

ONTOLOGICAL PURITY FOR FORMAL PROOFS

ROBIN MARTINOT

Department of Philosophy and Religious Studies, Utrecht University

Abstract. Purity is known as an ideal of proof that restricts a proof to notions belonging to the ‘content’ of the theorem. In this paper, our main interest is to develop a conception of purity for formal (natural deduction) proofs. We develop two new notions of purity: one based on an ontological notion of the content of a theorem, and one based on the notions of surrogate ontological content and structural content. From there, we characterize which (classical) first-order natural deduction proofs of a mathematical theorem are pure. Formal proofs that refer to the ontological content of a theorem will be called ‘fully ontologically pure’. Formal proofs that refer to a surrogate ontological content of a theorem will be called ‘secondarily ontologically pure’, because they preserve the structural content of a theorem. We will use interpretations between theories to develop a proof-theoretic criterion that guarantees secondary ontological purity for formal proofs.

§1. Introduction. In this paper, we are interested in developing a conception of purity for formal proofs. We form an ontological as well as a structural conception of purity, and develop proof-theoretic criteria that guarantee these types of purity for formal proofs, while also specifying their manifestation with respect to informal proofs. For this, we actively make use of existing intuitions that mathematicians have about purity, but also extend these once we arrive in the formal setting. This will contribute to a better formal understanding of informal notions, and can be seen as a case study for how proof-theoretic criteria can help identify philosophically meaningful derivations. The latter is an area that has not yet enjoyed many substantial results, but which relates, for example, to ‘Hilbert’s 24th problem’ for formal proofs: the problem of finding criteria for simple proofs [17]. The following section will introduce the notion of purity, followed by an outline of the structure of the paper.

1.1. Purity of proof. Purity has a long history as an ideal of proof for mathematicians, tracing back to early writings of Aristotle and Archimedes [12]. The notion concerns itself with certain restrictions in the way we are allowed to prove a theorem, and has several interpretations in the literature. Generally, we take a pure proof to only draw upon notions that belong to the content of the theorem. Here, ‘content’ should be taken as synonymous to the term ‘topic’ or ‘subject matter’. Impure proofs distinguish themselves from pure proofs by making use of concepts that are somehow extraneous

Received: January 17, 2022.

2020 *Mathematics Subject Classification*: Primary 00A30, 03A05, 03F03.

Key words and phrases: purity of proof, formalization, interpretations, structuralism, philosophy of proof theory.

© The Author(s), 2023. Published by Cambridge University Press on behalf of The Association for Symbolic Logic. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.



to what a theorem is about. To provide an intuitive sense of what purity involves, we mention two common types of (im)pure proofs found in the literature.

EXAMPLE 1.1 (Extraneous disciplines of mathematics). *A contrast is often made between arithmetical proofs of the Infinitude of Primes (IP), and a topological proof of the theorem by Furstenberg [14]. A proof of IP from the axioms of Peano Arithmetic is usually considered pure, since, for example, “it is reasonable to think of these axioms together with the [...] definitions of primality and divisibility as at least approximating the topic of IP” [3]. On the other hand, the proof by Furstenberg is classified as impure, since it requires several set-theoretic as well as topological commitments that, according to Arana and Detlefsen, lead outside the topic of IP. Similarly, Dawson [11] mentions that the search for purity can be recognized in attempts to exclude topological elements from proofs of the Fundamental Theorem of Algebra. We recognize this type of purity more generally, as well, for example, in the idea that “[i]n many cases our understanding is not satisfied when, in a proof of a proposition of arithmetic, we appeal to geometry, or in proving a geometrical truth we draw on function theory” (a quote from Hilbert cited in [16]). This illustrates a common view that a theorem belongs to a particular discipline of mathematics, and that the particular discipline used to prove the theorem in affects purity results.*

EXAMPLE 1.2 (Extraneousness within a discipline of mathematics). *A different shade of purity statements suggests that impurity can occur when the theorem and proof still share a general conception of a mathematical sort (such as ‘number’ or ‘set’, where we here take a discipline of mathematics to concern itself with a specific sort). For instance, although all numerical domains can be thought to concern numbers, Dawson [11] implies that impurity may occur when allowing imaginary values to occur in a real-valued power series. Additionally, consider Planar Desargues’s Theorem, which assumes that two triangles ABC and $A'B'C'$ lie in the same plane, and that they “are so arranged that the lines AA' , BB' , CC' meet at a point. Desargues’s Theorem then says that the three points of intersection generated by the three pairs of straight lines AB and $A'B'$, BC and $B'C'$, AC and $A'C'$ themselves lie on a straight line” [16]. A common proof of this theorem uses a point outside the plane, and draws upon all spatial axioms of Hilbert’s incidence and order axioms (Groups I and II). Although the theorem and the proof share their focus on geometrical primitives such as points and lines, the theorem appears to only concern a two-dimensional version of these notions. Therefore, the point outside the plane can be seen as a source of impurity (see also [4]). A proof from just Hilbert’s linear and planar axioms (Group I 1-2 and Group II 1-5), however, does not exist.*

(Im)purity of proof has been considered to provide various values for mathematicians over time. Purity was originally taken to indicate reliability of proof. For one, a pure proof may be more appropriate to the theorem than an impure proof, as the latter “strays at points from its proper topic and is not in all its parts relevant” [12]. Additionally, pure proofs might match more the generality of the theorem they prove, an argument put forward originally by Bolzano [7]. This relates to the idea that different domains of mathematics should not be mixed, because there is a “natural order of priority among truths in mathematics and logic”, a belief for instance supported by Frege [8]. In this traditional conception, a pure proof does not rely on a subject matter that violates the right ‘hierarchy’ of truths.

While this idea has largely disappeared nowadays, remaining values of purity focus more on epistemic benefits. For example, pure proofs allow us to “become familiar with the specific details of the subject of the theorem” [26]. In other words, we acquire a ‘deeper’ understanding of the truth of a theorem based only on local, direct properties of the relevant objects. Lehet observes, however, that contemporary mathematics has lost much of its interest for pure proofs. Instead, the traditional conception of impurity fits today’s mathematical endeavours better. Although impure proofs might divert the attention to notions outside the topic of the theorem, they are valued for their ability to unify and generalize mathematical results. For example, by proving analytic theorems algebraically, the distinct mathematical domains of analysis and algebra are unified by showing that they have intrinsic conceptual relations. Further, Lehet recognizes that explanatoriness can accompany impurity, for example, when multiple objects from different mathematical disciplines are represented by the same structure in category theory. The relation between impurity and other epistemic values such as simplicity and explanatoriness is also explored by Arana [2], Iemhoff [19], and Lange [24].

Purity is by origin characterized by the intuitions of mathematicians concerning *informal proofs*, by which we mean proofs as mathematicians conduct them in practice and present them in natural language interspersed with formal symbols (called ‘social proofs’ in [9] (p. 2), though the term ‘informal proof’ is also used, e.g., in [5]). The main notion of purity in the literature is *topical purity* of Arana and Detlefsen [3]. For Arana and Detlefsen, the topic of a theorem is the collection of commitments that determine the understanding of this theorem, relative to an agent α . These are the definitions, axioms and inferences such that if α stopped accepting one of them, then she would no longer understand this theorem. So, a proof is *topically pure* if it only draws on what belongs to the topic of a theorem. This is an epistemic notion of purity, in the sense that it considers purity to arise with respect to one’s knowledge and understanding of a theorem. Baldwin [6] describes a context-relative variant of topical purity, that checks whether, given a certain formalization and context of acceptable concepts, the proof introduces any notion outside this context by explicit definition. Kahle and Pulcini [22] propose a notion of purity that compares the strength of operations occurring in the proof with those occurring in the theorem. The latter forms an ontological notion of purity, by basing purity evaluations on the mathematical objects and/or operations that a theorem and its proof concern.

We will introduce a notion of ontological purity, where purity is generally achieved if any notion in a proof can be made sense of in terms of the mathematical ontology that a theorem concerns. Our main interest is to provide a notion of full and secondary ontological purity for *formal proofs*, by which we mean formal derivations in a proof system—although we will also specify these types of purity for informal proofs. We specifically focus on formal proofs in the classical first-order natural deduction proof system (as found in, for instance, [9]). A specific investigation into purity for formal proofs has to the best of our knowledge only been conducted by Arana [1], who concludes that a syntactic approach to purity is not desirable (although such an approach is also discouraged by others, e.g., Baldwin [6] and Kahle and Pulcini [22]). In this paper, we propose a new way to connect purity to formal derivations by taking an ontological and structural conception of content (one that is weaker, i.e., more tolerant, than previous conceptions).

1.2. Outline of the paper.

1.2.1. Aims of the paper. We provide two main contributions to the literature. First, we develop two new conceptions of purity: one based on an ontological notion of the content of a theorem, and one based on the notions of surrogate ontological content and structural content. In doing this, we depart from the notion of purity as occurring in mathematical practice. Ontological, but especially surrogate ontological content, broaden the notion of traditional purity and blur some distinctions that are made in purity evaluations in practice, most clearly distinctions between disciplines of mathematics. Such a conception of purity has been mentioned before in the literature (see, e.g., [3]), but has never been made precise.

The second aim of this paper is to characterize which (classical) first-order natural deduction proofs of a mathematical theorem are pure. Given an informal mathematical theorem, we provide a way of determining its ontological content. Formal proofs that refer to the content of theorem will be called ‘fully ontologically pure’. Formal proofs that refer to a surrogate version of the ontological content of a theorem will be called ‘secondarily ontologically pure’, because they preserve the structural content of a theorem. We will use interpretations between theories (as in [34]) to develop a proof-theoretic criterion that guarantees secondary ontological purity for formal proofs.

1.2.2. Structure of the paper. The notion of full ontological purity is developed in Section 2. First, Section 2.1 describes the ontological content of a theorem. The way this type of content can subsequently be captured by a formal theory (the ‘context’ theory) is given in Section 2.2. These efforts culminate in a first presentation of a criterion for full ontological purity (of formal proofs, as well as informal proofs) in Section 2.4. We then argue that definitional extensions of the context theory capture the same ontological content as the context theory in Section 2.5. This results in an extension of the criterion for full ontological purity of formal proofs, which is presented in Section 2.6.

Section 3 is devoted to extending full ontological purity to secondary ontological purity. First, ontological content is extended to surrogate ontological content (see Section 3.1.1) and structural content (see Section 3.1.2). We propose that this is a worthwhile extension in Section 3.2. Section 3.3 introduces interpretations between theories, and we describe how an interpretation can give rise to a restriction on syntax that allows for reference to surrogate and structural content in Section 3.4. We characterize the initial formulas and inference rule applications of natural deduction proofs that satisfy this syntax restriction in Section 3.5, leading finally to the criterion for secondary ontological purity of formal proofs in Section 3.6. Secondary ontological purity for informal proofs is considered in Section 3.7. Finally, we comment on the interaction between full and secondary ontological purity in Section 3.8, and give (partial) criteria for impurity of proof in Section 3.9. We conclude in Section 4.

The main criteria for full and secondary ontological purity of formal proofs are given by Definition 2.7 and Definition 3.4.

1.3. Formal preliminaries. As several sections of this paper will draw on rigorous definitions, we describe our main technical conventions here. First, we will use the standard classical first-order language, containing \wedge , \vee , \rightarrow , \forall , \exists , and \perp , where $\neg A$ is defined as $A \rightarrow \perp$. A signature is a set of predicate symbols and function symbols

(where constants are nullary function symbols), and we use \mathcal{L} to refer to the first-order language of a certain signature. First-order theories T are sets of axioms in a language \mathcal{L}_T .

Furthermore, $\varphi, \psi, \chi, \dots$ range over formulas in a language \mathcal{L} , and s, t, u, \dots over terms. For every formula φ in which x does not occur as a bound variable, we write $\varphi[x \setminus t]$ for the result of substituting t for x everywhere in φ . We write $\Gamma \vdash_T \varphi$ to mean that φ is derivable from assumptions Γ and axioms of T in the standard first-order natural deduction calculus (see, e.g., [9]). To clarify the language of a formula, we write φ_T to mean that φ is a formula in language \mathcal{L}_T .

Additionally, a derivation in the natural deduction proof system of theorem φ is a tree (V, E) labeled with formulas. Given a natural deduction proof and its tree representation (V, E) , we say that any node v in the tree is *instantiated* by the corresponding formula in the natural deduction proof. The root is instantiated by theorem, and the formulas instantiating the leaves are axioms, or assumptions that are later discarded. Two nodes v_{n+1} and v_n are connected by an edge $(v_{n+1}Ev_n)$ just in case the instantiation of v_n is obtained in the proof from the application of a single inference rule to a set of premises including the instantiation of v_{n+1} . Then a *branch* of the proof is any sequence $v_nEv_{n-1}E \dots Ev_1Ev_0$, where v_n is a leaf, and v_0 the root. A branch is *open* when its leaf is instantiated by an assumption (that is either later discarded or not), while a branch is *closed* when its leaf is instantiated by an axiom.

Finally, we will also refer to a natural deduction proof by \mathcal{D} . Similarly to Visser [34], given a formula δ with one free variable, δ_φ will stand for $\bigwedge\{\delta(x) \mid x \text{ free in } \varphi\}$. We then use $\lambda_{\mathcal{D}}$ to stand for $\bigwedge\{\varphi \mid \varphi \text{ instantiates a node in the tree representation of a proof } \mathcal{D}\}$. Thus, $\delta_{\lambda_{\mathcal{D}}}$ will be the conjunction of δ 's applied to all free variables in a proof \mathcal{D} . This is not to be confused with our additional notation of $\delta_{\bar{x}}$. We will write \bar{x} for the finite sequence of elements x_1, \dots, x_n . Then $\delta_{\bar{x}}$ will stand for the conjunction $\delta(x_1) \wedge \dots \wedge \delta(x_n)$.¹ Thus, we maintain a difference between using a formula or a sequence of terms as a subscript.

We will add smaller formal specifications in the relevant sections where necessary.

§2. Full ontological purity. In order to consider purity statements for formal derivations, we want to be able to tell when a formal proof belongs to the content of a theorem. In this section, we describe how, given an informal theorem, one can first describe the ontological content of the theorem. Next, a formal theory should be picked (the ‘context theory’) that is a formal counterpart to ontological content. We additionally provide a notion of equivalence for formal theories that capture the same content. This leads to a first criterion for full ontological purity in Section 2.4, and an extension of this criterion in Section 2.6.

2.1. An ontological understanding of content. Generally stated, we interpret the content of a theorem as what the theorem is about. More specifically, we stick to the conception that a theorem is “about those things to which the terms appearing in it refer” [12]. This is one of several possible conceptions of purity, and puts the emphasis on mathematical material itself. We can think of content as the ‘ontological realization’ of the topic or subject matter of the theorem, i.e., the range of mathematical objects and operations that the theorem speaks about. By not yet introducing any axiomatic

¹ We use this notation for conciseness, especially for the Appendix.

systems, syntax or semi-formal definitions, we aim to capture an intuitive view of mathematical material. For example, we may think informally about an ontology of ‘the natural numbers’, but not yet associate it with specific primitives or underlying principles. We think an intuitive way of considering mathematical material captures the essence of mathematical content, and this is what purity is naturally based on.

This view also ensures that (for now) we leave open the formal characterization of the content, emphasizing that a theorem is not ultimately about defining principles, but rather the things they define. On the other hand, the epistemological notion of purity of Arana and Detlefsen [3] takes as a basis for content (topic) “the elements that determine our grasp or understanding of mathematical problems”, such as definitions, axioms and inferences. Such content is given to a certain problem \mathcal{P} , consisting of an interrogative attitude, a propositional content, and a formulation of the content. So, while their approach is also sensitive to different possibilities of formalization, Arana and Detlefsen accommodate this in an early stage—and additionally include particular formalization choices in the topic of the problem. For now (as far as possible) we will stay away from any specific formalization choices, and focus on the ontology itself.

The focus on ontology is also motivated by our intent to look at purity for formal proofs. How formal proofs correspond exactly to informal proofs, and how they can preserve epistemic values of informal proofs, is difficult to determine. It is far from evident that a proof system can be genuinely close to how mathematicians think in practice. This means that any truly epistemic notion of purity seems (for now) unsuitable for formal proofs. We think a more accepted premise is to let the syntax of formal proofs correspond to a mathematical ontology.

