Formal and Natural Proof - A phenomenological approach

Merlin Carl

Abstract

It is frequently claimed (see e.g. [Rav]) that the formalization of a mathematical proof requires a quality of understanding that subsumes all acts necessary for checking the proof and that, consequently, automatic proof checking cannot lead to an epistemic gain about a proof. We present a project developing what is sometimes called a ‘fortified formalism’ and argue, taking a phenomenological look at proof understanding, that proofs can be (and often are) given in a way that allows a formalization sufficient for producing an automatically checkable write-up, but does not subsume checking.

1 Introduction

It is a striking consequence of Gödel’s completeness theorem (see e.g. chapter 3 of [Rau]) that, whenever there is a correct mathematical proof of a certain sentence ϕ from any fixed set S of axioms stated in first-order predicate calculus, there is also a formal derivation of ϕ in the sense of a system of formal deduction.[1] It is this force of the completeness theorem that makes the study of formal proofs relevant to mathematical practice, as it demonstrates a certain kind of adequacy of formal proofs as a model of normal mathematical arguments. This adequacy, however, is rather weak: The only guarantee is that the set of correctly provable assertions coincides with the set of formally derivable assertions. No claim is made on the relation between mathematical arguments and formal derivations.

[1] The reasoning here is roughly this: If ϕ is not formally derivable from S, then there is a model of $S + \neg\phi$, i.e. a way to interprete the occuring notions in such a way that S becomes true, but ϕ becomes false. In the presence of such an interpretation, no argument claiming to deduce ϕ from S can be conclusive. A similar point is made by Kreisel in [Kr].

There are in fact various reasons to assume that this relation cannot be too close: For example, Boolos ([Bo]) constructed an example of a statement ϕ that is easily seen to be derivable in first-order logic from a set S of premises, but all such derivations are provably of such a vast length that they cannot possibly be actually 'written down' in any sense of the word. On an even more fundamental level, as most interesting theories like Peano arithmetic or ZFC set theory are incomplete and there are statements rather canonically associated with them (i.e. consistency) that might be used in a natural argument, one might even doubt how far carrying out natural mathematics or even merely particular areas as number theory in a fixed axiomatic system can work in principle.

On the other hand, there is the program of formal mathematics, where mathematics is actually carried out in a strictly formal framework. This has now been done for a huge variety of important and non-trivial theorems, including e.g. the Gödel completeness theorem (see [FM1]), Brouwer's fixed point theorem ([FM2]) and the prime number theorem ([FM3]). There is even a 'Journal of Formalized Mathematics', entirely dedicated to completely formal proofs. Such proofs are usually done by formalization, i.e. a translation or re-formulation of mathematical arguments in a formal system. The success of this approach suggests that, in spite of the objections mentioned above [2], formal proofs can be adequate to mathematical arguments in a stronger sense; namely, that correct arguments can be translated into derivations. However, this process of translation is often highly nontrivial and in many cases, the essence of an argument seems to get lost in translation. For any non-formalist view of mathematics, this seems inevitable: If mathematical arguments have content and are about 'objects', then this essential relation to 'objects' must be lost when one passes to formal derivations, which are void of content. It is hence a crucial question for the philosophy of mathematics to determine the relation between arguments and derivations.

In [Az1] and [Az2], Jody Azzouni considers this question while he searches for an explanation for what he calls the 'benign fixation of mathematical practice', namely the fact that mathematics, considered as a social practice, is remarkably stable when compared to other social practices such as art, religion, politics, philosophy and even the natural sciences. The notion of

[2] For example, Boolos' proof has recently been formalized, see [BB]. The matter appears to be more a question of the choice of the formal system than one of formal vs. informal proof or of first-order vs. higher-order logic: When a proof is 'possibe, but far too long' in a certain first-order axiomatic system, there are usually natural stronger systems that allow for the necessary abbreviations.

a mathematical proof appears to be particularly invariant: While the standards for what can count as evidence in, say, physics or biology have considerably changed over the last 2000 years, we can still evaluate an argument from e.g. Euclidean geometry and agree on its correctness. Even where the practice splits, for example into a classical and an intuitionistic branch, this agreement is not lost: For the intuitionist mathematician, a valid classical argument may seem invalid from his standpoint, yet he will usually be able to distinguish it from a classically invalid argument. Similarly, one doesn't need to become an intuitionist to see whether an argument is intuitionistically valid.

Azzouni's explanation, which he labels the 'derivation indicator view', or DI-view, or mathematical practice, goes roughly as follows: There is a notion of proof, namely formal proofs in one or another setup, that allows for a purely mechanical proof-check. That is, the correctness of a proof given in this form can be evaluated by simply processing the symbols of which it consists according to a certain algorithm. Since any two persons (and, in fact, a trained monkey or even a computer) applying this algorithm will obtain the same result, this explains the broad agreement at least for formal proofs.

