

The Content of Inference

Peter Kuhn, peterkuhn92@gmail.com

DRAFT, 30.10.2024

Abstract

Inferentialism is the view that representational content is explained by lingual or mental states interacting according to inferential rules. Mendelovici and Bourget have argued against inferentialism that rules of inference do not sufficiently constrain content. This paper argues that their argument can be further strengthened such that its conclusion yields that content and inferential roles are strictly independent. It will then be argued that this conclusion is untenable and that the argument, rather than undermining inferentialism, corrodes the model theoretic foundations of standard philosophical semantics. The paper concludes with some hints towards an alternative semantic paradigm.

Keywords: Representational content, inferentialism, conceptual role semantics, model theory, type theory.

No funding was received to assist with the preparation of this manuscript. There are no conflicts of interest to disclose.

Introduction

The insight that the meaning of words is not, or at least not wholly, conferred to them by mental acts, but by the rules that govern their usage, is one of the most important discoveries of modern philosophy. The meaning of a speech act is comparable to the status of a move in the game of chess. If you proclaim that you believe that blue roses exist this warrants certain speech acts and prohibits others, just as when you move your queen to a certain board position this warrants certain moves and prohibits others. And importantly, what your proclamation *means* seems somehow to be connected to what else it warrants and prohibits to say. With the advent of the computer age similar views became popular as a theories of mental in addition to linguistic content. Insofar as the brain is a computational system and the mind is in some way connected to its functioning, it was natural to speculate that not only the meaning of words is connected to the rules of language use, but the meaning of thoughts is connected to the computational rules implemented in the brain. Views of the kind described can be gathered under the label of *inferentialism* or *conceptual role semantics*.

In their contribution to the Oxford handbook on consciousness, Mendelovici and Bourget have formulated an argument against inferentialism. Their *swapping argument* demonstrates that arbitrary representational contents can go together with any kind of inferential relations between the states representing them, thus making inferentialism untenable. (Mendelovici and Bourget, 2020) My first goal in this paper will be a spirited defense of the validity of their argument. Mendelovici and Bourget have indeed shown that, given the model theoretic assumptions standardly used in philosophical semantics, inferential relations and content are independent. Furthermore, the argument can be applied to *all* kinds of inferentialism, be it modest or ambitious, concerned with mental or lingual content, and so on. Its generality, I will argue, is

also its downfall. As inferentialism, at least in some limited variety, is extremely plausible, the swapping argument, rather than weakening inferentialism, should make us critical of the model theoretic foundations it is based on. Philosophical semantics is in need of an alternative framework. And just as standard model theoretic semantics goes back to the meta-mathematical thought of Frege, philosophers might look toward the meta-mathematical framework of intuitionistic type theory for new semantic foundations.

The following section will give an overview of the various forms inferentialism can take. The section after that introduces the swapping argument and shows why it can indeed be used to refute the whole breadth of inferentialisms. Section four will use two thought experiments to give intuitive support for inferentialism in spite of the swapping argument. This will also lay the foundation for a discussion of alternative views on the nature of content in the final section, views where content is *identical to* certain inferential roles rather than being constituted from referential relations.

Inferentialisms

Inferentialism holds that the content of representational states is determined by their inferential roles. In its most general form this means that representational states behave in accordance with rules that cohere with the inferential relations between the contents they bare. The simplest example for such a theory is one for the meaning of logical operators. Gentzen has discovered that logical operators can, without loss of generality, be defined in terms of the rules that govern their usage in a language. (Gentzen, 1935) Thus, to urge an example, a very simple inferentialist theory might claim that a state represents logical conjunction in virtue of the rules that govern its usage. A symbol 'a' signifies conjunction precisely if, if we have two statements 'A' and 'B' in our language, we can introduce 'AaB', and further, if we have

‘CaD’, we can write ‘C’ and we can write ‘D’. An inferentialist theory of the meaning of logical conjunctions is one that holds that this is all there is for a symbol to signify conjunction.

The inferentialisms we are interested in here are those that go beyond the meaning of logical operators and try to capture non-logical content. These inferentialisms come in many guises. First, they differ in their *explananda* in that some cover *mental*, some cover *lingual* content. Secondly, inferentialism can be *pure* or *impure*, where the former holds that inferential relations are sufficient for explaining content, (for instance Brandom, 1996) while the latter holds that they are merely co-determinative of content. (Chalmers, 2021) Pure inferentialists hold that bearing content is solely a matter of inferential roles, impure ones hold that inferentialism needs to be conjoined with other theories of representation. For instance, one might hold that the information bearing co-variation or *tracking relations* of representational states together with their inferential roles explain content. Or one might hold that inferential relations need to be linked up to an irreducibly mental phenomenal intentionality, i.e. the content-bearing nature of conscious mental states, to explain content. Philosophers sometimes also differentiate between *short-armed* and *long-armed* inferentialism, where the former holds that inferential roles cover merely the interaction of representational states, while the latter holds that inferential relations sometimes include non-representational states of affairs. (Harman, 1987) But of course also long-armed inferentialists agree that inter-representational relations non-trivially contribute to content determination. The distinction between long-armed pure inferentialist and short-armed impure inferentialists that think inferentialism needs to be combined with a tracking theory of representational content can get blurry at times.

