Can Computers Reason Like Medievals? Building ‘Formal Understanding’ into the Chinese Room[1]

LASSI SAARIO-RAMSAY

What do the Middle Ages and modern computers have in common? Nothing, you may say: the Middle Ages are synonymous with darkness, stagnation, and ignorance, whereas computers are the crown jewels of enlightenment, progress, and innovation. I would reply that they actually have many things in common, one of which is their intimate relationship with logic, the study of valid reasoning—and yet the medieval conception of valid reasoning turns out to be a tough nut to crack for a computer, as we shall see.

The question of whether computers can distinguish valid from invalid arguments was first asked before there were any actual computers. The idea of implementing an algebra of logic on a computing machine had been envisioned as early as the 17th century by G. W. Leibniz (Davis, 2018, 5). In 1928, David Hilbert and Wilhelm Ackermann posed the so-called Decision Problem (*Entscheidungsproblem*), leading to the development of various models of computation that were then incarnated by actual computers as we know them. We can paraphrase this problem, given certain now obvious conventions and results, as follows:[2]

[1] This paper has undergone peer review.

[2] The problem, as it was originally formulated, was to find “a procedure that allows one to decide through finitely many operations whether a given logical expression is universally valid or, alternatively, satisfiable” (Hilbert and Ackermann, 1928, 73). We may equate such a procedure with an actual computer, given the so-called Church–Turing thesis (see Copeland, 2024) and the fact that all prominent models of computation coincide with what actual computers can do. A formula A is “universally valid” (i.e., a logical truth) if and only if the argument from the empty set of premises to A is formally valid, and an argument from $A_1, A_2, \dots, A_n$ to C is formally valid (i.e., C is a logical consequence of $A_1, A_2, \dots, A_n$) if and only if the formula $A_1 \land A_2 \land \cdots \land A_n \to C$ is universally valid.

> (The Decision Problem) Is it possible for a computer to *decide* for any given argument in first-order predicate logic whether it is formally valid?

The problem was to find a 'decision procedure', or an 'effective method', that would take an argument of first-order predicate logic as its input, terminate after a finite number of steps, and produce the output 'Yes' if the argument is valid and 'No' if it is not. The corresponding problem about propositional logic had already been solved by the truth-table method, which allows one to decide the validity of any given argument in finitely many steps. It had also been proven that decision procedures exist for certain fragments of first-order predicate logic. However, in 1936, Alonzo Church and Alan Turing proved that such a procedure could not be found for full predicate logic and that the Decision Problem was thus unsolvable (Church, 1936; Turing, 1936). In spite of the many amazing things that computers are capable of, the full depth of valid reasoning is something that evades them.

This, of course, is not to say that computers would not be useful for inferential purposes. Even if the Decision Problem is unsolvable, we may find consolation in the existence of various complete proof systems that can be straightforwardly implemented on a computer. Although not all invalid arguments are refutable in these systems in a finite number of steps, all valid arguments *are* provable in these systems in a finite number of steps. We can simply write a computer program that takes an argument as its input and looks for a proof of the conclusion from the premises by systematically applying the rules of the chosen proof system. If the argument is invalid, there is no proof to be found and the execution of the program will never end, but if the argument is valid, there is a proof to be found, and the computer is guaranteed to find it eventually. We have thus solved a weakened version of the Decision Problem:

> (The Recognition Problem) Is it possible for a computer to *recognize* any given formally valid argument in first-order predicate logic?

The solution of the Recognition Problem depends on *formalization*, namely, the fact that the conditions of valid inference can be formulated as formal rules that only consist in manipulating strings of symbols (the *syntax*) without regard to their meanings (the *semantics*). The computer can, for instance, be told that the rule of universal instantiation,

$$\frac{\forall x \varphi(x)}{\varphi(c)}$$

is valid regardless of what the open formula $\varphi(x)$ is about or what object the constant term c denotes. Thus a computer can, in principle, confirm that the argument:

(1) Valtteri is a human being; every human being is an animal; therefore, Valtteri is an animal

is valid; it only has to be formalized into, e.g.,:

(1') $H(v), \forall x(H(x) \rightarrow A(x))$; therefore, $A(v)$,

and then the computer can search for a proof using the rules of the system. Nothing in this process requires that the computer 'understand' what the premises and the conclusion *say* or what the terms that occur in them *mean*.

In our everyday life, however, we are interested in all sorts of arguments, only a part of which can be validated in a formal system of the kind described above. Consider the following arguments, for example:

(2) Valtteri is a human being; therefore, Valtteri is an animal.

(3) Turku is west of Helsinki; therefore, Helsinki is east of Turku.

(4) Daffy is black; therefore, Daffy is not white.

These arguments, as natural as they may seem, do not have a valid form in standard first-order predicate logic. Nowadays we are accustomed to classifying such arguments as merely *materially* valid and leaving them outside the scope of logic proper, but perhaps that is just a prejudice due to our received philosophy of logic (cf. Read, 1994; Halbach, 2020, n. 1). If only we could travel back to the Middle Ages, we would find that most logicians at that time would classify these kinds of arguments as *formally* valid because their conclusions are contained in the 'formal understanding' of their premises, or because the negations of their conclusions are 'formally incompatible' with their premises. Let us call arguments of this kind 'conceptual-formally' valid (to be specified shortly).

To be sure, I am not saying that we have to travel as far as the Middle Ages to find anything like conceptual-formal validity. It lives on in the notion of *analyticity*, traditionally attributed to Kant (cf. Sundholm, 2013), famously refuted by Quine (1951), and recently reconsidered by Boghossian and others (see e.g., Boghossian, 1996; Russell, 2008; Boghossian and Williamson, 2020). The related notion of 'analytic validity' or 'analytic consequence' is sporadically mentioned in the literature (e.g., Kment, 2006; Etchemendy, 2008; Read, 2015); a mechanism called "Analytic Consequence (Ana Con)" has actually been implemented on a computer, in a restricted and incomplete form, as part of the software accompanying the textbook *Language, Proof and Logic* (Barker-Plummer, Barwise and Etchemendy, 2011).[3] Even so, it goes without saying that the medieval account of these issues is in many ways different from ours: our concern with artificially constructed languages, for one thing, is even more alien to the medievals than their Aristotelian worldview of definitions, accidents, and other 'predicables' is to us. Not only is a systematic study of the medieval account an interesting task in its own right, but it also casts a new light on our contemporary considerations on formal validity.

The problem that I pose and try to solve in this paper, then, is this:

> (The Medieval Recognition Problem) Is it possible for a computer to recognize any given *conceptual*-formally valid argument?

As it turns out, this problem is considerably harder than the ordinary Recognition Problem. The notion of conceptual-formal validity, unlike that of ordinary validity, is inexact to begin with. In order to solve the Medieval Recognition Problem, we must first rephrase the problem in more precise terms, or rather, whether conceptual-formal validity can be precisely defined is part of the problem itself. I will begin by sketching a preliminary account of conceptual-formal validity within the framework of modern logic.