But proofs as they appear in mathematics are virtually never formal proofs in a certain proof system. In fact, formal proofs but for the most trivial facts tend to become incomprehensible to the human reader. What we find in textbooks are arguments presented in natural language, mixed with formal expressions, diagrams, pictures etc. Checking those is not a mechanical procedure; rather, it requires careful concentration in carrying out the indicated mental steps in one's mind, while questioning every step, sustaining it if possible and rejecting it otherwise. The question hence arises how we account for the broad agreement on proofs presented in this manner.

Azzouni's answer is that such proofs, while not formal themselves, 'indicate', 'point to' formal proofs. They are to be considered as recipes for producing a fully formal version of an argument. This indication is clear to us in the same way it is, e.g., clear to us how a cooking recipe is to be transformed into a series of muscle movements in our kitchen. The notion of a formal proof is here independent from the choice of one or another concrete system of representation; rather, it is a form of proof in which every step is a single inference according to some valid deduction rule. This concept is prior to the development of actual representations for formal proofs and could well have been intended by mathematicians in the era before formal logic systems were introduced. In particular, in such a proof, every reference to an imagination of the concepts used can be put aside. We can see that 'If all zunks are zonks, and Jeff is a zunk, then Jeff is a zonk.' is true without knowing what zunks and zonks are or who Jeff is. The subjective component of the argument is hence eliminated as far as possible (all that remains is observing finite sign configurations) and this is the reason for the wide agreement.

In [Rav], Yehuda Rav objects to this view with an argument that I want to summarize as follows: Formal proofs cannot provide a basis for the explanation of our agreement on the correctness of proofs. This agreement is based on understanding. Once a proof is transformed into a form in which it is algorithmically checkable, it must be void of content: all contributions of our understanding must have entered the formalization as additional symbol strings. To do this, the argument must have been clarified to the last extent. Hence, at the moment where an algorithmically checkable proof is obtained, the 'battle is over', i.e. the checking is already finished as far as human understanding is concerned: The interesting work is done exactly along the way of formalizing the argument, and this process is non-algorithmical. It is based on an understanding of the occuring concepts, it has an 'irreducible semantic content'. Therefore, carrying out the algorithmic checking for the (formalized) argument will not result in any epistemic gain concerning the (original informal) argument. In particular, it does not strengthen the position that the (original) argument is valid. It might show us that we made some mistake in the 'exercise' of rewriting the proof in a formal system, but that tells us nothing about the proof itself, just as, in programming, an implementation mistake tells us nothing about the correctness of the algorithm we had in mind. Concerning the derivation-indicator view, this implies that it fails to explain the consensus about mathematics: For the consensus about formal proofs via algorithmic checking procedures is of no help unless we explain consensus about the relation between the natural argument and its 'formalization', for which no algorithmic checking procedure is at hand.

This argument has certainly a good degree of persuasive power. We want to evaluate this criticism closer. First, an epistemic gain through automatic proof checking is indeed possible when the latter is used as a means of communication: Even if the inventor of a proof would not learn anything new about the proof by having it automatically checked, the automatic checking can serve as a certificate for the correctness of the result for others. An attempt to communicate an otherwise hardly accessible proof by this means is the work on the formalization of Thomas Hales' proof of the Kepler conjecture in the Flyspeck project, which was recently announced to be completed (see e.g. [FS]). Focusing on the inventor of a proof himself, it is clear that indeed a lot can be learned about a proof through the process of formalization; being forced to work according to the outermost standards of precision, one is more likely to spot mistakes that otherwise evaded one's attention.

But neither of these ways to obtain an epistemic gain from automatic proof checking concerns the point made by Rav. The question is then, more precisely: Can the actual process of automatic checking itself (in contrast to the production of an automatically checkable format) lead to an epistemic gain about a proof that one already knows (e.g. by being its originator)? This is what Rav's criticism is about.

On the surface, we claim that the image of an 'algorithmic system' underlying Rav's argument is too narrow: It falls short of taking into account e.g. methods for automatic language processing or the possibility of using an automatic theorem prover for bridging gaps. But our main intention is deeper: We want to examine when and how an algorithmic system may lead to an epistemic gain about a natural mathematical argument.

When we talk about gaining trust in a proof, we have obviously left the realm where one can consider a proof as mere text or string of symbols; we have to take into account our attitude towards the proof, the way it is given to us or it presents itself to us. The question concerning the epistemic gain should then be reformulated as follows: 'Is there a state of mind towards a proof that allows the construction of an automatically checkable write-up, but is undecided about the correctness of the proof?'

This formulation makes it obvious that the question can't be decided by merely considering mathematical texts of different degree of formalization. Rather, the representation of a proof in its reader's consciousness has to be taken into account: It will e.g. be relevant whether the reader only briefly skimmed through it or studied it thoroughly, whether he worked the missing steps out or merely granted them, whether a cited result is applied with understanding or merely as a 'black box' etc. Such differences are idealized away in most approaches to logic. One of the rare approaches to logic which seriously takes into account such aspects is found in Husserl's 'Formal and transcendental logic' ([Hu1]). This motivates us to chose this work as a starting point for our investigation.