Inferentialisms also differ in their *explanans*. There are both different conceptions about what constitutes accordance with a rule and different conceptions about what the relevant inferential relations between contents are. First, one can think of rules *descriptively* or *normatively*.

Descriptive inferentialism, what one might also call *computational* inferentialism, conceptualizes rules roughly in the way as computer programs are conceptualized in computer science. They are factual descriptions of transitions between states. A descriptive inferentialist regarding logical vocabulary might hold that a certain symbol employed by a computational formalism represents conjunction because the system uses the symbol in the way specified above in its calculations. (Block, 1986) On the other hand, a normative inferentialist has a stronger conception of what a rule is. Here a rule requires that there is a certain norm in place that can be violated. Norms might exist in virtue of a societal practice of enforcing them, (Brandom, 1996) but one might also imagine weaker normative conceptions similar to those employed by the teleosemanticists, (for instance Dretske, 1995) i.e. something behaves according to a rule where it fulfills its biological function by doing so.

Furthermore, different theorists differ in what the relevant rules of inference holding between contents are. A natural first assumption would be to use the rules of first-order predicate logic. However, one might also use a probabilistic calculus to cover a wider range of inferential forms, (Kuhn, 2022) or one might hold that inferences that can be justified *a priori* are relevant for content determination. (Chalmers, 2021) Some inferentialists also hold that there are special meaning-constituting *material* inferences, where these are roughly the typical inferences that language users will learn to make when they acquire a concept. (Sellars, 1953; Brandom, 2000)

All mentioned views are views about how representational states are imbued with content. These *meta-semantic* views have to be firmly kept apart from *semantic inferentialisms* that claim that content *is* to stand in certain inferential relations. For the time being I will assume what I shall call *standard philosophical semantics*, the view according to which to have content is to have conditions of satisfaction and conditions of satisfaction can be spelled out in

terms of set-theoretic constructions, i.e. objects that bear predicates, relations, and functions that take objects as arguments. The content of sub-sentential (or whatever the mental equivalent might be) representational states is then thought to be determined by their role in determining the conditions of satisfaction of full-fledged representations. Probably the best worked out version of this view is possible worlds semantics as described by Chalmers. (Chalmers, 2006) Later, I will argue that the combination of meta-semantic inferentialism and standard philosophical semantics is unstable. The meta-semantic inferentialist should wholeheartedly embrace *semantic* inferentialism, too. More on that later.

Meta-semantic inferentialism constitutes a family of diverse views rather than a single coherent doctrine. It is all the more surprising that there is strong argument that seems to refute them wholesale.

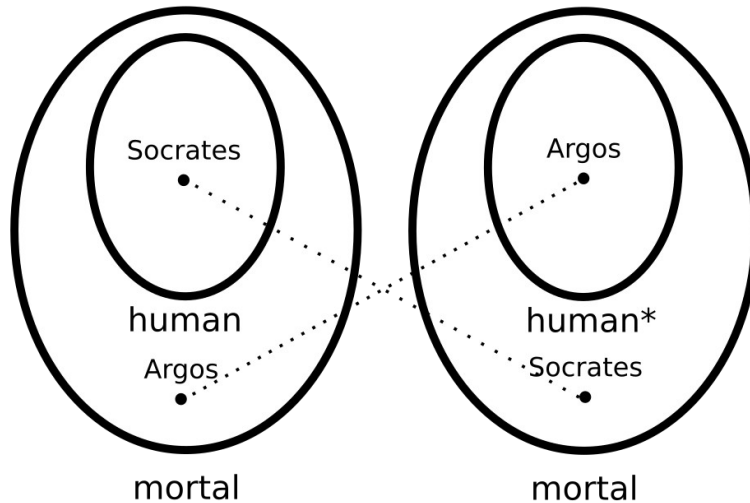
The Swapping Argument

Mendelovici and Bourget formulate their swapping argument against inferentialism in a context of a discussion of the relation of phenomenal consciousness and mental content. Their concern is mainly to question whether naturalist accounts of mental content can capture the distinctive content of conscious states. One such account would be a pure, descriptive, short-armed mentalist inferentialism that is, as I read the authors, built on the inferential rules of first-order predicate logic. The swapping argument seems to show that this view massively under-determines content. In this section I will first present the swapping argument and then show why it refutes inferentialism across the board, not just a limited mentalist variety.

The swapping argument is inspired by Putnam's model theoretic argument against realism. (Putnam, 1977) The central claim is that, even if we grant that the inferentialist can pick out

inferential relations between contents by the causal relations between brain states, the inferential relations will still not serve to fix the content of the mental states. The reason is the following. Imagine three mental states a, b and c, where a and b together cause c in some way that suits your inferentialist account. Let us further assume that a represents proposition A 'Socrates is human.', b represents B 'Humans are moral' and c represents C 'Socrates is mortal'. Furthermore, the inferentialist will hold that the contents are determined by the mirroring of inferential by causal relations.

The swapping argument then takes the following form: The inferential relations between the contents of the mental states do not sufficiently fix the contents of those states. For the same inferential relations can go together with many different contents. In model theoretic terms we can express this as the well known fact that the inferential relations between statements of a theory will not help to fix the intended model of the theory. In the case of states a, b and c referring to A, B and C, we can easily construct an alternative model that would go together with the same inferential relations. The following image shows two equivalent models that satisfy the same inferential relations between contents. The dashed lines signify the swapping operation for constructing an equivalent model from an intended model.