1. The Medieval Account of Conceptual-Formal Validity

14th-century theories of formal consequence (*consequentia formalis*) are by far the most remarkable precedent of our contemporary idea of logical consequence or formal validity (cf. Dutilh Novaes, 2020). These theories dethroned the Aristotelian syllogism, which had held pride of place among valid arguments for over a thousand years, in favour of a general account of valid arguments of all sorts. In this section, I will formulate a schematic definition of formal validity for

---

[3] Something reminiscent of 'conceptual-formal' validity may also be found in contemporary information-theoretic logic (Corcoran, 1998), intensional logic (Fitting, 2022), and varieties of necessity (e.g., Plantinga, 1978; Fine, 2005).

an arbitrary first-order language, inspired by what I take to be some of the key tenets of the medieval theories.

The medievals typically first define what it is for an argument to be valid in the first place and then distinguish between formally and materially valid arguments. In our terminology, their definition of validity (consequence) may be stated as follows:

> (Validity) An argument is valid if and only if it is necessarily the case that, if the premises are true, then the conclusion is also true.[4]

The definition clearly hinges on the sense of "necessarily". In contrast to contemporary philosophers of logic who try to neutralize the appeal to necessity in one way or another, medieval authors typically presuppose a primitive notion of necessity and understand it in a very broad sense, including merely temporarily or 'as of now' (*ut nunc*) necessary cases that we would regard as contingent (see Read, 2012; Strobino, 2017). In the above definition, I presuppose a narrower notion of necessity that does not include such cases; in so doing, I am deliberately glossing over irrelevant aspects in the medieval sources to facilitate the systematization of the relevant ones.[5]

In order to define *formal* validity, we need an account of what the form of an argument amounts to. There were roughly two competing notions of form in the 14th century, which I call *structural form* and *conceptual form*.[6] For our purposes, the structural form of an argument is simply what we get when we translate the argument into the language of predicate logic and abstract away from the 'intended' interpretation of the non-logical constants. Structural form is

[4] Equivalently: an argument is valid if and only if it is *impossible* that the premises be true and the conclusion false. This definition is ultimately based on Aristotle's definition of a syllogism (Aristotle, 2009, Ch. 1).

[5] In other words, I restrict the class of valid consequences to what many medieval authors call 'absolute' or 'simple' (*simplex*, as opposed to *ut nunc*) consequences. This restriction is of no importance to us, given that we are solely concerned with formal consequences. Some medieval authors (e.g., Peter of Mantua) subsumed the formal vs. material distinction under simple consequences; others (e.g., John Buridan) subsumed the simple vs. as of now distinction under material consequences; either way, formal consequences turn out to be necessary in the narrow (simple) sense.

[6] My distinction between structural and conceptual form is a systematization of passages found in Walter Burley (2000, 173), Adam Wodeham (1990, 82–83), and several subsequent authors in the 14th century. I have accounted for this distinction elsewhere (Saario-Ramsay, revised and resubmitted). Similar distinctions have previously been suggested by Klima (2016, 335) and Strobino (2017, 184).

located, as it were, at the level of the object language and involves no explanation of the non-logical constants in the metalanguage. For instance, the structural form of the argument (2) is:

(2') $H(v)$; therefore, $A(v)$,

where the predicate symbols H and A and the individual constant v may receive any interpretations of the appropriate type. The conceptual form, by contrast, is given in the metalanguage: in addition to the structural form, it imposes some constraint on the mutual relationships of the interpretations of the non-logical constants. In the minds of the medievals, the argument (2) is underwritten by the understanding that, in the Aristotelian hierarchy of universal terms, the concept 'animal' is part of the essence of the concept of 'human being'. Correspondingly, the conceptual form of (2) consists of (2') *and* the requirement that the extension of H be included in that of A:

(2") '$H(v)$; therefore, $A(v)$' and $ext(H) \subseteq ext(A)$.

Thus, for instance, 'Bucephalus is a horse; therefore, Bucephalus is a mammal' and 'Aristotle is Greek; therefore, Aristotle is Egyptian' are both of the structural form (2'), but only the first is of the conceptual form (2'').

With these notions of form at hand, we can proceed to define formal validity based on the following scheme:

> (___-formal validity) An argument is ___-formally valid if and only if every instance of its ___ form is valid.

Each of the two notions of form, when applied to the scheme above, results in a corresponding notion of formal validity. An argument is *structural-formally valid* if every instance of its structural form is valid, and *conceptual-formally valid* if every instance of its conceptual form is valid.[7] Structural-formal validity roughly

[7] The two kinds of validity correspond to the "FO Con" and "Ana Con" procedures of the *Language, Proof and Logic* package mentioned above. "In theory", Ana Con should be able to verify any argument that is valid in virtue of the meanings of logical constants *and the non-logical constants* of the software's specific object language, while in practice, its implementation disregards the meanings of two predicates as too complex and even then fails to find some proofs (Barker-Plummer, Barwise and Etchemendy, 2011, 60–61, 577–578; see also Cohen, 2004). I take

corresponds to the so-called Parisian account of formal consequence, represented by John Buridan and a few others (see Archambault, 2017; Ciola, 2018). Conceptual-formal validity, in turn, renders a line of thought gradually developed in the so-called British tradition of formal consequence (see Saario-Ramsay, revised and resubmitted). My version of it is mostly inspired by Peter of Mantua (Strobino, 2017) and, to some extent, by Paul of Venice (1990). Structural-formal validity is analogous to our modern notion of logical consequence and has in fact been formalized using the resources of modern formal logic (Dutilh Novaes, 2007), but it is conceptual-formal validity that more closely corresponds to the dominant views in the 14^{th} century (cf. Saario-Ramsay, revised and resubmitted). On this account of formal validity, arguments like (2), (3), and (4) are not merely materially valid, as they are on the structural account, but formally valid. According to a common formulation in the British school, such arguments are formally valid because the conclusion is *contained* in the signification or 'in the formal understanding' (*de formali intellectu*) of the premises, so that anyone who understands that the premises are true also understands that the conclusion is true (see e.g., Seaton, 1973; Green-Pedersen, 1981), or because the contradictory opposite of the conclusion 'is formally incompatible with' (*formaliter repugnare*) the premises (Paul of Venice, 1990).

Let us now return to the Medieval Recognition Problem. Are computers capable of recognizing conceptual-formally valid arguments? To get things going, let us adopt for now the set-up of John Searle's Chinese Room Argument (Searle, 1980, 1984). Imagine that a computer is like a person—let us call her Siri—sitting in a closed room with two letterboxes, one for input and the other for output. She has a book of formal rules that tell her what conclusions she is allowed to derive from what premises—the rules are "formal" in the sense that they only concern the shapes of symbols—and apart from these rules, she knows absolutely nothing about the symbols. She is given a finite set of premises and a conclusion as an input and she is supposed to produce an output "Yes" if the conclusion is derivable from the premises by the rules. Accept, for the sake of argument, that this set-up is an adequate allegory of a computer. The question is, can you write a book of rules for conceptual-formal validity so that Siri would be able to correctly recognize all and only conceptual-formally valid arguments?

---

it that a full implementation for the particular language in question would be possible along the lines I discuss in Sections 0 and 0.