We see an ontology as a ‘domain of discourse’, or a ‘realm of mathematical objects’, as in [30]. Given an informal theorem, its content is obtained by deciding what basic mathematical objects it is making a claim about. These should be objects in their intuitive form such as numbers, lines, classes, and sets. They can also be particular variants of these sorts (even numbers, finite sets, etc).² For purity purposes, it is important to have a subjective feeling of the nature of these objects; and one should be able to describe the size of this domain (e.g., if a theorem talks about all numbers, the ontology should be infinite). For instance, the Infinitude of Primes informally stated as ‘for all natural numbers a , there exists a natural number $b > a$ such that b is prime’ concerns an ontology of all natural numbers. Additionally, the relevant complexity of these objects should be considered. What version of the chosen mathematical objects does the theorem concern, i.e., what main operational machinery are the objects equipped with? Arana and Detlefsen [3] tell us that the content of the Infinitude of Primes is made up of axioms or definitions of successor, induction, an ordering, primality, and divisibility and multiplication, and that “the first-order Peano axioms for the natural numbers provide a reasonable formulation of these commitments, augmented by the definition of primality and divisibility”. Namely, these ingredients can all be seen as necessary in order to properly understand IP. We agree that the intuitive operations behind these axioms or definitions are part of the ontology of IP, as they show the way the objects are related to each other and how they may be manipulated. Our notion subtly differs by leaving open definitional dependencies

² An ontology can also be made up of abstract objects, such as groups or lattices, if they are considered as proper objects of a mathematical domain themselves, or as concrete instances of another mathematical sort: for instance, ‘all finite sets that are groups’.

between operations, which fits the idea that an ontology can be described in multiple ways.

REMARK 2.1 *As remarked by a reviewer, we are essentially taking the notion of an (intended) standard model for ontological content. This is a helpful alternative description we have in mind, and it also aligns with the way Shapiro [30] uses the term ‘domain of discourse’. However, we will stick to the terminology ‘ontology’ throughout this paper, to emphasize that the intuition for an ontology precedes any formal theory, and that for any reasonably complicated theory it becomes problematic to choose a standard model. We will sometimes, however, use the ‘standard model understanding’ in places where a formal theory has already been chosen, and where it reinforces our argument, for instance, in the next section.*

2.2. A formal counterpart for an ontology. In order to let a formal proof correspond to ontological content, we take the notion of a formal (first-order) mathematical theory (as in Section 1.3) as the formal counterpart for content. Given an ontology, we will call this theory the *context theory*. We recognize that, in principle, there is no necessary relation between a theory and a mathematical ontology in two ways: first, the subjective nature of the ontology of a theory is open (in an informal sense, PA can be taken to refer to numbers, but also to binary strings, sets, and so on). Secondly, a theory can have multiple potential ontologies (just like it can have different models), and an ontology can also be formalized by different theories, that prove different subsets of all the true sentences in the ontology. However, the decision to accept a relation between an ontology and a theory is what transfers the meaning of purity to the formal setting. Thus, we require such a relation to be fixed, in order to talk about purity for formal proofs. We think the incorporation of a theory choice is natural, as formal theories are commonly designed with the purpose of describing intuitive mathematical material. Consider for instance: “Geometry began with the informal ideas of lines, planes and points [...] Gradually, these were massaged into Euclidean geometry: a *mathematical theory* of these notions [...] [Peano Arithmetic] was intended to be a theory of our intuitive notion of number, including the basis of counting” [33]. And purity judgements in practice already include drawing on sets of axioms, e.g., “[...] there are proofs, like Furstenberg’s topological proof of the infinitude of primes, whose axioms are widely agreed to be irrelevant to the conclusion [...]” [1].

In [3], notions like axiomatic theories, definitions, inference rules, and so on, mark certain epistemic commitments and make up content (a ‘topic’) itself, instead of providing a formal counterpart for content. Baldwin [6] picks a formal vocabulary and theory to describe the topic, but only explicit definitions are compared to the intuitive content. For us, the context theory is itself a complete formal counterpart to an ontology, and we do not impose any restrictions on what the theory may define (note in particular that the theory also does not need to be able to prove the theorem, which allows for the prevention of purity results with respect to that theory).

Purity of a formal proof will then depend on whether the syntax in the proof indeed refers only to the right ontology. We think of a first-order theory as referring to an ontology through two aspects: the signature and the axioms. The signature of the context theory (constants, function symbols and relation symbols, and we here think of variables as well) denotes the basic objects and operations: terms will pick out objects,

and function symbols and predicates operations and properties.³ The referents of the primitives should correspond to basic elements and properties of the ontology, that intuitively determine its nature. However, objects and operations may also be referred to by descriptions in terms of these symbols, where their definition will determine their ontology. For example, in PA, the constant 0 denotes the particular object we think of as ‘zero’, and S an intuitive successor relation. Then $S0$ will not denote an isolated, separate object that we think of as ‘one’, but it will really correspond to ‘one’ as equivalent to ‘the successor of zero’. Hence, the ontology of the primitives will determine the ontology of more complex syntax. Furthermore, after Quine, quantifiers will signal ontology by indicating reference to an object in the domain: “[a]n object exists, or is in our ontology, just in case it is in the range of a bound variable” [30]. What kinds of intuitive objects exactly make up this ontology, however, is left up to the interpreter of the signature of a theory.

Secondly, the axioms of a theory play a role in referring to the ontology. Theories with the same signature can have different ontologies, simply because their axioms differ. For example, depending on the axioms, set theories with just the signature $\{\in\}$ can have the cumulative hierarchy T as their ontology, or T including urelements, T plus an inaccessible cardinal, and so on. If a theorem speaks about all sets, the ontology in this case depends on what the interpreter considers to be the ‘right’ universe of sets. If a theory is considered to refer to a proper part of the ontology of a theorem, however, it can refer to nothing extraneous, and it will also preserve purity of proof. That is, although one can maintain one collection of mathematical entities as the ‘real’ ontology a theorem speaks about, purity of proof should be preserved for restrictions of that ontology.

The exact way that theories refer to mathematical objects and operations is a field of research of its own, with various problems described, for instance, by Shapiro [30] and Lavine [25], such as whether a formal theory always has enough syntactic labels for ontologies with very large domains. Again, we will not assume any necessary ontological commitments of theories, however, but we ask some relation between an ontology and a theory to be fixed. Whether this relation is reasonable can be verified by the mathematical community.

2.3. More on the selection of a context theory. The previous section has generally clarified how to select an ontology for a theorem, and how a theory can subsequently refer to this ontology. We here highlight two aspects of choosing a context theory, that give some more insight into how the method works in practice.

First, the selection of a context theory can be dependent on who you ask: while most of us may find a theorem that concerns the natural numbers to correspond to an arithmetical theory like PA, a set theorist may really think of the numbers as sets, and connect the ontology of natural numbers immediately to a set theory restricted to the domain of set-theoretic natural numbers. Similarly, mathematicians may pick theories with different strengths. To illustrate, take the simple theorem ‘there are no two consecutive even numbers’, where the notion of being even can involve division by two or being the sum of two equal numbers. One person may prefer the first conception,

³ The reference to ontology can thus be seen as similar to model-theoretic interpretation of a signature—while emphasizing that predicates, function symbols and terms should be seen as referring to the ‘actual’, intuitive objects, operations and properties we have in mind, instead of just tuples of domain elements that satisfy them.

and pick PA as context theory, while someone who prefers the second conception may as well pick the weaker theory PrA (Presburger Arithmetic), which does not define multiplication. Finally, we may only have weak intuitions about the ontology of some theorems. For example, a simple set-theoretic theorem as Cantor’s Theorem (for all sets A , $|\mathcal{P}(A)| > |A|$) concerns all sets, but does not clearly tell us which universe of sets (described by which axioms) it is about. Rather, this seems dependent on what one considers the main set-theoretical universe (this is mirrored by the unclearness on what the standard model of set theory should be). These situations all show that individual preferences play a role, and that the resulting notion of purity should be seen as relative to the individual choosing the context theory.

Second, a theory capturing a certain ontology may be able to encode other content. This relates to what Isaacson [21] calls ‘hidden’ content of a theory, which may represent potentially extraneous elements, such as in the example below.

EXAMPLE 2.1 (Convergence of two Taylor series). *Described in [24] is the theorem that the Taylor series of $1/(1 - x^2)$ and $1/(1 + x^2)$ have the same convergence behaviour. Lange notes that the mathematical objects that the theorem refers to are Taylor series that involve real numbers. In particular, complex numbers can be considered impure elements, while their introduction provides a natural and explanatory proof. Any theory of the real numbers (such as Tarski’s first-order axiomatization of real closed fields, RCF⁴), will be able to represent complex numbers as ordered pairs of real numbers.*

We maintain that such a coding of extraneous elements does in fact not reduce the suitability of a context theory for capturing an ontology. Namely, we fix a relation between an ontology and the context theory (say, RCF). When RCF represents complex numbers, we have a choice in how to interpret these representations ontologically: as pairs of real numbers, or as a separate ontology of complex numbers. The fixed relation of RCF to an ontology of real numbers justifies that we interpret them as real numbers, as we have not assumed any connection between complex numbers and RCF. Thus, it is characteristic of the ontological approach that if we consider the ontology of real numbers acceptable, we find pairs of real numbers acceptable, too (possibly in contrast to an epistemic approach to purity). In Section 2.5, we repeat this point by making an extension of context theories. Complex numbers can then only be considered impure if they are thought to not ‘really’ be pairs of real numbers, but to be separate objects of their own. In this case, we may still recognize that the pairs of real numbers can ‘accurately approximate’ complex numbers. In Section 3.1.1, we will say more about such approximations, and we will attribute a secondary level of purity to them.

2.4. First criterion for full ontological purity of proof. We now present a first criterion for full ontological purity of formal as well as informal proofs. The criterion for formal proofs will be extended in Section 2.6. We do not claim any relation between an informal proof and a particular formal one, and so the purity results for formal and informal proofs should be seen as separate, though of course, compatible.

Given an informal theorem, the previous sections can be seen to provide the components of what we refer to as the *ontological context* of that theorem, denoted as

⁴ A reviewer notes that it is more accurate to say that RCF refers to the real algebraic numbers, or even another ontology—and that RCF cannot prove all intuitive properties of the real numbers. RCF should thus be seen as an imperfect context theory choice, but still as one of the most suitable first-order theories we have to refer to the ‘real numbers’.

a tuple (O, φ_T, R) . In this context, O indicates the choice of ontology for the theorem and T the choice of context theory, where φ_T stands for the theorem formalized in \mathcal{L}_T . We introduce R as a specification (with a reasonable level of detail) of how the signature of T naturally captures the basic elements of the ontology, as elaborated on in Section 2.2.

2.4.1. Purity for formal proofs. We restrict to the standard first-order natural deduction calculus. We need to be convinced that any inference rule that we use cannot introduce concepts outside of our ontology. As we consider the inference rules of the first-order natural deduction system to be truly logical, we consider them as satisfying this requirement. This then justifies the following criterion for full ontological purity.

DEFINITION 2.2 (First criterion for full ontological purity of formal proofs). *Given an informal mathematical theorem corresponding to the ontological context (O, φ_T, R) , any formal proof $\Gamma \vdash_T \varphi_T$ is fully ontologically pure for that theorem.*

2.4.2. Purity for informal proofs. For proofs as actually carried out by mathematicians, we need a slightly different approach. Like formal proofs, an informal proof should be fully ontologically pure if it only draws on the ontology of the theorem. This requires a relation between the notions described in an informal proof, and their interpretation in an ontology. Given an ontological context (O, φ_T, R) for a theorem, we may judge this by checking how the notions in an informal proof are formalizable into T . What exactly is formalization is a question outside the scope of this paper, but we suffice here in saying that it cannot just be any mapping from informal notions to syntactic elements—it will have to satisfy certain criteria that convince us it is really a particular notion that is being formalized. Next, we assume that there is a difference between a formalization in general, and a ‘natural formalization’. We will say that a ‘natural’ formalization of an informal notion into T requires the notion to intuitively be made up of basic elements of the ontology of the signature of T (just like the connection of an ontology to a formal theory in Section 2.2)⁵, and to syntactic descriptions in \mathcal{L}_T that are relatively efficient and/or elegant. Thus, we propose the following definition for full ontological purity, and discuss an example below.

DEFINITION 2.3 (First criterion for full ontological purity of informal proofs). *Given an informal mathematical theorem corresponding to the ontological context (O, φ_T, R) , an informal proof of the theorem is fully ontologically pure if there exists a natural formalization of any notion in the proof into T .*

EXAMPLE 2.2 (Infinitude of Primes). *Consider Euclid’s proof (as described, e.g., in [3]) and the topological proof of IP [14]. Let the ontology of IP be the natural numbers with arithmetical operations including addition and multiplication, and let PA be the context theory. Any notion in Euclid’s proof is formalizable in PA, and it should be relatively uncontroversial to say that any notion is even naturally formalizable into PA. We therefore consider this proof to be pure. The topological proof contains some elements that are by definition not formalizable in PA: for instance, the proof includes a topology of arithmetical sequences that has uncountably many elements. This is something that PA*

⁵ We thus assume that besides associating a theorem to an ontology, one can intuitively associate notions in a proof to elements of an ontology.

cannot define. However, it is still something that PA can represent, by letting an individual element stand for the uncountable set. We consider it likely for mathematicians to agree that this representation is still a formalization of the topological proof in PA. Although the representation of an infinity is simplified, the result is still recognizable as the notion occurring in Furstenberg’s proof.

Whether the proof is fully ontologically pure, however, depends on whether the formalization in PA is natural enough. For a natural formalization, notions like ‘arithmetic sequence’ and ‘integer’ should intuitively be made sense of in terms of natural numbers, and correspond naturally to primitives or formulas of PA. In PA, an arithmetic sequence is represented by the coding of one of its elements $a + bn$. An integer a or b is represented, for instance, by an even natural number if it is negative, and an odd natural number if it is positive. But it should be clear that integers, or arithmetic sequences, are not intuitively made sense of this way, in terms of the ontology of natural numbers. This is enough to conclude that Furstenberg’s proof is not fully ontologically pure.

This shows that full ontological purity behaves relatively similar to traditional notions of purity. If the proof contains notions that are only naturally made sense of in an ontology that the context theory does not capture, this prevents full ontological purity.

2.5. Equivalence of context theories. The rest of this chapter is devoted to extending the first criterion of full ontological purity for formal proofs. We would like to consider formal proofs modulo differences that do not affect their capturing of content. Thus, we are looking for a notion of equivalence for context theories. We will argue that definitional extensions provide such a notion. First, we introduce definitional extensions as in [18].

DEFINITION 2.4 (Explicit definition). Let \mathcal{L} and \mathcal{L}^+ be languages with $\mathcal{L} \subseteq \mathcal{L}^+$, and let R be a relation symbol, c a constant and f a function symbol of \mathcal{L}^+ . Then explicit definitions of R , c , and f in terms of \mathcal{L} , respectively, are sentences of the form

$$\begin{aligned} \forall \bar{x}(R(\bar{x}) \leftrightarrow \varphi(\bar{x})), \\ \forall y(c = y \leftrightarrow \psi(y)), \\ \forall \bar{x}, y(f(\bar{x}) = y \leftrightarrow \chi(\bar{x}, y)), \end{aligned}$$

where φ, ψ and χ are formulas of \mathcal{L} .

DEFINITION 2.5 (Definitional extension). Let T be a theory of language \mathcal{L} . A definitional extension of T to \mathcal{L}^+ is a theory $T \cup \{\theta_S \mid S \text{ a symbol in } \mathcal{L}^+ \setminus \mathcal{L}\}$ where for each symbol S in $\mathcal{L}^+ \setminus \mathcal{L}$:

- θ_S is an explicit definition of S in terms of \mathcal{L} .
- If S is a constant defined by ψ then $\vdash_T \exists_{=1} y \psi(y)$. If S is a function symbol defined by χ , then $\vdash_T \forall \bar{x} \exists_{=1} y \chi(\bar{x}, y)$.

Definitional extensions thus allow us to define abbreviations in a theory, by “replacing complicated formulas by simple ones” [18]. For example, in set theory we may denote the set z such that $\forall w(w \in z \leftrightarrow w \in x \vee w \in y)$ by $x \cup y$. We will say that two theories are equivalent, if they are both definitional extensions of the context theory.

2.5.1. Referring to the same content. We claim that formal proofs from a definitional extension V of the context theory T are fully ontologically pure, because V

refers to the same ontology as T . Given the ontological context (O, φ_T, R) , this is easiest to see if we interpret O as the (intended) standard model of T . The explicit definitions of an added symbol S in V tell us exactly how to interpret S model-theoretically in O : if S is a constant, the explicit definition of S will point to an object already in O . If S is a predicate or function symbol, its explicit definition will point to relations between objects in O that T was already aware of. Indeed, nothing new is added to O , only some elements of O that were already there are given a new name.