So instead of A, B and C, the three mental states could just as well represent A* ‘Argos is a human*.’, B* ‘Humans* are mortal.’ and C* ‘Argos is mortal.’ So in other words, whether the inferential relations pick out the concept human or human*, where a human* picks out all humans except Socrates but instead Odysseus’s dog Argos, is not fixed by the inferential relations. Accordingly, the conditions of satisfaction aren’t fixed by the inferences either. Of course philosophers would prefer models that capture the intensions of concepts as functions from possible worlds to extension and contents as functions from possible worlds to truth values, however it should be clear that the argument works just as well if there are many Socrateses and Argoses spread out across many different possible worlds. Mendelovici and Bourget conclude: “Our Putnam-style procedure constructs minimally differing extensions and intensions, but it is easy to see that we can also construct massively different extensions and intensions, since many worlds contain large numbers of objects, which can be swapped. Even if

some of our contents in fact are a little bit indeterminate, it is implausible that they are massively indeterminate in this way.” (Mendelovici and Bourget, 2020, p. 572)¹

The first important question is *how* indeterminate are contents if the swapping argument is valid? Is there space for a substantial inferentialism that holds that contents are fixed up to some threshold of indeterminacy where the swapping argument becomes relevant? There is not. It turns out that if the swapping argument is valid then inferentialism can at most fix the meaning of logical constants. To see this, we can use what Walsh and Button have called a *push-through construction* that shows that for two theories (or inferential systems) with two different intended models (or interpretations in terms of contents) we can swap the models between the two theories (inferential systems) entirely. (Walsh and Button, 2018, p. 35)

So let us assume that the inferential systems α and β are composed of internal representations that, unlike in the simple example given by Mendelovici and Bourget, have a combinatorial structure to them. Representations will consist of logical constants, name-, relation- and function terms. Furthermore, α and β will have an intended model that assigns interpretations for these terms that are part of the domains α and β are referring to. In some instances, as in the example of human and human*, the domains of the inferential system’s of the intended interpretations may be identical. We now need some bijective function s from the domain α and to the domain of β . We will call this function the *swapping function* and we will call the result of its application a *swap*, the application of its inverse a *re-swap*. In the example above the swap takes us from Socrates in α to Argos in β and from Argos in α to Socrates in β . We can then construct an interpretation (an alternative model) for α where the representations in α will end

1 Under plausible assumptions the swapping argument can be generalized to cover *interpretativism*, very roughly the view that a system bears representational content if the behavior of the system can be predicted well, if we assume that it represents that content. (for instance Lewis, 1984) For, it seems what is rational are *inferences*. But swapped inferences are just as rational as non-swapped ones.

up referring to objects in the domain of β : Names of α refer to the object their referent has been swapped with and relations terms hold between the swapped relata. Functions are a little complicated to handle as, on the swapped interpretation, they take an argument from the swapped input value to the re-swapped output value. So for instance, if we introduced a looks-at function that outputs the object someone is looking at and suppose that Socrates is looking at Argos, then under a swapped interpretation we want the function to take us from Argos to Socrates, evidently picking out some other function looks-at*.

This generalized swapping argument is almost completely general. It works for all content beyond the meaning of logical constants and it shows that content can be swapped between two inferential systems as long as there exists a bijective swapping function between their domains. Of course, such a function is guaranteed to exist as long as the domains have equal cardinality. So it seems inferentialism can fix nothing beyond the cardinality of a domain.

One might ask whether the swapping argument can be weakened by an appeal to other theories of inferential relations than first-order predicate logic. This seems unlikely. As the original formulation of the swapping argument seems to show, mere propositional logic seems just as vulnerable as its first-order generalization. Second-order logic, where we are allowed to quantify over relations and thus make statements like ‘There exists a relation such that...’, is unlikely to be of much use either. The same procedure that constructs human* from human will give us alternative higher-order sets for a second-order calculus to quantify over.

A probabilistic calculus also does not seem to be of help. Imagine replacing the rigid circles in the above diagram with cloud like structures, such that an object can fall under a certain predicate with a certain probability. Or alternatively, instead of representations assigning binary truth values to sets of possible world, image them assigning continuous values between zero and one. Arguably, the right kind of inferential relations would then described by a kind of

Bayesian conditionalization, where an inferential system would assign probability values to propositions, and the state transitions of the system would be isomorphic to those dictated by Bayes' rule. (Kuhn, 2022) Still it seems that, on the level of objects, a swapping operation could take place, thus constructing arbitrary alternative interpretations of the same probabilistic inferential system.

It might also be wondered whether informally described material inferences might do the trick. From the perspective of the current discussion it is unclear how this should happen. The trouble originates from the way we are used to interpreting an inferential calculus, namely by the assignments of interpretations or models, rather than from the properties of any particular calculus. So if we conceive of interpretations as assignments of objects, sets of object etc. to terms, and if we think of valid inferences as derivations that are valid under those interpretations, then we can always construct arbitrary countervailing interpretations by swapping referents.