If Siri is given (1) as an input and asked if it is structural-formally valid, she can easily confirm that it is: all she needs to do is derive $H(s) \rightarrow A(s)$ from the second premise by the rule of universal instantiation and then derive $A(s)$ from that formula and the first premise by the rule of Modus Ponens. By contrast, if she is given (2) as an input and asked if it is conceptual-formally valid, it is far from obvious what kind of rules could have been written to allow her to handle cases like this. At face value, it seems that she would need to *understand* what the terms H (human being) and A (animal) are about, but it was assumed at the outset that she does not understand the underlying language beyond what is written in the rulebook. It might be thought that encoding all the complex connections between concepts understood by the human intellect into formal rules to be mechanically followed by a non-intellect is very challenging if not impossible. Even if they could be so encoded, we might wonder whether that would amount to "formal understanding" in the medieval sense: as Searle himself says, "whatever purely formal principles you put into the computer, they will not be sufficient for understanding" (Searle, 1980, 238).[8]

I will say that a notion of validity is *formalizable* if it can be implemented by a Chinese Room in the way outlined above. Importantly, validity must be reduced to formal rules that Siri can follow without understanding their content, and Siri will follow these rules in a completely mechanic manner: she is not allowed to make any exceptions or change the rules along the way. As noted, there are *prima facie* reasons to suspect that conceptual-formal validity cannot be reduced to formal rules of the kind described. These considerations give rise to the following argument:

> Validity can be recognized by a computer only if it is formalizable;
> Conceptual-formal validity is not formalizable;
> Therefore, conceptual-formal validity cannot be recognized by a computer.

[8] One of my referees objected that Searle's Chinese Room Argument is not really relevant for this paper. I agree that I could very well study the question of modelling conceptual-formal validity without any reference to the Chinese Room. I just think that the Chinese Room provides us with a nice illustration of the idea of formalization: in this sense, it is nothing but a rhetorical device used to enliven the discussion. That said, I do think that there is an interesting analogy between the Searlean juxtaposition of syntax vs. semantics and the medieval juxtaposition of structure vs. conceptual containment. At the very least, the medieval way of combining "form" and "understanding" clearly challenges the Searlean way of contrasting these two notions.

That is to say, the Medieval Recognition Problem is unsolvable. Add the premise that conceptual-formal validity is essential to everyday reasoning, and you get the corollary that computers are unable to accommodate everyday reasoning. The crucial question is, do we need to accept this argument? In the following sections, I will attack the second premise—the claim that conceptual-formal validity is not formalizable—from various directions, drawing on different sources of inspiration in modern logic.[9] My aim is not so much to argue for any particular view as to bring together alternative approaches, relate them to the present problem, and map out paths for further inquiry. Having identified problems with each of the considered attacks on the second premise, I devote the final section to questioning the first.

2. Enriching the Object Language

I am certainly not the first to think about ways of formalizing the medieval account of 'conceptual-formal' validity within modern logic. Peter King does so while discussing William of Ockham's rule that a consequence "from a superior distributed term to an inferior distributed term", such as "Every animal is running; therefore, every human being is running", is valid (Ockham, 1974, 591). King notes that this rule:

> treats the relation between the terms 'animal' and ['human being'] as a formal feature. Modern first-order logic does not normally respect such relationships, but could do so in a number of ways: indexing or sorting the term-variables; adding semantic rules along the lines of meaning-postulates; and the like. (King, 2001, n. 35)

I will first consider the suggestion of indexing or sorting the term variables and postpone the discussion of meaning postulates. Given that term variables like 'animal' and 'human being' are usually formalized in first-order predicate logic by predicate symbols, we might follow King's lead by enriching the object language with subscripted predicate symbols, where the subscripts are used to impose certain limitations on the interpretation of the symbols in question. Let us say that

[9] I am thus taking up the challenge formulated by Stephen Read at the end of his paper (Read, 1993, 259) as follows: "What I have not tackled is the open question whether a non-classical account of logical consequence, perhaps based on the recurring idea of containment or inclusion of consequent in antecedent, can be developed, inspired by the medieval phrase *de formali intellectu*. But that will have to wait for another occasion."

for any predicate symbols P and Q, there is a subscripted predicate symbol $P_{\subseteq Q}$, the extension of which must be included in that of Q. Ockham's example can then be formalized as:

$$\forall x(A(x) \rightarrow R(x)); \text{ therefore, } \forall x(H_{\subseteq A}(x) \rightarrow R(x)),$$

which comes out as valid in the semantics described.

A weakness of this strategy is revealed by considering another example:

(5) Valtteri is a human being; therefore, Valtteri is a living thing,

where the formalization of 'human being' by $H_{\subseteq A}$ and 'living thing' by L does not produce a valid form. We have two options: either we introduce an alternative symbol for 'human being', such as $H_{\subseteq L}$, and use that symbol rather than $H_{\subseteq A}$ in the formalization of this argument, or we conflate the two subscripts within one predicate symbol, as in $H_{\subseteq A,L}$, and use that symbol in the formalization of both arguments.

Further examples along these lines make the notation even more complicated. The problem is not just that all relations of mutual inclusion have to be integrated into the syntax; mutual inclusion is but one of many different types of term relationships, each of which needs its own notation. In argument (3), for instance, the relationship between the terms 'is west of' and 'is east of' is not that one is superior to the other; rather, they are each other's inverses in the sense that x is west of y if and only if y is east of x. Following the common practice in mathematical logic, this relationship would be formalized by something like:

$W(t,h)$; therefore, $W^{-1}(h,t)$.[10]

The more terms are included in the object language, the more complex is the syntax required to encode all their mutual relationships. And yet, as long as there is only a finite number of terms and only a finite number of different types of relationships among terms, all the possible relationships between the terms are in principle encodable by subscripted symbols. Once we have set down all the different types of term relationships that we want to be able to encode, we can tell the computer how to decode them. The question is, will this strategy provide us with a solution to the Medieval Recognition Problem?

---

[10] The terms of argument (4) exhibit yet another kind of relationship, namely, mutual incompatibility. They could be formalized by, say, B and $W_{\cap B=\emptyset}$, where the subscript requires that the extensions of the predicates be disjoint; and so on for other types of term relationships.

I argue that it does not. A core intuition underlying the problem is that the computer should already 'know' how particular terms relate. It should, for instance, 'know' that the predicate 'animal' applies to every individual to which the predicate 'human being' applies. By the 'enriched language' strategy, however, the computer would not 'know' any such relationships; it would merely 'know' abstract patterns of the sort: 'if there are any predicates of the form Q and $P_{\subseteq Q}$, then Q applies to every individual to which $P_{\subseteq Q}$ applies'. It would be the task of the user of the computer to fill in all relevant information on term relationships when feeding an argument to the computer, and this would have to be done separately for every argument ever fed to the computer. For instance, in order to check the validity of argument (5), the user should formalize not only the predicates as H and L, but also the inclusion of H in L by adding the appropriate subscript. The user would, in effect, do all the interesting work on behalf of the computer. We must therefore look for a solution elsewhere.

3. The Suppressed Premise Strategy

Perhaps the most obvious way to try to formalize conceptual-formal validity is what Read (1994) calls "the suppressed premise strategy". This strategy is based on the idea that the condition on conceptual connections expressed in the metalanguage could be written out in a sentence of the object language and added as a premise to the argument to make it structural-formally valid. In the conceptual form of (2), for instance, the metalanguage condition $ext(H) \subseteq ext(A)$ could be translated into the object language sentence $\forall x\big(H(x) \rightarrow A(x)\big)$. When we add this sentence as a premise to (2), we get the argument (1), which is structural-formally valid if and only if (2) is conceptual-formally valid.