One might raise the objection that adding an abbreviation can change the content: for example, suppose we extend PA by the definition for a membership-like symbol \in from ZF_{fin}^+ (see [23]), or suppose we extend RCF by a relation symbol and definition for ‘being a complex number’. In these cases, it seems we only introduced the abbreviation in order to talk about actual set-theoretic membership or actual complex numbers—and this appears to introduce new objects and properties we did not have before. For both Arana and Detlefsen [3] and Baldwin [6], it is the case that, for instance, adding an explicit definition of ‘membership’ to PA can introduce impurity in a proof of, say, IP. Something new is introduced that goes beyond the topic of a theorem, a concept that we ourselves can only really understand as set-theoretic membership.

However, for ontological purity of formal proofs, we emphasize that an ontological context (O, φ_T, R) gives an ontological interpretation of anything that T can prove. This contrasts with an epistemic conception of purity, where we may think the axioms of T belong to the content of a theorem, but it does not follow that we can easily understand (and thus accept) anything that T can prove. Ontological purity of a formal proof tells us that any notion in the formal proof has an interpretation made up of the basic ontological elements of the T -primitives. Any definitional extension certainly satisfies this, as its added symbols can be made sense of ontologically exactly the way T could make sense of their definitions. It is this interpretation of the added symbols that we claim is ontologically pure, and so this relies strongly on the connection of introduced symbols to their explicit definitions. Thus, when we add \in to PA as in [23], we are not ‘really’ adding set-theoretic membership to PA: we are abbreviating a complex number-theoretic property of PA, and we are simply affirming that this property is pure. The fact that this property simultaneously is a way of representing the membership symbol of ZF_{fin}^+ , does not take away the purity of the property in PA. The symbol \in will only stand for actual set-theoretic membership, then, when it has a definition in, or is incorporated in the axioms of, a set theory that we associate with an ontology of sets. On the contrary, a definitional extension of PA by \in gives an arithmetic definition of membership; and the syntax of this definition is made sense of by the PA-axioms, which were associated to an ontology of natural numbers.

This point is reinforced by the form of natural deduction proofs in a definitional extension. The formal proofs of the context theory are all preserved by the definitional extension, but we gain ones that use a definitional axiom as in Definition 2.4. A proper use of a definitional axiom can only serve to introduce the new symbol S in the proof, which is initially embedded in a universal operator and a bi-implication. In order to actually use S in a proof, first a proof of its definitional formula is required, and this will be given from the context theory axioms. This emphasizes that the definitional axiom is not a fully independently functioning axiom, and the extension cannot fashion the (ontological) meaning of S out of thin air. This ensures purity of formal proofs from definitional extensions of the context theory.

2.5.2. *Natural formalizations into definitional extensions.* The situation for informal proofs again deserves a separate elaboration. We claimed that informal proofs that are naturally formalizable in the context theory, are fully ontologically pure. We here claim further that a proof is naturally formalizable in the context theory just in case it is naturally formalizable in a definitional extension of the context theory. Thus, while the inclusion of definitional extensions renders a larger number of formal proofs fully ontologically pure, it does not change which informal proofs are fully ontologically pure.

We illustrate our argument by taking again the example of the Infinitude of Primes. Let PA be the context theory, and consider the definitional extension $PA + \forall x \forall a \forall b (E(x, a, b) \leftrightarrow \varphi(x, a, b))$. Here, let $E(x, a, b)$ be the relation symbol added to the signature of PA, that codes $x \in B_{a,b}$ for an arithmetic sequence $B_{a,b}$ as occurring in the topological proof of IP.⁶ Intuitively, it feels like the notion of arithmetic sequences is more naturally formalizable in the definitional extension than in PA: it can now be efficiently formalized as a single predicate. At first sight, the topological proof of IP thus appears to become more pure with respect to the definitional extension.

However, note again that the predicate E is tied to its explicit definition in the language of PA. That is, when we consider how to formalize $x \in B_{a,b}$ in the definitional extension, we cannot ‘just pick’ $E(x, a, b)$. We only know to pick this symbol as the formalization because it stands for the right arithmetical coding. The syntax of this coding, subsequently, is made sense of through their use in the axioms of PA, which were associated with an ontology of natural numbers. Thus, in the end it is always the fundamentally primitive syntax which determines a formalization and an ontology. The primitives in $\varphi(x, a, b)$ still refer naturally to arithmetic notions, and not to topological concepts. In other words, we still cannot say that Furstenberg’s proof is fully ontologically pure in the definitional extension. So, we make the extension just to capture more broadly which formal proofs are fully ontologically pure.⁷

2.5.3. *Other notions of equivalence.* There exist many different notions of equivalence for theories, which can all be seen to potentially affect the nature of purity in a different way. Notions like synonymy, mutual or bi-interpretability [13], which typically arose to transfer formal results between fields, are some other options. However, these notions would equate PA and ZF_{fin}^+ , and many more theories that we intuitively think of as capturing different ontologies. Another possibility is requiring that a theory is a conservative extension of the context theory. This choice, too, is more controversial: for instance, NBG, a set theory that includes classes in its ontology, is a conservative extension of ZF. Hence, for our notion of purity, we stick to the weaker notion of definitional extensions.⁸

⁶ Here, a and b should also be seen as codes, for integers.

⁷ This deserves emphasis to avoid misunderstanding: while formal proofs in definitional extensions of the context theory are pure because of their referral to a pure ontology, this does *not* mean that informal proofs using the notion that is intuitively abbreviated in the definitional extension, also become pure, as the intuitive notion may naturally concern a different ontology.

⁸ Of course, one may still think that definitional extensions of a theory intuitively capture a different ontology as well—but we argued in this section that it is reasonable to abandon that intuition, given an ontological conception of content. For stronger notions of equivalence, we think this is no longer reasonable.

2.6. Extended criterion for full ontological purity. We end the chapter by presenting the extended criterion for full ontological purity of proof. The requirement for informal proofs here will remain the same as in Section 2.4.2, so the extension only has an effect on purity for formal proofs.

2.6.1. Purity for formal proofs. We restrict again to the first-order natural deduction calculus, and first present a definition.

DEFINITION 2.6 (Definitionally equivalent formulas). *Let \mathcal{V} be a definitional extension of \mathcal{T} , extended by the symbol S defined by the \mathcal{T} -formula ψ .⁹ Then $\varphi_{\mathcal{V}}$ is definitionally equivalent to $\varphi_{\mathcal{T}}$ if:*

- $\varphi_{\mathcal{T}}$ does not contain any instances of ψ , and $\varphi_{\mathcal{V}} = \varphi_{\mathcal{T}}$.
- $\varphi_{\mathcal{T}}$ does contain instances of ψ , which are possibly replaced by the abbreviation S . That is, we have the following cases:¹⁰
 - S is a relation symbol with explicit definition $\forall \bar{x}(S(\bar{x}) \leftrightarrow \psi(\bar{x}))$, and $\varphi_{\mathcal{V}} = \varphi_{\mathcal{T}}[\psi(\bar{x}) \setminus S(\bar{x})]$.
 - S is a function symbol with explicit definition $\forall \bar{x}, y(S(\bar{x}) = y \leftrightarrow \psi(\bar{x}, y))$, and $\varphi_{\mathcal{V}} = \varphi_{\mathcal{T}}[\psi(\bar{x}, y) \setminus (S(\bar{x}) = y)]$.
 - S is a constant with explicit definition $\forall y(S = y \leftrightarrow \psi(y))$, and $\varphi_{\mathcal{V}} = \varphi_{\mathcal{T}}[\psi(y) \setminus (S = y)]$.

Now for the extended criterion for full ontological purity of formal proofs.

DEFINITION 2.7 (Extended criterion for full ontological purity for formal proofs). *Suppose we are given an informal mathematical theorem corresponding to the ontological context $(\mathcal{O}, \varphi_{\mathcal{T}}, R)$. Let*

$$[\mathcal{T}] := \{\mathcal{V} \mid \mathcal{V} \text{ definitionally extends } \mathcal{T}\}.$$

Then, for $\varphi_{\mathcal{V}}$ definitionally equivalent to $\varphi_{\mathcal{T}}$, any formal proof $\Gamma \vdash_{\mathcal{V}} \varphi_{\mathcal{V}}$ for $\mathcal{V} \in [\mathcal{T}]$ is fully ontologically pure for that theorem.

§3. Secondary ontological purity. So far, we have claimed that the mathematical ontology of a theorem can be captured by a theory and its definitional extensions, and that natural deduction proofs starting from the axioms of these theories (and informal proofs naturally formalizable in them) are fully ontologically pure. We are now interested in extending this notion of purity in a way that has been mentioned in the literature several times. For example, Arana [1] mentions in a footnote that we might think that for Furstenberg’s proof of IP, “the allegedly extraneous topological elements are really just reconceptualizations of what was already the concern, namely sets of natural numbers, and hence are in fact relevant to the infinitude of primes”. Arana and Detlefsen [3] note that Colin McLarty, as well as a Bourbakiste tradition of arithmetical research, consider theorems like IP to contain *both* arithmetical and topological content—although Arana and Detlefsen do not follow this line of thought, because on their epistemological account, a notion that takes into account the understanding of the theorem has priority. Additionally, McCarthy [28] discusses the possibility that

⁹ This definition can easily be extended to work for extensions by multiple symbols.

¹⁰ We describe here the case where all instances of ψ are replaced by S , but alternatively, $\varphi_{\mathcal{V}}$ may also replace only some or no instances of ψ .

“[w]hether a proof of a number-theoretic assertion counts as “pure” depends on the conceptual background against which it is formulated. If it is framed in a wide context, for example, second-order analysis or ZF, then it may be that the most natural notion of purity is that applying to the wider context and not the strictly number-theoretic one”.

Generally, these proposals are rejected by their authors, because they fail to retain the traditional values of purity. Hence, it seems clear that if we make such a generalization of purity, the original values of (im)pure proofs change. One may then have concerns about the motivation for considering this notion of purity. In the next section, we develop the notions of surrogate and structural content, after which we provide several reasons why the type of purity that deals with these forms of content is still worth studying. We will then introduce the notion of interpretations, which will serve to preserve the structural content of a theorem. This will enable a proof from a theory that is neither the context theory nor a definitional extension, when restricted by a derivation criterion (defined in Section 3.6), to still correspond to the structural content of a theorem and gain a secondary level of purity. We end the chapter by considering impurity of proof and the interaction between full and secondary ontological purity.

3.1. Extending content. Here, we extend the notion of ontological content to surrogate content, which will be the ontology that secondarily ontologically pure proofs refer to. Next, we suggest the ontology of a context theory and its surrogate versions have structural content in common. This is what will justify the attribution of a secondary level of purity to proofs referring to surrogate content.

3.1.1. Surrogate content. Suppose that we relate a particular mathematical ontology to a context theory. Subsequently, we consider a second theory that concerns a very different ontology. Intuitively, we may ignore a large part of the latter informal objects and forget some of their properties, i.e., we may conceptually ‘trim’ and weaken the entities that make up the content for us. That is, (1) we can ignore the part of the domain of the ontology of a second theory, that will not take part in representing the ontology of the context theory. We may (2) pair up or equate remaining objects if that is necessary for representing individual context theory objects. And (3) we can ignore properties of the objects that the second theory can prove, but that do not take part in representing properties of the ontology of the context theory.¹¹

Arguably, we change the nature of the ontology of the second theory by carrying out these steps, since they do not form anymore the intuitive material that corresponds to the full formal theory. In fact, the ontology of the second theory can be (informally) restricted in such a way that the things that remain function as *surrogates* of the intended mathematical entities of the context theory. This makes up a large part, for example, of the foundational role of set theory:

(Maddy [27]) “To say that ‘the universe of sets is the ontology of mathematics’ amounts to claiming that the axioms of set theory imply the existence of (surrogates for) all the entities of classical mathematics – a simple affirmation of set theory’s role as Generous Arena”.

¹¹ On the view of an ontology as an intended standard model, surrogate content can be given a more precise description. We will comment on that in Section 3.3.

For instance, if we consider a universe of sets to be the material that ZFC refers to, we can restrict this universe to just the finite ordinals, that satisfy only set-theoretically realized arithmetical properties. The resulting intuitive objects can be seen as surrogates for the intuitive natural numbers that PA refers to—and we say that the restricted content of ZFC consists of surrogates of the content of PA. We emphasize that the extracted surrogate content in such theories really loses some of its original nature. That is, the collection of sets that simulate numbers is not by itself (without a connection to the full content) anymore the content of ZFC. Besides foundational theories, examples of ‘surrogative reasoning’ (a term also used in [32]) can be found in mathematics in practice: for instance, in analytic geometry, where Cartesian coordinates are used to represent points in space. Additionally, the relevance of restricting to surrogates can be recognized in Maddy’s [27] description of Essential Guidance: “[the universe of sets] includes hordes of useless structures and [...] no way of telling the mathematically promising structures from the rest”.

In a footnote (pp. 3–4), Arana [1] suggests ‘reconceptualizations’ (similar to the notion of surrogates) may not be relevant for purity: first, “not every reconceptualization of a subject matter is necessarily relevant to that subject matter”. Specifically, “the infinitude of primes does not seem to concern sets at all; so I do not see why sets, even simple ones, are relevant to the problem”. We agree that surrogate sets are not necessarily relevant epistemically to IP: they do not seem to contribute to the usual understanding of the problem. However, we propose a different kind of relevance: surrogates are relevant for ontological purity, because they are connected through an underlying structure with the ontology of the context theory (see Section 3.1.2, and again, the motivation for such a notion is elaborated on in Section 3.2). Another point made in the footnote is that “if Furstenberg’s proof were formalized in a different way, say in a theory in which the notion of open set was taken as primitive [...] the response [that set theory shows the topology used to prove IP is just limited to simple sets] would not work. Why should set-theoretic formalization take precedence over these alternatives?” We think the emphasis should not be so much on the idea that the topology is limited to simple sets, but rather on the idea that the topology is limited to some representation that we know relates to arithmetic (such as ‘simple’, or surrogate, sets). In a theory that takes open sets as primitive, we could again find that the open sets used to prove IP are just simple ones, by seeing that they are still surrogates of arithmetical notions. That is, the aim is not to get rid of topology, but to make sure its use is restricted to number-surrogates, whichever ones.

The type of purity we will attribute to proofs referring to surrogate content has a ‘secondary level’ compared to full ontological purity. Namely, unlike definitional extensions of the context theory, for theories referring to surrogate content we cannot talk about strict equality to original content anymore, as they may well correspond to a different ontology. Thus, full ontological purity cannot be attained here. Secondary ontological purity then entails that, if a theorem concerns an ontology of natural numbers, its proof should only draw on this ontology, or alternatively its set-theoretic surrogates, geometric surrogates, and so on—but not anything else.

3.1.2. Structural content. We here propose that for a certain ontology, each type of surrogate content has structural content with it in common. That is, we suggest each theorem has structural content in the form of a mathematical structure, which has as its instances the ontological content of a theorem, as well as surrogate versions of the

ontological content. Structural content has been proposed before as “the instantiation of a particular fundamental mathematical structure by the entities intuitively involved in the statement” [29], but for us the structural content will refer exactly to the structure itself. There are several variants of structuralism, with a distinction between eliminative structuralism (there are possible structures, but not actual ones) and non-eliminative structuralism (there are actual structures) [30]. Within non-eliminative structuralism, we can distinguish between *in re* structuralism and *ante rem* structuralism. In short, *in re* structuralism says that there is no more to structures than their instances, while *ante rem* structuralism claims that structures exist independently of the systems that realize them. Our approach to purity is neutral with respect to the debate on structuralism, but will adopt *ante rem* structuralism for the notion of structural content. This is because it shows that, by existing independently from (non-mathematical) exemplifications, structures are themselves something that can be preserved when switching between ontological content and its surrogates. A structure independently shows the properties that different versions of surrogate content have in common with each other.