The case against the inferentialist can be further generalized. As the swapping argument attacks not the presumed connection between inferential relations and the interactions of representational states, but rather questions whether inferential relations can fix content, the argument is completely neutral on the varieties of inferentialism discussed above. It does not matter whether we conceive of inferential relations as normative or descriptive, of the relevant states as mental or lingual in nature. But the most surprising fact here is that the swapping argument is just as well applicable to impure inferentialism as it is applicable to pure inferentialism. How can this be, as an impure inferentialist has vastly greater resources at her disposal, because she only claims that inferential roles play *some* part in the explanation of content?

The issue is, as we shall see, that when one claims that inferential relations plus X explain content, the swapping argument can be used to show that in fact X does all the explanatory

work while the inferential relations do none. To make this plausible, some additional context is required. In the philosophy of science there has long been a debate very similar the current discussion. Realists in the philosophy of science want to hold that science teaches us about the objective structure of reality and is not merely instrumentally useful. *Epistemic structural realism*, or *structuralism* for short, claims that the required non-observational content of scientific theories lies in their capturing the abstract or logical structure of reality. It is well acknowledged that these theories suffer from serious strain from the so-called Newman's objection, originally raised against Russel's structuralism. (Newman, 1928) And the problem here is precisely that the part of a theory that is not referring to observational reality, but only consists of statements that bear certain inferential relations, can be made to cohere with arbitrary interpretations or models. (Ainsworth, 2009; Walsh and Button, 2018, chapter 3) Similar debates have sprung up around the validity of Putnam's original model theoretic argument against realism. (Button, 2013)

The analogy I want to draw here is that, just as the structuralist holds that a theory has observational plus mere structural content, the impure inferentialist holds that some content of representational states is determined by some non-inferential X and some content is determined by mere inferential relations. But just as in the case of structuralism we can argue as follows. We can separate the representational states of an inferential system into those that have some non-inferential tethering, and those that have not. This tethering might consist in irreducibly mental phenomenal intentionality, it might consist in tracking relations to external states of affairs, or what not. By hypothesis, the non-inferentially tethered representations that tether the inferential network to reality cannot be subject to arbitrary reinterpretation. Assume, for example, that we have some non-inferential relation that tethers 'Socrates is human.' to state a. However, we can still give re-interpretations for all other states in the inferential system for

precisely as long as we respect the content of the tethered representations: All swappings that are not ruled out by non-inferential tethering are still on the table. Thus, all the explanatory work will be done by whatever tethering we invoke and impure inferentialism collapses into non-inferentialism.

It might seem that the only kind of inferentialism not ruled out by the swapping argument is the long-armed variety. However, as mentioned above, typically even long-armed inferentialists will assume that inter-representation inferential relations enter into content determination, otherwise inferentialism becomes strictly equivalent to a tracking theory. But we have by now ruled out the representational import of inter-representational interactions. Thus, we have also ruled out a long-armed inferentialism that is meaningfully different from a tracking theory.

Before concluding my defense of the swapping argument there is one more view to consider. Chalmers' inferentialism paper is inspired by a certain take on conceptual analysis, where concept users are capable, under idealized conditions, to see whether a given concept applies in a given condition. (a framework established in Chalmers, 2012) As you will see later, I am sympathetic to this part of his theory. The relevant conditions will be given to the concept user in form of a sub-conceptual representations that bears content that can not be further explained in terms of inferential roles. According to Chalmers, these contents are conferred by an irreducibly phenomenal form of representation, though this is imaterial to the discussion at hand. The interesting suggestion is that the relevant inferential relations are those that can be *a priori* justified, given the content of non-inferentially tethered states. (Chalmers, 2021) For instance, we might imagine some phenomenally conscious representations of certain experiences of a glass of water which justify *a priori* that there is water in the glass. Then a further representational state that is properly caused by these conscious representations will bear the content of there being a glass of water. The restriction to *a priori* inferences gives us no free

hand to swap the referents of the state representing that there is water in the glass, with something else. It thus seems, the swapping argument none-withstanding, we have given a non-vacuous account of concept application in terms of inferential roles.

Chalmers' theory is fine as far as it goes, however it is important to see that really all the explanatory work is done by the non-inferentialist elements of the theory. We might say that Chalmers' inferentialism is *epiphenomenal* in that content ends up correlating with inferential roles, but they really have no explanatory powers of their own. The reason we cannot construct swapped interpretations for the content of the glass-of-water representation is that whatever determines the content of the non-inferential representations that tether it will end up making it true *a priori* that the glass-of-water representation is appropriate. Similar accounts can be constructed in a tracking paradigm: We might say that state a in the original example tracks water, b tracks glasses. Then c, which is caused by a and b together, will track water in glasses. Such an epiphenomenal inferentialism does not allow for true content determination by inferential roles. The inferential elements can be canceled out. Also, it should be noted that epiphenomenal inferentialisms can put additional explanatory strain on the non-inferential elements of the relevant theories. For instance, in Chalmers' case, his phenomenal intentionality theory has to explain how the content that there is water in the glass is already fully implicit in the phenomenally conscious representational states.

The swapping argument pushes us to the conclusion that inferential roles can *contribute nothing at all* to content determination. Is this an acceptable conclusion?

Two Thought Experiments

In this section I will motivate the view that the radical anti-inferentialism implied by the swapping argument is unacceptable. To do so, I will present two thought experiments. The first one motivates that mere computational processes sometimes suffice for moderate content attribution and is thus more in line with a mentalistic descriptive inferentialism. The second thought experiment is more aimed at motivating linguistic inferentialism. It will also give us a hint of what goes wrong in the swapping argument.