Something like this can be already found in medieval authors. John Buridan, while defining formal validity in terms of structural form and thereby classifying consequences like (2) as merely materially valid, remarks that they can nevertheless be reduced to formally valid ones by adding a necessarily true premise (Buridan, 2015, 68). For instance, (2) could be reduced to a formal consequence by adding the premise 'Every human being is an animal', which expresses a necessary truth. Similarly, (3) and (4) could be reduced to formal consequences by adding, e.g., the necessary truths 'Turku is west of Helsinki if and only if Helsinki is east of Turku' and 'Daffy is not both black and white', respectively. As it turns out, all conceptual-formally valid arguments can be reduced to structural-formally valid arguments in this way (cf. Read, 1994).

The Achilles heel of the strategy lies in the requirement that the added premise be *necessarily true*. The mere fact that an argument can be made structural-formally valid by adding a premise is insufficient for its conceptual-formal validity: indeed, *any* argument 'φ; therefore, ψ' can be trivially made structural-formally valid by adding the premise $\varphi \rightarrow \psi$. This leaves us with a dilemma. Either we require that the user supplies the suppressed premise, if there is one, for every argument that they feed to the computer, or we devise a computer that can supply the suppressed premise by itself. The first option would be cheating, for the user would then be doing the work of identifying conceptual-formally valid arguments while the computer would only recognize structural-formally valid ones.[11] The other option would require that the computer not only formulates the missing premise, but also recognizes that it is necessarily true, and that is easier said than done. Even if we found a way to teach the computer to recognize necessary truths, the act of adding a premise is not the critical step; on the contrary, the whole weight of the problem is shifted to recognizing the necessary truth of that premise (cf. Read, 1994, 258). Plausibly, a computer that can recognize necessary truth can just as well recognize conceptual-formal validity directly without recourse to structural-formal validity. We are back at square one.

4. The Shorthand Strategy

Another obvious way to formalize conceptual-formal validity is based on the idea that defined terms be short for their definitions (cf. Quine, 1951; Fine, 2005, 236). This insight would allow us to reduce conceptual-formally valid arguments to structural-formally valid ones by replacing the defined terms with their definitions. For instance, given that a 'human being' is defined as a 'rational animal', we could replace 'human being' with 'rational animal' in the argument (2) and thereby reduce it to:

Valtteri is a rational animal; therefore, Valtteri is an animal.

---

[11] Recall that we are looking for a formalization that is 'static' like that of structural-formal validity. The computer is not allowed to store suppressed premises from the user's input into its memory and use them for processing future input. We will give up this requirement in the final section.

In formalized language, this corresponds to applying the definition $H(x) \equiv_{df} R(x) \wedge A(x)$ to the argument (2'), resulting in:

$R(v) \wedge A(v)$; therefore, $A(v)$.

This argument is clearly structural-formally valid, and hence we would conclude that the original argument (2) is conceptual-formally valid.

Certain qualifications have to be made to circumvent certain difficulties analogical to the problems of the suppressed premise strategy. First, the fact that a given argument can be made structural-formally valid by expanding suitable 'definitions' of the primitive terms is insufficient for conceptual-formal validity. For instance:

Plato is a philosopher; therefore, Plato is a potato

reduces to a structural-formally valid argument under the substitution of 'mashed potato' for 'philosopher', but this 'definition' is of course incorrect. It is an obvious requirement that only correct definitions be used. Moreover, we cannot expand the correct definitions before we feed the argument to the computer, for it would, again, be cheating.

It remains to be shown how to encode all the definitions to a computer. This does not seem too difficult: cannot we just list all the definitions and append them to Siri's rulebook, like a dictionary where she can look up the definitions of primitive predicate symbols? So long as there is only a finite number of non-circular definitions formalizable in the object language, it is fairly easy to design an algorithm for expanding all of them in any given argument. Suppose, for instance, that the dictionary included the following definitions (here given in the vernacular for the sake of illustration):

'bachelor' $=_{df}$ 'unmarried man'

'man' $=_{df}$ 'male human'

'human' $=_{df}$ 'rational animal'.

We could then reduce the argument:

Plato is a bachelor; therefore, Plato is rational to a structural-formally valid one, applying each of the three definitions in turn:

Plato is an unmarried man; therefore, Plato is rational

Plato is an unmarried male human; therefore, Plato is rational

Plato is an unmarried male rational animal; therefore, Plato is rational.

This approach applies to all cases of conceptual-formal validity that are grounded on definitions of terms. Argument (2) is another example of this kind. Arguments (3) and (4), however, are different.[12] In fact, the examples (2), (3), and (4) have been deliberately modelled upon three medieval 'degrees of formality' (Paul of Venice, 1990, 90–93; see also Strobino, 2017), only the first of which is based on definitions. It is impossible that Valtteri would be a human being but not an animal, because the definition of the term 'human being' includes the term 'animal'.

The two other degrees of formality are grounded on necessary truths of a weaker kind, where the meanings of the terms are necessarily connected but not through their definitions. To accommodate these degrees of formality, we should include in Siri's dictionary all such necessary truths, which are typically more complicated than definitions. Take (4), for instance. Black and white are not defined in terms of each other but rather, say, in terms of wavelengths of light or primitive phenomenal qualities—neither is more primitive than the other (Read, 1994, 250–251)—and yet it is a necessary truth that nothing is both black and white. The conceptual form of (4) is:

'$B(d)$; therefore, $\neg W(d)$' and $ext(B) \cap ext(W) = \emptyset$,

where $ext(B) \cap ext(W) = \emptyset$ expresses a non-definitional necessary truth. Such necessary truths cannot be reduced to dictionary definitions, but perhaps they can be encoded to a computer in some alternative way.

5. Meaning Postulates

Rudolf Carnap (1952) defends the notion of analyticity against Quine's (1951) famous attack by introducing semantic rules known as *meaning postulates*.[13] In

[12] Boghossian would say that they are examples of "a priori statements that are not Frege-analytic" (Boghossian, 1996, 368).

[13] Cf. Boghossian's (1996) idea of "implicit definition".

the framework of the present paper, Carnap's meaning postulates can be cashed out as the object language translations of the metalanguage conditions pertaining to conceptual form. The difference with respect to the suppressed premise strategy is that these sentences are not added as premises to any particular arguments but rather constitute a body of 'presuppositional knowledge' that is common to all arguments and belongs to the underlying system as a whole. The difference with respect to the shorthand strategy is that these sentences include not only definitions but also weaker necessary truths that are supposed to constrain the interpretation of the predicate symbols. To continue with our previous metaphor, we could say that meaning postulates are added as a special appendix to Siri's rulebook. This appendix is not made up of definitions like a traditional dictionary, but rather of non-logical axioms that enable Siri to deal with non-logical symbols according to their intended meanings.

To guarantee the validity of arguments (2), (3), and (4), we may introduce the following meaning postulates:

(MP2) $\forall x\big(H(x) \rightarrow A(x)\big)$
(MP3) $\forall x \forall y\big(W(x,y) \leftrightarrow E(y,x)\big)$
(MP4) $\neg\exists x\big(B(x) \wedge W(x)\big)$.

Such postulates are arguably similar to 'intrinsic middles' or 'intrinsic topics' that were appealed to by medieval logicians like Walter Burley and William of Ockham (cf. King, 2001, 130; Archambault, 2018; Crimi, 2018). For instance, the formal consequence 'A human being is running; therefore, an animal is running' was apparently thought to hold through the intrinsic middle 'Every human being is an animal' (Crimi, 2018, 261).

