the duration*. Due to the representational possibilities of the signed modality, these iconic modulations are just far richer in sign than in speech. In addition, it has been suggested that for *GROW* and *long* alike, iconic modulations make an at issue contribution. Thus *loooong* means very much the same thing as *very long*, and this is the content that gets denied by a negation, as in: *The talk wasn't loooong* (see Schlenker 2018b for ASL *GROW*).

Iconicity doesn't just interact with lexical material: grammatical expressions such as pronouns can be modulated in a highly iconic fashion as well. Just like the feminine specification of *she* triggers a presupposition that the denoted person is female, pronouns in ASL (American Sign Language) and LSF (French Sign Language) can have high specifications (realized by pointing upwards) that trigger the presupposition that the denoted person is tall, powerful or important. Furthermore, these modulations can display 'iconicity in action': one typically points towards the part of the representation corresponding to the head, with the result that when the denoted individual is presented as rotated in various position, the loci get rotated as well. The conclusion is that sign language loci can simultaneously be variables and simplified pictures of their denotations (see Liddell 2003, Schlenker et al. 2013, Schlenker 2014, Schlenker 2018a).

A striking case of interaction between grammar, logical semantics and iconicity involves repetition-based plurals. In ASL, one can optionally realize a plural, such as that of the word *TROPHY*, by repeating the noun. This yields standard readings of English-style plurals. But simultaneously, the shape of the repetitions can provide iconic information about the arrangement of the denoted objects (Schlenker and Lamberton 2019a): one may repeat *TROPHY* in a horizontal or triangular shape to signify that the trophies are arranged on a line or as triangle. But there is more: the repetitions can be easy to count, with clear separations between them ('punctuated'), or hard to count, without clear separations ('unpunctuated'). In the former case (*TROPHY TROPHY TROPHY)*, they typically denote as many objects as there are iterations. In the latter case, three unpunctuated (and thus hard to count) iterations, transcribed as *TROPHY-rep3,* may mean *several* (typically at least three) *trophies*. Strikingly, the existence of repetition-based plurals as well as the distinction between punctuated and unpunctuated ones exists in several sign languages, and was described in a homesigner (deaf person who grew up without access to sign language) who appeared to have invented this device, as it was absent from the gestures of his hearing mother (Coppola et al. 2013). Hearing non-signers might also understand related distinctions in repeated gestures (Schlenker, to appear a).

What should we conclude from the existence of repetition-based plurals in such diverse situations? One possibility is that, in this area, 'Universal Grammar' doesn't just specify the syntactic and semantic behavior of plurality, but also its phonological realization (to wit, as repetitions). An alternative is that repetitions are an iconic device, which is exploited for grammatical purposes (Schlenker and Lamberton 2019b). Punctuated repetitions are easy to analyze within an iconic semantics: *TROPHY TROPHY TROPHY* arranged on a line are a simplified pictorial representation of the denoted group, as represented in (15). Importantly, the projection rule is more complex than in Greenberg's pictorial semantics because the conventional word *TROPHY* need not resemble a trophy: the arrangement of the iterations is iconic, but the individual words need not be.

(15) TROPHY TROPHY TROPHY denotes 

The case of unpunctuated repetitions is harder: why can *TROPHY-rep3* (= three hard to count iterations) come to have an 'at least several trophies' meaning? This is surprising in that an 'approximately three' reading would make more iconic sense. Schlenker and Lamberton 2019b propose a theory that combines three components: they start from a Greenberg-style pictorial semantics but add a vague component (the 'blurry' component of unpunctuated repetitions), and then allow for a pragmatic mechanism to work on top of it. Importantly, their analysis applies both to words and to purely iconic, gestural representations within ASL. In simple cases, a blurry picture such as that on the left in (16) may be vague between the two representations on the right, thus being true of situations in which there are three or four trophies.

(16)  can correspond to  or to 

There are multiple ways to resolve the vagueness, though some may be more likely than others. On one, the three blurry iterations can stand for three or four objects (as in (16)), on another, three or four or five, etc. There is also one on which the blurry iterations can stand for three or four or five or six or…, in other words for at *least three objects*. While this is an extremely unlikely possibility, it is exploited for communicative purposes because there is a strong pressure to find some representation to express an 'at least three' reading, and no iconic representation does it better than the three blurry iterations. In the end, pictorial semantics, pictorial vagueness and communicative rationality conspire to give rise to an 'at least three' reading. Irrespective of the final analysis, this is a case in which a detailed semantic analysis has no choice but to have an explicit iconic component.

In sum, the interplay between compositional semantics and iconicity is a key issue in sign language semantics/pragmatics (aspects of spoken language also benefit from iconic analyses, e.g. Dingemanse 2013). Strikingly, since sign languages have the same general grammatical and logical resources as spoken languages, but make greater use of iconicity, sign is along some dimensions more expressive than speech.

4.3 Logical and grammatical structures in gestures[5]

Gestures offer a prime example of iconicity, but they also have structure. While they have nothing like the sophisticated grammar of sign languages, they sometimes have a proto-grammar that is reminiscent of it: non-signers appear to know constraints that track some sign language rules standardly classified as 'grammatical', and play a particularly important role in linguistic semantics.[6] In order to abstract away from the special semantic issues raised by co-speech gestures, which are in some ways parasitic on the spoken expressions they enrich (a point we revisit in Section 5.1), we focus for the moment on gestures that fully replace some words (henceforth 'pro-speech gestures').

Two examples will make this line of research concrete. In (17), The expression *a mathematician* is pronounced with an open hand (palm up) on the right (transcribed as *IX-hand-a*, and preceding in the transcription the co-occurring expression, which is boldfaced), while the expression *a sociologist* co-occurs with an open hand on the left (transcribed as *IX-hand-b*). With these loci in place, a pointing gesture can fully replace a pronoun that would be expected after *pick*: an index towards the right (transcribed as *IX-a*) serves to refer to the mathematician, associated with locus a.

(17) Whenever I can hire IX-hand-a **[a mathematician]** or IX-hand-b **[a sociologist]**, I pick IX-a.
Meaning: whenever I can hire a mathematician or a sociologist, I pick the former.

Importantly, in this case *him* or *her* could be ambiguous between the two antecedents, whereas the pointing gesture isn't: the gesture is not just a code for a pronoun.

A second case, involving grammatical agreement, was studied with experimental means. In ASL, some verbs include loci in their realization and are for this reason called 'agreement verbs'. For instance, *I give you* could be realized with a movement going from the signer to the addressee (it is transcribed as *1-GIVE-2*); *I give him* starts from the signer's position and targets a third person locus, for instance *a* on the right – in which case it is transcribed as *1-GIVE-a*. These loci have been argued to display the behavior of agreement markers, although alternative analyses have been offered as well (Liddell 2003; Lillo-Martin and Meier 2011; Pfau et al. 2018 and Schembri et al. 2018). Irrespective of the final analysis, these agreement verbs appear to have gestural counterparts (Schlenker and Chemla 2018).

To have a point of comparison, let us consider the ASL examples in (18), constructed around the agreement verb *1-GIVE-2* or *1-GIVE-a*.

(18) POSS-2 YOUNG BROTHER MONEY IX-1 **1-GIVE-a**.
'Your younger brother, I would give money to.
a. IX-2 IX-1 NOT.
You, I wouldn't.'
b. *IX-2 IX-1 NOT **1-GIVE-a**.

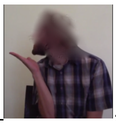
In the first sentence of (18), the verb *GIVE* displays object agreement with a third person locus *a*, corresponding to the younger brother, hence: *1-GIVE-a*. Now the continuation in (18)a involves a missing Verb Phrase, which in spoken and sign language alike usually entails copying an antecedent. But something happens in the copying process: the agreement markers can be disregarded, which explains why the continuation (18)a is acceptable even though that in (18)b (with overt copying) isn't. Similar rules of partial copying under ellipsis are attested in English, as in (19), where the third person features and the feminine features of *her* are disregarded under ellipsis, and thus *I did too* has a reading that *I too did her homework* lacks.

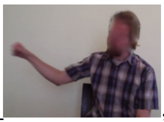
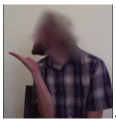
[5] Our discussion follows Schlenker, to appear a.

[6] Needless to say, this implies in no way that sign languages are 'merely' gestural: their sophisticated grammars have been described in great detail, and share multiple properties with those of spoken languages (see Sandler and Lillo-Martin 2006). Rather, the argument is that despite their expressive limitations, gestures have a proto-grammar reminiscent of sign language.