Our approach here is hence to analyze to a certain (humble) extent the phenomenology of proof understanding. We will distinguish two qualities of the way how a proof can be mentally represented, applying the approach of Husserl's analysis of judgements, in particular his notions of 'distinctiveness' and 'clarity', to proofs. We will argue that, if a mental representation of a proof has both of these qualities to a maximal extent, then Rav is right in claiming that the 'battle is over' and an algorithmic proof check cannot lead to an epistemic gain. On the other hand, we claim that only distinctiveness is necessary for putting an argument into a form that can be subjected to an automatic proof check. Therefore, we obtain a margin in which automatic proof checking can indeed give substantial information on the correctness of an argument: namely if the proof is mentally present in a distinct, but unclear manner. Considering several examples from the history and the folklore of mathematics, we demonstrate that this tends to occur frequently in mathematical practice.

To sustain our claims and make them more concrete, we will, in the course of this paper, refer to Naproche, a system for the automatic checking of natural mathematical arguments. Its aim is exactly, as in Rav's words, 'to do the work of automatic checking even an informal proof' and already in its current form it gives a vivid picture of the surprisingly natural form an automatically checkable proof can take. We will therefore start by shortly introducing the Naproche system in the next section.
In section 3, we explain the distinction between distinctiveness and clarity, using several examples. In section 4, we demonstrate that automatic checking requires the former, but not the latter quality, again giving examples. In section 5, we argue that clarity and distinciveness correspond in a certain way to a 'complete', 'gapless' derivation as they are represented in formats like natural deduction or the sequent calculus. The goal is to show which features of a mental representations of a proof are expressed in such a derivation. In section 6, we analyze a famous historical example of a false proof in these terms, considering whether or not and how a Naproche-like system might have helped to spot the mistake. Section 7 contains a critical review of our account, suggesting various ways in which the use of automatic proof checking is limited. Finally, we give in section 8 our conclusions and plans for future considerations on the topic.

2 Naproche

Naproche is an acronym for NAtural language PROof CHEcking, a joint project of mathematical logicians from Bonn, formal linguists from Duisburg-Essen and computer scientists from Cologne. It is a study of natural mathematical language with the goal to bridge the gap between formal derivations and the form in which proofs are usually presented. For this, the expressions of natural mathematical language are interpreted as indicators for certain operations, like introducing or retracting an assumption, starting a case distinction, citing a prior result, making a statement etc. We will describe only very roughly how this system works, as the details are irrelevant for our purpose. The interested reader may e.g. consult [CKKS] for a detailed description. Also, more information and a web interface are available at [NWI].

In the course of the project, a controlled natural language (CNL) for mathematics is developed, which resembles natural mathematical language and is constantly expanded to greater resemblance. This Naproche CNL contains linguistic triggers for common thought figures of mathematical proofs. Texts written in the Naproche CNL are hence easy to write and usually immediately understandable for a human reader. If one was presented with a typical Naproche text without further explanation, one would see a mathematical text, though one in a somewhat tedious style.
Here is an excerpt from a short text about number theory in the Naproche CNL, accepted by the current Naproche version 0.47:

Definition 29: Define m to divide n iff there is an l such that $n = m \cdot l$.

Definition 30: Define $m|n$ iff m divides n.

Lemma DivMin: Let $l|m$ and $l|m+n$. Then $l|n$.
Proof: Assume that l and n are nonzero. There is an i such that $m = l \cdot i$. Furthermore, there is a j such that $m + n = l \cdot j$.
Assume for a contradiction that $j < i$. Then $m+n = l \cdot j < l \cdot i = m$. So $m \leq m+n$. It follows that $m = m + n$. Hence $n = 0$, a contradiction. Thus $i \leq j$.
Define k to be $j - i$. Then we have $(l \cdot i) + (l \cdot k) = (l \cdot i) + n$. Hence $n = l \cdot k$. Qed.

Via techniques from formal linguistics, namely an adapted version of discourse representation theory (see [KR]), the content of such texts can be formally represented in a format that mirrors its linguistical and logical structure. This format is called a proof representation structure (PRS). In particular, whenever a statement is made, it can be computed from the PRS whether this is supposed to be an assumption or a claim and, in the latter case, under what assumptions this claim is made. In this way, the text is converted into a series of proof goals, each asking to deduce the current claim made in the proof from the available assumptions. The Naproche system then uses automatic theorem provers to test whether the claim indeed follows in an obvious way from the available assumptions. This allows the system to close the gaps that typically appear in natural proofs, one of the crucial features in which natural proofs differ from formal derivations. In this way, every claim is checked and either deduced (and accepted) or not, in which case the checking fails and returns an error message indicating the first claim where the deduction could not be processed.