While this strategy is in many ways superior to the previous ones, it should be noted that the available meaning postulates are limited by the expressive power of the object language. There are cases in which the requisite meaning postulate simply cannot be expressed in the object language (cf. Etchemendy, 1999, 71). Consider, for instance, arguments such as:

Ada is a parent of Bob; therefore, Bob is not a parent of Ada.

Ada is a parent of Bob; Bob is a parent of Chris; therefore, Ada is an ancestor of Chris.

> Ada and Bob have no ancestor in common; Chris is a parent of Ada; therefore, Chris is not a parent of Bob.

In order to be able to deal with all possible arguments involving the binary predicates 'is a parent of' and 'is an ancestor of', we would need particularly strong meaning postulates. Let us assume, for simplicity, that there are no other (non-logical) predicates in our language besides these two. The predicate 'is a parent of' might be defined by axioms of irreflexivity ('no one is a parent of oneself'), asymmetry ('for all x and y, if x is a parent of y, then y is not a parent of x'), and so on. However, to establish the semantic connection between 'is a parent of' and 'is an ancestor of', we would need to add a postulate along the following lines:

> For all x and y, x is an ancestor of y if and only if there are $z_1, z_2, \dots, z_n$ such that x is a parent of z_1, z_1 is a parent of z_2, …, and z_n is a parent of y.

This postulate can only be expressed in higher-order logic. While we can choose our object language to be of any order we like, limitations of expressive power are bound to reappear at the higher level. In general, the Carnapian approach limits the conceptual form to what can be expressed in the object language, but plausibly, the realm of meaning exceeds any such technical limitations: the medievals, in particular, knew no distinction between object and metalanguage. In the next section, I will consider the possibility of transferring the Carnapian approach to the level of metalanguage.

6. From Rules to Models

The three previous strategies have all tried to embed conceptual-formal validity to structural-formal validity through some restrictive conditions expressed at the object level, be it in the form of premises, definitions, or semantic rules. In this section, I show how the notion of conceptual-formal validity can be directly formalized at the metalevel, taking up the informal characterization that we started with in the first section and making it more precise by modifying the standard model-theoretic semantics.

Suppose that our object language contains the predicate symbols H ('human being') and A ('animal'). In ordinary metalogic, we would characterize these symbols as non-logical constants and vary their interpretation quite freely from model to model, while keeping the interpretation of logical constants essentially fixed across models. I say "quite freely" and "essentially fixed" because, on one hand, the predicate symbols cannot be assigned just any sets as extensions: they have to be subsets of the domain. The domain, on the other hand, is the set that the quantifiers range over, and this set varies with the model. In Etchemendy's (1999) terminology, there is a "cross-term restriction" that fixes a certain connection between the predicate symbols and the quantifiers, namely, that the extensions of the predicate symbols are subsets of the domain of quantification. However, "there is no significant difference between this sort of restriction and the use of so-called meaning postulates" (Etchemendy, 1999, 71), such as the requirement that the extension of H be included in that of A. Why, then, is the former restriction accepted in the standard semantics while the latter is rejected? Etchemendy's response is to reject the former along with the latter; our response, by contrast, is to incorporate the latter in the metalogic along with the former.

Gillian Russell (2024) makes a similar point by distinguishing between fixed ('conditional') and variable ('environmental') meanings of expressions. The key insight, as I see it, is this. The classical demarcation of expressions into logical and non-logical is, at best, a rough approximation; to make it more accurate, we must formalize each expression as consisting of a fixed (logical) and a variable (non-logical) part. As we saw above, not even in standard metalogic is there really a sharp boundary between logical and non-logical constants. Take the interpretation of the universal quantifier, for instance. Its intuitive meaning 'everything' can literally be split up into the fixed part 'every' and the variable part 'thing' (cf. Etchemendy, 1999, 67). The meaning of the quantifier requires satisfiability by *every* member of the domain, and this requirement invariably applies in any model, while the domain itself—the extension of 'thing'—varies with the model. But if quantifiers are already cut in half like this, why not let the cut extend to constants traditionally held as non-logical? Indeed, we may similarly divide the interpretation of our predicate symbols H and A into two parts: the requirement that the extension of H be a subset of that of A is the fixed part, and the elements of the extensions, which can be any elements of the domain as long as the required relation of subsethood holds, is the variable part. Logicality is thus not a binary classification but a matter of degree: some constants are more logical than others, depending on the relative 'sizes' of their fixed parts (cf. Sagi, 2018).

With this demarcation of fixed and variable parts in mind, we get a precise idea of what an "instance" of a conceptual form is: it is a model, the definition of which has been amended by imposing certain restrictions on the interpretations of the predicate symbols. Just like with ordinary structural form, formal validity is truth-preservation in all models, only now the set of relevant models has been narrowed down. We have thus incorporated the Carnapian meaning postulates into the metalogic in a way that is a natural continuation of the usual semantics. On this account, as Russell (2024, 233) puts it, "there is no principled discontinuity between logical and analytic consequence", or as we would say, *between structural-formal and conceptual-formal validity*. Argument (2) is conceptual-formally valid because, by the refined definition of a model, all the models of its premise are also models of its conclusion.

To make things a bit more concrete, consider models $M_0 = \langle D_0, I_0 \rangle$ and $M_1 = \langle D_1, I_1 \rangle$ with the following domains:

$D_0 = \{$Valtteri, Socrates, Daffy, Helsinki, Turku, Big Ben, $\sqrt{2}\}$
$D_1 = \{$Plato, Aristotle, Alexander, Browny, Bucephalus, Marathon$\}$.

The predicates H and A and the individual constant v are interpreted as follows:

$I_0(H) = \{$Valtteri, Socrates$\}$
$I_0(A) = \{$Valtteri, Socrates, Daffy$\}$
$I_0(v) =$ Valtteri
$I_1(H) = \{$Plato, Aristotle, Alexander$\}$
$I_1(A) = \{$Plato, Aristotle, Alexander, Browny, Bucephalus$\}$
$I_1(v) =$ Marathon.

In both domains, the extension of H is a subset of that of A; otherwise the construction would not qualify as a model. Model M_0 is a model of both $H(v)$ and $A(v)$, whereas the model M_1 is a model of neither $H(v)$ nor $A(v)$. There could just as well be a model M_2 which is not a model of $H(v)$ but is a model of $A(v)$. There cannot, however, be a model which is a model of $H(v)$ but not of $A(v)$, and hence the argument (2) is conceptual-formally valid.

Similar adjustments to the definition of a model could be made to accommodate the conceptual-formal validity of arguments (3) and (4). An example of an argument that is merely materially valid might be something like 'There are no irrational numbers; therefore, Valtteri is running', formalized as:

$\neg\exists x I(x)$; therefore, $R(v)$.

Here, the premise is thought to be necessarily false but only 'as a matter of fact', because of its particular content.[14] In our formalization, this is to say that not all models of the premise are also models of the conclusion.

Let me use this opportunity to make a slight digression and sketch a systematic hierarchy of different types of models. For the sake of illustration, I will conceive of these models as possible worlds as in Etchemendy's (1999) 'representational semantics'.[15] In place of the scheme of '___-formal validity' proposed in the first section, let us now adopt the following, which is closer to the contemporary (Tarskian) way of doing things:[16]

> (___ validity) An argument is ___ly valid if and only if in all ___ly possible worlds (models) where its premises are true, its conclusion is also true.