(19) [Uttered by a male speaker] In my study group, Mary did her homework, and I did too.
can mean: I too did **my** homework (Schlenker and Chemla 2018)

The ASL contrasts can be replicated with gestural verbs in English. As in ASL, a movement towards the side has to correspond to a third person object, and thus the second clause of (20)b is deviant because the object is second person but the object agreement is third person (the kisses ought to be sent towards the addressee). Strikingly, this problem doesn't really arise in (20)a: as the missing verb is copied, its third person object agreement is disregarded, just as is the case for object agreement in ASL. Importantly, there are no comparable cases of object agreement in English: subjects had to infer an ASL-style rule on the basis of no or extremely limited evidence.

(20) a. Your brother, I am gonna SEND-KISSES-3_, then you, too.

b. *Your brother, I am gonna PUNCH-3_, then you, I am gonna SEND-KISSES-3_
.

It has been argued that several further properties of sign language grammar can be replicated with pro-speech gestures (Schlenker, to appear a). If confirmed, these results might be important to understand the historical origins of sign languages. It is noteworthy that homesigners, who grow up without access to sign language, develop gestural languages that share some properties of sign languages, but are expressively far more limited (e.g. Abner et al. 2015, Goldin-Meadow 2003). In some cases, the reason homesigners discover such properties on their own might be that, more generally, non-signers 'know' them.

Second, an important question for future research is to determine how these instances of 'zero-shot' grammatical learning can arise. One possible view is that Universal Grammar does not just specify the abstract form of grammatical rules, but also part of the mapping between forms and grammatical/semantic content: a pointing sign/gesture might thus be intrinsically endowed with pronominal properties. Another possible view is that some signs/gestures are naturally associated with a fixed grammatical/semantic component for deeper cognitive reasons. This debate is currently open.

5 Semantics beyond words I: iconic elements within the inferential typology

We turn to the place of iconic meanings in the typology of linguistic inferences. Here an important distinction must be drawn between iconic enrichments, which modulate the form and meaning of a word, and iconic replacements, which fully replace a word. Iconic enrichments have a special status due to their parasitic nature as add-ons to the main message. Iconic replacements make a different but powerful point: they fall within the very same typology as normal words, which provides new insights about the cognitive origin of this typology.

To introduce the typological issue and the terminology, let us consider the examples in (21). (21)a involves a slapping gesture that co-occurs with the verb *punish*. In (21)b, a post-speech slapping gesture appears after the Verb Phrase it modifies.[7] In (21)c, the slapping gesture fully replaces the verb and is thus a pro-speech gesture. Finally, in (21)d, a conventional word, *long*, is modified in an iconic fashion by way of an 'iconic modulation' (which by definition is always the modification of a conventional form). The same terminology is extended to sign language by replacing -*speech* with -*sign*.

[7] See Esipova 2019 for a broader discussion.

(21) a. Co-speech gestures (co-occur with the word they modify [boldfaced])

Asterix will  **punish** his enemy.

b. Post-speech gestures (follow the word they modify)

Asterix will punish his enemy – .

c. Pro-speech gestures (replace a word)

His enemy, Asterix will .

d. Iconic modulations (modify the form of a conventional word)

The talk was loooooong.

We start by considering co-speech gestures, iconic modulations and post-speech gestures, which are iconic enrichments.[8] We then turn to pro-speech gestures.

5.1 Iconic enrichments in the typology of linguistic inferences: sign with iconicity vs. speech with gestures

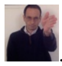
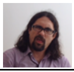
One motivation for investigating semantic enrichments is to assess in greater detail our earlier conclusion that sign languages are, along some dimensions, more expressive than spoken languages because they make greater use of iconicity. This conclusion might be premature because it fails to take into account the means of iconic enrichment afforded to speech by gestures: sign with iconicity should be compared to speech with gestures rather than to speech alone (Goldin-Meadow and Brentari 2017). But *even* when gestures are taken into account, systematic differences arguably remain between sign with iconicity and speech with gestures. The reason has to do with the inferential typology discussed in Section 2. The most salient means of iconic enrichment in speech lies in co-speech gestures. But it was argued from the start (notably in pioneering work by Ebert and Ebert 2014) that co-speech gestures do not typically make at issue contributions. By contrast, several iconic modulations of signs seem to make at issue contributions. We already mentioned this point in connection with *GROW* above, and the shape modulations of repetition-based plurals apparently make at issue contributions as well (Schlenker and Lamberton 2019a).

Granting that gestures have an iconic component reminiscent of iconic signs, the key question is *how* these iconic enrichments are effected, and distributed across the inferential typology. As we will now see, none of the gestures in (21)a-c have quite the same properties as iconic modulations. *Co-speech gestures* trigger presuppositions of particular sort (called 'cosuppositions'), and thus fail to be at issue, unlike many iconic modulations found in sign language. *Post-speech gestures* trigger supplements, i.e. the same type of meaning as non-restrictive relative clauses (Potts 2005), and thus they too fail to be at-issue. *Pro-speech gestures*, for their part, make at-issue contributions, but unlike signs (including ones with iconic modulations) they are not conventional words at all, and are thus expressively limited for other reasons. The typology is illustrated in (22), and it turns out to be crucial to answer (in the negative, we think) the question whether speech with gestures 'equals' sign with iconicity.

---

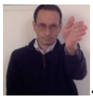
[8] As we will see, post-speech gestures behave semantically as non-restrictive relative clauses, and could thus be taken to be gestures that replace certain words that optional.

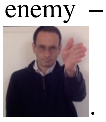
(22) **Typology of iconic enrichments** (after Schlenker 2018b)

	Co-speech/co-sign gestures	**Post-speech/post-sign gestures**	**Iconic modulations**	**Pro-speech/pro-sign gestures**
Meaning	**cosuppositions** (= presuppositions of a special sort)	**supplements** (like non-restrictive relative clauses)	**at-issue or not**, depending on the case	**at-issue**, with an additional non-at-issue component in some cases
Speech	Asterix will  **punish** his enemy.	Asterix will punish his enemy – .	The talk was loooong.	His enemy, Asterix is going to .
Sign	IX-arc-b NEVER  **[SPEND MONEY]**	IX-arc-b NEVER SPEND MONEY]$_b$ – .	POSS-1 GROUP GROW_  	[currently unclear]

To illustrate this typology, we note that the iconic enrichments in the positive sentences in (21) display radically different behaviors under negation, as seen in (23).

(23) a. Asterix won't  **punish** his enemy.
=> if I were to punish my enemy, slapping would be involved

b. #Asterix won't punish his enemy – .

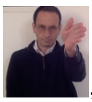
c. His enemy, Asterix won't .

d. The talk wasn't loooong.

(i) First, the co-speech gesture in (23)a triggers an inference that gets inherited across negation, to the effect that *if I were to punish my enemy, slapping would be involved*. Crucially, further tests suggest that this inference behaves like a presupposition; it has received a special name (cosupposition) because the inference is conditionalized on the meaning of the modified expression (here: *punish*). Experimental data (Tieu et al. 2017, 2018a) have buttressed this conclusion, using two methods: truth value judgment tasks, and (more clearly) inferential judgment tasks.
(ii) Second, the post-speech gesture in (23)b is deviant after a negated Verb Phrase. Recent literature (Schlenker 2018b) has argued that this is because the post-speech gesture behaves like a non-restrictive relative clause and contributes a supplement, which is often deviant in such negative environments, as in (24)a.

(24) #Asterix won't punish his enemy, which will involve slapping him.

Consideration of further examples strengthens the similarity with non-restrictive relative clauses: (25)a behaves like (7)a in suggesting that Asterix's action will involve slapping, unlike the control with a conjunction in (25)b.[9] These inferential contrasts were established with experimental means in Tieu et al. 2019, and they extend to cases in which a visual animation replaces the post-speech gesture – a point to which we return in Section 5.2.3.

(25) a. If Asterix punishes his enemy – , I might scream.
=> if Asterix punishes his enemy, slapping will be involved

[9] These inferential facts matter because the deviance of (23)b could be analogized to (i) rather than to (24). But in view of the difference between (25)a and (25)b, the analogy with (i) seems wrong-headed.

(i) #Asterix won't punish his enemy, and this will involve slapping him.

b. If Asterix punishes his enemy and this involves slapping him, I might scream.
≠> if Asterix punishes his enemy, slapping will be involved

(iii) Third, the pro-speech gesture in (23)c makes an at-issue contribution and yields neither a cosupposition nor an implicature. Importantly, however, pro-speech gestures are expressively limited because they are not based on conventional words, unlike the iconic modulations found in sign language: in LSF, the meaning of *UNDERSTAND* or *REFLECT* can be modulated by altering the speed with which part of the sign is realized (e.g. to indicate a quick or difficult understanding or reflection, Schlenker 2018a). It seems hopeless to represent modulations of such abstract concepts with pure gestures.
(iv) Fourth, the iconic modulation in (23)d makes an at-issue contribution and triggers no conditionalized inference akin to cosuppositions.