3 Intentions, Fulfillment, Clarity and Distinctiveness

In this section, we introduce Husserl's notions of the distinctiveness and clarity of a judgement as a starting point for our transfer of these concepts to proofs. Crucial for the difference between clarity and distinctiveness is the notion of 'fulfillment' of an intention. We therefore start with a brief introduction to this notion.

3.1 Fulfilled and unfulfilled intentions

A central notion of phenomenology is the notion of intention, i.e. the directedness towards something. Whatever this something is, it must, according to Husserl, correspond to a possible way of presenting itself in some kind of experience. Here, 'experience' is taken in a very broad sense, including sensual experience in the usual sense of seeing, hearing etc., but not limited to it: E.g. remembering, imagining, reading a mathematical proof etc. count as legitimate forms of experience. The intention towards X is hence associated with a system of experiences in which X appears. What kind of experience is relevant for a certain intention depends on - or rather strictly corresponds to - the type of the intended object: A piece of music will present itself in hearing, a phantasm will present itself in imagination, a sensual object - say, a tree - will present itself in organized visual, tactile etc. perception. Fulfillment of an intention now simply means that experiences presenting the intended object are made: The piece is heard, the phantasm imagined, the tree seen and felt. Fulfillment may be partial, and in fact, for many types of objects, it will be necessarily so: For example, the intention towards a tree includes anticipations that it can be seen from all sides, including rough anticipations what it will look like. Thus, seeing only one side of a tree, the fulfillment of the intention is only partial: There are anticipations corresponding to the object type that remain unfulfilled. Even if one has walked completely around the tree, this does not change, as now the backside is not given in visual or tactile experience, but only in memory thereof. Proceeding along these lines, it is not hard to see that intentions towards physical objects are necessarily only partially fulfilled.

In the case of a jugdement, the intention is directed towards a categorical object, i.e. the state of affairs expressed by the judgement, i.e. in the case 'The rose is red' the fact that the rose is red. In this case, some of the appearing partial intentions from which the judgement is build up (e.g. that towards a rose or towards redness) may be fulfilled while others remain unfulfilled:

We may e.g. experience a rose, but not its redness (say under bad lighting). In the case of a work of fiction, we may form a vivid imagination of some of the described objects, but merely skip over the others, leaving the intention 'signitive', merely indicated by a symbol.
We want to apply the concept of fulfillment in the context of mathematical proofs. Our main concern, then, are intentions towards mathematical objects and their fulfillment. One might worry at this point that this will force us to accept some mysterious supernatural faculty for seeing abstract objects, but this worry falls short of taking into account the flexibility of the phenomenological treatment described above: Fulfillment of an intention means having the corresponding experiences. What these experiences are would be the topic of a phenomenological investigation of mathematical objects. Luckily, our treatment does not require the prior execution of such a monumental task, as we will only be concerned with mathematical objects in a very special context. Still, we indicate here two examples of possible interpretations that are hardly 'mysterious': The first would be the intuitionistic standpoint, taken e.g. by Becker and Heyting, according to which the intention towards a mathematical object is fulfilled by a construction of that object in the sense of intuitionism. Another interpretation is that of mathematical objects as 'inference packages': The intention towards a mathematical object is directed towards a system of techniques how to deal with inferences that contain this notion. The fulfillment of such an intention - e.g. in the course of a proof - would hence be the application of these techniques for the full explication of a proof step. An account of mathematical objects in this spirit, though apparently not with Husserl in mind, can be found in [Az3]; the treatment on p. 106-111 of [Lo] is of a similar spirit.[3]

This indicates how the notion of fulfillment can be applied to mathematical proofs. Consider the inferential snippet 'As $1 < a < p$ and p is prime,

[3] Another related perspective is that of Martin-Löf given in [ML2] (in particular on p. 7): He distinguishes 'canonical' from 'noncanonical' or 'indirect' proofs, where a 'noncanonical proof' is a 'method' or a 'program' for producing a 'canonical' proof; as an example, an indirect proof that $123^5 + 5^{123} = 5^{123} + 123^5$ would consist in first proving the general law of commutativity for addition and then instantiating it accordingly rather than carrying out the constructions described by both sides of the equation and checking that they actually lead to the same result. In a similar way, we may view an informal high-level argument as a recipe for obtaining a proof in which every formerly implicit inferential step is actually carried out. Of course, we don't need to go along with Martin-Löf's constructivist approach concerning mathematics here: The checking of non-constructivistic proofs is - regardless of how one views their epistemological value - a cognitive act which is, along with the underlying notion of correctness and its relation to automatization, accessible to a phenomenological analysis.

a does not divide p'. We can skim through it, leaving the empty intention to use the primeness of p to see why the result holds. Fulfilling the intention would amount to actually seeing it, i.e. completing the proof.