At the first level of the hierarchy, we have the standard definition of a model. The corresponding models are identified as the *structural-formally* (or *logically*) possible worlds, and the resulting set of valid arguments is that of structural-formally valid ones. At the second level, we have the definition of a model amended in the way described above, resulting in *conceptual-formally* (or semantically) possible worlds and conceptual-formally valid arguments. Conceptual-formally possible worlds are a subset of structural-formally possible worlds, and conversely, conceptual-formally valid arguments are a superset of structural-formally valid arguments. At the third level, the definition of a model is further restricted so that it excludes metaphysical impossibilities, such as there being no irrational numbers, but still includes physical impossibilities, such as Valtteri running faster than the speed of light. Such *metaphysically* possible worlds are, again, a subset of conceptual-formally possible worlds, and the resulting set of metaphysically valid arguments is a superset of the set of conceptual-formally valid arguments. In this way, the medieval idea of conceptual form can be seen as introducing an intermediate sphere of necessity between the familiar categories of

---

[14] The medievals' favourite example of such a falsehood was 'God does not exist' (see e.g., Strobino, 2017).

[15] Here I diverge from Russell's (2024) account, wherein models are cashed out not as possible worlds but as combinations of worlds and languages. In this respect, I also diverge from other 'hybrid' accounts of logical consequence that have been developed in response to Etchemendy, such as those of Sher (1996) and Shapiro (2007).

[16] Note the analogy with "Tarski's thesis" (McGee, 1992) and the "Generalised Tarski Thesis" (Beall and Restall, 2005).

logical necessity and *metaphysical* (or broadly logical) necessity (cf. Plantinga, 1978; Fine, 2005). The three spheres of necessity are illustrated in Fig. 1.[17]

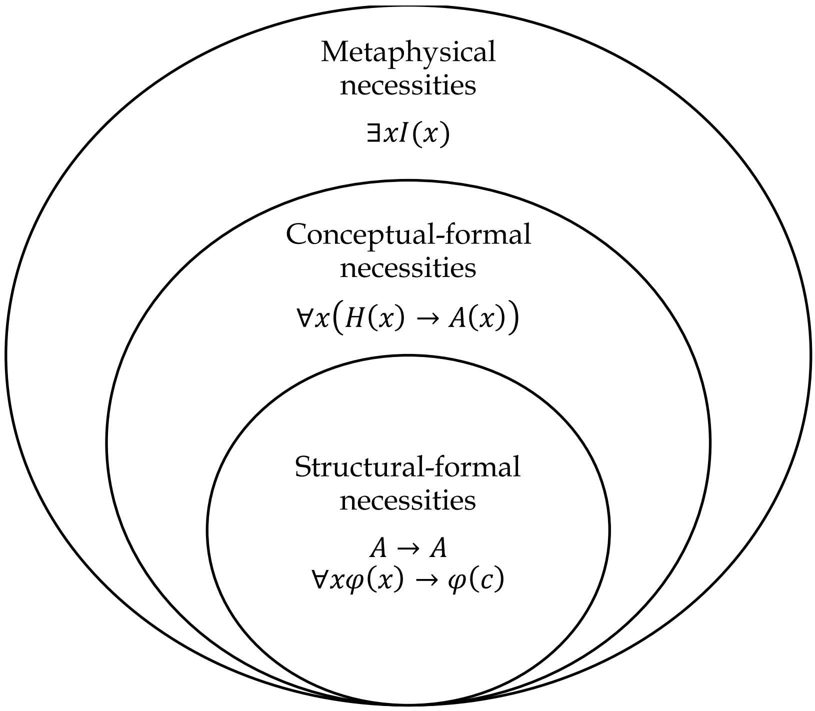

Fig. 1. Spheres of metaphysical, conceptual-formal (semantic), and structural-formal (logical) necessity.

This formalization of conceptual-formal validity not only connects to the contemporary discussion on the nature of necessity; it is also, I think, more faithful to the medieval account of degrees of formality than the alternative formalizations. This formalization also fares better than Carnap's meaning postulates in making room for conceptual relationships that cannot be expressed in first-order logic. Yet the problem was whether this formalization makes it possible for Siri to recognize conceptual-formally valid arguments, and that, unfortunately, seems somewhat

[17] The hierarchy could be extended in the obvious way by recognizing *physically* (causally, naturally) possible worlds as a subset of the metaphysically possible ones (cf. Plantinga, 1978, 14). In the opposite end, we could likewise recognize 'paralogically' possible worlds as a superset of logically possible ones. Such worlds would attach no fixed meanings to any constants beyond the minimal requirement that they be of the appropriate type: for instance, the logical connectives could be interpreted as any truth-functions of the appropriate arity, so that contradictions like $p \land \neg p$ would come out true in some possible worlds.

doubtful. The problem is that Siri could no longer restrict herself to the first-order rules of inference; instead, she would have to work with the definition of a model, and that definition is given in the higher-order metatheory. As is well-known, there is no complete proof system even for second-order logic with standard semantics. The details are too complicated to be discussed here, but the general worry is clear: basically, the metalogic may be too thick for Siri to swallow.

7. Beyond Formalization

All the strategies for solving the Medieval Recognition Problem discussed above ended up in difficulties of one kind or another. Should we give up and conclude that the problem is unsolvable? Perhaps it is indeed the case that conceptual-formal validity is not fully formalizable; perhaps there simply is no place for conceptual form in the Chinese Room. Even so, that does not yet show that conceptual-formal validity cannot be recognized by a computer. The claim that conceptual-formal validity is not formalizable was merely the second premise of our argument to that conclusion. We are still left with the first premise, namely, the claim that conceptual-formal validity is recognizable by a computer *only if* it is formalizable. Let me conclude by briefly outlining ways of debunking this first premise.

The first premise was originally based on the assumption that the Chinese Room is an adequate allegory of a computer. If a computer is nothing but an automaton manipulating symbols according to pregiven instructions, then it seems indeed that the only way for a computer to recognize conceptual-formally valid arguments would require a list of instructions for recognizing such arguments by their syntactic appearance. We do not, however, need to accept that the Chinese Room is an adequate allegory of a computer. Developments in artificial intelligence and related fields (see e.g., Bringsjord and Govindarajulu, 2024) suggest that a machine-learning model could be trained to recognize conceptual-formally valid arguments and outperform any traditional formalization we could possibly think of. As opposed to our original ‘top-down’ Chinese Room, where Siri was applying ready-made rules to given cases, our renovated ‘bottom-up’ Chinese Room would have Siri learn from concrete examples of conceptual-formally valid/invalid arguments and write the rules by herself, as it were, on that basis. The rules would amount to probabilistic calculations rather than, say, complete axiomatizations, and they would occasionally get it wrong, but perhaps we should give up on the requirement of complete accuracy. After all, human beings are fallible, too, and the class of conceptual-formally valid arguments may be too vague anyway to ever be given a fully definitive formalization. The accuracy of the model might never reach 100%, but at least we could approximate the accuracy by testing the model with different data than it was trained with, and we could always increase the accuracy by retraining the model with more data or by fine-tuning its inner architecture.