Importantly, with the exception of pro-sign gestures (i.e. sign-replacing gestures, whose existence and status is still somewhat unclear), the same typology might hold in sign (Schlenker 2018b). A disgusted facial expression co-occurring with the Verb Phrase *SPEND MONEY* (as on the last line of (22)) yields the same cosuppositional behavior as the slapping gesture co-occurring with *punish*:('if x spends money, this is disgusting').[10] The same facial expression could also follow the Verb Phrase, in which case it arguably behaves as a post-speech gesture and plausibly triggers a supplement. Finally, as argued above, the iconic modulations of *GROW* are best compared to those of *loooong*: at issue contributions are made in both cases.

Three conclusions are worth highlighting. First, iconic enrichments can be handled by established semantic mechanisms, although one case (cosuppositions) requires adjustments to presupposition theory. Second, the difference between iconic enrichments in speech and in sign is not one of type: the same abstract typology is found in both modalities. Third, gestural enrichments do not make the same type of (at issue) contribution as iconic modulations. But iconic modulations are arguably rich in sign, and impoverished in speech. This yields systematic differences between sign with iconicity and speech with gestures.

Importantly, the typology in (22) could be expected to apply to further types of enrichments, such as 'vocal gestures' (e.g. Schlenker 2018b). Pasternak 2019 and Pasternak and Tieu 2020 argue that sounds that co-occur with speech behave like co-speech gestures and yield an inferential profile characteristic gestural cosuppositions. Perhaps surprisingly, we will see in Section 6.3 that music that co-occurs with films or cartoons might behave like co-speech gestures in triggering cosuppositions.

5.2 Iconic replacements in the typology of linguistic inferences: replicating the typology without words

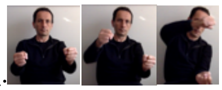
The semantic difference between co-, post- and pro-speech gestures is certainly due to the *manner* in which they are realized, namely as co-occurring, following or replacing a word (Schlenker 2018b and Esipova 2019 for possible explanations). But in addition, recent research suggests that pro-speech gestures alone neatly fill established categories of the inferential typology of language, which includes not just at issue entailments and supplements, but also implicatures and (standard, non-cosuppositional) presuppositions, among others: depending on their informational content, they may make *additional* contributions that reflect inferential types (and probably algorithms) that are found in normal words.

These gestural findings were obtained with experimental means in Tieu et al. 2019. But this paper goes one step further and replicates the typology in paradigms in which gestures are replaced with visual animations that the subjects could not have seen in a linguistic context before. This suggests that subjects divide 'on the fly' new semantic content among established categories of the inferential typology. This, in turn, argues for the existence of productive algorithms that make it possible to do so. For brevity, we discuss just two cases: iconic implicatures and iconic presuppositions.

[10] Importantly, this facial expression is not a grammatical marker, unlike other sign language facial expressions: it seems genuinely gestural.

5.2.1 *Iconic implicatures*

In some cases, the existence of such inferences is expected by current theories. Consider the case of scalar implicatures. In (26), a gesture representing a partial wheel-turning is contrasted with a complete wheel-turning. It can be checked in separate examples that *not* TURN-WHEEL can mean 'not turn the wheel at all' (rather than 'not turn the wheel exactly as depicted', for instance). This suggests that the partial wheel-turning (i.e. TURN-WHEEL) can have a weak meaning, akin to 'turn the wheel'. But as soon as this gesture evokes (thanks to the context) a more informative alternative TURN-WHEEL-COMPLETELY, an implicature is derived: *John will* TURN-WHEEL is understood to mean that John will turn the wheel but not completely, as shown in the target inference in (27)a, which is endorsed significantly more than the control inference in (27)b.

(26) *Context:* John is training to be a stunt driver. Yesterday, at the first mile marker, he was taught to TURN-WHEEL-COMPLETELY_. Today, at the next mile marker, he will TURN-WHEEL _. (Tieu et al. 2019)

(27) a. Target inference: John will turn the wheel, but not completely.
b. Baseline inference: John will turn the wheel completely.

Tieu et al. 2019 show in separate conditions that the gestures involved are unlikely to be mere codes for words because they have iconic implications that mere words would lack, for instance about the size of the wheel. In addition, similar results are obtained when gestures are replaced with artificial visual animations that subjects couldn't have seen before, with similar results. As expected, then, implicature derivation is a fully productive process.

5.2.2 *Iconic presuppositions*

In contrast with scalar implicatures, presuppositions are typically thought to be encoded in the lexical meaning of words (e.g. Heim 1983), although there have been various attempts to propose 'triggering algorithms' that *deduce* the presupposition of an expression from its informational content (see for instance Abrusán 2011 and Schlenker, to appear b for discussion). Strikingly, pro-speech gestures can trigger presuppositions, as can be illustrated by a modification of our TURN-WHEEL example: the question in (28) triggers the inference that Sally is behind the wheel, hence a significantly stronger derivation of the inference in (29)a than in (29)b. Many further tests have been used in the literature to argue that this and other gestural examples genuinely trigger presuppositions (e.g. Schlenker 2019a).

(28) Jake and Lily are watching their four children ride bumper cars at the carnival. Each bumper car has two seats. As one of the bumper cars nears a bend in the track, the parents wonder:

Will Sally TURN-WHEEL_ ?

(29) a. Target inference: Sally is in the driver's seat.
b. Baseline inference: Sally is in the passenger seat, not the driver's seat.

Strikingly, Tieu et al. 2019 show that the generation 'on the fly' of presuppositions from iconic material extends to stimuli that subjects couldn't possibly have seen before. For instance, the construction in (30), which represents an alien changing color from their normal state (green) to their meditating state (blue), triggers an inference that they were not initially meditating. This inference is preserved in the question, a characteristic property of presuppositions (illustrated in (6)a). Further tests and further pictorial representations representing changes of state were used to buttress the conclusion that visual animations too can trigger presuppositions.

(30) **Pictures from Tieu et al.'s videos testing presuppositions generated by visual animations**
(here: a change of state animation pertaining to an alien's antenna turning from green to blue; original video: https://youtu.be/U6dfs-XI2-4)

(green) (green+blue) blue

5.2.3 Further inferential types

The productive nature of implicature and presupposition generation extends to further inferential types, such as supplements: the contrasts illustrated in (25) were obtained with gestures and visual animations in Tieu et al.'s experiment. Further inferential types follow the same logic (Schlenker 2019a, Tieu et al. 2019), suggesting that pictorial content can be productively divided among a rich inferential typology. Migotti and Guerrini 2019 further argue that this result can be extended to auditory stimuli, specifically to pro- and post-speech onomatopoeias and musical content.

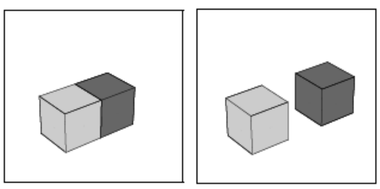
A key question for future research is whether this productive division of semantic content among established slots of the inferential typology only arises when the stimuli are embedded in sentences, or whether some of them might arise in purely visual scenes – a tantalizing if remote possibility.

6 Semantics beyond words II: visual and musical narratives

6.1 Visual narratives

A gulf separates the projection-based semantic of pictures introduced in Section 4.1 from actual pictorial narratives, from comics to films: formal pictorial semantics is in its infancy. But important progress is being made on several fronts. First, pioneering work on the visual language of comics is getting the recognition it deserves, from sophisticated analyses of visual morphology (e.g. reduplication used to evoke movement) to a theory of narrative structure in comics (Cohn 2013). Second, Cumming et al. 2017 investigate in formal detail constraints on viewpoint shift in film, i.e. permissible changes of camera angle. Third, Abusch and Rooth 2017 have upgraded Greenberg's projection-based semantics 'from the ground up' to analyze visual narratives. The simplest example appears in (31), which represents "a short comic of two cubes moving apart" (Abusch and Rooth 2017).

(31) Picture P1 Picture P2



The fact that the two pictures are arranged as a narrative sequence provides information about the situations described, as well as their ordering. In essence, a series of n pictures $<P_1, \dots, P_n>$ is true of n situations $<s_1, \dots, s_n>$ just in case the s_i's are temporally ordered in the right way, and each s_i projects onto the corresponding P_i, as stated in (32), which is a temporal version of (13).

(32) **Picture sequences true of n situations (after Abusch)**
A picture sequence $<P_1, \dots, P_n>$ is true of situations $<s_1, \dots, s_n>$ **relative to viewpoint v** along the system of projection S just in case:
(1) temporally, $s_1 < \dots < s_n$;
(2) $proj_S(s_1, v) = P_1$ and … and $proj_S(s_n, v) = P_n$.