More specifically, “an *ante rem* structure is, or is akin to, an *ante rem* universal, in that it is a one-over-many. The same structure can be exemplified in multiple systems, and the structure exists independent of any exemplifications it may have in the non-mathematical realm. The difference between structures and the more usual kind of universal, such as properties, is that structures are the forms, not of individual objects, but of systems, collections of objects organized with certain relations” [31]. Shapiro calls the structures studied in mathematics “free-standing”, i.e., anything at all can occupy their places. Here, we are interested in all instantiations of the places and relations of *ante rem* structures by mathematical ontologies. Given the ontology of a context theory, its underlying *ante rem* structure can be seen as consisting of ‘abstract places’ for each object in the domain, connected to each other by ‘relations’ that determine the fundamental connections of the structure. Like Shapiro [31], we think that “understanding the (formal) languages of mathematics is sufficient to understand the places and relations of at least some structures”. For instance, the familiar natural number structure consists of an initial object, and a successor relation satisfying the induction principle, connecting the initial object to a successor object, and so on.

A(n) (surrogate) ontology can let their objects occupy the structural places, and the structural relations can be occupied by their concrete instantiations selecting objects from an ontology. By allowing instantiation from any ontology, the structural content underlies ontological content. Its preservation when moving between surrogate versions of content is what justifies a secondary level of purity: it ensures that, while a proof can draw on various ontologies and disciplines of mathematics, it is at least restrained to the particular structure that a theorem concerns.

3.2. The value of extending ontological purity. We now have a conception of the types of content we want to extend purity results to. We here provide some reasons for why this extension of purity is worth considering. Arana and Detlefsen [3] mention the *intervenient* value of traditional purity, as well as an *epistemic* value that they consider in more detail. The *intervenient* value of purity is broadly the ‘development of a thorough way of thinking’ (Bolzano, cited in [3]). We believe secondary ontological purity still encourages a variant of this benefit: it encourages one to, within a specific discipline of mathematics, focus ones thinking on specific surrogates only. The *epistemic* value,

however, arguably disappears: this value focuses on giving insights that have a certain simplicity and naturalness, and according to topical purity, reduces ‘specific ignorance’. The latter is something our extension does not preserve, as we will, for instance, allow the concept of set to be used in a proof of IP, whereas Arana and Detlefsen [3] do not, on account of its failure to reduce specific ignorance for IP. Then what other values can we attribute to our extension of purity?

First of all, the extension can be seen as the weakening to a ‘core’ notion of purity, one that tells us what pure proofs should satisfy at the very least. In other words, secondary ontological purity in a sense underlies other, more specific notions of purity that satisfy additional criteria—and perhaps this may help us understand the connections and dependencies between other types of purity better, by seeing them as particular cases of the extension. We also suggest that this makes it easier to distinguish *levels* of (im)purity, instead of the more blunt distinction between ‘pure’ and ‘not pure’. We suggest we have secondary ontological purity and full ontological purity, but stricter notions like Arana and Detlefsen’s [3] topical purity induce even stronger notions of purity, allowing for a more nuanced picture suggesting purity really consists of a family of notions.

Furthermore, the extension of ontological purity caters to the views of structuralists. For ante rem structuralists, traditional purity may not properly reflect what a theorem is about. Mathematical content may for them ultimately concern structures, and we suggest there is still a notion of purity for them to value. Similarly, the extension of purity also allows for the view that an informal theorem does not have one ‘true’ ontology, and that surrogate content is a relevant subject matter of a theorem.

We additionally claim that the extension of purity will still have an epistemic value. Traditionally impure proofs have been said to lead to new insights, by connecting mathematical notions that at first seem separate (see, e.g., [26]). Our extension of purity brings nuance to this value: it suggests that a pure proof can also have this value, but only by connecting notions from different mathematical ontologies that represent the same placeholder in a structure. On the other hand, impure proofs can still lead to new insights, but in a different way: by showing what intricate notions really go beyond representing ‘pure’ content, yet are still useful for proving a theorem. That is, the new distinction between purity and impurity allows us to see what parts of a theory can be seen as ‘purer’ than others.

Finally, it should be observed that the extension of purity is valuable for characterizing purity for formal proofs: separate from the motivation for introducing surrogate content, the setting of formal proofs itself already encourages an extension. For example, if we insist on saying something about the formal proofs of PA, but the context theory for our theorem is Presburger Arithmetic (PrA), then it is reasonable to think that a restricted version of PA can provide pure proofs of the theorem. Intuitively, after all, both PA and PrA refer to the natural numbers, only PrA assigns fewer properties to its numbers than PA. If we can exclude the properties that PA can prove and that PrA cannot, proofs of PA can only draw upon relevant notions. Formal proofs can only satisfy such a restriction if we extend what we have done so far.

We thus consider the extension to surrogate (and structural) content a reasonable one. In what follows, we introduce a way to use interpretations between theories to restrict formal proofs, so that they refer only to surrogate (and structural) content and thereby induce a secondary sense of purity.

3.3. Interpretations. Interpretations between first-order theories provide a way for theories to represent each other’s language and provable statements. They will be important in developing formal guarantees for proofs to refer to surrogate content. We take a slightly altered version of the definition of relative interpretations in [34]. Consider first-order theories T_1 and T_2 with respective languages \mathcal{L}_{T_1} and \mathcal{L}_{T_2} that have identity. An *interpretation* i of T_1 into T_2 ($i : T_1 \rightarrow T_2$) has two ingredients, which will determine the interpretation translation $(.)^i$:

1. A function F mapping the relation symbols R and the function symbols f of \mathcal{L}_{T_1} to formulas of \mathcal{L}_{T_2} . If the arity of R (respectively f) is k , then $F(R)$ (respectively $F(f)$) has k free variables.
2. A formula δ of \mathcal{L}_{T_2} , with one free variable, giving the domain of the interpretation.¹²

A well-known example of an interpretation is that of arithmetic into set theory, for instance, of PA into ZFC—where the domain formula δ restricts the universe of sets to the finite ordinals. Then, for instance, the PA-constant 0 can be translated as the empty set, the successor operation $S(x)$ as $x \cup \{x\}$, and so on.

A first minor adaptation that we make of Visser’s [34] definition concerns variables. Visser extends \mathcal{L}_{T_2} with fresh variables in order to avoid variable clashes. Instead, we will make the simplified assumption that we can always map variables x of \mathcal{L}_{T_1} to identically named variables in \mathcal{L}_{T_2} ($x \mapsto x$). In case this does cause variable clashes for a translated variable x , we will take x to ‘stand for’ a suitable fresh variable of \mathcal{L}_{T_2} . By keeping the variable names the same, we avoid focusing on purely formal aspects of interpretability, and keep its definition a bit more intuitive.¹³

Now i gives our translation $(.)^i$ of \mathcal{L}_{T_1} into \mathcal{L}_{T_2} in the following way. First, \perp is interpreted as itself. Further, the translation of a formula φ will have the same free variables as φ itself—while the translation of a term t will have the free variables of t plus one more fresh variable, standing for the *value* of t . However, as we send variables to themselves, term translations disappear when the term is a variable. In the definition, let \overline{x}_k stand for a sequence of k variable terms, and let \overline{t}_l stand for a sequence of l non-variable terms. (See Section 1.3 for the notation $\delta_{\overline{x}}$.)

- $R(\overline{t}_l, \overline{x}_k)^i := \exists \overline{y}_l (\delta_{\overline{y}_l} \wedge F(R)(\overline{y}_l, \overline{x}_k) \wedge (t_1)^i(y_1) \wedge \dots \wedge (t_l)^i(y_l))$
- $f(\overline{t}_l, \overline{x}_k)^i := \exists \overline{y}_l (\delta_{\overline{y}_l} \wedge F(f)(\overline{y}_l, \overline{x}_k) \wedge (t_1)^i(y_1) \wedge \dots \wedge (t_l)^i(y_l))$
- $(.)^i$ commutes with the propositional connectives.
- $(\forall x \varphi)^i := \forall x (\delta(x) \rightarrow \varphi^i)$, $(\exists x \varphi)^i := \exists x (\delta(x) \wedge \varphi^i)$

The last item conveys that the quantifiers of translated formulas become relativized by the domain formula δ . In the interpretation translation of predicates and function symbols (the first two items above), Visser introduces unrelativized existential quantifiers. As shown above, our second adaptation is that we relativize even these quantifiers. This will ensure that we can consistently restrict a proof to ‘good’ syntax later on.

¹² Note that we are here only allowing one-dimensional interpretations. This is not essential: surrogates can be made up of single elements or of tuples from the ontology of the interpreting theory. However, one-dimensionality provides for notational simplicity in further definitions and in the Appendix.

¹³ It will, however, have the consequence that we need to distinguish between variable and non-variable terms in the Appendix, although this does not take too much effort.

Finally, we state the preservation of provability of an interpretation. T_2 interprets T_1 via i if: for all theorems φ of T_1 , $\vdash_{T_2} \delta_\varphi \rightarrow \varphi^i$.

We also note that, given a model \mathcal{M} of T_2 , an interpretation $i : T_1 \rightarrow T_2$ gives a way to induce a model of T_1 on δ [34]. Essentially, the objects of \mathcal{M} are equated according to interpreted identity, after which the objects that satisfy δ are selected. Interpreted predicates and function symbols then work on these equivalence classes of objects (see for full details [34]). This can be seen as a formal way of making sense of surrogate content, analogous to how we can make sense of the ontology of a theory generally as a standard model.

3.4. Referring to extended content. This section will elaborate on how syntax restricted in a way inspired by interpretations can refer to surrogate and structural content. We discuss how interpretations can induce a natural notion of surrogate content of an ontology, and how they provide a way of preserving structural content.

3.4.1. Referring to surrogate content. The interpretation translation of $i : T_1 \rightarrow T_2$ gives a clear indication of how T_2 -syntax should refer to surrogate content. The domain formula δ of i characterizes the collection of surrogate objects in T_2 . In particular, for any T_1 -formula φ that indicates an object or operation in its ontology, φ^i will indicate the surrogate version. And similarly to the transformation of a model of T_2 into a model of T_1 described above, we may view two objects in the ontology of T_2 as exactly the same surrogate if they are equal under $F(=)$.

We suggest that the properties of the interpretation translation are suitable for inducing a notion of surrogate content as we informally mean it. An important feature of a translation is (a first-order variant of) *schematicity* as in [20], where “the translation of a complex formula [should be] a fixed schema of the translation of its parts”. This is handled by the interpretation by its translation of a formula based on the individual translations of constants, function symbols and predicates occurring in it, and by its commutativity with propositional connectives. The translation is also injective, and the arity of function symbols and predicates is preserved, so that the translation in T_2 makes the same distinctions between objects and operations as T_1 does. This is confirmed by the fact that the interpretation translation allows for well-known definitions of surrogates in practice: for instance, both the Von Neumann ordinals and the Zermelo ordinals can easily be defined through δ (see for various other examples of interpretations [34]).

In addition to accepting the interpreted \mathcal{L}_{T_1} -formulas, however, we are looking to characterize the part of the \mathcal{L}_{T_2} -syntax that we can in practice restrict a T_2 -proof to, so that the restricted part of the proof only refers to surrogate content. This cannot simply be the set of interpreted \mathcal{L}_{T_1} -formulas, as not all inference rules preserve ‘being of interpreted form’. Instead, we propose to restrict a proof to certain ‘good’ \mathcal{L}_{T_2} -formulas, of which the interpreted formulas will be a subset, and which comes down to an extension by instances of δ and by interpreted terms. Since instances of δ and interpreted terms only highlight specific elements within the surrogate ontology, this extension is harmless for referring to surrogate content. The definition of ‘good formulas’ is as follows.

DEFINITION 3.1 (Good \mathcal{L}_{T_2} -formulas). *Let $i : T_1 \rightarrow T_2$ be an interpretation with domain formula δ . First, an \mathcal{L}_{T_2} -term t is good if it is either a variable, or a function symbol f applied to terms t_1, \dots, t_n such that:*

- Each t_j is good ($1 \leq j \leq n$).
- $\vdash_{T_2} \delta_{\bar{t}} \rightarrow \delta(f(\bar{t}))$.

Now we define the set of \mathcal{L}_{T_2} -formulas \mathcal{S} by the following ingredients.

$$\mathcal{S} := \{\varphi^i \mid \varphi \in \mathcal{L}_{T_1}\} \cup \{t^i \mid t \text{ a term of } \mathcal{L}_{T_1}\} \cup \{\delta(t) \mid t \text{ a good term of } \mathcal{L}_{T_2}\}.$$

We then define the good \mathcal{L}_{T_2} -formulas as follows:

- Each $\varphi \in \mathcal{S}$ is good.
- If φ and ψ are good, then $\varphi \circ \psi$ is good ($\circ \in \{\wedge, \vee, \rightarrow\}$).
- If φ is good, $\exists x(\delta(x) \wedge \varphi)$ is good.
- If φ is good, $\forall x(\delta(x) \rightarrow \varphi)$ is good.

Thus, the set of good \mathcal{L}_{T_2} -formulas will be exactly what we will restrict a T_2 -proof to for secondary ontological purity. Finally, we note here that Arana [2] has argued that interpretations are not sufficient to preserve topical purity, as translations do not preserve understanding. Specifically, Arana rejects the idea that “if two theories T_1 and T_2 are mutually interpretable, then their semantic parts (terms, statements) have identical meanings”, because then topical purity does not capture mathematical practice anymore. We agree with both points, and emphasize that we only claim a correspondence between \mathcal{L}_{T_1} -syntax and the set of good \mathcal{L}_{T_2} -formulas. Nor do we claim that these two sets have a fully identical ontology, but we will argue in the next section that they have an ante rem structure in common, which suffices for our sense of secondary ontological purity.

3.4.2. Referring to structural content. We see a theory as referring to an ante rem structure through the reference to its ontology. Thus, each formula φ describing an object in the ontology T_1 can be seen to additionally refer to the structural place underlying this object in the ante rem structure. It may differ per ontology what relations the structure preserves (in ‘placeholder’ form). Given the relations of a structure, however, their ontological versions will certainly be captured by the formulas of T_1 . If anything, the structure will have less (detailed) properties than a full ontology, so that T_1 should always be able to refer (by an ontological instance) to what the structure is made up of.¹⁴

The preservation of provability of interpretations now ensures that reference to this structure is preserved by the interpreted formulas in T_2 . That is, for each \mathcal{L}_{T_1} -formula φ that refers to an element of its ante-rem structure through its ontology, φ^i in \mathcal{L}_{T_2} can be seen to refer to the same element of the structure through its surrogate ontology. If we take T_1 to refer to some ontology, and T_2 to be able to refer to surrogates of this ontology, then we should also accept that both of them can refer to what these ontologies have in common. Preservation of provability makes exactly the right connection between an ontology and a surrogate ontology, and the ante rem structure that both of them instantiate. The general situation is illustrated in Figure 1.

¹⁴ Shapiro [30] notes that the natural numbers with just a successor operation, or the natural numbers with, e.g., additionally an order relation, ideally describe the same structure. Here, we only claim a structure is preserved when all provable properties of a theory are preserved; and so we distinguish more structures than Shapiro ultimately intends. Secondary purity could thus be extended even further—but for now, we have at least ensured structure preservation.

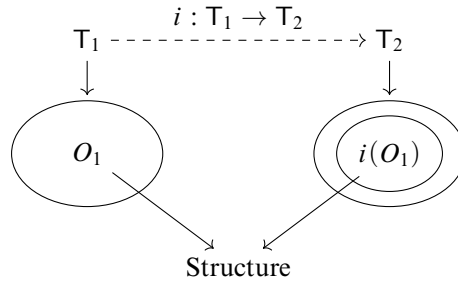


Figure 1. A visual representation of the reference of theories T_1 and T_2 to the ontologies O_1 and O_2 , and of the structure underlying O_1 and the surrogate ontology $i(O_1)$.

3.5. A restriction on proof rules. We are now set to define the two ingredients that will culminate in the derivation criterion in the next section, restricting formal proofs to ‘good’ formulas. The first ingredient selects particular instances of ‘good’ formulas (the interpreted T_1 -axioms, δ -instances, and instances of interpreted functionality and identity axioms). Considering each natural deduction proof as a tree (V, E) as in Section 1.3, part of the derivation criterion will be to require that one of these formulas occurs in each branch. Here, we will write $=^i$ for interpreted equality, and use $\bar{x} =^i \bar{y}$ for the conjunction $x_1 =^i y_1 \wedge \dots \wedge x_n =^i y_n$.