Arguably, conceptual-formal validity has already been realized in the said sense by the recent rise of large language models. Take ChatGPT, for instance, which is a chatbot based on such a model. When I enter the prompt "Valtteri is a human being. Does it follow that Valtteri is an animal?" into ChatGPT, I get the following response:

> Yes, it follows that Valtteri is an animal because humans are biologically classified as animals. Specifically, humans belong to the kingdom Animalia. Therefore, if Valtteri is a human being, Valtteri is also an animal. (OpenAI, 2024)[18]

Indeed, it seems that the Chinese Room has already undergone a thorough renovation. Thanks to the renovation, conceptual form appears more welcome in the Chinese Room, but we can go even further than that. The rise of externalist '4E' theories of mind and the concomitant advances in robotics provoke the thought that one day we could build a full-blown artificial mind (Telakivi and Arstila, 2021). Imagine, for instance, an artificial neural network that is embodied in a robot, embedded in and evolving in continuous interaction with its environment through sensors and actuators, pretty much like the human brain. Such a system would then be expected to *make sense* of its environment and come to learn *meanings* of terms like 'human being' and 'animal' based on its experience, emulating the way human beings learn the meanings of those terms (cf. Telakivi and Arstila, 2021). Just as we can give a conceptual-formally valid argument to a human being who then confirms that it is conceptual-formally valid, we could give such an argument to the artificial neural network, which would likewise confirm that it is conceptual-formally valid. The inner workings of the network might appear as a black box—we might not, for instance, be able to identify any part of the network as a representation of the term 'human being' or of the rule by which its meaning is governed—and again, there is no 100% guarantee of accuracy, but perhaps that is just the price we have to pay for

[18] This response was created using the preselected model, GPT-4o.

conceptual-formal validity. Perhaps conceptual form is too savage a beast to be tamed by any rules of our clean-cut modern logic; too widespread a jungle to be contained by any room. [19]

University of Helsinki

References

Archambault, J. (2017), *The Development of the Medieval Parisian Account of Formal Consequence*, PhD thesis, Fordham University. Available at: https://research.library.fordham.edu/dissertations/AAI10617109 (Accessed: 26 January 2023).

Archambault, J. (2018), "Consequence and Formality in the Logic of Walter Burley", *Vivarium* 56(3–4), 292–319. Available at: https://doi.org/10.1163/15685349-12341355.

Aristotle (2009), *Prior Analytics: Book I*, translated with an introduction and commentary by G. Striker, Oxford University Press, Oxford. Available at: https://doi.org/10.1093/actrade/9780199250400.book.1.

Barker-Plummer, D., J. Barwise and J. Etchemendy (2011), *Language, Proof and Logic*, 2nd edn., CSLI Publications, Stanford, CA.

Beall, J. C. and G. Restall (2005), *Logical Pluralism*, Oxford University Press, Oxford. Available at: https://doi.org/10.1093/acprof:oso/9780199288403.001.0001.

Boghossian, P. A. (1996), "Analyticity Reconsidered", *Noûs* 30(3), 360–391. Available at: https://doi.org/10.2307/2216275.

Boghossian, P. and T. Williamson (2020), *Debating the A Priori*, Oxford University Press, Oxford. Available at: https://doi.org/10.1093/oso/9780198851707.001.0001.

Bringsjord, S. and N. S. Govindarajulu (2024), "Artificial Intelligence", in E. N. Zalta and U. Nodelman (eds.) *The Stanford Encyclopedia of Philosophy*, Summer 2024, Metaphysics Research Lab, Stanford University. Available at: https://plato.stanford.edu/archives/sum2024/entries/artificial-intelligence/ (Accessed: 19 June 2024).

Buridan, J. (2015), *Treatise on Consequences*, translated with an introduction by S. Read, Fordham University Press, New York, NY. Available at: https://doi.org/10.5422/fordham/9780823257188.001.0001.

[19] I am grateful to Gabriel Sandu, José Filipe Silva, and two anonymous referees for their helpful comments on the manuscript. The ideas developed in Section 0 have originally been inspired by Gillian Russell's presentation "Rethinking Logical Consequence" and the subsequent discussion in the Metaphysics and Logic Seminar at the Arché Research Centre, University of St Andrews, in May 2023.

Burley, W. (2000), *On the Purity of the Art of Logic: The Shorter and the Longer Treatises*, translated by P. V. Spade, Yale University Press, New Haven, CT. Available at: https://doi.org/10.12987/yale/9780300082005.001.0001.

Carnap, R. (1952), "Meaning Postulates", *Philosophical Studies* 3(5), 65–73. Available at: https://doi.org/10.1007/BF02350366.

Church, A. (1936), "A Note on the Entscheidungsproblem", *Journal of Symbolic Logic* 1(1), 40–41. Available at: https://doi.org/10.2307/2269326.

Ciola, G. (2018), "Marsilius of Inghen on the Definition of *Consequentia*", *Vivarium* 56(3–4), 272–291. Available at: https://doi.org/10.1163/15685349-12341359.

Cohen, S. M. (2004), "The Logic of Boolean Connectives", in *Introduction to Logic: Lecture Notes*, University of Washington. Available at: https://faculty.washington.edu/smcohen/120/Chapter4.pdf (Accessed: 19 June 2024).

Copeland, B. J. (2024), "The Church-Turing Thesis", in E. N. Zalta and U. Nodelman (eds.) *The Stanford Encyclopedia of Philosophy*, Spring 2024, Metaphysics Research Lab, Stanford University. Available at: https://plato.stanford.edu/archives/spr2024/entries/church-turing/ (Accessed: 4 April 2024).

Corcoran, J. (1998), "Information-Theoretic Logic", in C. Martínez, U. Rivas, and L. Villegas-Forero (eds.) *Truth in Perspective*, Ashgate, Aldershot, 113–135. Available at: https://doi.org/10.4324/9780429433450.

Crimi, M. (2018), "Formal and Material Consequences in Ockham and Buridan", *Vivarium* 56(3–4), 241–271. Available at: https://doi.org/10.1163/15685349-12341360.

Davis, M. (2018), *The Universal Computer: The Road from Leibniz to Turing*, 3rd edn., CRC Press, Boca Raton, FL. Available at: https://doi.org/10.1201/9781315144726.

Dutilh Novaes, C. (2007), *Formalizing Medieval Logical Theories:* Suppositio*,* Consequentiae *and* Obligationes, Springer, Dordrecht. Available at: https://doi.org/10.1007/978-1-4020-5853-0.

Dutilh Novaes, C. (2020), "Medieval Theories of Consequence", in E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*, Fall 2020, Metaphysics Research Lab, Stanford University. Available at: https://plato.stanford.edu/archives/fall2020/entries/consequence-medieval/ (Accessed: 19 May 2023).

Etchemendy, J. (1999), *The Concept of Logical Consequence*, CSLI Publications, Stanford, CA.

Etchemendy, J. (2008), "Reflections on Consequence", in D. Patterson (ed.) *New Essays on Tarski and Philosophy*, Oxford University Press, Oxford, 263–299. Available at: https://doi.org/10.1093/acprof:oso/9780199296309.003.0011.

Fine, K. (2005), "The Varieties of Necessity", in K. Fine (ed.) *Modality and Tense: Philosophical Papers*, Oxford University Press, Oxford, 235–260. Available at: https://doi.org/10.1093/0199278709.003.0008.

Fitting, M. (2022), "Intensional Logic", in E. N. Zalta and U. Nodelman (eds.) *The Stanford Encyclopedia of Philosophy*, Winter 2022, Metaphysics Research Lab, Stanford University. Available at: https://plato.stanford.edu/archives/win2022/entries/logic-intensional/ (Accessed: 14 June 2024).