3.2 Husserl's Notions of Distinctiveness and Clarity

In [Hu1] and [Hu2], Husserl offers a phenomenological analysis and foundation for logic. As the notions in question have their systematic place in this analysis, we give a short recapitulation.

At the beginning, logic is taken in the traditional sense as the study of the forms of true judgements. It soon becomes apparent that, in the way this is traditionally done, numerous implicit idealizations are presupposed concerning the judgement and the modus in which it is given. These idealizations are made explicit. In the course of this explication, logic quite naturally splits into several subsections depending on the stage of idealization assumed. It turns out that most of these subsections are not considered by traditional logic, which is concerned with what finally turn out to be distinctly given judgements to which we are directed with epistemic interest. Furthermore, in the study of the abstract forms of judgements, the extra assumption is made that the referents occuring in the jugdement forms considered are to be interpreted in a way making the statement meaningful. [4] [5]

The first subsection is what Husserl calls the purely logical grammar ('rein logische Grammatik'), i.e. the mere study of forms that can possibly be a judgement at all in contrast to arbitrary word sequences like 'and or write write', ignoring all connections with truth. This part will not concern us further.

A terminological distinction made in various places in Husserl's work is that between 'distinctiveness' and 'clarity'. We find it e.g. in chapter 20 of [Hu4] with respect to concepts: While 'distinctiveness' means having explicated what one means with a certain concept, 'clarity' brings about an intuition of the intended object. This is explained on page 101 of [Hu4]:

'Die Verdeutlichung des Begriffs, des mit dem Wort Gemeinten als solchen, ist eine Prozedur, die sich innerhalb der bloßen Denksphäre abspielt. Ehe der mindeste Schritt zur Klärung vollzogen ist, während keine oder eine ganz unpassende und indirekte Anschauung mit dem Worte eins ist, kann

[4] E.g. the statement 'The theory of relativity is green' is arguably neither true nor false, which nevertheless doesn't contradict the principle of the excluded middle.

[5] See [Lo] for a further discussion of this point.

> überlegt werden, was in der Meinung liegt, z.B. in 'Dekaeder': ein Körper, ein Polyeder, regelmäßig, mit zehn kongruenten Seitenflächen. (...) Bei der Klärung überschreiten wir die Sphäre der bloßen Wortbedeutungen und des Bedeutungsdenkens, wir bringen die Bedeutungen zur Deckung mit dem Noematischen der Anschauung (...)' [6]

A parallel distinction concerning judgements instead of concepts is then introduced in [Hu1]:
A judgement is given in a `distinct` manner when its parts and their references to each other are made explicit. The intentions indicated by its parts may remain unfulfilled, but the compositional structure of the partial intentions is apparent. Given a jugdement of the form 'S is p', we can, going along with the formulation, 'carry out' the judgement by explicitely setting S and applying p to it without fulfilling the intentions corresponding to S or p. This explication shows that certain intentions such as those of the form $P \wedge \neg P$ are inexecutable in principle. The subdiscipline of logic concerned with this kind of givenness is 'consequence logic' which considers the executability of a judgement in principle, based on its mere structure, without regard to 'facts'.
`Clarity`, on the other hand, is obtained when the indicated intentions are 'fulfilled', e.g. the objects under consideration are brought to intuition. This may still lead to falsity and absurdity, but these are then of a semantic nature, not apparent from the mere form of the judgement. Of course, both clarity and distinctiveness come in degrees and can be present for certain parts of a judgement, but not for others.
A crucial point of the analysis is that the inexecutability of intentions indicated by certain distinctively given judgements already makes certain assumptions on the objects under consideration which are tacitly presupposed in logical considerations (see above).
Our aim is to apply this classification from single judgements to arguments, particularly mathematical proofs. For example, like single judgements, arguments have a hierarchical intentional structure which can be given in a vague or in a distinct way and also can be partly or completely fulfilled or unfulfilled. The everyday experience with the process of understanding mathematical arguments suggests that something corresponds to these notions in

[6] 'Making a concept distinct, i.e. making distinct that which is meant by the word by itself, is a procedure which takes place within the sphere of pure thinking. Before the least step of clarification is performed, while no or a completely inadequate intuition is associated with the word, we can reflect upon that which is meant, i.e. in 'decahedron': a solid, a polyhedron, regular, with ten congruent faces. (...) With a clarification, we transcend the sphere of mere meanings and thinking concerning meanings, we match the meanings with the noematic content of intuition.' [Translation by the author]

the realm of such arguments. In particular, the difference between grasping the mere meaning ('Vermeinung') of an argument or actually mentally following it is probably well-known to readers of mathematical texts.