First, consider an advanced chess computer. The most sophisticated chess computers to date will use reinforcement learning to learn from past mistakes. Furthermore, they are able to improve their own game by playing against themselves. *Prima facie* it seems extremely plausible that there is some non-vacuous sense in which the internal states of such a system represents aspects of the game of chess just in virtue of the interactions of internal states being subject to certain rules.²

Now it might be argued against the intuition that the chess computer bears non-trivial content by holding that content attribution is really to be explained in terms of its contextual embedding rather than in terms of its intrinsic structure. For instance, it might be claimed that the only sense in which the chess computers' internal states represent chess moves is to be explained solely in terms of us using it to play chess against. If taken to the extremes I find this view quite implausible. It would imply that we could confer arbitrary contents to the chess computer's internal states by varying its contextual embedding. We might construct a trans-

² I have the suspicion that, how intuitive readers will find this argument, will depend directly on their practical experience with programming. Among some philosophers without any programming experience there is a certain tendency to regard the actions of computers as 'just code' or 'just syntax', while for those with such experience the world of computer code is one of interesting and non-trivial structures.

ducer layer that maps the inputs and outputs of the chess computer to the inputs and outputs of a simple tick-tack-toe program. As there are vastly more board positions and moves in a game of chess than in a game of tick-tack-toe, such a mapping is quite possible. We can now use the chess computer to play tick-tack-toe. It seems highly implausible that under such conditions, we should describe the figure corresponding to the queen in the computer's internal processing as *solely* representing some aspect of the game of tick-tack-toe. Rather, the thought experiment of the chess computer gives at least *prima facie* evidence that we need some conception of representational content that can explain the representational import of purely inter-representational inferential roles.

Let us now consider the content of language. Instead of looking at communication between humans, let us consider a Martian anthropologist trying to learn the human language. If the concept of content has any relevance at all, then learning a language should involve grasping the content of the statements made in it. On an abstract level, what I have called standard philosophical semantics would imply that such grasping would involve finding an interpretation of the language by figuring out the right mapping of terms to objects, relations and functions. This would arguably not happen by figuring out the mapping directly, but by figuring out the mapping to an already properly interpreted Martian language.

It is a staple of inferentialist thought going back to Wittgenstein, (Wittgenstein, 2016) to point out that a view of language that takes reference as primary does not fit well with how language is actually learned. Instead, it seems one first learns the proper rules governing the use of certain terms, which will essentially involve rules about how certain terms relate to other terms. Any mapping of human language to an already interpreted Martian language will proceed *after* the relevant inferential rules are learned. Unlike someone who already has a dictionary for a foreign language, the Martian anthropologist would first figure out some substantial

part of the rules of human language and learn how to apply them before trying any inter-lingual mapping. The task of starting by figuring out the reference of utterances before understanding how they fit roughly into the network of language practices is hopeless. Note for instance that pointing at objects in order to obtain their names is quite out of the question, as this already presupposes some understanding of the pointing gesture in relation to the concept of a name and a referent. As least as far as learning a language is concerned, inferential rules come first, referents comes second. If the swapping argument is valid however, it is very hard to account for this priority of rules in language learning as it seems to show that rules are immaterial to content.³

The thought experiment of the Martian anthropologist adds some intuitive support to the importance of inference in the explanation of content, however it might hold the promise of showing us where the swapping argument went wrong. If learning the content of expressions is not primarily the learning of a correct interpretation but of learning the rules connecting it to other expressions and to extra-lingual states of affairs then this seems to indicate that content is primarily a matter inferences rather than of referents and truth conditions.

Semantic Alternatives

This section will explore how an inferentialist semantic framework can save meta-semantic inferentialism from the swapping argument. It will describe such a semantic framework in informal terms, show how many features of standard semantics can be recovered by using a

3 One might here also argue that the success of large language models adds further weight to the inferentialist case. I think this is so, however it hinges on the controversial issue whether the utterances of these machines bear content independently of our interactions with us.

framework inspired by intuitionistic type theory, and finally, will briefly go through some potential advantages and disadvantages as compared to standard philosophical semantics.

Semantic inferentialism is the view that the content of a representational state *consists* in the rules that govern its relations to other representational states, and maybe also, but not exclusively, not non-representational states of affairs. We have discussed a limited semantic inferentialist view already when introducing the Gentzen-style inferentialist analysis of logical constants. The idea there was that being governed by the relevant rules is all there is to signifying logical conjunction. A more general semantic inferentialism would hold that this approach can be generalized to non-logical contents.

Before continuing outlining a more general semantic inferentialism, I first want to argue for the coherence of such a doctrine. My argument will be that we already possess a formally worked out semantic inferentialist account of a certain subset of language, namely constructive mathematics, in the form *intuitionistic type theory*. (Martin-Löf, 1980; see Dybjer and Palmgren, 2024 for an accessible introduction) Constructive mathematics is a field of mathematics that is skeptical of mathematical realism where mathematical statements are presumed to be true independently of our in-principle capacity to ascertain their truth value. Instead, the idea here is that the truth of a mathematical statement just *is* its provability and the existence of a mathematical object just *is* constructability. The notion of construction here depends on the relevant mathematical object, of instance we can construct natural numbers by counting, adding etc.