DEFINITION 3.2 (Pure \mathcal{L}_{T_2} -formulas). *Let $i : T_1 \rightarrow T_2$ be an interpretation with domain formula δ . Then an \mathcal{L}_{T_2} -formula φ is pure if one of the following cases holds:*

1. φ is an interpreted (non-)logical axiom of T_1 .
2. $\varphi = \delta(t)$, for t a good term of \mathcal{L}_{T_2} .
3. φ is a relativized uniqueness or totality axiom for interpreted function symbols. This means that φ can be:

(a) (Uniqueness) For any \mathcal{L}_{T_1} -term t ,

$$\forall x \forall y (\delta(x) \wedge \delta(y) \rightarrow (t^i(x) \wedge t^i(y) \rightarrow x =^i y))^{15}$$

(b) (Totality) For any \mathcal{L}_{T_1} -term t ,

$$\exists y (\delta(y) \wedge t^i(y))$$

4. φ is a relativized first-order equality axiom for interpreted equality.¹⁶ This means that φ can be:

(a) (Reflexivity)

$$\forall x (\delta(x) \rightarrow x =^i x)$$

(b) (Function substitution) For any \mathcal{L}_{T_1} -function symbol $f(\bar{x})$,

$$\forall \bar{x} \forall \bar{y} (\delta_{\bar{x}} \wedge \delta_{\bar{y}} \rightarrow (\bar{x} =^i \bar{y} \rightarrow (F(f)(\bar{x}, z) \rightarrow F(f)(\bar{y}, z))))$$

¹⁵ The usual definition of Uniqueness uses an implication between $\delta(x)$ and $\delta(y)$. Here we use the equivalent definition by conjunction, so that we can use the abbreviation $\delta_{\bar{x}}$ in the Appendix in long natural deduction proofs. The same holds for the substitution axioms in 4.

¹⁶ Note that the equality axioms technically fall under 1 as the interpretation of a logical T_1 -axiom. We add them for clarity, as they will be used regularly in the Appendix.

(c) (Formula substitution) For any $\mathcal{L}_{\mathcal{T}_1}$ -formula φ ,

$$\forall \bar{x} \forall \bar{y} (\delta_{\bar{x}} \wedge (\delta_{\bar{y}} \rightarrow (\bar{x} =^i \bar{y} \rightarrow (\varphi^i(\bar{x}) \rightarrow \varphi^i(\bar{y}))))).$$

The second ingredient specifies which applications of the inference rules of the natural deduction proof system preserve ‘goodness’ of formulas. Thus, we are not defining a new proof system, but we emphasize for a rule R to which instances (denoted by R^i) it should be restricted in proofs that are to refer to surrogate (and structural) content.

DEFINITION 3.3 (Pure rule applications). *We provide four restricted inference rules, so that if the premises of the rule are good, then the conclusion of the rule is good as well. It is easily verifiable that the other remaining rules already preserve goodness—we will call their instances together with those of the restricted rules pure rule applications.*

- Disjunction introduction. *In order to ensure that the conclusion of this rule is good, we require that the introduced disjunct is also good (denoted by ψ_g).*

$$\frac{\varphi}{\varphi \vee \psi} \forall I \quad \frac{\varphi}{\varphi \vee \psi_g} \forall I^i$$

- Universal introduction. *In order to ensure that the conclusion of this rule is good, we require that the form of the premise must be restricted to implications with antecedent δ .*

$$\frac{\varphi(y)}{\forall x \varphi(x)} \forall I \quad \frac{\delta(y) \rightarrow \varphi(y)}{\forall x (\delta(x) \rightarrow \varphi(x))} \forall I^i$$

- Existential introduction. *In order to ensure that the conclusion of this rule is good, we require that the form of the premise is restricted to conjunctions, with δ as one of the conjuncts.*

$$\frac{\varphi[x \setminus t]}{\exists x \varphi(x)} \exists I \quad \frac{(\delta \wedge \varphi)[x \setminus t]}{\exists x (\delta(x) \wedge \varphi(x))} \exists I^i$$

- Universal elimination. *In order to ensure that the conclusion of this rule is good, we require that the instantiation with term t in the conclusion is such that t is good (denoted by t_g).*

$$\frac{\forall x \varphi(x)}{\varphi[x \setminus t]} \forall E \quad \frac{\forall x \varphi(x)}{\varphi[x \setminus t_g]} \forall E^i$$

3.6. Criterion for secondary ontological purity of formal proofs. We combine the previous definitions and state when a formal proof is secondarily ontologically pure.

DEFINITION 3.4 (Criterion for secondary ontological purity of formal proofs). *Suppose we are given an informal theorem corresponding to the ontological context $(O, \varphi_{\mathcal{T}_1}, R)$. Suppose we are given a natural deduction proof in \mathcal{T}_2 of φ , where φ is a formalization of the theorem into $\mathcal{L}_{\mathcal{T}_2}$. Let φ instantiate the node v in the corresponding tree (V, E) . Then the proof is secondarily ontologically pure (denoted $\Gamma \vdash_{\mathcal{T}_2}^P \varphi$) if there exists an interpretation $i : \mathcal{T}_1 \rightarrow \mathcal{T}_2$ such that:*

- Any assumption instantiating a node a in the proof is good, and the open branch $aE \dots Ev$ contains only pure rule applications.
- Any closed branch contains a node p that is instantiated by a pure formula, such that the final branch part $pE \dots Ev$ contains only pure rule applications.

This criterion ensures that, starting from the leaves, at some point in the proof each branch is restricted to good formulas only. We will sometimes refer to this criterion as ‘the derivation criterion’ for conciseness.

3.6.1. Robustness of the criterion. In order to show that this criterion is robust, we show that the criterion has instances. Namely, we show that T_2 has a secondarily ontologically pure formal proof of every interpreted theorem of T_1 . For this, we use what we call ‘simulation’. Recall that $\delta_{\lambda_{\mathcal{D}}}$ is the conjunction of δ ’s applied to all the free variables of formulas occurring in the proof \mathcal{D} . The instances of δ are needed to guarantee provability of interpreted formulas (see Section 3.3). Furthermore, let Γ^i be the set $\{\gamma^i \mid \gamma \in \Gamma\}$.

DEFINITION 3.5 (T_1 -simulation). *Let $i : T_1 \rightarrow T_2$ be an interpretation. Let \mathcal{D} refer to $\Gamma \vdash_{T_1} \varphi$, and let it correspond to a tree (V, E) . Then a simulation of \mathcal{D} in T_2 is a proof $\Gamma^i, \delta_{\lambda_{\mathcal{D}}} \vdash_{T_2} \varphi^i$, corresponding to (V', E') , with the following requirements.*

- For every branch step $v_{n+1}Ev_n$ in (V, E) such that χ instantiates v_{n+1} and ψ instantiates v_n , there exists a sequence $v_mE \dots Ev_k$ ($m > k$) in (V', E') such that χ^i instantiates v_m and ψ^i instantiates v_k .
- Take any sequence of nodes $v_{n+2}Ev_{n+1}Ev_n$ in (V, E) . Suppose $v_{n+2}Ev_{n+1}$ corresponds to the sequence $v_mE \dots Ev_k$ in (V', E') ($m > k$), and $v_{n+1}Ev_n$ corresponds to the sequence $v_lE \dots Ev_q$ in (V', E') ($l > q$). Then $v_k = v_l$.

Intuitively, this is a simulation, as each inference step is replaced by a proof from its interpreted premises to the interpreted conclusion. More formulas can be added in between the interpreted formulas, however, to secure provability. Thus, there are various ways of simulating a proof. We now consider the theorem, which shows there are pure simulations.

THEOREM 3.6 (Existence of pure simulations). *Let $i : T_1 \rightarrow T_2$ be an interpretation. Let \mathcal{D} be the proof $\Gamma \vdash_{T_1} \varphi$ in the classical first-order natural deduction calculus. Then $\Gamma^i, \delta_{\lambda_{\mathcal{D}}} \vdash_{T_2}^P \varphi^i$ by simulation.¹⁷*

Proof. The proof idea is as follows: we provide the pure T_2 -simulations of each inference rule in T_1 . Then a complete T_1 -proof is simulated in T_2 by pasting together the simulations of the individual rule applications. For propositional rules, and the rules $\forall I$ and $\exists E$, it is easy to see there are pure simulations of any use of the rule. For $\forall E$ and $\exists I$, we need a lemma to deal with interpreted formulas of the form $(\varphi(t))^i$, which cannot easily be rewritten, as the term translation t^i appears only at the atomic level of translations of predicates and function symbols inside φ . To provide a simulation of $\forall E$ and $\exists I$, then, we take a detour through the formula $\exists x(\delta(x) \wedge t^i(x) \wedge \varphi^i(x))$. The full proof can be found in the Appendix. □

¹⁷ Although it is very likely that the other direction of this theorem also holds (if φ^i can be proven in T_2 in a pure way, then T_1 proves φ), we do not show it here. Note that, although this is a desirable property, a counterexample would merely show that a notion can be made sense of in terms of surrogate ontology, in a more complex way than T_1 finds acceptable.

3.6.2. *A (very) small working example.* Take the interpretation $i : Q \rightarrow C_{FO}^2$ of Robinson's Q into a theory of concatenation that has signature $(*, a, b)$. The interpretation is defined in [15]. We provide the domain formula and the interpretation of the constant zero (which suffices for the example). The interpretation is identity-preserving.

- $\delta(x) := T(a, x) \vee x = b$. Here, $T(a, x)$ is an abbreviation for saying x is a string made up entirely of a 's. Thus, a natural number $n > 0$ is represented in C_{FO}^2 as a string of a 's of length n .
- $0^i := x = b$.

Now take the very simple Q-proof:

$$\frac{\forall x(x = x)}{0 = 0} \forall E$$

And consider a similar proof in C_{FO}^2 .

$$\begin{array}{c} C_{FO}^2 \\ \vdots \\ \frac{\forall \mathbf{x}((\mathbf{T}(\mathbf{a}, \mathbf{x}) \vee \mathbf{x} = \mathbf{b}) \rightarrow \mathbf{x} = \mathbf{x})}{(T(a, b) \vee b = b) \rightarrow b = b} \forall E \quad \frac{\frac{\forall x(x = x)}{b = b} \forall E}{\mathbf{T}(\mathbf{a}, \mathbf{b}) \vee \mathbf{b} = \mathbf{b}} \forall I \\ \hline b = b \rightarrow E \end{array}$$

Marked bold are the pure formulas occurring in each branch: from that moment on, the proof is secondarily ontologically pure. Note that this proof satisfies the derivation criterion, but is *not* a simulation, as we reach $b = b$ instead of the literal translation $(0 = 0)^i$. The Appendix provides a way to properly simulate $\forall E$ that does end at $(0 = 0)^i$, but it is one of the tedious cases of simulation, and not helpful for intuitions in an example. Still, $b = b$ is a natural translation of $0 = 0$, and its proof as shown here is also an intuitive imitation of $\forall E$. This suggests that the notion of interpretation, and so that of simulation, could be extended to include different translations, in case the languages \mathcal{L}_{T_1} and \mathcal{L}_{T_2} are quite alike (e.g., where constants can be interpreted directly as constants).

3.6.3. *Remarks on the criterion.* We here discuss some aspects of the derivation criterion worth mentioning. First, it matches the notion of equivalence for context theories well. Suppose T' definitionally extends T by a symbol S defined by φ . Suppose also that we have an interpretation $i_1 : T \rightarrow V$. Then we can naturally define $i_2 : T' \rightarrow V$ by setting $S^{i_2} = \varphi^{i_2} = \varphi^{i_1}$. Note that the definitional axiom of T' then becomes an interpreted tautology in V of the form $\forall \bar{x}(\delta_{\bar{x}} \rightarrow ((\varphi(\bar{x}))^{i_2} \leftrightarrow (\varphi(\bar{x}))^{i_1}))$. And, as $\varphi^{i_1} = \varphi^{i_2}$, this is also an interpreted tautology with respect to i_1 . This means that the derivation criterion resulting from i_2 is indistinguishable from the one resulting from i_1 , and shows some consistency between full ontological purity and secondary ontological purity: if we consider two theories to be equivalent as context theories for a certain theorem, then we see that they can naturally induce the same derivation criterion for purity of formal proofs in a third theory.¹⁸

¹⁸ Note, however, that there may be choices of S^i that do not exactly coincide with φ^i . For example, we could add to PA a symbol $E(x)$ that says that x is even. In PA, the defining

Furthermore, we remark that any formal proof that satisfies the derivation criterion for an interpretation $i : T \rightarrow V$ needs to prove the pure formulas from V itself. Additionally, strictly speaking, \mathcal{L}_V -quantifiers (even when relativized by δ) should be taken as ranging over the entire V -domain. Thus, the pure formulas of \mathcal{L}_T cannot refer exclusively to the surrogate content—rather, they illuminate or highlight the surrogate content within the content of V . Ideally, perhaps, we would like to have formal proofs of V refer to nothing but the surrogate content, to fully eliminate any extraneousness. However, the embedding in the content of V may well be the only way in which surrogate content is given meaning. Namely, the ‘pure formulas’ are built from primitives whose ontological meaning is properly given by the axioms of V ; and it is unclear whether they can be thought of as independently corresponding to the surrogate ontology (without the connection to V). On a slightly different note, because formal proofs satisfying the derivation criterion start with the fully powered V -axioms, they are allowed to refer to extraneous content before the derivation of the pure formulas. When calling these proofs pure, then, we restrict that statement to the part of such proofs that comes after the pure formulas. Still, perhaps we may also view the part before the pure formulas as clarifying the way in which the interpreting theory approaches the subject matter of the context theory, and showing how this material fits into the interpreting theory.

3.7. Criterion for secondary ontological purity of informal proofs. We presented an elaborate criterion that characterizes which formal proofs are secondarily ontologically pure. It tells us that any interpretation induces a sense of surrogate content and secondarily ontologically pure formal proofs. Like formal proofs, an informal proof should be secondarily ontologically pure if it only draws on the surrogate ontology of the theorem. As before, this will require a connection between the notions in an informal proof, and their interpretation in a surrogate ontology. Given an ontological context (O, φ_{T_1}, R) for a theorem, we may judge this by checking *whether* the notions in a proof are formalizable in T_1 or equivalently, given an interpretation $i : T_1 \rightarrow T_2$, *whether* they are formalizable in terms of the good formulas of \mathcal{L}_{T_2} . Since the notions in the informal proof can intuitively concern a very different ontology than O , the formalization in T_1 does not anymore have to be ‘natural’. We do require that there exists a formalization in T_2 that is natural, so that T_2 refers to the right ontology associated with the notions in the proof. However, given an interpretation, the formalization in terms of the good formulas in T_2 does *not* need to be natural (although of course it can be). This is because the good formulas will code the notions of an informal proof in a way that T_2 may not itself.

Still, we will know that the informal proof is concerned with the ontology of T_2 , and that it can be made sense of in terms of T_1 -surrogates, a version of the pure content. That is, we just need to know, given an informal proof, that it is possible to make sense of its notions in terms of the restricted surrogate ontology. If this is the case, secondary ontological purity holds. Thus, we will maintain the following definition.

formula could be $\varphi_1(x) = \exists y(x = y + y)$, or $\varphi_2(x) = \exists y(x = SS0 \cdot y)$. When a theory interprets PA by i , it is clear that the translations φ_1^i and φ_2^i will not be the same. Hence, one could theoretically pick φ_1 to define the definitional extension of PA, but pick φ_2^i as the interpreting translation of $E(x)$.

DEFINITION 3.7 (Criterion for secondary ontological purity of informal proofs). *Given an informal mathematical theorem corresponding to the ontological context (O, φ_{T_1}, R) , an informal proof of the theorem is secondarily ontologically pure if there exists a formalization of any notion in the proof in T_1 , if there exists a natural formalization of any notion in the proof in some theory T_2 , and if there exists an interpretation $i : T_1 \rightarrow T_2$.*

EXAMPLE 3.1 (Infinitude of Primes). *Consider again the arithmetical proof of IP, which we have established is fully ontologically pure. Trivially, this will be secondarily ontologically pure with respect to PA, for the identity interpretation $i : PA \rightarrow PA$. Thus, full ontological purity implies secondary ontological purity.*

Now consider Furstenberg’s topological proof of IP, which we established before as not fully ontologically pure. The notions used in the informal proof arguably correspond to an ontology of sets, which are used to construct a topology, arithmetic sequence, and even the integers and natural numbers. Thus, a suitable set theory could provide a natural formalization of these notions, say ZFC. The context theory for IP remains PA as before, and as argued in Section 2.4.2, Furstenberg’s proof has a formalization in PA (although not a natural one). Now it is easy to come up with an interpretation $i : PA \rightarrow ZFC$, where for instance $\delta(x) := x \in \omega$. As the topology used in IP is reducible to an ontology of natural numbers (recall the considerations of Example 2.2), it is also reducible to the surrogate ontology of set-theoretic numbers—and since a set-theoretic ontology is associated with the notions in the proof, this establishes secondary ontological purity.