3.3 Proofs, Arguments and Understanding - a Clarification

Is it possible to understand a false proof? Certainly. We can be convinced by it, explain it to others (and convince them), translate it to another language, reformulate it, recognize it in its reformulations etc. Even if we know it is false, this does not necessarily hinder our understanding, and it is even often possible (and sometimes takes some effort to avoid) to re-enter a state of mind in which it is still convincing. This for example seems for some people to be the case with the 'goat problem' ([GP]).

Of course, in the usual understanding and despite common manners of speech, a 'false proof' is not a proof. It merely shares some features with a proof on the surface. Anyway, the word is often used in such a way that a proof can be false. This use seems to resemble closer the way we internally think of proofs. We could replace the word 'proof' e.g. by 'argument' to avoid this ambiguity, but we prefer to keep it. Hence, we use the word proof in the sense of a proof attempt. Otherwise, we could never know if something is a proof, for in principle, we could always have been mistaken in checking it.

In this section, we make a humble approach to the study of the ways how a proof can appear to us. We take a phenomenological viewpoint: Hence, instead of asking what proofs might be in themselves - like platonic ideas, patterns of brain activity, mere sequences of tokens or of thoughts - we focus on the question how they give themselves when we encounter them in our mental activity. Mental activities directed towards proofs are e.g. creating it, searching for it, explaining it, remembering it, checking it etc. In such acts, we can experience a proof in different qualities. It is these qualities of proof experience that we consider here, focusing on two, namely clarity and distinctiveness. These are hence not properties of proof texts, but of our perception of proofs. The only way to point to such qualities is hence to create the corresponding experience and then naming it. This is what the following is about. Importantly, we will consider examples of proofs that are likely to lead to an experience with the quality in question, yet one must keep in mind that it is the experience, not the proof text, we are talking about, and that the same text may well be perceived in different ways. The point of this is to find out what is needed for our perception of a proof to make it checkable and compare it to what we need to formalize it. By Rav's claim, the qualities necessary for formalization presuppose those necessary for checking. We aim at demonstrating the contrary.

We now proceed to apply distinctiveness and clarity to proofs (in the sense above). In analogy to the case of judgements, a proof is distinctively given when its parts and their relations are made explicit. At this stage, we hence pay no attention to the correctness of the proof, only to its 'structure'. Distinctiveness about a proof means consciousness of what exactly is claimed and assumed at each point, from what a claim is supposed to follow, where assumptions are needed, which objects are currently relevant, which of the objects appearing are identical, how they are claimed to relate to and depend on each other etc. In analogy with distinct jugdements, a distinct proof does not need to be correct. Not even its logical structure must be sound: A distinctively given argument can well be circular. However, from a distincive perception of the argument, it will be apparent that it is. What we have with a distinct perception of a proof can be seen as a 'proof plan', a description of its logical architecture. In particular, distinctiveness includes consciousness of the sequential structure of the argument. To some extent, it is necessary whenever we even attempt to formulate it. [7] Naturally, distinctiveness comes in degrees. An argument can be distinct in certain parts but not in others. We frequently experience aquaintance with a proof without being able to state exactly where each assumption enters the argument, where each auxiliar lemma is used etc. Also, quite often in understanding natural proofs, we encounter some mixture of distinct deductive steps and imaginative thought experiments.

Concerning clarity, consider the following well-known 'proof' that $2 = 1$:

Let $a, b \in \mathbb{R}$, $a = b$. As $a = b$, we have $a^2 = ab$, hence $a^2 - b^2 = ab - b^2$. Dividing by $(a - b)$, we get $a + b = b$. With $a = b$, it follows that $2b = b$. Dividing by b, we obtain $2 = 1$.

Is this proof - seen as a train of thought - lacking distinctiveness? Not at all. It is completely apparent which of these few steps is supposed to follow from which assumption or fact earlier obtained. In fact, it is in a form that closely resembles a formal derivation (in particular, it could easily be

[7] Indeed, as a working mathematician, one occasionally experiences the perception of a vague proof idea which seems quite plausible until one attempts to actually write it down. When one finally does, it becomes apparent that the argument has serious structural issues, e.g. being circular. This particularly happens when one deals with arguments and definitions using involved recursions or inductions.

processed by Naproche), and not much would be necessary to make it completely formal.
Anyway, it is of course invalid, yet many people, including clever ones, at first don't see why. The problem here is obviously not that one does not really know what is stated in each step, or that one doesn't know what is supposed to follow or how; the problem is a misperception of division. Following the habit that 'you may cancel out equal terms', the semantic layer is left for the sake of a symbolic manipulation. On this level of consideration, one easily forgets about the condition imposed on such a step. If one takes the effort of really going back to what division is and why the rule that is supposedly applied here works, i.e. if one sharpens the underlying intuition of division, and if one additionally goes back to the meaning of the syntactical object '$a-b$', the mistake - division by 0 - is easily discovered. What is now added and was missing in the beginning is hence a more precise, adequate perception of the objects and operations appearing in the argument, in other words, a step towards the fulfillment of the intention given by the expression $a-b$. We call this degree of adequacy to which the notions are perceived the 'clarity' of the proof perception.