Assuming an equivalence between truth and provability naturally should make one skeptical of the notion of semantic and model theoretic validity. Instead, proofs are understood as constructions *within* a language. Consequently, the truth-makers of claims within a language are conceived as other fragments of the same language, the fragments that constitute proof for the

relevant claims. For instance, the proof of a certain claim about natural numbers is not made true by a proper match under an interpretation relation between the lingual tokens and the natural numbers, but by constructing the relevant natural number that serves as a witness to the claim we want to prove. The construction is done in accordance with the rules of the language and so truth-maker will be inner-lingual.

The main differences between intuitionistic type theory and the first-order logic mostly employed by philosophers are its different logical principles and its employment of the notion of a type. The difference in logical principles directly results from the equivalence of provability and truth: It is not generally true that, for any P, either we can prove P or we can prove non-P. Thus, the law of the excluded middle will not hold in intuitionistic type theory. The second important principle of intuitionistic type theory, the notion of a *type*, is that of certain terms for which there are decidable rules for checking whether some lingual object falls under the type, and thus rules for introducing such types. For instance, natural numbers, sets, and proofs themselves are all types within intuitionistic type theory. We can think of the rules for determining types *meaning-constitutive* rules that supply meaning for the relevant type-term.

As an aside, it should be noted that the possibility of intuitionistic type theory itself puts strain on standard philosophical semantics. According to it, one would have to spell out the content of a type theoretic account of constructive mathematics in terms of interpretations in terms of set theoretic constructions. The strain comes from the fact that, the way type theory is set up, giving such an interpretation seems strictly unnecessary as interpretations can be constructed within the theory itself in terms of proofs and constructions. There is thus at least some tension between how type theorists claim their theory handles semantics, and how their expressions get imbued with meaning according to standard model theoretic semantics.

I do not wish to dive deeper into intuitionistic type theory here. It merely serves as an illustration of the coherence of an inferentialist semantics. The remainder of the paper will now deal with the question whether this should make us positive that an inferentialist semantics can work across the board.

The most obvious rejoinder to an analogizing of broadly inferentialist semantics and intuitionistic type theory is that the practice of constructive mathematics is disanalogous to almost all other representation endeavors. While for the former it may be plausible to capture content solely in terms of the language itself, this certainty is not valid generally, as most representational systems represent something *outside* of themselves.

Of course, we have already seen that inferentialism, meta-semantic or semantic, does not condemn us to solipsism. For the inferentialist can claim that long-armed inferential roles make representations answerable to non-representational states of affairs. To defend inferentialism against the solipsism objection we do not have to settle on any particular ‘length of inferential arms’. One might hold that inferential roles can be defined purely externalistically, in terms of actual object in the environment, or more internalistically, in terms of sensory stimulation or sensory experiences. These will be connected to representational states that play a similar role as primitive types in intuitionistic type theory and that will be connected to special introduction rules, that involve reference to non-representational states of affairs. Action taking effect in non-representational reality can be covered similarly. In fact, views of this kind have been elaborated within formal linguistics for some time now. (Ranta, 1994; see Chatzikiyiakidis and Cooper, 2016 for an accessible introduction) Still, the deeper philosophical relevance of these type theoretic treatments seems to have been lost on most analytic philosophers outside the field of formal philosophical logic.

In a philosophical inferentialist semantics, the notion of type would arguably be replaced with that of a *concept*, where a concept will be defined in terms of its introduction rules, i.e. in terms of the rules for checking whether some state of affairs justifies applying the concept. This is of course quite similar to Chalmers' account of concept application inspired by conceptual analysis: Concept users, under ideal conditions, are thought of as being able to decide when to apply a concept they have mastered.

The idea of constitutive rules is under Quinean strain: Quine famously argued against a distinction between meaning-constitutive analytic rules and synthetic ones. (Quine, 1951) Without such a difference, inferentialism collapses into a form of holism where the only way two states of two representational systems can have the same inferential role, and thus the same content, is for the two inferential systems to be identical wholesale – an unacceptable conclusion. However, I agree with Chalmers that, as long as we are sufficiently careful in describing the relevant inferential rules, we can answer Quine straightforwardly. (Chalmers, 2012, chapter 5) We just have to define our complex concepts in terms of basic observational ones, or in terms of primitive ones (like that of an integer). Given such definitions, it is obviously not the case that meaning-constitutive inferential connections automatically spread through the entire inferential network. Furthermore, any accusation of a *petitio* on part of the Quinean, due to the fact that the inferentialist account already *assumes* that there are certain differences between meaning-constitutive and non-constitutive inferences would be misplaced: Quine merely questions whether we can give a satisfying definition of the notion of equality of meaning, going on from there to question the notion of analytically valid inferences. If we turn semantics on its head and define the notion of meaning identity in terms of basic meaning-constitutive inferences, the Quinean argument becomes mute.

So let us summarize the emerging perspective. I have been arguing that a coherent meta-semantic inferentialism needs to be semantically inferentialist as well. Semantic inferentialism treats contentful representations as constructed from concepts, and concepts as entities ruled by meaning-constitutive rules. Truth conditions of representations would be replaced by rules that establish what would count as evidence for or against a certain claim. This evidence gathering might involve long-armed inferences of certain types, thus establishing a solid relation between representation and reality. This view is neutral on the question of pure vs. impure meta-semantic inferentialism. Other forms of meta-semantic theories might matter in explaining content determination, their semantic import being accounted for in terms of long-armed inferential roles. Semantic inferentialism is of course also neutral on the issue of whether inferences should be characterized normatively or descriptively.