Still, Abusch convincingly argues that this framework is insufficient, in particular because narrative sequences give rise to ambiguities of cross-reference. Concretely: the second picture of (31) is most naturally interpreted as involving the same cubes as the first, but nothing fully excludes the possibility that the dark cube disappeared and was replaced by another dark cube a bit further away. A projection-based semantics doesn't suffice to derive the most salient reading, nor the ambiguity. Abusch 2013 suggests that the salient interpretation is obtained because visual representations contain abstract variables (which she relates to Pylyshin's (2003) indexes in vision): a projection-based semantics is combined with variables to derive the meaning of visual narratives (see Abusch and Rooth 2017 for further pictorial operations possibly reminiscent of language).

6.2 Musical narratives

6.2.1 Basic music semantics

There are iconic effects not just in the visual but also in in the auditory modality, and a simple projection-based approach won't do in this case. On a more abstract level, iconic representations exploit properties of perception, which seek to recover information about the causal sources of a percept. Visual iconic representations produce information about a (real or imagined) causal source through its interaction with light in the environment (hence Greenberg's projection-based semantics). But this more abstract perspective applies to auditory iconic representations as well. The onomatopoeia *pshhhhh* can for instance be used to refer to an object (a sound source) that produces this kind of sound, for instance a can of beer when it is opened. One could in principle investigate auditory narratives that are based on natural sounds, but this doesn't seem to be a salient genre (sounds effects are normally used in conjunction with another medium). On the other hand, musical narratives are common, and there have been recent attempts to investigate their semantics.

The first question, however, is whether music has a semantics, i.e. whether it can convey information about a music-external reality. An initial list of systematic effects appears in (33):

(33) Examples of inferential effects (Schlenker 2019b Appendix II, with links to examples)
a. Lower pitch may indicate that a virtual source (i) is larger, or (ii) is less excited/energetic.
b. Lower loudness may indicate that the source is (i) less energetic, or (ii) further away.
c. Lower speed may indicate that the source is slower.
d. Silence may indicate that an event is interrupted.
e. Lesser harmonic stability may indicate that the source is in a less stable (i) physical or (ii) emotional position.
f. A change of key may indicate that the source is moving to a new environment.

To explain these effects, it has been proposed (following Bregman 1994) that hearers posit 'virtual sources' behind the music: not the real sound sources (e.g. the instruments), but objects whose properties and behavior are constrained by the music. Some inferences about these virtual sources are lifted from normal auditory cognition, as for (33)a-d. To illustrate (33)a,b, in normal auditory cognition, lower pitch may be indicative of a larger source (e.g. a larger animal producing a sound), or if the source is known, it may be indicative of a lower level of excitement of the source; a sound that is becoming softer may indicate that the source is losing energy or moving away. One particularly fruitful line of research has focused on properties of animal and human vocal signals that can used in music to trigger powerful emotional effects (e.g. Aucouturier et al. 2016, Liuni et al. 2020). Other inferential rules are more specifically musical in nature; this particularly applies to harmonic notions, which pertain to the consonance or dissonance of chords and melodies, as in (33)e,f. In Western classical music and in jazz, an important notion is that of a tonal pitch space, with parts that are more stable than others, areas that correspond to 'keys', and non-trivial relations of distance among notes or chords.[11]

If music indeed triggers inferences about virtual sources, can we define the semantic content of music? This was done in a simplified framework by positing that a series of n musical events is true of

[11] Lerdahl 2001, 2019 hints at an analysis of musical meaning in terms of a "journey through tonal pitch space", while Ganroth-Wilding and Steedman 2014 provide an explicit semantics for jazz sequences in terms of motion in tonal pitch space. The notion of a 'tonic center' (a note or chord of greatest stability) might play a role across musical traditions (Mehr et al. 2019).

an object undergoing n events in the (real or imagined) extra-musical world just in case the inferences triggered by the music are all satisfied by the corresponding events. Here it matters that all the inferential effects in (33) are of the following form: *If musical events M_1 and M_2 stand in relation R, their respective denotations s_1 and s_2 stand in relation R**. A semantics can for this reason be defined by requiring that certain relations among musical properties be preserved by the events depicted by the music. This makes it possible to define a notion of truth for musical sequences, partly similar to the notion of truth for pictorial sequences in (32) (here we use the term 'events' rather than 'situations' because it is more intuitive, but the structure of the account is similar):

(34) **Musical sequences true of n events (after Schlenker 2017, 2019)**
A musical sequence $<P_1, \dots, P_n>$ is true of an object undergoing events $<s_1, \dots, s_n>$ relative to auditory point v just in case:
(1) temporally, $s_1 < \dots < s_n$;
(2) certain preservation conditions are satisfied, for instance ones corresponding to (33)a,e:
a. If P_i is less loud than P_k, then (i) s_i has less energy than s_k; or (ii) s_i is further from the auditory point v than s_k is.
b. If P_i is less harmonically stable than P_k, then s_i is less (i) physically, or (ii) emotionally stable than s_k.

In this simple proof of concept, it can already be seen that the inferences triggered will be very general, hence the abstract character of musical meaning. It can be highlighted by revisiting Leonard Bernstein's celebrated discussion of Richard Strauss's Don Quixote (Variation II) (link: https://youtu.be/XFZ7wORtj2A). Bernstein sought to convince his audience that *even* a piece that has such an explicitly descriptive character (an instance of 'program music') in no way describes a real or imagined world: the true meaning of music is "the way it makes you feel when you hear it". To make his point, Bernstein told the *wrong* story to fit the music, and showed that it worked just as well as the intended one. The original story is about Don Quixote (i) departing on his horse to conquer the world, (ii) approaching a flock of sheep baaing, which he mistakes for an army, (iii) charging at them and creating chaos in the process, and finally (iv) feeling proud about his knightly deed. Bernstein's *wrong* story pertained to Superman (i') departing on his motorcycle to free an unjustly imprisoned friend of his, (ii') approaching the jail, in which the prisoners are snoring, (iii') charging into the prison and wreaking havoc in the process, and finally (iv') triumphantly hurling his friend back to freedom. It is striking that Bernstein's two stories are, with minor exceptions, isomorphic to each other: there is a close correspondence to the events described in (i)-(i'), (ii)-(ii'), (iii)-(iii') and (iv)-(iv').

This is not for lack of imagination: the details of the music trigger inferences that must be preserved by both stories. For instance, the triumphant character of the beginning is in part due to the upwards melodic movement, as heard in (35); when the music is rewritten in accordance with composition rules so as to invert the melodic movement, this triumphant character disappears, as expected by (33)a(ii). The sheep baaing and the prisoners snoring create a somewhat chaotic effect in part because of dissonances in the music, as heard in (36); when the dissonant chords are replaced with consonant ones, the chaotic effect largely disappears, as expected by (33)e(i). Similarly, the sheep and prisoners appear to be approaching the perspectival center because of a crescendo (increasing loudness) in the music, as heard in (37); when the crescendo is replaced with a diminuendo (decreasing loudness), the effect disappears (if anything, the virtual sources appear to move away).

(35) a. Original: upwards melodic movement at the beginning (simplified MIDI)
https://www.dropbox.com/s/dn9cwtc0zf7ji9m/Strauss-Don%20Quixote%20cello%20original.aiff?dl=0
b. Rewritten: downwards melodic movement at the beginning (A. Bonetto)
https://www.dropbox.com/s/xffokt0huvkg1fl/Strauss-Don%20Quixote%20cello%20vers%20le%20bas.aiff?dl=0

(36) a. Original: dissonant chords (simplified MIDI)
https://www.dropbox.com/s/memrvak6ewewosi/Strauss-Don%20Quixote%20moutons%20original%20v2.aiff?dl=0
b. Rewritten: consonant chords (A. Bonetto)
https://www.dropbox.com/s/f76s0acuv3mwu26/Strauss-Don%20Quixote%20moutons%20sans%20dissonances%20v2.aiff?dl=0

(37) a. Original: crescendo (simplified Midi) https://www.dropbox.com/s/memrvak6ewewosi/Strauss-Don%20Quixote%20moutons%20original%20v2.aiff?dl=0
b. Rewritten: diminuendo (A. Bonetto) https://www.dropbox.com/s/gyrpkbjts0f2f9p/Strauss-Don%20Quixote%20tout%20en%20dim%20v2.aiff?dl=0

In sum, one's ability to tell the 'wrong' story to fit the music doesn't show that the latter has no semantics, just that it has a fairly abstract one, i.e. that its meaning singles out rather diverse situations.

6.2.2 *Adding variables*

While in simple cases one considers excerpts with a single musical voice corresponding to a single virtual source, in more complex cases one may countenance several virtual sources, which raises a question analogous to the one Abusch asked about visual narratives: what are the cross-reference relations them? Here it may be fruitful to enrich musical representations with variables, just as was argued for visual narratives by Abusch.