By using interpretability results, secondary ontological purity results broaden our understanding of how disciplines of mathematics represent each other. They encourage us to see connections between formalizations of the same informal proof in different theories, and even within a theory. And, they reassure us that the notions we use in a proof have an ontologically (though not necessarily epistemically) harmless formalization. Instead of separating, for instance, the arithmetical and the topological informal proofs of IP, a secondary ontological purity result encourages us to delve more into their connections. Aside from ontological similarities, we may continue the comparison by looking at their proof strategies, such as in [10], and possibly connect these findings again to epistemic properties of the proofs.

3.8. Interaction between full and secondary ontological purity. We shortly discuss the interaction between the two types of purity. For formal proofs, it should be noted that fully ontologically pure and secondarily ontologically pure proofs of the same theorem intersect if a proof satisfies the derivation criterion within the context theory or a definitional extension. However, there are also fully ontologically pure proofs that do not satisfy the derivation criterion, as well as secondarily ontologically pure proofs that are not proven in the context theory or a definitional extension. Thus, the criteria for full and secondary ontological purity are independent, but may happen to overlap.

For informal proofs, we saw that full ontological purity always implies secondary ontological purity, as guaranteed by the identity interpretation from and to the context theory. The informal side just asks whether the theory under consideration naturally formalizes the proof, and subsequently asks for the existence of a formalization of the proof into the context theory and an interpretation of the context theory. The latter is just a box to check, after which we know the proof is formalizable in terms of a sense of surrogate content. Then an informal proof is secondarily ontologically pure. We see

that this differs on the formal side, because we can there more easily distinguish between different types of surrogate content, and emphasize that it is the specific formalization in terms of the surrogate content that is secondarily ontologically pure. However, just like on the formal side, there also are secondarily ontologically pure proofs that are not fully ontologically pure, if they are not naturally formalizable into the context theory.

3.9. Impurity of proof. Lastly, based on the criteria for purity that we have formulated so far, this section mentions some criteria for impurity of proof. Secondary impurity results show that there are ingredients of a proof that cannot be, in an acceptable way, represented by an ontology. Full impurity results show that a proof really properly leaves a (surrogate) ontology, and must go beyond representing the context theory.

3.9.1. Secondary impurity. Full ontological purity is based on a context theory (or multiple), definitional extensions and their ontology. Its negation gives rise to a secondary sense of impurity.

DEFINITION 3.8 (Criteria for secondary impurity). *Suppose we are given an ontological context (O, φ_{T_1}, R) for an informal theorem. An informal proof is secondarily impure if some notion A in the proof is not naturally formalizable in T_1 . A formal proof $\Gamma \vdash_{T_2} \varphi_{T_2}$ is secondarily impure if $T_2 \notin [T_1]$.*

Thus, secondary impurity tells us that the proof does not concern itself with the true ontological content of a theorem, but that it still might be representable by a surrogate version of this content. Note that the condition for formal proofs relies on the idea that $[T_1]$ contains *all* theories that capture the right ontology. This will not always be the case. Thus, secondary ontological impurity for formal proofs simply says we do not have the purity guarantee.

3.9.2. Full impurity. For full impurity, suppose again we have an ontological context (O, φ_{T_1}, R) for an informal theorem. Full impurity of informal proofs then negates the main aspect of secondary ontological purity.

DEFINITION 3.9 (Criterion for full impurity of informal proofs). *An informal proof is fully impure if some notion in the proof cannot be formalized in T_1 .*

Thus, full impurity tells us the ontology of the theorem is simply insufficient for capturing a proof, and subsequently any surrogate ontology is, too. Now, for a formal proof, we could say that it is fully impure if it does not satisfy the derivation criterion for any interpretation. But this does not fully capture it: while the derivation criterion guarantees purity, it does not exclude other pure proofs from existing. Thus, an all-encompassing guarantee for full impurity is hard to specify; it requires us to figure out when an interpreting theory uses its full strength ‘only to reach the interpreted theory’, but not to do anything meaningful in the proof itself. Instead, we give one case that we can make clear.

DEFINITION 3.10 (Criterion for full impurity of formal proofs). *Suppose we are given an informal theorem corresponding to the ontological context (O, φ_{T_1}, R) . Let $i : T_1 \rightarrow T_2$ be an interpretation. Then a natural deduction proof in T_2 of $\varphi_{T_1}^i$, where $\varphi_{T_1}^i$ instantiates the node v in the corresponding tree (V, E) , is fully impure if:*

- Any assumption instantiating a node a in an open branch is good.
- Any closed branch in (V, E) contains a node p instantiated by a pure formula.
- There is a final branch part $aE \dots Ev$ or $pE \dots Ev$ that uses an inference rule which violates the restriction of Definition 3.3.

This says that at some point in the proof, we do reach the point where we restrict to the interpreted theory. However, by violating one of the restrictions on proof rules, we leave this interpretation again during the proof, and thus leave a description of surrogate content by our notion of good formulas. We finish this section with two examples of informal proofs.

EXAMPLE 3.2 (Planar Desargues’s Theorem). *Consider Planar Desargues’s Theorem, with an ontology of planar geometrical notions (such as points and lines), and let the spatial axioms of Hilbert’s incidence and order axioms (Group I and II; see [16]) be the context theory (of course, other choices can be made). The argument here is simple, although it should be noted that it uses the unproven but highly likely other direction of Theorem 3.6, as mentioned in footnote 17: as the theorem is unprovable in the context theory, we expect it is unprovable in any other theory that interprets the context theory, when this theory is restricted to just the surrogate content. This would mean that any informal proof of the theorem that we do have, is not formalizable in the context theory, and is necessarily fully impure.*

EXAMPLE 3.3 (Infinitude of Primes). *Consider the topological proof of IP, which we saw before has secondary ontological purity. However, we see here that it also has secondary impurity, as a consequence of the result in Section 2.4.2 that it is not fully ontologically pure: we explained in Example 2.2 why we do not think that there is a natural formalization of all the notions occurring in the proof into PA. This emphasizes that ‘secondary ontological purity’ really lowers the purity level, as it can co-occur with an unnatural ontology.*

Thus, this nuances the ‘pure/impure’ distinction: we suggest there is full ontological purity, full impurity, and a level in between that exhibits properties of both purity and impurity. The incorporation of other notions of purity should be able to induce even more levels of purity on top of these.

§4. Conclusion. In this paper, we have supplied formal and informal proofs with a notion of full ontological purity based on ontological content, and a notion of secondary ontological purity, based on surrogate ontological content and structural content. For formal proofs, we suggest that (definitional extensions of) the context theory guarantee(s) full ontological purity, and that the satisfaction of a derivation criterion motivated by an interpretation of the context theory into another theory guarantees secondary ontological purity. For informal proofs, full ontological purity requires ‘natural’ formalizability in the context theory, while for secondary ontological purity, formalizability in the context theory together with natural formalizability into an interpreting theory suffices. Full and especially secondary ontological purity induce a more general notion of purity than traditional conceptions, matching more a structuralist way of thinking, and encouraging a more nuanced distinction between levels of purity.

Future research can take several directions. For one, the formal guarantee for secondary ontological purity relies on the idea that the full power of a theory is

only used to ‘get to’ the interpreted theory, and not for any notion essential to the proof. For this, our approach relies on the order of inference steps in natural deduction proofs. However, there may be other ways of determining whether ‘extraneousness’ of a theory is relevant in a meaningful way. This would lead to more encompassing guarantees for formal (im)purity.

Furthermore, there are several possibilities for extending the approach of this paper. We have not so far taken into account second- or higher-order theories, which may also provide a natural counterpart for content. Already within first-order theories, however, extensions of the derivation criteria may be found. An interpretation consists of a domain relativization and a translation function—but for theories that already have a very similar domain to the context theory, it would be more elegant to leave out a (trivial) relativization (e.g., when $\delta(x) := x = x$). Alternatively, when a theory differs only from the context theory in that it captures more objects (of a similar ontology), a domain relativization might be all that we need. This might also give rise to variants of the derivation criterion that induce ‘domain purity’ and ‘operational purity’, instead of purity in both aspects.

Finally, it has been suggested previously that impurity relates to simplicity and explanatoriness of proofs [2, 19, 24]. It would be interesting to see how full and secondary ontological purity interact with these properties.

§A. Appendix. We here show rigorously by induction on the length of a derivation that, given an interpretation $i : T_1 \rightarrow T_2$, any proof in T_1 can be simulated in a pure way in T_2 :

THEOREM A.1 *Let $i : T_1 \rightarrow T_2$ be an interpretation. Let \mathcal{D} be the proof $\Gamma \vdash_{T_1} \varphi$ in the classical first-order natural deduction calculus. Then $\Gamma^i, \delta_{\lambda_{\mathcal{D}}} \vdash_{T_2}^P \varphi^i$ by simulation.*

We will use the notation $(\varphi(t))^i$ when t is a non-variable term, and the notation $\varphi^i(t)$ when t is a variable. This emphasizes that, in the first case, t itself needs to be translated into a separate formula, while in the second case, we will map t to t itself. Furthermore, in order to make big natural deduction proofs readable we sometimes split them up: asterisks $*_j$ will indicate that this spot needs to be filled by the formula labeled by $(*_j)$, attaching two proofs together. Symbols \approx_j will indicate that this spot needs the formula labeled by (\approx_j) (and its proof) to be repeated there.

Before embarking on the proof of the full theorem, however, we need two lemmas for the case where φ was obtained by an application of $\forall E$, and where it was obtained by $\exists I$. These cases involve formulas $(\varphi(t))^i$, which cannot easily be manipulated since t^i occurs only at the depth of atomic formulas. We distinguish between two cases: for non-variable terms, we take a detour here for their simulation through $\exists x(\delta(x) \wedge \varphi^i(x) \wedge t^i(x))$. The case of variable terms is included at the end of the full proof, which is easier as it has no term translations.

LEMMA A.2 *Given an interpretation $i : T_1 \rightarrow T_2$, we have the following proofs for any T_1 -formula φ and non-variable term t :*

$$\begin{aligned} & (\forall x\varphi(x))^i \vdash_{T_2}^P \exists x(\delta(x) \wedge \varphi^i(x) \wedge t^i(x)) \\ & \exists x(\delta(x) \wedge \varphi^i(x) \wedge t^i(x)) \vdash_{T_2}^P (\exists x\varphi(x))^i \end{aligned}$$

Proof. The first is given by the following derivation, which uses totality. Let assumption [1] be $[\delta(y) \wedge t^i(y)]$.

$$\frac{\frac{\frac{[\forall x(\delta(x) \rightarrow \varphi^i(x))]}{\delta(y) \rightarrow \varphi^i(y)} \forall E \quad \frac{[1]}{\delta(y)} \wedge E}{\varphi^i(y)} \rightarrow E \quad \frac{[1]}{t^i(y)} \wedge E}{\varphi^i(y) \wedge t^i(y)} \wedge I \quad \frac{[1]}{\delta(y)} \wedge E}{\delta(y) \wedge t^i(y)} \wedge I}{\frac{\exists x(\delta(x) \wedge \varphi^i(x) \wedge t^i(x))}{\exists x(\delta(x) \wedge \varphi^i(x))} \exists I \quad \frac{\exists x(\delta(x) \wedge t^i(x))}{\exists x(\delta(x) \wedge \varphi^i(x) \wedge t^i(x))} \exists E^1}$$

The second proof is given by the derivation below. We use totality, uniqueness, and an equality axiom. Let [1] be $[\delta(w) \wedge t^i(w)]$, and let [2] be $[\delta(z) \wedge \varphi^i(z) \wedge t^i(z)]$.

$$\frac{\frac{[2]}{t^i(z)} \wedge E \quad \frac{[1]}{t^i(w)} \wedge E}{t^i_w} \wedge I \quad \frac{\frac{[2]}{\delta(z)} \quad \frac{[1]}{\delta(w)}}{\delta_{z,w} (\approx)} \quad \frac{\forall xy(\delta_{x,y} \rightarrow (t^i_{x,y} \rightarrow x =^i y))}{\delta_{z,w} \rightarrow (t^i_{z,w} \rightarrow z =^i w)} \forall E}{z =^i w} \rightarrow E \quad \frac{\forall xy(\delta_{x,y} \rightarrow (x =^i y \rightarrow (\varphi^i(x) \rightarrow \varphi^i(y))))}{\delta_{z,w} \rightarrow (z =^i w \rightarrow (\varphi^i(z) \rightarrow \varphi^i(w)))} \forall E}{z =^i w \rightarrow (\varphi^i(z) \rightarrow \varphi^i(w))} \rightarrow E^* \approx$$

$$\frac{* \quad \frac{[2]}{\varphi^i(z)} \wedge E}{\varphi^i(w)} \wedge E \quad \frac{[1]}{\delta(w)} \wedge E}{\delta(w) \wedge \varphi^i(w)} \wedge I}{\frac{\exists x(\delta(x) \wedge \varphi^i(x))}{\exists x(\delta(x) \wedge \varphi^i(x))} \exists I \quad \frac{\exists x(\delta(x) \wedge t^i(x))}{\exists x(\delta(x) \wedge \varphi^i(x) \wedge t^i(x))} \exists E^1} \exists E^2$$

□

The next lemma will complete the simulation of $\forall E$ and $\exists I$ for non-variable terms, but requires a more elaborate proof.

LEMMA A.3 *Given an interpretation $i : T_1 \rightarrow T_2$, we have the following proofs for any T_1 -formula $\chi(x)$ and non-variable term t :*

$$\begin{aligned} & (\chi(t))^i \vdash_{T_2}^P \exists x(\delta(x) \wedge t^i(x) \wedge \chi^i(x)) \\ & \exists x(\delta(x) \wedge t^i(x) \wedge \chi^i(x)) \vdash_{T_2}^P (\chi(t))^i. \end{aligned}$$

Proof. We use induction on χ .¹⁹ ‘IH’ will refer to relevant applications of the induction hypothesis.

- **Base case.** $\chi(t)$ is a predicate $R(t)$. By the (adapted) definition of an interpretation, $(R(t))^i$ is equal to $\exists x(\delta(x) \wedge t^i(x) \wedge F(R)(x))$. Since x is a variable, $F(R)(x)$ is exactly $R^i(x)$. So, by definition $(R(t))^i \leftrightarrow \exists x(\delta(x) \wedge t^i(x) \wedge R^i(x))$, which becomes a tautology instance.
- **Conjunction.** $\chi(t)$ is a formula $\varphi(t) \wedge \psi(t)$. We show both directions of the pure proof of $((\varphi(t))^i \wedge (\psi(t))^i) \leftrightarrow \exists x(\delta(x) \wedge t^i(x) \wedge \varphi^i(x) \wedge \psi^i(x))$.

Left-to-right direction. We use first-order substitution in the proof. The proof is split up into three parts. Let [1] be $[\delta(w) \wedge t^i(w) \wedge \varphi^i(w)]$, and let [2] be

¹⁹ Of the propositional connectives, for succinctness we only treat \wedge and \rightarrow , which together also cover \neg (which is defined as $\rightarrow \perp$) and \vee . Note that we do have \forall in our language, but we choose to not present this case here because its classical definition is covered by the other connectives, and this paper restricts to the classical first-order natural deduction system.