Importantly, long-armed inferentialism is no longer threatened to collapse into a mere tracking view, no matter how long the arms of our inferential system are. For there is room for purely inner-representational relations determining content. Think back of the thought experiment of the chess computer linked up to the world in terms of a tick-tack-toe transducer: The internal states of the chess computer might track states in tick-tack-toe, but through their internal inferential relations they also represent aspects⁴ of the game of chess. This meta-semantic purport cannot be captured by a semantics that thinks of meanings in terms of interpretations and referents.

Meta-semantic inferentialism in conjunction with semantic inferentialism is impervious to the swapping argument. When the meaning of a representation is conceived primarily in terms of

4 It only grasps *aspects*, because arguably the full concept of chess involves knowledge that chess is a social game, that winning increases social status, and so on. These aspects of course elude the machine described above.

the inferences we are allowed to make with it, and only secondarily in terms of reference to non-representational matters of fact, then the idea of reference-swapping makes no sense anymore. There can still be interpretations of languages in terms of other languages. However, the interpretation is secondary to inferential structure. Interpretation is something done *within* languages. It is not the thing connecting language or thought to extra-lingual reality in the first place.

A possible objection is that the given account leads to idealism or anti-realism.⁵ For it seems clear that there can be no facts (true contentful states) that are in-principle outside of the web of inferential and evidential relations. I myself am sympathetic to this conclusion. Given that the resulting view would not force one to draw a clear and tight boundary between what can and what cannot reasonably be said and thought the resulting idealism would not be one that ties the nature of facts to some contingent bound of human intellection. That being said, the argument of this paper is also compatible with realism, namely if we accept the right kind of long-armed inferential roles into our theory. If inferential roles stretch out into the world and touch the objects of discourse, idealism might be avoided. When long-armed inferential roles involve recognition transcendent matter of fact, it will sometimes recognition transcendent whether a state plays some long-armed inferential role. Thus, any long-armed semantic theory will entail the existence of partly recognition transcendent contents. I suspect however that this will be a *general* consequence of realist theories. Thus, the semantic issue seems to be or-

5 I use the terms anti-realism and idealism as more or less equivalent. Standardly, anglophone philosophy tries to differentiate clearly between views where reality is mental, and views where all facts are in-principle intelligible. However, there is reason to assume that idealism, as the term was used by mainly by German idealists, was much closer to what is today called anti-realism than to the view that the world is made from mind-stuff. (Brandom, 2001) As I think the terminological confusion has lead to an under-appreciation of the idealist tradition and Hegel in particular within analytic philosophy, I stick with the ‘classical’ terminology.

thogonal to the realism-idealism one. The differences arise merely from the fact that standard semantics has an intrinsically realist feel to it, a feel that evaporates as soon as we pay sufficient attention to its model theoretic foundations. (Button, 2013; Haukioja, 2017; Walsh and Button, 2018)

Having discussed some supposed disadvantages of semantic inferentialism, I will also mention two possible advantages, one of them admittedly speculative. First, semantic inferentialism can deal better with *hyperintensionality*. Hyperintensionality refers to differences in content that cannot easily be rendered as a difference in truth conditions. For instance, the thought that the angles in a triangle sum up to 180° and the thought that four minus four isn't four seem to have very different contents. However, they seem to have the same truth conditions, when conceived in possible worlds semantics: Both are true in *all* possible worlds. Defenders of possible worlds semantics have given various treatments of the issue. Lewis has said that, what we learn when we learn a hyperintensional truth, it is really just a truth about how certain words are used in the respective language community. That " $4+4\neq 4$ " is considered false by some given language community is *not* true in all possible worlds. Chalmers has amended this idea with a treatment in terms of two-dimensional semantics, but fundamentally he seems to agree with Lewis. (Chalmers, 2021) And Priest has suggested that we need to consider a paraconsistent logical paradigm, so that we can say that hyperintensional contents rule out certain *impossible* worlds, i.e. worlds where four plus four *is* four, but still the angles in triangles sum to 180° . (Priest, 2008) None of these strike me as terribly convincing. Intuitively, mathematicians don't mainly study a language community. And it does not seem to be the case that in order to grasp the difference between hyperintensional claims I have to grasp any paraconsistent logical principles.

From the perspective of inferentialist semantics hyperintensionality can be handled gracefully. The meaning of representations will be connected to the scenarios that give evidence for or against them. In case of the above propositions, these will be arithmetical and geometrical proofs. And the proofs will be quite dissimilar for both propositions.

Now let me end on a speculative note. Mendelovici and Bourget have formulated the swapping argument in the context of a discussion of the content of conscious states. In their view, naturalist accounts of these content are lacking because neither inferential roles nor tracking relations can explain the primitive ‘blueness’ of a visual experience, like the experience of seeing a blue rose. Inferential roles cannot explain it because, in light of the swapping argument, they explain nothing at all. Tracking relations cannot explain it because there are there aren’t instances of primitive blueness in our environment to track. (an argument worked out in detail in Mendelovici, 2018) In light of the discussion of the paper we have to ask: Is this failure of meta-semantic naturalism just the result of an over-reliance on the standard philosophical semantic, and could an inferentialist semantics do better? Whether we can explain the primitive content of conscious experience in terms of introducing inferentially primitive observational concepts must be discussed in another paper however.