An example will make this issue clear. The beginning of Chopin's Mazurka Op. 33 No 2 has an extremely simple structure of the form AB A'B' with A' = A and B' = B. If this beginning is given a flat realization with constant loudness, as in (38)a, it is compatible with multiple interpretations: (i) a single virtual source for the whole, or (ii) one for [AB] and another for [A'B'], or possibly (iii) one for A and A', and another one for B and B'. What we have, in effect, is an ambiguity in cross-reference, just as was the case with Abusch's cube in (31), or for that matter with (3)a above. When we add the dynamics, as in Chopin's score (https://drive.google.com/file/d/15Vz7lKgRDKWawUTxW9ZKWsmoaDT912A8/view?usp=sharing), possibly (iii) becomes far less likely. The reason is that [AB] is played *forte*, while [A'B'] is played *pianissimo*. This can be made sense of if one source is energetic or close and corresponds to [AB], while the other is less energetic or further away and corresponds to [A'B']. But one can also create an 'anti-Chopin' dynamics, in which A and A' are realized forte, while B and B' are realized pianissimo: if anything, this suggests that A and A' correspond to one source, while B and B' correspond to another, as can be heard in (38).

(38) a. Flat realization [AB] [A'B'] https://www.dropbox.com/s/xbmlw1nszjwok1b/Chopin-Mazurka%2033-2%28sans%20nuances%29.aiff?dl=0
b. Chopin's dynamics $[AB]_{f}$ $[A'B']_{pp}$ https://www.dropbox.com/s/qq9c61eoaviafxv/Chopin-Mazurka%2033-2%28original%29.aiff?dl=0
c. Anti-Chopin dynamics $A_{f}B_{pp}$ $A'_{f}B'_{pp}$ https://www.dropbox.com/s/vu8tzaugin7wk2l/Chopin-Mazurka%2033-2%28nuances%20inversees%29.aiff?dl=0

In several orchestrations written for a ballet by Michel Fokine, the identity of the virtual sources is made more salient by the use of different timbres: Chopin's dynamics $[AB]_{f}$ $[A'B']_{pp}$ (*forte*, then *pianissimo*) is reflected by the use of different timbres for [AB] and [A'B'], typically the orchestra for [AB] and individual wind arguments for [A'B']. An example is Britten's orchestration in (39) (see Schlenker 2019c for further examples of orchestration of the same piece that make related choices).

(39) Benjamin Britten's orchestration (1941)[12]: $[AB]_{\text{orchestra}}$ $[A'B']_{\text{oboe+flute}}$ https://www.youtube.com/watch?v=DQLde6QJXvM

The existence of variables becomes even more salient when a ballet co-occurs with the music. In Fokine's piece, [AB] corresponds to a movement of the main ballerina, [A'B'] to that of the other dancers.

(40) Fokine's *Les Sylphides* (originally called *Chopiniana)*, movement on Chopin's Mazurka Op. 33 No 2
Performance from 1984, American Ballet Theatre, on an orchestration close to Britten's version
https://www.youtube.com/watch?v=LBJNc3h7Hp8&t=10m46s
$[AB]_{\text{main ballerina}}$ $[A'B']_{\text{other dancers}}$

In sum, an abstract notion of iconic semantics (based on virtual sources) may be used to develop a semantics for music, and it too might benefit from the use of variables to enrich iconic representations, just as was argued by Abusch for visual narratives. An obvious question is how these developments will interact with the theory of dance, in particular pioneering work on dance syntax and dance semantics by Charnavel (2016, 2019) and Patel-Grosz et al. (2018) respectively.

6.3 Interaction between visual and musical narratives

The exploration of the interaction between visual and musical narratives suggests that the notion of cosuppositions, originally used for co-speech gestures, might have fully non-linguistic applications as well. This is illustrated in (41) by combining music with a simple gif representing Asterix drinking the magic potion, hitting a Roman soldier and leaving the room.

[12]Retrieved from https://www.youtube.com/watch?v=DQLde6QJXvM on December 13, 2019 Credits: Chopin: Les Sylphides (Arr. for Orchestra By Benjamin Britten, Mono Version) Joseph Levine, American Ballet Theatre Orchestra 1 January 1954.

(41) *Context:* Asterix had an earlier encounter with a Roman soldier. Now he is faced with him once again.[13] What will happen next? Will Asterix…

a. ♬ WHISTLE https://www.dropbox.com/s/ysx313tuds310j3/Asterix-co-gif-a1.mp4?dl=0

?

b. ♬ WHISTLE https://www.dropbox.com/s/cpsgjid59iohofx/Asterix-co-gif-a2.mp4?dl=0

?

A light-hearted whistling tune accompanies either the entire scene (=(41)a) or just Asterix's departure (=(41)b). The entire excerpt is embedded in a question so as to tease apart at issue content from presuppositions.

Unsurprisingly, the visual narrative provides information about the action and thus its main content is at issue. Strikingly, the music adds inferences to the scene, but different ones depending on where the music appears. When the whistling tune co-occurs with the entire scene, the whole sequence of actions is presented as light-hearted. When the whistling scene only co-occurs with the section in which Asterix departs, the inference is different, along the lines of: *if Asterix leaves the room after drinking the potion and hitting the Roman soldier, his departure will be light-hearted*. Crucially, this inference exists despite the presence of a question, and it is conditionalized relative to the scene it modifies. In these two respects, it behaves just like cosuppositions triggered by co-speech gestures. In the end, this behavior can be derived from a semantics for visual narratives along the lines of (32), combined with hypothesis that music semantics provides cosuppositional information about the part of the scene it co-occurs with (Schlenker 2019c).

These examples raise two questions for future research. First, is it in general the case that, when embedded in sentences, film and cartoon music makes the same kind of semantic contribution as co-speech gestures? Second, does this behavior only arise when film and cartoon excerpts are embedded in sentences, or also in normal (unembedded) films and cartoons? The latter possibility is tantalizing but remote – as was the case for the hypothesis that visual content on its own gets productively divided among some established slots of the inferential typology.

7 Semantics without phonology

The systems we have reviewed so far all have an overt realization: human language, primate calls, music, and visual narratives are instantiated in utterances we can perceive and parse, music we can hear, drawings we can see. We close by showing that the same primitives can be found in systems without an overt realization.

Thought and reasoning allow us to manipulate representations, transforming models of what we know into models of conclusions we can draw from them, with crucial consequences for action. The tools of semantics can help illuminate this extraordinary human faculty. One starts with the description of a "language of thought" and a "logic of thought". That is, we specify what the elementary building blocks ("words") are and how they may be combined, and we can then derive new "sentences" from existing ones while preserving useful properties like truth. We outline concrete examples of insights that can be gained from this line of inquiry shortly. This semantics of thought doesn't just radically extend the program of linguistic semantics; it might contribute to an understanding of its cognitive roots.

Still, there is an important methodological difference between the semantics of language and of thought. In the latter case, the absence of an overt realization forces us to reverse-engineer these systems through more indirect evidence: behavioral consequences must be used to infer what the atomic elements (lexical items) are. Developing language-free diagnostics makes it possible to export this enterprise to many systems, such as the reasoning and concept-formation faculties of non-human animals. In this sense, the semantics of thought and reasoning has been pursued for a long time. We argue in favor of regimenting these results with the formal tools of semantics.

[13] Music: Billy Mowbray, Uke and Whistle.

7.1 Content concepts and logical concepts

Linguistic semantics starts with constraints on minimal lexical meanings: similar concepts are found across languages, while some concepts are never found. As we saw in Section 2, one such constraint is Gärdenfors's connectedness, which has the result that no word may mean *mushroom or table*. Similarly, no word in no language is known to mean *none or all*, *few or many*, and a logical version of connectedness can derive this constraint as well, and it generalizes the old insights that primitive words are either positive or negative.

Can similar constraints be displayed for language-free concepts? Consider the following thought experiment. You see sets of 5 objects each, and observe that groups with 2, 3 or 5 red objects behave in some way, while groups with 0 or 1 red object do not. It is natural to infer that groups with 4 red objects will behave like the former: connectedness constrains generalizations drawn on the fly. Similarly, a connectedness-compatible generalization involving groups with at least 2 red objects will be easier to infer than a connectedness-incompatible generalization involving groups with 0, 1 or 4 red objects. Using tasks of this sort, we can arguably investigate people's language of thought without relying on their languages (Piantadosi et al. 2016), with interesting conclusions – including that the language of thought might make use of bound variables (e.g. Overlan et al. 2017). Still, how can we ensure that subjects are not relying on an inner linguistic monologue – which would make the final analysis about words? Researchers have typically made participants work with concepts that are not expressed in their language: even in such cases, concepts that obey connectedness are inferred more rapidly than ones that don't.