$[\delta(z) \wedge t^i(z) \wedge \varphi^i(z)]$.

$$\frac{\frac{\forall xy(\delta_{x,y} \rightarrow (t^i_{x,y} \rightarrow x =^i y))}{\delta_{z,w} \rightarrow (t^i_{z,w} \rightarrow z =^i w)} \forall E \quad \frac{\frac{[1]}{\delta(w)} \wedge E \quad \frac{[2]}{\delta(z)} \wedge E}{\delta_{z,w} (\approx)} \wedge I \quad \frac{\frac{[1]}{t^i(w)} \wedge E \quad \frac{[2]}{t^i(z)} \wedge E}{t^i_{z,w}} \wedge I}{\frac{t^i_{z,w} \rightarrow z =^i w}{z =^i w (*_1)} \rightarrow E} \rightarrow E$$

$$\frac{\frac{\forall xy(\delta_{x,y} \rightarrow (x =^i y \rightarrow \varphi^i(x) \rightarrow \varphi^i(y)))}{\delta_{z,w} \rightarrow (z =^i w \rightarrow \varphi^i(z) \rightarrow \varphi^i(w))} \forall E \quad \frac{\frac{[2]}{\varphi^i(z)} \wedge E}{\varphi^i(z) \rightarrow \varphi^i(w)} \rightarrow E}{\frac{z =^i w \rightarrow (\varphi^i(z) \rightarrow \varphi^i(w))}{\varphi^i(z) \rightarrow \varphi^i(w)} \rightarrow E} \rightarrow E \quad \frac{[1]}{\delta(w) \wedge t^i(w) \wedge \varphi^i(w) \wedge \psi^i(w) (*_2)} \wedge I$$

$$\frac{\frac{\frac{\frac{(\varphi(t))^i \wedge (\psi(t))^i}{(\varphi(t))^i} \wedge E}{\exists x(\delta(x) \wedge t^i(x) \wedge \varphi^i(x) \wedge \psi^i(x))} \exists I \quad \frac{\frac{(\varphi(t))^i \wedge (\psi(t))^i}{(\psi(t))^i} \wedge E}{\exists x(\delta(x) \wedge t^i(x) \wedge \varphi^i(x))} \exists E^1 \quad \frac{\frac{(\varphi(t))^i \wedge (\psi(t))^i}{(\psi(t))^i} \wedge E}{\exists x(\delta(x) \wedge t^i(x) \wedge \psi^i(x))} \exists E^1}{\frac{\exists x(\delta(x) \wedge t^i(x) \wedge \varphi^i(x) \wedge \psi^i(x))}{\exists x(\delta(x) \wedge t^i(x) \wedge (\varphi^i(x) \wedge \psi^i(x)))} \exists E^1} \exists I$$

Right-to-left direction. Let [1] be $[\delta(z) \wedge t^i(z) \wedge \varphi^i(z) \wedge \psi^i(z)]$. In the proof, the vertical dots indicate that this derivation is analogous to that of $(\varphi(t))^i$ in the left branch.

$$\frac{\frac{\frac{[1]}{\delta(z) \wedge t^i(z) \wedge \varphi^i(z)} \wedge E}{\exists x(\delta(x) \wedge t^i(x) \wedge \varphi^i(x))} \exists I \quad \frac{[1]}{\vdots} \exists I}{\frac{(\varphi(t))^i}{(\varphi(t))^i} \wedge I \quad \frac{(\psi(t))^i}{(\psi(t))^i} \wedge I} \wedge I \quad \frac{\exists x(\delta(x) \wedge t^i(x) \wedge \varphi^i(x) \wedge \psi^i(x))}{(\varphi(t))^i \wedge (\psi(t))^i} \exists E^1$$

- **Implication.** $\chi(t)$ is a formula $\varphi(t) \rightarrow \psi(t)$. We show both directions of the pure proof of $((\varphi(t))^i \rightarrow (\psi(t))^i) \leftrightarrow \exists x(\delta(x) \wedge t^i(x) \wedge (\varphi^i(x) \rightarrow \psi^i(x)))$.

Left-to-right direction. The proof uses totality. Let [1] be $[\varphi^i(w)]$, [2] be $[\delta(w) \wedge t^i(w)]$ and [3] be $[\delta(z) \wedge t^i(z) \wedge \psi^i(z)]$.

$$\frac{\frac{\frac{[2]}{\delta(w)} \wedge E \quad \frac{[1]}{\varphi^i(w)} \wedge E}{\exists x(\delta(x) \wedge t^i(x) \wedge \varphi^i(x))} \exists I \quad \frac{(\varphi(t))^i}{(\varphi(t))^i} \wedge I \quad \frac{(\psi(t))^i}{(\psi(t))^i} \wedge I}{\frac{(\varphi(t))^i \rightarrow (\psi(t))^i}{\exists x(\delta(x) \wedge t^i(x) \wedge \psi^i(x)) (*_1)} \rightarrow E} \rightarrow E$$

$$\frac{\frac{[2]}{t^i(w)} \wedge I \quad \frac{[3]}{t^i(z)} \wedge I}{\frac{\frac{[2]}{\delta(w)} \wedge E \quad \frac{[3]}{\delta(z)} \wedge E}{\delta_{w,z} (\approx)} \wedge I \quad \frac{\forall xy(\delta_{x,y} \rightarrow (t^i_{x,y} \rightarrow x =^i y))}{\delta_{w,z} \rightarrow (t^i_{w,z} \rightarrow w =^i z)} \forall E \quad \frac{\forall xy(\delta_{x,y} \rightarrow (x =^i y \rightarrow (\psi^i(x) \rightarrow \psi^i(y))))}{\delta_{w,z} \rightarrow (w =^i z \rightarrow (\psi^i(z) \rightarrow \psi^i(w)))} \forall E}{\frac{t^i_{w,z}}{w =^i z} \wedge I \quad \frac{t^i_{w,z} \rightarrow w =^i z}{w =^i z} \rightarrow E \quad \frac{w =^i z \rightarrow (\psi^i(z) \rightarrow \psi^i(w))}{\psi^i(z) \rightarrow \psi^i(w) (*_2)} \rightarrow E} \rightarrow E$$

$$\frac{\frac{\frac{\frac{\frac{\frac{[3]}{\psi^i(z)} \wedge E}{\psi^i(w)} \rightarrow E}{\psi^i(w)} \exists E^3}{\varphi^i(w) \rightarrow \psi^i(w)} \rightarrow I^1}{\frac{\delta(w) \wedge t^i(w) \wedge (\varphi^i(w) \rightarrow \psi^i(w))}{\exists x(\delta(x) \wedge t^i(x)) \wedge (\varphi^i(x) \rightarrow \psi^i(x))} \exists I} \frac{[2]}{\wedge I} \frac{\exists x(\delta(x) \wedge t^i(x))}{\exists x(\delta(x) \wedge t^i(x)) \wedge (\varphi^i(x) \rightarrow \psi^i(x))} \exists E^2$$

Right-to-left direction. Let [1] be $[\delta(w) \wedge t^i(w) \wedge \varphi^i(w)]$, [2] be $[\delta(z) \wedge t^i(z) \wedge (\varphi^i(z) \rightarrow \psi^i(z))]$, and [3] be $[(\varphi(t))^i]$.

$$\frac{\frac{\frac{\forall xy(\delta_{x,y} \rightarrow (t^i_{x,y} \rightarrow x =^i y))}{\delta_{w,z} \rightarrow (t^i_{w,z} \rightarrow w =^i z)} \forall E^*}{t^i_{w,z} \rightarrow w =^i z} \frac{\frac{[1]}{\delta(w)} \quad \frac{[2]}{\delta(z)}}{\delta_{w,z} (\approx)} \quad \frac{[1]}{t^i(w)} \quad \frac{[2]}{t^i(z)} \wedge I}{t^i_{w,z} \rightarrow E} \frac{w =^i z (*_1)}{w =^i z (*_1)}$$

$$\frac{\frac{\frac{\forall xy(\delta_{x,y} \rightarrow x =^i y \rightarrow (\varphi^i(x) \rightarrow \varphi^i(y)))}{\delta_{w,z} \rightarrow (w =^i z \rightarrow (\varphi^i(w) \rightarrow \varphi^i(z)))} \rightarrow E}{w =^i z \rightarrow (\varphi^i(w) \rightarrow \varphi^i(z))} \rightarrow I}{\varphi^i(w) \rightarrow \varphi^i(z)} \rightarrow I \approx \frac{[1]}{\varphi^i(w)} \wedge E}{\varphi^i(z)} \rightarrow E \frac{[2]}{\varphi^i(z) \rightarrow \psi^i(z)} \wedge E}{\psi^i(z) (*_2)} \rightarrow E$$

$$\frac{\frac{\frac{[2]}{\delta(z) \wedge t^i(z)} \wedge E}{\delta(z) \wedge t^i(z) \wedge \psi^i(z)} \wedge I}{\exists x(\delta(x) \wedge t^i(x) \wedge \psi^i(x))} \exists I}{(\psi(t))^i} \exists I \frac{\exists x(\delta(x) \wedge t^i(x) \wedge (\varphi(x))^i \rightarrow (\psi(x))^i)}{(\psi(t))^i} \exists E^2}{(\varphi(t))^i \rightarrow (\psi(t))^i} \rightarrow I^3 \frac{[3]}{\exists x(\delta(x) \wedge t^i(x) \wedge \varphi^i(x))} \exists I}{(\psi(t))^i} \exists E^1$$

- **Universal quantifier.** $\chi(t)$ is a formula $\forall x\varphi(x, t)$. We show both directions of the pure proof of $(\forall x(\delta(x) \rightarrow (\varphi(x, t))^i) \leftrightarrow \exists x(\delta(x) \wedge t^i(x) \wedge \forall y(\delta(y) \rightarrow \varphi^i(y, x)))$. The induction hypothesis will hold for each instance φ .

Left-to-right direction: The proof uses totality, uniqueness and first-order equality. [1] will stand for $[\delta(c)]$, [2] for $[\delta(h) \wedge t^i(h) \wedge \varphi^i(c, h)]$, [3] for $[\delta(e) \wedge t^i(e) \wedge \varphi^i(a, e)]$, and [4] for $[\delta(a) \wedge t^i(a)]$ (all the lowercase letters introduced here stand for variables).

$$\frac{\frac{[3]}{t^i(c)} \wedge E}{t^i_{c,h}} \wedge E \quad \frac{[2]}{t^i(h)} \wedge E}{t^i_{c,h}} \wedge E \quad \frac{\frac{\forall xy(\delta_{x,y} \rightarrow (t^i_{x,y} \rightarrow (x =^i y)))}{\delta_{c,h} \rightarrow (t^i_{c,h} \rightarrow (e =^i h))} \frac{[3]}{\delta(e)} \quad \frac{[2]}{\delta(h)}}{\delta_{c,h} (\approx)} \frac{\forall xy(\delta_{x,y} \rightarrow (x =^i y \rightarrow (\varphi^i(c, x) \rightarrow \varphi^i(c, y))))}{\delta_{c,h} \rightarrow (e =^i h \rightarrow (\varphi^i(c, h) \rightarrow \varphi^i(c, e)))} \approx}{e =^i h} \frac{t^i_{c,h} \rightarrow (e =^i h)}{e =^i h} \rightarrow E \quad \frac{\varphi^i(c, h) \rightarrow \varphi^i(c, e) (*_1)}{\varphi^i(c, h) \rightarrow \varphi^i(c, e) (*_1)} \rightarrow E$$

$$\begin{array}{c}
 \frac{\forall x(\delta(x) \rightarrow (\varphi(x, t))^i)}{\delta(c) \rightarrow (\varphi(c, t))^i} \forall E \quad \frac{[1]}{\rightarrow E} \quad \frac{[2]}{\wedge E} \\
 \frac{(\varphi(c, t))^i}{\exists v(\delta(v) \wedge t^i(v) \wedge \varphi^i(c, v))} \text{IH} \quad \frac{*_1 \quad \varphi^i(c, e)}{\varphi^i(c, e)} \rightarrow E \\
 \frac{\varphi^i(c, e)}{\delta(c) \rightarrow \varphi^i(c, e)} \rightarrow I^1 \\
 \frac{\delta(c) \rightarrow \varphi^i(c, e)}{\forall y(\delta(y) \rightarrow \varphi^i(y, e)) (*_2)} \forall I \\
 \\
 \frac{*_2 \quad \frac{[3]}{\delta(e) \wedge t^i(e)} \wedge E \quad \frac{\forall x(\delta(x) \rightarrow (\varphi(x, t))^i)}{\delta(a) \rightarrow (\varphi(a, t))^i} \forall E \quad \frac{[4]}{\delta(a)} \wedge E}{\delta(e) \wedge t^i(e) \wedge \forall y(\delta(y) \rightarrow \varphi^i(y, e))} \wedge I \quad \frac{(\varphi(a, t))^i}{\exists b(\delta(b) \wedge t^i(b) \wedge \varphi^i(a, b))} \text{IH} \\
 \frac{\exists x(\delta(x) \wedge t^i(x) \wedge \forall y(\delta(y) \rightarrow \varphi^i(y, x)))}{\exists x(\delta(x) \wedge t^i(x) \wedge \forall y(\delta(y) \rightarrow \varphi^i(y, x)))} \exists I \quad \frac{\exists b(\delta(b) \wedge t^i(b) \wedge \varphi^i(a, b))}{\exists x(\delta(x) \wedge t^i(x) \wedge \forall y(\delta(y) \rightarrow \varphi^i(y, x)))} \exists E^3 \\
 \frac{\exists x(\delta(x) \wedge t^i(x))}{\exists x(\delta(x) \wedge t^i(x) \wedge \forall y(\delta(y) \rightarrow \varphi^i(y, x)))} \exists E^5
 \end{array}$$

Right-to-left direction: Let [1] be $[\delta(b) \wedge t^i(b) \wedge \forall y(\delta(y) \rightarrow \psi^i(y, b))]$ and [2] be $[\delta(z)]$.

$$\begin{array}{c}
 \frac{[1]}{\delta(b) \wedge t^i(b)} \wedge E \quad \frac{\frac{[1]}{\forall y(\delta(y) \rightarrow \psi^i(y, b))} \wedge E}{\delta(z) \rightarrow \psi^i(z, b)} \forall E \quad \frac{[2]}{\rightarrow E} \\
 \frac{\delta(b) \wedge t^i(b)}{\exists x(\delta(x) \wedge t^i(x) \wedge \psi^i(z, b))} \exists I \\
 \frac{\exists x(\delta(x) \wedge t^i(x) \wedge \psi^i(z, b))}{(\psi(z, t))^i} \text{IH} \\
 \frac{(\psi(z, t))^i}{\delta(z) \rightarrow (\psi(z, t))^i} \rightarrow I^2 \\
 \frac{\delta(z) \rightarrow (\psi(z, t))^i}{\forall y(\delta(y) \rightarrow (\psi(y, t))^i)} \forall I \\
 \frac{\exists x(\delta(x) \wedge t^i(x) \wedge \forall y(\delta(y) \rightarrow \psi^i(y, x)))}{\forall y(\delta(y) \rightarrow (\psi(y, t))^i)} \exists E^1
 \end{array}$$

- **Existential quantifier.** $\chi(t)$ is a formula $\exists y\varphi(y, t)$. We show both directions of the pure proof of $\exists x(\delta(x) \wedge (\varphi(x, t))^i) \leftrightarrow \exists y(\delta(y) \wedge t^i(y) \wedge \exists x(\delta(x) \wedge (\varphi(x, y))^i)$.

Left-to-right direction: Let [1] be $[\delta(a) \wedge (\varphi(a, t))^i]$ and [2] be $[\delta(b) \wedge t^i(b) \wedge \varphi^i(a, b)]$.

$$\begin{array}{c}
 \frac{[1]}{\delta(a)} \wedge E \quad \frac{[2]}{\varphi^i(a, b)} \wedge E \\
 \frac{\delta(a) \wedge (\varphi(a, t))^i}{\exists y(\delta(y) \wedge \varphi^i(y, b))} \exists I \\
 \frac{\exists y(\delta(y) \wedge \varphi^i(y, b))}{\exists x(\delta(x) \wedge (\varphi(x, t))^i)} \exists E^1 \\
 \frac{[1]}{(\varphi(a, t))^i} \wedge E \quad \frac{[2]}{\delta(b) \wedge t^i(b)} \wedge E \\
 \frac{\exists x(\delta(x) \wedge t^i(x) \wedge \varphi^i(a, x))}{\exists x(\delta(x) \wedge t^i(x) \wedge \exists y(\delta(y) \wedge \varphi^i(y, x)))} \text{IH} \quad \frac{\delta(b) \wedge t^i(b) \wedge \exists y(\delta(y) \wedge \varphi^i(y, b))}{\exists x(\delta(x) \wedge t^i(x) \wedge \exists y(\delta(y) \wedge \varphi^i(y, x)))} \exists I \\
 \frac{\exists x(\delta(x) \wedge t^i(x) \wedge \exists y(\delta(y) \wedge \varphi^i(y, x)))}{\exists x(\delta(x) \wedge t^i(x) \wedge \exists y(\delta(y) \wedge \varphi^i(y, x)))} \exists E^2
 \end{array}$$

Right-to-left direction. Let [1] be $[\delta(c) \wedge \varphi^i(c, a)]$ and [2] be $[\delta(a) \wedge t^i(a) \wedge \exists y(\delta(y) \wedge \varphi^i(y, a))]$.

$$\begin{array}{c}
 \frac{[1]}{\varphi^i(c, a)} \wedge E \quad \frac{[2]}{\delta(a) \wedge t^i(a)} \wedge E \\
 \frac{\delta(a) \wedge t^i(a) \wedge \varphi^i(c, a)}{\exists x(\delta(x) \wedge t^i(x) \wedge \varphi^i(c, x))} \exists I \\
 \frac{\exists x(\delta(x) \wedge t^i(x) \wedge \varphi^i(c, x))}{(\varphi(c, t))^i} \text{IH} \quad \frac{[1]}{\delta(c)} \wedge E \\
 \frac{\delta(c) \wedge (\varphi(c, t))^i}{\exists y(\delta(y) \wedge (\varphi(y, t))^i)} \exists I \\
 \frac{\exists y(\delta(y) \wedge (\varphi(y, t))^i)}{\exists y(\delta(y) \wedge (\varphi(y, t))^i)} \exists I \\
 \frac{\exists y(\delta(y) \wedge (\varphi(y, t))^i)}{\exists y(\delta(y) \wedge (\varphi(y, t))^i)} \exists E^1 \\
 \frac{\exists x(\delta(x) \wedge t^i(x) \wedge \exists y(\delta(y) \wedge \varphi^i(y, x)))}{\exists y(\delta(y) \wedge (\varphi(y, t))^i)} \exists E^2
 \end{array}$$

□

Now we are ready to give the full proof of the theorem, which we repeat here.