Conclusion

This paper made the claim that in order to deal with the swapping argument, we have to turn our entire semantic framework on its head. According to standard philosophical semantics, given some representational system, we have to ask how the states within that system are to be connected to referents in order to acquire meaning. But this way of thinking about meaning leaves no role to play for the manifestly meta-semantically relevant inferential structure of representational systems. Instead, we should start out from basic rules governing the interac-

tions of representations, bootstrapping concepts like interpretation and truth-conditions from there. Inferences come first, reference comes along for the ride.

Literature

Ainsworth, P.M. (2009) ‘Newman’s objection’, *British Journal for the Philosophy of Science*, 60(1), pp. 135–171. Available at: <https://doi.org/10.1093/bjps/axn051>.

Block, N. (1986) ‘Advertisement for a semantics for psychology’, *Midwest Studies in Philosophy*, 10, pp. 615–678. Available at: <http://dx.doi.org/10.1111/j.1475-4975.1987.tb00558.x>.

Brandom, R.B. (1996) *Making it Explicit*. Harvard University Press.

Brandom, R.B. (2000) *Articulating Reasons, Articulating Reasons*. Harvard University Press. Available at: <https://doi.org/10.2307/j.ctvjghvz0>.

Brandom, R.B. (2001) ‘Holism and Idealism in Hegel’s Phenomenology’, *Hegel-Studien*, 36, pp. 61–95.

Button, T. (2013) *The Limits of Realism*. Oxford University Press UK.

Chalmers, D. (2006) ‘Two-Dimensional Semantics’, in E. Lepore and B. Smith (eds) *The Oxford Handbook to the Philosophy of Language*. Oxford University Press.

Chalmers, D. (2012) *Constructing the World*. Oxford University Press.

Chalmers, D. (2021) ‘Inferentialism, Australian Style’, *Proceedings and Addresses of the American Philosophical Association*, 92.

Chatzikyriakidis, S. and Cooper, R. (2016) ‘Type Theory for Natural Language Semantics’, *Oxford Reference Encyclopedia* [Preprint].

Dretske, F. (1995) *Naturalizing the Mind*. MIT Press.

Dybjer, P. and Palmgren, E. (2024) ‘Intuitionistic Type Theory’, in E.N. Zalta and U. Nodelman (eds) *The Stanford Encyclopedia of Philosophy*. Winter 2024. Metaphysics Research Lab,

Stanford University. Available at: <https://plato.stanford.edu/archives/win2024/entries/type-theory-intuitionistic/>.

Gentzen, G. (1935) 'Untersuchungen Über Das Logische Schließen. I.', *Mathematische Zeitschrift*, 35, pp. 176–210.

Harman, G. (1987) '(Nonsolipsistic) Conceptual Role Semantics', in E. LePore (ed.) *New Directions in Semantics*. London: Academic Press, pp. 55–81.

Haukioja, J. (2017) 'Postscript: Recent Work on Putnam's Model-Theoretic Argument', in B. Hale (ed.) *A Companion to Philosophy of Language*. Blackwell, pp. 730–733.

Kuhn, P. (2022) *The World From Within: An Investigation Into the Hard Problem of Consciousness From the Perspective of Bayesian Cognitive Science*. PhD Thesis. Goethe University Frankfurt.

Lewis, D. (1984) 'Putnam's Paradox', *Australasian Journal of Philosophy*, 62(3), pp. 221–236.

Martin-Löf, P. (1980) *Intuitionistic Type Theory*. Bibliopolis.

Mendelovici, A. (2018) *The Phenomenal Basis of Intentionality*. Oxford University Press.

Mendelovici, A. and Bourget, D. (2020) 'Consciousness and Intentionality', in U. Kriegel (ed.) *The Oxford Handbook of the Philosophy of Consciousness*. New York, USA: Oxford University Press, pp. 560–585.

Newman, M.H.A. (1928) 'Mr. Russell's Causal Theory of Perception', *Mind*, 37(146), pp. 26–43. Available at: <https://doi.org/10.1093/mind/xxxvii.146.137>.

Priest, G. (2008) *An Introduction to Non-Classical Logic: From If to Is*. New York: Cambridge University Press.

Putnam, H. (1977) 'Realism and Reason', *Proceedings and Addresses of the American Philosophical Association*, 50(6), pp. 483–498.

Quine, W.V.O. (1951) 'Two Dogmas of Empiricism', *Philosophical Review*, 60(1), pp. 20–43. Available at: <https://doi.org/10.2307/2266637>.

Ranta, A. (1994) *Type-Theoretical Grammar*. Oxford University Press on Demand.

Sellars, W. (1953) 'Inference and Meaning', *Mind*, 62(247), pp. 313–338. Available at: <https://doi.org/10.1093/mind/LXII.247.313>.

Walsh, S. and Button, T. (2018) *Philosophy and Model Theory*. Oxford, UK: Oxford University Press.

Wittgenstein, L. (2016) *Werkausgabe Band 1*. 22nd edn. Suhrkamp Verlag.