Green-Pedersen, N. J. (1981), "Nicolaus Drukken de Dacia's Commentary on the *Prior Analytics* – with Special Regard to the Theory of Consequences", *Cahiers de l'Institut Du Moyen-Âge Grec et Latin* 37, 42–69.

Halbach, V. (2020), "The Substitutional Analysis of Logical Consequence", *Noûs* 54(2), 431–450. Available at: https://doi.org/10.1111/nous.12256.

Hilbert, D. and W. Ackermann (1928), *Grundzüge der theoretischen Logik*, Springer, Berlin.

King, P. (2001), "Consequence as Inference: Mediaeval Proof Theory 1300–1350", in M. Yrjönsuuri (ed.) *Medieval Formal Logic: Obligations, Insolubles and Consequences*, Springer, Dordrecht, 117–145. Available at: https://doi.org/10.1007/978-94-015-9713-5_5.

Klima, G. (2016), "Consequence", in C. Dutilh Novaes and S. Read (eds.) *The Cambridge Companion to Medieval Logic*, Cambridge University Press, Cambridge, 316–341. Available at: https://doi.org/10.1017/CBO9781107449862.014.

Kment, B. (2006), "Counterfactuals and the Analysis of Necessity", *Philosophical Perspectives* 20(1), 237–302. Available at: https://doi.org/10.1111/j.1520-8583.2006.00108.x.

McGee, V. (1992), "Two Problems with Tarski's Theory of Consequence", *Proceedings of the Aristotelian Society* 92(1), 273–292. Available at: https://doi.org/10.1093/aristotelian/92.1.273.

Ockham, W. of (1974), *Opera Philosophica. Vol. I: Summa Logicae*, edited by P. Boehner, G. Gál, and S. Brown, The Franciscan Institute, St. Bonaventure, NY.

OpenAI (2024), *ChatGPT*. Available at: https://chatgpt.com/share/670b99b8-42d4-8005-825a-f8203151af08 (Accessed: 13 October 2024).

Paul of Venice (1990), *Logica Magna: Part II Fascicule 4*, edited by G. E. Hughes, Oxford University Press, Oxford.

Plantinga, A. (1978), *The Nature of Necessity*, Oxford University Press, Oxford. Available at: https://doi.org/10.1093/0198244142.001.0001.

Quine, W. V. (1951), "Two Dogmas of Empiricism", *The Philosophical Review* 60(1), 20–43. Available at: https://doi.org/10.2307/2181906.

Read, S. (1993), "Formal and Material Consequence, Disjunctive Syllogism and Gamma", in K. Jacobi (ed.) *Argumentationstheorie*, Brill, Leiden, 233–259. Available at: https://doi.org/10.1163/9789004450448_019.

Read, S. (1994), "Formal and Material Consequence", *Journal of Philosophical Logic* 23(3), 247–265. Available at: https://doi.org/10.1007/BF01048482.

Read, S. (2012), "The Medieval Theory of Consequence", *Synthese* 187(3), 899–912. Available at: https://doi.org/10.1007/s11229-011-9908-6.

Read, S. (2015), "Proof-Theoretic Validity", in C. R. Caret and O. T. Hjortland (eds.) *Foundations of Logical Consequence*, Oxford University Press, Oxford, 136–158. Available at: https://doi.org/10.1093/acprof:oso/9780198715696.003.0005.

Russell, G. (2008), *Truth in Virtue of Meaning: A Defence of the Analytic/Synthetic Distinction*, Oxford University Press, Oxford. Available at: https://doi.org/10.1093/acprof:oso/9780199232192.001.0001.

Russell, G. (2024), "Logical Consequence (Slight Return)", *Aristotelian Society Supplementary Volume* 98(1), 233–254. Available at: https://doi.org/10.1093/arisup/akae012.

Saario-Ramsay, L. (revised and resubmitted), "The Two Schools of Formal Consequence in the 14th Century: Redrawing the Distinction".

Sagi, G. (2018), "Logicality and Meaning", *The Review of Symbolic Logic* 11(1), 133–159. Available at: https://doi.org/10.1017/S1755020317000247.

Searle, J. R. (1980), "Minds, Brains, and Programs", *Behavioral and Brain Sciences* 3(3), 417–424. Available at: https://doi.org/10.1017/S0140525X00005756.

Searle, J. R. (1984), *Minds, Brains and Science*, Harvard University Press, Cambridge, MA.

Seaton, W. K. (1973), *An Edition and Translation of the* Tractatus de Consequentiis *by Ralph Strode, Fourteenth Century Logician and Friend of Geoffrey Chaucer*, PhD thesis, University of California, Berkeley.

Shapiro, S. (2007), "Logical Consequence, Proof Theory, and Model Theory", in S. Shapiro (ed.) *The Oxford Handbook of Philosophy of Mathematics and Logic*, Oxford University Press, Oxford, 651–670. Available at: https://doi.org/10.1093/oxfordhb/9780195325928.003.0021.

Sher, G. Y. (1996), "Did Tarski Commit 'Tarski's Fallacy'?", *Journal of Symbolic Logic* 61(2), 653–686. Available at: https://doi.org/10.2307/2275681.

Strobino, R. (2017), "What Is Form All About? A 14th-Century Discussion of Logical Consequence", in L. Cesalli, F. Goubier, and A. De Libera (eds.) *Formal Approaches and Natural Language in Medieval Logic*, Brepols, Turnhout, 173–194. Available at: https://doi.org/10.1484/M.TEMA-EB.4.2017095.

Sundholm, G. (2013), "Containment and Variation; Two Strands in the Development of Analyticity from Aristotle to Martin-Löf", in M. van der Schaar (ed.) *Judgement and the Epistemic Foundation of Logic*, Springer, Dordrecht, 23–35. Available at: https://doi.org/10.1007/978-94-007-5137-8_3.

Telakivi, P. and V. Arstila (2021), "Onko mahdollista rakentaa keinotekoinen mieli?", in P. Raatikainen (ed.) *Tekoäly, ihminen ja yhteiskunta: Filosofisia näkökulmia*, Gaudeamus, Helsinki, 87–101.

Turing, A. M. (1936), "On Computable Numbers, with an Application to the Entscheidungsproblem", *Proceedings of the London Mathematical Society* s2-42(1), 230–265. Available at: https://doi.org/10.1112/plms/s2-42.1.230.

Wodeham, A. de (1990), *Lectura secunda in librum primum sententiarum: Distinctiones II–VII*, edited by G. Gál and R. Wood, St. Bonaventure University, St. Bonaventure, NY.

Reports from the Department of Philosophy
Vol. 52

TRUE COLORS, TIME AFTER TIME

Essays Honoring Valtteri Arstila

Edited by
Alexander D. Carruth,
Heidi Haanila,
Paavo Pylkkänen,
Pii Telakivi

University of Turku
Finland

Cover Image: Polaris Koi

SERIES EDITORS:

Olli Koistinen
Juha Räikkä

Department of Philosophy
University of Turku
FI-20014 Turku
Finland

ISSN 1457-9332
ISBN 978-951-29-9959-0 (PRINT)
ISBN 978-951-29-9960-6 (PDF/Online)

Painosalama Oy, Turku 2024