Importantly, such language-free tasks can be extended to non-linguistic participants, such as pre-linguistic infants and non-human animals, with striking results. To cite but one, baboons have been argued to obey a connectedness constraint just like human adults do: they infer connectedness-compatible generalizations faster than connectedness-incompatible ones (Chemla et al. 2019b). Thus a pervasive constraint in human content words and quantifiers arguably has roots in the conceptual behavior of rather distant non-human primates.

Going full circle, remember that in view of the complexity of field experiments, analyses of primate call meaning are greatly underspecified by available data. But constraints on animal concepts are plausibly constraints on call meanings as well, and they might help adjudicate among competing theories (e.g. the meanings that tend to replace general calls in Steinert-Threlkeld et al.'s results may not be connected, in which case the hypothesis of general calls might be saved in the end).

7.2 Alternatives and attention

There are further interesting restrictions on cognitive modules without a phonology like thought, reasoning, or vision. We focus here on the role of **attention** as a facilitator of tractability for hard problems and its ramification for the typology of inferences discussed above.

A particularly clear case is mental-model theory (Johnson-Laird 1983), one of the most influential theories of human reasoning. Focusing on deductive reasoning, mental-model theory proposes a representational system that distinguishes between categorical premises that correspond to *exactly one mental model*, and another class of premises that instead propose a set of *alternative mental models* for consideration. Alternative mental models are generated by premises involving natural language disjunction (e.g. English *or*), and by reasoning-internal processes that allow humans to flesh out the full space of possibilities given a set of premises, at a potentially considerable computational cost.

For present purposes, the crucial idea behind alternatives in mental-model theory is that they represent not the possible states of affairs that we *know* to be compatible with the premises, but those that we happen to *be attending to*. Imagine you had the following two pieces of information as premises in a deductive problem: P1: *Either John speaks English and Mary speaks French, or Bill speaks German;* P2: *John speaks English*. Can you conclude with certainty that *Mary speaks French*? In a study of structurally identical examples, Walsh and Johnson-Laird (2004) show that over 85% of participants draw this conclusion. Yet it is a fallacy, for the following possible state of affairs makes the premises true and the conclusion false: John speaks English, Mary *does not* speak French, Bill speaks German. On mental model theory, the fallacious conclusion is a result of the alternative possibilities we are attending to. We *know* that the counterexample just presented is possible and invalidates the fallacious conclusion, but we are not *attending* to that possibility. The only possibilities we consider by default are the two possibilities expressed by the disjunctive premise, and those possibilities suggest that the only way to make the second premise *John speaks English* true is by making *Mary speaks French* true as well.

Attention determines what alternative possibilities we consider when reasoning, and linguistic operators like disjunction drive attention. This is both a boon and a liability. On the one hand, it allows us to draw conclusions on the basis of a small space of alternative possibilities, reducing cognitive costs. On the other hand, it makes us vulnerable to fallacious reasoning.

What can semantics contribute to this research program? As mentioned, mental-model theory distinguishes between categorical premises and disjunctive premises that give rise to alternatives. Koralus and Mascarenhas (2013) show that this distinction is best regimented by incorporating semantic insights: semantic analyses of disjunction and several other operators have long made the same conceptual distinction as mental-model theory, and they thus offer off-the-shelf formal accounts of the representations of sentence meanings suitable to feed a mental-model-like reasoning module (e.g. the alternative semantics of Kratzer and Shimoyama, 2002; the inquisitive semantics of Groenendijk, 2008, and Mascarenhas, 2009; the truth maker semantics of Fine, 2012).

This discovery illustrates the fruitfulness of the interplay between semantics and reasoning. First, the formal landscape of theories of reasoning is enriched by this connection with linguistic semantics: semantic theories are characterized by their careful use of a broad palette of sophisticated formal systems, including non-standard logics developed in mathematics and computer science. The connection between semantics and reasoning allows us to import into reasoning brand new formal systems, offering new, useful, and insightful tools for theory building in the psychology of reasoning.

For example, Koralus and Mascarenhas's (2013) reimagining of mental-model theory offers a proof system that is associated with well-studied logical systems from the semantics literature. This allowed them to prove meta-logical results with cognitive import. For example, they show that there is a derivation strategy (i.e. a reasoning strategy) in their system that guarantees logically valid conclusions, at the cost of exponential blowup of alternatives under consideration in the worst case. While the versions of mental-model theory due to Johnson-Laird and collaborators have typically come accompanied by open-source computer implementations, full-fledged soundness proofs of the kind just mentioned did not exist. The computer implementations of traditional mental-model theory did not constitute logics in the traditional sense, making soundness and completeness theorems hard to formulate and prove. Formal methodologies from semantics opened the door to proper meta-logical results in mental-model theory broadly conceived.

Second, the empirical scope of existing theories of reasoning increases in interesting ways. For example, the effects of attention and alternatives in the representations of disjunctive sentences discovered by mental-model theorists have been replicated in empirical domains where semantic theories have posited disjunction-like alternatives: indefinites like *some* (Mascarenhas and Koralus, 2017) and epistemic operators like *might* (Mascarenhas and Picat, 2019; Johnson-Laird and Ragni, 2019). As the broader extent of the empirical phenomenon reveals itself, new and better constraints on our theories of reasoning emerge.

Finally, these results highlight the importance of cross-fertilization between theories of language and of thought, with the tantalizing possibility of broad unifications, as has been sketched for the role of alternatives. Moreover, further cognitive models can be brought into the fold in new ways. For instance, if effects of interpretation and reasoning with alternatives are at a deeper level about attention manipulation, then the need emerges to apply semantic methods to our study of visual attention as well.

It is interesting in this connection to return to the notions of implicatures, presuppositions, and supplements that we introduced before. One dimension of variation here concerns how attention is directed towards different parts of the world, or of a message in this case. We have argued that the division between these notions naturally arises beyond words (Section 5). In some cases at least, their pervasiveness and productivity in and beyond language might be due to how cognition is structured: the semantics of thought might yield the key to foundational issues in linguistic semantics.

8 Typology of semantic operations in nature

The generalized semantic approach advocated here makes it possible to replace overly general questions about languages and meanings in nature with more specific and enlightening ones. *Does species X have language? Are gestures part of language? Is music a language?* These are all hard questions to answer without a definition of language, and human language is so unique that multiple clusters of properties could be used to 'define' it. On the other hand, we can start from the observation that multiple systems convey information, and ask what their formal properties are, how integrated they are with human language, and what properties, if any, they share with it.

Focusing on semantic phenomena, we have provided preliminary answers summarized in (42), organized around four questions: (i) How do elementary parts of the system produce meaning? (ii) How are the meanings of elementary parts combined? (iii) Are there variable-like elements in the system? (iv) Does the system have a non-trivial inferential typology?

(42)

	Meaning of elementary parts	**Composition of meanings**	**Variable-like elements?**	**Inferential typology?**	
Human speech and signs	Lexical (+ iconic modulations, especially in signs)	Function application, …	Yes	Rich: implicatures, presuppositions, …	
Human gestures	Iconic + grammatical (e.g. loci)	Modification or replacement of words	Yes	Co-occurring with words: cosuppositions	Replacing words: rich typology: implicatures, presuppositions, …
Primate calls	Innate	Conjunctive juxtaposition?	No	Implicatures?	
Ape gestures	Innate (+ iconic?)	Conjunctive juxtaposition?	No?	?	
Pictures	Iconic (projection-based semantics)	Iconic (projection-based semantics)	Yes?	?	
Music	Iconic (abstract)	Iconic (abstract)	Yes?	Co-occurring with film/gifs/words: cosuppositions?	Pure music: ?
Thought	Primitive content concepts and logical concepts	Mental models with attention/alternatives	Yes ?	? (attention-related?)	

References

Abner N., Cooperrider K. and Goldwin-Meadow S.: 2015, Gesture for Linguists: A Handy Primer, *Lang. Linguist. Compass*, 9/11, 437–449

Abrusán, Márta. 2011. Predicting the presuppositions of soft triggers. *Linguistics and Philosophy*, **34**(6), 491–535.

Abusch, Dorit: 2013, Applying discourse semantics and pragmatics to co-reference in picture sequences. In *Proceedings of Sinn und Bedeutung* 17: 19–25.

Abusch, Dorit: 2015, Possible Worlds Semantics for Pictures. Manuscript, Cornell University.

Abusch, Dorit and Rooth, Mats: 2017, The formal semantics of free perception in pictorial narratives. In Alexandre Cremers, Thom van Gessel & Floris Roelofsen (eds), *Proceedings of the 21st Amsterdam Colloquium*. Available online at https://semanticsarchive.net/Archive/jZiM2FhZ/AC2017-Proceedings.pdf.

Aucouturier, J. J., Johansson, P., Hall, L., Segnini, R., Mercadié, L., & Watanabe, K. (2016). Covert digital manipulation of vocal emotion alter speakers' emotional states in a congruent direction. *Proceedings of the National Academy of Sciences*113(4):948-53. doi: 10.1073/pnas.1506552113.