The above proof is hence distinct, yet not clear. Another famous example is the following:

$$1 = \sqrt{1} = \sqrt{(-1)(-1)} = \sqrt{-1}\sqrt{-1} = i \cdot i = i^2 = -1$$

Here, the mistake is obviously a misperception of the complex square root that probably comes from a prior intuition about square roots in the positive reals.

Already these primitive examples show that distinctiveness can be present without clarity.

Of course, we can have both. Every well-understood proof from a thorough textbook is an example. We can also have neither, and if one teaches mathematics, one will occasionally find examples in homework and exams. Further examples are most of the many supposed constructions for squaring the circle, most attempts at an elementary proof of Fermat's Last Theorem (see [FF]), disproofs for Gödel's Incompleteness Theorem, 'proofs' for the countability of $\mathbb{R}$ etc.

Clarity of a proof hence means fulfillment of the occuring intentions. These intentions involve those directed to mathematical objects as well as those directed towards logical steps, i.e. inferential claims: For a distinct understanding of a proof, it suffices to understand that, at a certain point, A is supposed to follow. This supposed inference is an empty intention directed at a derivation of A from the information given at this point. The fulfillment of this intention consists in carrying out this derivation. If a proof is given in perfect clarity, one knows exactly how to carry out each claimed inferential step, using the inferences associated with the occuring objects.

Let us briefly discuss the relation of the preceeding account to the view that proofs serve as fulfillments of mathematical intentions, as e.g. advocated by P. Martin-Löf or R. Tieszen (see Chapter 13 of [Ti1]). At first sight, our claim that proofs can be present in an unclear manner seems to conflict with the claim that proofs themselves are 'fulfillments' - for that would make them 'unfulfilled fulfillments'. This apparent conflict is, however, merely terminological: In Tieszen's account of Martin-Löf's position, proofs are cognitive processes of 'engaging in mental acts in which we come to 'see' or 'intuit' something' serving as evidence for mathematical claims. The intention towards a mathematical object would then be fulfilled by evidence for its existence, i.e. a - possibily preferably constructive - proof of its existence. From the point of view of a proof checker that we are interested in, on the other hand, a purported proof is, at first, a linguistic object: It can only serve as evidence once it has been worked through and understood, a process that may require considerable amounts of time, patience and cognitive involvement. In particular, it must be correct in order to provide evidence; in contrast, in our use of the word 'proof' as discussed at the start of the current section, we follow the usual manner of speech to allow for 'false proofs'.[8] The notion of proof that Tieszen seems to aim at is the way a proof is present after this process has been carried out. This is explicated on page 277 of [Ti1]:

> 'One can get some sense of the concept of 'evidence' that I have in mind by reflecting on what is involved when one does not just mechanically step through a 'proof' with little or no understanding, but when one 'sees', given a (possibily empty) set of assumptions, that a certain proposition must be true. Anyone who has written or read proofs has, no doubt, at one time or another experienced the phenomenon of working through a proof in such a merely mechanical way and knows that the experience is distinct from the experience in which one sees or understands a proof. (...) To give a rough

[8] This point is also discussed in [ML], p. 418: 'And it is because of the fact that we make mistakes that the notion of validity of a proof is necessary: If proofs were always right, then of course the very notion of rightness or rectitude would not be needed.'

> description, one might say that some form of 'insight' or 'realization' is involved, as is, in some sense, the fact that the proof acquires 'meaning' or semantic content for us upon being understood.'

It is exactly this process of acquiring 'meaning or semantic content for us' that we are concerned with. In the first sentence of the quoted passage by Tieszen, our notion of proof is present as 'proof' in quotation marks.

4 Distinctiveness, Clarity and automatic proof checking

Having established the two qualities relevant for our approach, we now want to link them to natural and automatic proof checking. Our thesis is that, while a full formalization requires a distinct and clear presentation of a proof, which means that there's nothing left to do for an automatic checker, distinctiveness is sufficient for producing an automatically checkable text, but not for checking the proof 'by hand'.

4.1 Distinctiveness is sufficient for automatic checkability

We start by comparing a short and basic natural (though very thoroughly written) mathematical argument written up for human readers to its counterparts in the Naproche language. The following is a passage from the English translation of Edmund Landau's 'Grundlagen der Analysis' ([La]) (all variables denote natural numbers):