THEOREM A.4 *Let $i : T_1 \rightarrow T_2$ be an interpretation. Let \mathcal{D} be the proof $\Gamma \vdash_{T_1} \varphi$ in the classical first-order natural deduction calculus. Then $\Gamma^i, \delta_{\lambda_{\mathcal{D}}} \vdash_{T_2}^P \varphi^i$ by simulation.*

Proof. We prove this by induction on the length of the derivation of $\Gamma \vdash_{T_1} \varphi$.

- **Base case.** φ is a T_1 -axiom or -assumption. By definition of an interpretation, there exists a (trivially secondarily ontologically pure) proof $\delta_\varphi \vdash_{T_2}^P \varphi^i$.
- **Case $\wedge I$.** Suppose φ equals $\chi \wedge \psi$ and the last step of \mathcal{D} was $\frac{\chi \quad \psi}{\chi \wedge \psi} \wedge I$. This means there are subderivations \mathcal{D}_1 referring to $\Gamma \vdash_{T_1} \chi$, and \mathcal{D}_2 referring to $\Gamma \vdash_{T_1} \psi$.²⁰ Then, by the induction hypothesis, we have pure simulations $\Gamma^i, \delta_{\lambda_{\mathcal{D}_1}} \vdash_{T_2}^P \chi^i$, and $\Gamma^i, \delta_{\lambda_{\mathcal{D}_2}} \vdash_{T_2}^P \psi^i$. Then we obtain the following pure simulation of the whole proof \mathcal{D} :

$$\frac{\frac{\frac{[\Gamma^i, \delta_{\lambda_{\mathcal{D}_1}}]}{\Gamma^i, \delta_{\lambda_{\mathcal{D}_1}}} \wedge E \quad \frac{[\Gamma^i, \delta_{\lambda_{\mathcal{D}_2}}]}{\Gamma^i, \delta_{\lambda_{\mathcal{D}_2}}} \wedge E}{\vdots} \quad \frac{\vdots}{\chi^i} \quad \frac{\vdots}{\psi^i}}{\chi^i \wedge \psi^i} \wedge I$$

This works by definition of the interpretation, as $\chi^i \wedge \psi^i = (\chi \wedge \psi)^i$. The case $\wedge E$ works similarly, which we omit.

- **Case $\vee I$.** Suppose φ equals $\chi \vee \psi$ and the last step of \mathcal{D} was $\frac{\chi}{\chi \vee \psi} \vee I$. This means there is a subderivation \mathcal{D}_1 referring to $\Gamma \vdash_{T_1} \chi$. By the induction hypothesis, we have a pure simulation $\Gamma^i, \delta_{\lambda_{\mathcal{D}_1}} \vdash_{T_2}^P \chi^i$. Then we obtain the following pure simulation of the whole proof \mathcal{D} :

$$\frac{\frac{[\Gamma^i, \delta_{\lambda_{\mathcal{D}_1}}]}{\Gamma^i, \delta_{\lambda_{\mathcal{D}_1}}} \wedge E}{\vdots} \quad \frac{\vdots}{\chi^i} \vee I^i}{\chi^i \vee \psi^i}$$

This works because of the interpretation, as $\chi^i \vee \psi^i = (\chi \vee \psi)^i$, and because certainly the introduction of ψ^i is valid according to the restricted rule $\vee I^i$.

- **Case $\vee E$.** Suppose φ was obtained by the $\vee E$ -rule, i.e.,

$$\frac{\frac{[\chi] \quad [\psi]}{\vdots} \quad \frac{\vdots}{\chi \vee \psi} \quad \frac{\vdots}{\varphi} \quad \frac{\vdots}{\varphi}}{\varphi} \vee E$$

. This means there are subderivations \mathcal{D}_1 referring

²⁰ \mathcal{D}_1 and \mathcal{D}_2 may only use a subset of Γ , but for easy notation we will use the entire set Γ , which will only require a conjunction elimination in the worst case, and we can easily transform a pure simulation using a subset of Γ^i into one that starts from Γ^i as a whole.

to $\Gamma \vdash_{T_1} \chi \vee \psi$, \mathcal{D}_2 referring to $\chi \vdash_{T_1} \varphi$ and \mathcal{D}_3 referring to $\psi \vdash_{T_1} \varphi$. By the induction hypothesis, we have pure simulations $\Gamma^i, \delta_{\lambda_{\mathcal{D}_1}} \vdash_{T_2}^P (\chi \vee \psi)^i$, $\varphi^i, \delta_{\lambda_{\mathcal{D}_2}} \vdash_{T_2}^P \varphi^i$ and $\psi^i, \delta_{\lambda_{\mathcal{D}_3}} \vdash_{T_2}^P \varphi^i$. Then we obtain the following pure simulation of the whole proof \mathcal{D} .

$$\frac{\frac{\frac{[\Gamma^i, \delta_{\lambda_{\mathcal{D}}}] \wedge E}{\Gamma^i, \delta_{\lambda_{\mathcal{D}_1}}} \quad \frac{\frac{[\delta_{\lambda_{\mathcal{D}}}] \wedge E}{\delta_{\lambda_{\mathcal{D}_2}}, \chi^i} \wedge I}{\delta_{\lambda_{\mathcal{D}_2}}, \chi^i} \quad \frac{\frac{[\delta_{\lambda_{\mathcal{D}}}] \wedge E}{\delta_{\lambda_{\mathcal{D}_3}}, \psi^i} \wedge I}{\delta_{\lambda_{\mathcal{D}_3}}, \psi^i} \wedge I}{\frac{(\chi \vee \psi)^i}{\varphi^i} \vee E} \quad \frac{\vdots}{\varphi^i} \quad \frac{\vdots}{\varphi^i} \vee E}{\varphi^i} \vee E$$

Again, this works because we may see $(\chi \vee \psi)^i$ in the premise of the rule as $\chi^i \vee \psi^i$, so that it is ready for a disjunction elimination rule to be applied to it. The cases of $\rightarrow I$ and $\rightarrow E$ work similarly, which we omit.

- **Case $\forall I$.** Suppose φ equals $\forall x \chi$ and the last step of \mathcal{D} was $\frac{\chi[x \setminus y]}{\forall x \chi} \forall I$. This means there is a subderivation \mathcal{D}_1 referring to $\Gamma \vdash_{T_1} \chi[x \setminus y]$. By the induction hypothesis, we have the pure simulation $\Gamma^i, \delta_{\lambda_{\mathcal{D}_1}} \vdash_{T_2}^P \chi^i(y)$. Then we obtain the following pure simulation of the whole proof \mathcal{D} . In this case, we can see that $\delta_{\lambda_{\mathcal{D}}} = \delta_{\lambda_{\mathcal{D}_1}}$.

$$\frac{\frac{\frac{[\Gamma^i, \delta_{\lambda_{\mathcal{D}}}] \quad \vdots}{\chi^i(y) \quad [\delta(y)]^1} \wedge I}{\chi^i(y) \wedge \delta(y)} \wedge E}{\frac{\chi^i(y)}{\delta(y) \rightarrow \chi^i(y)} \rightarrow I^1} \rightarrow I^1 \quad \frac{\delta(y) \rightarrow \chi^i(y)}{\forall x(\delta(x) \rightarrow \chi^i(x))} \forall I}{\chi(y)} \forall I$$

- **Case $\exists E$.** Suppose φ was obtained by the $\exists E$ -rule, i.e., $\frac{\frac{\vdots}{\exists x \chi(x)} \quad \varphi}{\varphi} \exists E$.

This means there are subderivations \mathcal{D}_1 referring to $\Gamma \vdash_{T_1} \exists x \chi(x)$ and \mathcal{D}_2 referring to $\chi(y) \vdash_{T_1} \varphi$. By the induction hypothesis, there exist pure simulations $\Gamma^i, \delta_{\lambda_{\mathcal{D}_1}} \vdash_{T_2}^P \exists x(\delta(x) \wedge \chi^i(x))$ and $\chi^i(y), \delta_{\lambda_{\mathcal{D}_2}} \vdash_{T_2}^P \varphi^i$. Then, in T_2 ,

$$\begin{array}{c}
 [\Gamma^i, \delta_{\lambda_{\mathcal{D}}}] \\
 \vdots \\
 \frac{\chi^i(t) \quad \frac{[\Gamma^i, \delta_{\lambda_{\mathcal{D}}}]}{\delta(t)} \wedge E}{\delta(t) \wedge \chi^i(t)} \\
 \hline
 \exists x(\delta(x) \wedge \chi^i(x))
 \end{array}$$

For *non-variable* terms, the simulation of the rule consists of:

$$\begin{array}{c}
 [\Gamma^i, \delta_{\lambda_{\mathcal{D}}}] \\
 \vdots \\
 (\chi(t))^i \\
 \hline
 \exists x(\delta(x) \wedge t^i(x) \wedge \chi^i(x)) \quad \text{Lemma A.3} \\
 \hline
 (\exists x \chi(x))^i \quad \text{Lemma A.2}
 \end{array}$$

□

Acknowledgements. I am grateful to Rosalie Iemhoff and Luca Incurvati for extensive discussions and comments on earlier versions of this paper. I also thank Amir Tabatabai, Albert Visser, Francesca Poggiolesi and three referees for their useful comments, as well as the audience at several conferences, most notably ‘LOGICA 2021’ and ‘The Third Workshop on Proof Theory and its Applications’.

Funding. I gratefully acknowledge the support of the Netherlands Organisation for Scientific Research under grant 639.073.807.

BIBLIOGRAPHY

- [1] Arana, A. (2009). On formally measuring and eliminating extraneous notions in proofs. *Philosophia Mathematica*, **17**(2), 189–207.
- [2] ———. (2017). On the alleged simplicity of impure proof. In *Simplicity: Ideals of Practice in Mathematics and the Arts*. Cham: Springer, pp. 205–226.
- [3] Arana, A., & Detlefsen, M. (2011). Purity of methods. *Philosophers’ Imprint*, **11**(2), 1–20.
- [4] Arana, A., & Mancosu, P. (2012). On the relationship between plane and solid geometry. *Review of Symbolic Logic*, **5**(2), 294–353.
- [5] Avigad, J. (2021). Reliability of mathematical inference. *Synthese*, **198**(8), 7377–7399.
- [6] Baldwin, J. T. (2013). Formalization, primitive concepts, and purity. *Review of Symbolic Logic*, **6**(1), 87–128.
- [7] Bolzano, Bernard (1817). *Die drei Probleme der Rectifikation, der Complonation und der Cubierung, ohne Betrachtung des unendlich Kleinen, ohne die Annahme des Archimedes und ohne irgend eine nicht streng erweisliche Voraussetzung gelöst; zugleich als Probe einer gänzlichen Umgestaltung der Raumwissenschaft allen Mathematikern zur Prüfung vorgelegt*. Leipzig: Gotthelf Kummer.

[8] Burge, T. (2005). *Truth, Thought, Reason: Essays on Frege*, Vol. 1. Oxford: Oxford University Press on Demand.

[9] Buss, S. R. (1998). *Handbook of Proof Theory*. Amsterdam and New York: Elsevier.

[10] Carlson, N. A. (2014). A connection between Furstenberg's and Euclid's proofs of the infinitude of primes. *The American Mathematical Monthly*, **121**(5), 444.

[11] Dawson, J. W. (2006). Why do mathematicians re-prove theorems? *Philosophia Mathematica*, **14**(3), 269–286.

[12] Detlefsen, M. (2008). Purity as an ideal of proof. In *The Philosophy of Mathematical Practice* (online edition). Oxford: Oxford Academic, pp. 179–197.

[13] Friedman, H. M., & Visser, A. (2014). When bi-interpretability implies synonymy. *Logic Group Preprint Series*, **320**, 1–19.

[14] Furstenberg, H. (1955). On the infinitude of primes. *American Mathematical Monthly*, **62**(353), 286.

[15] Ganea, M. (2009). Arithmetic on semigroups. *Journal of Symbolic Logic*, **74**(1), 265–278.

[16] Hallett, M. (2007). Reflections on the purity of method in Hilbert's *Grundlagen der Geometrie*. In *Grundlagen der Geometrie* (online edition). Oxford: Oxford Academic, pp. 198–255.

[17] Hipolito, I., & Kahle, R. (2019). Discussing Hilbert's 24th problem. *Philosophical Transactions of the Royal Society A*, **377**(2140), 20180040.

[18] Hodges, W. (1993). *Model Theory*. Cambridge: Cambridge university press.

[19] Iemhoff, R. (2017). Remarks on simple proofs. In *Simplicity: Ideals of Practice in Mathematics and the Arts*. Cham: Springer, pp. 143–151.

[20] Incurvati, L., & Nicolai, C. (2020). On logical and scientific strength. Unpublished manuscript. Available from: <https://philarchive.org/archive/INCOLA>.

[21] Isaacson, D. (1987). Arithmetical truth and hidden higher-order concepts. In *Logic Colloquium '85: Proceedings of the Colloquium held in Orsay, France, July 1985*. Studies in Logic and the Foundations of Mathematics, Vol. 122. Amsterdam–New York–Oxford–Tokyo: North-Holland, pp. 147–169.

[22] Kahle, R., & Pulcini, G. (2017). Towards an operational view of purity. In *The Logica Yearbook*. London: College Publications, pp. 125–138.

[23] Kaye, R., & Wong, T. L. (2007). On interpretations of arithmetic and set theory. *Notre Dame Journal of Formal Logic*, **48**(4), 497–510.

[24] Lange, M. (2019). Ground and explanation in mathematics. *Philosophers' Imprint*, **19**(33), 1–18.

[25] Lavine, S. (2000). Quantification and ontology. *Synthese*, **124**, 1–43.

[26] Lehet, E. (2021). Impurity in contemporary mathematics. *Notre Dame Journal of Formal Logic*, **62**(1), 67–82.

[27] Maddy, P. (2019). What do we want a foundation to do? In *Reflections on the Foundations of Mathematics*. Cham: Springer, pp. 293–311.

[28] McCarthy, T. (2021). Induction, constructivity, and grounding. *Notre Dame Journal of Formal Logic*, **62**(1), 83–105.

[29] Ryan, P. J. (2021). Szemerédi's theorem: An exploration of impurity, explanation, and content. *Review of Symbolic Logic*, **16**(3), 700–739.

[30] Shapiro, S. (1997). *Philosophy of Mathematics: Structure and Ontology*. Oxford: Oxford University Press on Demand.

[31] ———. (2008). Identity, indiscernibility, and ante rem structuralism: The tale of I and -I. *Philosophia Mathematica*, **16**(3), 285–309.

[32] Swoyer, C. (1991). Structural representation and surrogate reasoning. *Synthese*, **87**, 449–508.

[33] Turner, R. (2010). Programming languages as mathematical theories. In *Thinking Machines and the Philosophy of Computer Science: Concepts and Principles*. Hershey: IGI Global, pp. 66–82.

[34] Visser, A. (1997). *An Overview of Interpretability Logic*. Logic Group Preprint Series, Vol. 174. Utrecht: Department of Philosophy, University of Utrecht

DEPARTMENT OF PHILOSOPHY AND RELIGIOUS STUDIES

UTRECHT UNIVERSITY

JANSKERKHOF 13

3512 BL UTRECHT, NETHERLANDS

E-mail: r.a.martinot@uu.nl