Barker, Chris and Jacobson, Pauline I.: 2007, *Direct Compositionality*. Oxford University Press.

Berthet, M., Mesbahi, G., Pajot, A., Cäsar, C., Neumann, C., & Zuberbühler, K. (2019a). Titi monkeys
combine alarm calls to create probabilistic meaning. *Science Advances*, *5*(5), eaav3991.
https://doi.org/10.1126/sciadv.aav3991

Berthet, Mélissa & Benjumea, Juan & Millet, Juliette & Cäsar, Cristiane & Zuberbühler, Klaus & Dunbar, Ewan. (2019b). Animal linguistics and the puzzle of titi monkeys alarm sequences. *Proceedings of the Amsterdam Colloquium 2019*. http://events.illc.uva.nl/AC/AC2019/Proceedings/

Barwise, Jon and Robin Cooper, 1981, "Generalized Quantifiers and Natural Language", Linguistics and Philosophy, 4(2): 159–219. doi:10.1007/BF00350139

Bregman, Albert S.: 1994, *Auditory Scene Analysis*. MIT Press.

Byrne RW, Cartmill E, Genty E, Graham KE, Hobaiter C, Tanner J.: 2017, Great ape gestures: Intentional communication with a rich set of innate signals. *Animal Cognition* 20(4):755–69.

Cäsar Cristiane, Klaus Zuberbühler, Richard J. Young, and Richard Byrne. 2013. Titi monkey call sequences vary with predator location and type. *Biology Letters* 9: 20130535. doi:10.1098/rsbl. 2013.0535.

Charnavel, Isabelle: 2016, Steps Towards a Theory of Dance Cognition. Manuscript, Harvard University.

Charnavel, Isabelle: 2019, Steps towards a Universal Grammar of Dance: Local Grouping Structure in Basic Human Movement Perception. *Frontiers in Psychology* (section Cognition), Volume 10, Article 1364.

Chemla, E.; Buccola, B.; and Dautriche, I. (2019a). Connecting content and logical words. Journal of Semantics.

Chemla, E.; Dautriche, I.; Buccola, B.; and Fagot, J. (2019b). Constraints on the lexicons of human languages have cognitive roots present in baboons (Papio papio). *Proceedings of the National Academy of Sciences*, 116(30): 14926-14930.

Chierchia, G.; Crain, S.; Guasti, M.T.; Gualmini, A.; Meroni, L.: 2001, The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. *Proceedings from the Annual Boston University Conference on Language Development*, 25, pp. 157-168.

Cohn, Neil: 2013, *The visual language of comics: Introduction to the structure and cognition of sequential images*. London: Bloomsbury.

Coppola, M., Spaepen, E., & Goldin-Meadow, S. (2013). Communicating about quantity without a language model: Number devices in homesign grammar. *Cognitive Psychology*, 67, 1–25

Cumming, Samuel, Greenberg, Gabriel & Kelly, Rory: 2017, Conventions of Viewpoint Coherence in Film. *Philosophers' Imprint* 17, 1.

Dingemanse, Mark: 2013, Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass* 6.654–72. DOI: 10.1002/lnc3.361

Douglas, P.H., and L.R. Moscovice. 2015. Pointing and pantomime in wild apes? Female bonobos use referential and iconic gestures to request genito-genital rubbing. Sci Rep DOI: 10.1038/srep13999

Ebert, Cornelia and Ebert, Christian: 2014, Gestures, Demonstratives, and the Attributive/Referential Distinction. Handout of a talk given at Semantics and Philosophy in Europe (SPE 7), Berlin, June 28, 2014.

Esipova, Maria: 2019, Composition and Projection in Speech and Gesture. PhD dissertation, New York University.

Fine, Kit. 2012. A difficulty for the possible world analysis of counterfactuals. *Synthese*.

Gärdenfors, Peter: 2004, *Conceptual Spaces: The Geometry of Thought*. MIT Press. Cambridge, MA.

Gärdenfors , Peter: 2014, *The Geometry of Meaning: Semantics Based on Conceptual Spaces*. MIT Press. Cambridge, MA.

Genty, Emilie and Zuberbühler, Klaus: 2014, Spatial Reference in a Bonobo Gesture, *Current Biology*, http:// dx.doi.org/10.1016/j.cub.2014.05.065

Goldin-Meadow, S.: 2003, *The Resilience of Language*. Taylor & Francis.

Goldin-Meadow, Susan and Brentari, Diane: 2017, Gesture, sign and language: The coming of age of sign language and gesture studies. *Behavioral and Brain Sciences*, doi:10.1017/S0140525X15001247

Granroth-Wilding, Mark and Steedman, Mark: 2014, A robust parser-interpreter for jazz chord sequences. *Journal of New Music Research* 43 (4), 355-374.

Greenberg, Gabriel: 2013. Beyond Resemblance. *Philosophical Review* 122:2, 2013

Greenberg, Gabriel: 2019, Semantics of Pictorial Space. Manuscript, UCLA.

Grice, Herbert Paul. 1975. Logic and conversation. In Cole, P., and Morgan, J. L. (eds.), *Syntax and Semantics*, Vol. III: Speech Acts. New York: Academic Press, pp. 41– 58.

Groenendijk, Jeroen. 2008. Inquisitive Semantics: Two possibilities for disjunction. In *Proceedings of the Seventh International Tbilisi Symposium on Language, Logic and Computation*.

Gutzmann, Daniel; Matthewson, Lisa ; Meier, Cécile; Rullmann, Hotze ; Zimmermann, Thomas Ede; Dina Voloshina (eds): to appear, *The Wiley Blackwell Companion to Semantics,* Wiley.

Heim, Irene, 1983, On the projection problem for presuppositions. In Barlow, M. and Flickinger, D. and Westcoat, M. (eds.), Second Annual West Coast Conference on Formal Linguistics, Stanford University, 114–126.

Hobaiter C, Byrne RW: 2017, What is a gesture? A meaning-based approach to defining gestural repertoires. *Neuro Biobehav Rev* 82:3–12

Johnson-Laird, Philip N. 1983. *Mental models: towards a cognitive science of language, inference, and consciousness*. Cambridge: Cambridge University Press.

Johnson-Laird, Philip N. and Marco Ragni. 2019. Possibilities as the foundation of reasoning. *Cognition*, 193.

Keenan, Edward L. and Dag Westerståhl: 2011, Generalized Quantifiers in Linguistics and Logic. In van Benthem and ter Meulen (eds), *Handbook of Logic and Language*, Amsterdam, Elsevier: 859–910.

Kersken, V., Gómez, JC., Liszkowski, U., Soldati, A., Hobaiter, C.: 2018, A gestural repertoire of 1- to 2-year-old human children: in search of the ape gestures. *Animal Cognition*. https://doi.org/10.1007/s10071-018-1213-z

Koralus, Philipp and Salvador Mascarenhas. 2013. The erotetic theory of reasoning: bridges between formal semantics and the psychology of deductive inference. *Philosophical Perspectives*, 27:312–365.

Kratzer, Angelika and Junko Shimoyama. 2002. Indeterminate pronouns: the view from Japanese. In *Third Tokyo Conference on Psycholinguistics*.

Kuhn, Jeremy: 2016, ASL Loci: Variables or Features? *Journal of Semantics*. doi: 10.1093/jos/ffv005

Lerdahl, Fred: 2001. *Tonal Pitch Space*. Oxford University Press.

Lerdahl, Fred: 2019, *Composition and Cognition: Reflections on Contemporary Music and the Musical Mind*. University of California Press.

Liddell, Scott K.: 2003. *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge: Cambridge University Press.

Lillo-Martin, Diane & Richard P. Meier. 2011. On the linguistic status of 'agreement' in sign languages. *Theoretical Linguistics* 37(3/4). 95–141. DOI: https://doi.org/10.1515/thli.2011.009 Maestripieri D.: 1997, Gestural communication in macaques: Usage and meaning of nonvocal signals. *Evolution of communication* 1: 193–222.
Liuni, M., Ponsot E., Bryant, G. A. and Aucouturier, J. J.: 2020, Sound context modulates perceived vocal emotion. *Behavioural Processes*, 172, 104042.
Maienborn, Claudia; von Heusinger, Klaus; Portner, Paul: 2011, *Semantics: An International Handbook of Natural Language Meaning*. De Gruyter.
Mascarenhas, Salvador. 2009. *Inquisitive Semantics and Logic*. Master's thesis, ILLC.
Mascarenhas, Salvador and Philipp Koralus. 2017. Illusory inferences with quantifiers. *Thinking and Reasoning*, 23(1):33–48.
Mascarenhas, Salvador and Léo Picat. 2019. Might as a generator of alternatives: the view from reasoning. In *Proceedings of SALT 29, UCLA*.
Mehr SA, Singh M, Knox D, Ketter DM, Pickens-Jones D, Atwood S, Lucas C, Jacobi N, Egner AA, Hopkins EJ, et al. Universality and diversity in human song. *Science* Vol. 366, Issue 6468, eaax0868 DOI: 10.1126/science.aax0868
Migotti, Léo and Guerrini, Janek: 2019, The music/language interaction: a semantic account. Talk given within the LINGUAE seminar, ENS, Paris, April 25, 2019.
Montague, Richard, 1970a. English as a Formal Language. In Visentini, B. et al. (eds.) Linguaggi nella società e nella tecnica. Milan: Edizioni di Comunità. 189-224.
Montague, Richard, 1970b. Universal grammar. *Theoria* 36: 373-398.
Morris, Charles W.: 1938, Foundations of the Theory of Signs. In Otto Neurath (ed.), *International Encyclopedia of Unified Science*, vol. 1 no. 2. Chicago: University of Chicago Press.
Overlan, M. C.; Jacobs, R. A.; Piantadosi, S. T.: 2017, Learning abstract visual concepts via probabilistic program induction in a Language of Thought. *Cognition* 168, 320–334.
Pasternak, Robert: 2019, The projection of co-speech sound effects. Manuscript, ZAS Berlin.
Pasternak, Robert and Tieu, Lyn: 2020, Co-linguistic content projection: From gestures to sound effects and emoji. Manuscript, ZAS Berlin and U. of Western Sydney.
Patel-Grosz, Pritty; Grosz, Patrick Georg; Kelkar, Tejaswinee & Jensenius, Alexander Refsum (2018). Coreference and disjoint reference in the semantics of narrative dance, In Uli Sauerland & Stephanie Solt (ed.), *Proceedings of Sinn und Bedeutung* 22, vol. 2, ZASPiL 61. Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS). Chapter in Vol. 2. s 199 - 216
Pfau, R., Salzmann, M., & Steinbach, M. (2018). The syntax of sign language agreement: Common ingredients, but unusual recipe. *Glossa: a journal of general linguistics*, *3*(1): 107. 1–46, DOI: https://doi.org/10.5334/gjgl.511
Piantadosi, S. T.; Tenenbaum, J.; Goodman, N.: 2016, *The logical primitives of thought: Empirical foundations for compositional cognitive models*, *Psychological Review*, 123:392-424
Potts, Christopher. 2005. *The Logic of Conventional Implicatures*. Oxford Studies in Theoretical Linguistics, Oxford: Oxford University Press.
Pylyshyn, Zenon: 2003, *Seeing and Visualizing: It's Not What You Think*. MIT Press.
Sandler, Wendy and Lillo-Martin, Diane: 2006, *Sign Language and Linguistic Universals*. Cambridge University Press.
Schembri, Adam, et al. 2018. Indicating verbs as typologically unique constructions: Reconsidering verb 'agreement' in sign languages. *Glossa: a journal of general linguistics* 3(1): 89. 1–40, DOI: https://doi. org/10.5334/gjgl.468
Schlenker, Philippe: 2011. Donkey Anaphora: the View from Sign Language (ASL and LSF). *Linguistics and Philosophy* **34**(4): 341-395.
Schlenker, Philippe: 2014. Iconic Features. *Natural Language Semantics* **22**(4): 299-356.
Schlenker, Philippe: 2017, Outline of Music Semantics. *Music Perception: An Interdisciplinary Journal* 35, 1: 3-37 DOI: 10.1525/mp.2017.35.1.
Schlenker, Philippe: 2018a, Visible Meaning: Sign Language and the Foundations of Semantics. *Theoretical Linguistics*.
Schlenker, Philippe: 2018b, Iconic Pragmatics. *Natural Language & Linguistic Theory* 36, 3:877–936
Schlenker, Philippe: 2019a, Gestural Semantics: Replicating the typology of linguistic inferences with

pro- and post-speech gestures. *Natural Language & Linguistic Theory* 37, 2: 735–784.
Schlenker, Philippe: 2019b, Triggering Presuppositions. Manuscript, Institut Jean-Nicod and New York University.
Schlenker, Philippe: 2019c, Musical Meaning within Super Semantics. Manuscript, Institut Jean-Nicod and New York University.
Schlenker, Philippe: to appear a, Gestural Grammar. Accepted with minor revisions, *Natural Language & Linguistic Theory*.
Schlenker, Philippe: to appear b, Iconic Presuppositions. *Natural Language & Linguistic Theory*.
Schlenker, Philippe, and Chemla, Emmanuel: 2018, Gestural Agreement. *Natural Language & Linguistic Theory*. 36, 2: 87–625587. https://doi.org/10.1007/s11049-017-9378-8
Schlenker, Philippe, Lamberton, Jonathan & Santoro, Mirko: 2013. Iconic Variables. *Linguistics & Philosophy* **36**(2): 91-149.
Schlenker, Philippe, Chemla, Emmanuel, Arnold, Kate, Lemasson, Alban, Ouattara, Karim , Keenan, Sumir , Stephan, Claudia, Ryder, Robin, Zuberbühler, Klaus: 2014, Monkey Semantics: Two 'Dialects' of Campbell's Monkey Alarm Calls. *Linguistics & Philosophy* 37, 6: 439-501. *DOI 10.1007/s10988-014-9155-7*
Schlenker, Philippe; Chemla, Emmanuel; Schel, Anne; Fuller, James; Gautier, Jean-Pierre; Kuhn, Jeremy; Veselinovic, Dunja; Arnold, Kate; Cäsar, Cristiane; Keenan, Sumir; Lemasson, Alban; Ouattara, Karim; Ryder, Robin; Zuberbühler, Klaus: 2016a, Formal Monkey Linguistics. *Theoretical Linguistics* 42,1-2:1–90, DOI: 10.1515/tl-2016-0001
Schlenker, Philippe; Chemla, Emmanuel; Zuberbühler, Klaus: 2016c, What do Monkey Calls Mean? *Trends in Cognitive Sciences,* 20, 12, 894–904.
Schlenker, Philippe; Chemla, Emmanuel; Cäsar, Cristiane; Ryder, Robin; Zuberbühler, Klaus: 2016d, Titi Semantics: Context and Meaning in Titi Monkey Call Sequences. *Natural Language & Linguistic Theory* doi:10.1007/s11049-016-9337-9
Schlenker, Philippe and Lamberton, Jonathan: 2019a, Iconic Plurality. *Linguistics & Philosophy* 42: 45-108 https://doi.org/10.1007/s10988-018-9236-0
Schlenker, Philippe and Lamberton, Jonathan: 2019b, Meaningful Blurs: the Sources of Repetition-based Plurals in ASL. Manuscript, Institut Jean-Nicod and New York University.
Steinert-Threlkeld, Shane; Schlenker, Philippe; Chemla, Emmanuel: 2019, Referential and General Calls in Primate Semantics. Manuscript, U. of Washington, IJN and LSCP.
Tarski, Alfred: 1935, The Concept of Truth in Formalized Languages. *Logic, Semantics, Metamathematics*, Indianapolis: Hackett 1983, 2nd edition, 152–278.
Tarski, Alfred: 1944, The Semantic Conception of Truth and the Foundations of Semantics. *Philosophy and Phenomenological Research* 4.
Thijsse, Elias: 1983, 'On some Proposed Universals of Natural Language', in A. ter Meulen, ed., Studies in Model-Theoretic Semantics, Foris, Dordrecht, 19-36
Thomason, R. H. (ed.), 1974, *Formal Philosophy. Selected Papers of Richard Montague*, New Haven: Yale University Press.
Tieu, Lyn; Pasternak, Robert; Schlenker, Philippe; Chemla, Emmanuel: 2017, Co-speech gesture projection: Evidence from truth-value judgment and picture selection tasks. *Glossa* 2(1).
Tieu, Lyn; Pasternak, Robert; Schlenker, Philippe; Chemla, Emmanuel: 2018a, Co-speech gesture projection: Evidence from inferential judgments. *Glossa* 3(1), 109. DOI: http://doi.org/10.5334/gjgl.580
Tieu, Lyn; Schlenker, Philippe; Chemla, Emmanuel: 2019, Linguistic Inferences Without Words. *Proceedings of the National Academy of Sciences* 116 (20) 9796-9801.
Walsh, Clare and Philip N. Johnson-Laird. 2004. Coreference and reasoning. *Memory and Cognition*, 32:96–106.
Zuberbühler, Klaus, 2009. Survivor signals: the biology and psychology of animal alarm calling. *Advances in the Study of Behavior* 40: 277-322.
Zuberbühler, Klaus, Jenny David, and Bshary Redouan. 1999. The predator deterrence function of primate alarm calls. Ethology 105(6): 477–490. doi:10.1046/j.1439-0310.1999.00396.x.