Theorem 9: For given x and y, exactly one of the following must be the case:
1) $x = y$.
2) There exists a u (exactly one, by Theorem 8) such that $x = y + u$
3) There exists a v (exactly one, by Theorem 8) such that $y = x + v$.
Proof: A) By Theorem 7, cases 1) and 2) are incompatible. Similarly, 1) and 3) are incompatible. The incompatibility of 2) and 3) also follows from Theorem 7; for otherwise, we would have $x = y + u = (x + v) + u = x + (v + u) = (v + u) + x$. Therefore we can have at most one of the cases 1), 2) and 3).
B) Let x be fixed, and let M be the set of all y for which one (hence, by A), exactly one) of the cases 1), 2) and 3) obtains.
I) For $y = 1$, we have by Theorem 3 that either $x = 1 = y$ (case 1) or $x = u' = 1 + u = y + u$ (case 2). Hence 1 belongs to M.
II) Let y belong to M. Then either (case 1) for y) $x = y$, hence $y' = y + 1 = x + 1$

(case 3) for y'); or
(case 2) for y) $x = y + u$, hence if $u = 1$ then $x = y + 1 = y'$ (case 1) for y'); but if $u \neq 1$, then, by Theorem 3, $u = w' = 1 + w$, $x = y + (1 + w) = (y + 1) + w = y' + w$ (case 2) for y'); or (case 3) for y) $y = x + v$, hence $y' = (x + v)' = x + v'$ (case 3) for y'). In any case, y' belongs to M. Therefore, we always have one of the cases 1), 2) and 3).

Now compare this to the following variant, a fragment of a text which is accepted by the current version of Naproche (taken from [CCK]):

Theorem 9: Fix x, y. Then precisely one of the following cases holds:
Case 1: $x = y$.
Case 2: There is a u such that $x = y + u$.
Case 3: There is a v such that $y = x + v$.
Proof: Case 1 and case 2 are inconsistent and case 1 and case 3 are inconsistent. Suppose case 2 and case 3 hold. Then $x = y + u = (x + v) + u = x + (v + u) = (v + u) + x$.
Contradiction. Thus case 2 and case 3 are inconsistent. So at most one of case 1, case 2 and case 3 holds.
Now fix x. Define $M(y)$ iff case 1 or case 2 or case 3 holds.
Let y such that $y = 1$ be given. $x = 1 = y$ or $x = u' = 1 + u = y + u$. Thus $M(1)$.
Let y such that $M(y)$ be given. Then there are three cases:
Case 1: $x = y$. Then $y' = y + 1 = x + 1$. So $M(y')$.
Case 2: $x = y + u$. If $u = 1$, then $x = y + 1 = y'$, i.e. $M(y')$. If $u \neq 1$, then $u = w' = 1 + w$, i.e. $x = y + (1 + w) = (y + 1) + w = y' + w$, i.e. $M(y')$.
Case 3: $y = x + v$. Then $y' = (x + v)' = x + v'$, i.e. $M(y')$.
So in all cases $M(y')$. Thus by induction, for all y $M(y)$. So case 1 or case 2 or case 3 holds. Qed.

We see some extra complications arise due to the fact that the Naproche language is a controlled language, so that formulations like ‘For given x and y’ or ‘incompatible’ need to be replaced by their counterparts ‘Fix x, y’ and ‘inconsistent’ in the Naproche language. This, of course, can easily be overcome by amending the language accordingly. The passage is given exactly in the way Naproche can currently read it to avoid the criticism of being speculative, but we are safe to assume that a line like ‘exactly one of the following must be the case’ can be processed by a slighly improved version. Now, what does it take to go from the natural to the Naproche version? Do we need to understand the proof in some depths or see its correctness? Certainly not. Rather, we reformulate the proof according to some linguistic restrictions. Hence, while a certain difference in the wording is obvious, these two texts are very similar in content and structure. Given a knowledge of the current Naproche language, passing from the first version to the second is trivial: One merely changes some formulations, permanently working along the original. One can do this with virtually no understanding of the original text, as long as one keeps the indicators for assuming, deducing and closing assumptions and uses the same symbol where the same object is meant. One does not even need to know the meaning of the symbols used. It seems that any state of mind allowing one to write the first text also allows one to write the second. In fact, even a faint memory of a vague understanding might suffice.

This basic example indicates what is necessary for producing an automatically checkable version of a proof: The argument must be given to us as a sequence of steps in such a way that we can see what is currently claimed and assumed, which objects are considered, when a new object is introduced and when something new is claimed about an object introduced earlier (so we will e.g. use the same symbol). A mere image of some mental movement, which is indeed often the way one remembers or invents an argument, is not sufficient. One needs an explicit consciousness of the way primitive intentions are build together to form judgements and then how these complex intentions are used to build up the argument. On the other hand, it is not necessary at all to reduce everything to formal statements and simple syllogisms. Whether or not a concrete checker will succeed in a particular case depends of course on how well the checker captures the semantics of natural mathematical language, but in principle, arguments at this stage of understanding are open to an automatic checking process. Hence, to produce an automatically checkable format, it suffices to have a distinct understanding of the proof.

We consider another example, which deserves special interest, as it is explicitely given by Rav in [Rav] as an example of an 'ordinary' proof, which he comments thus:

> The proof of this theorem as given above is fully rigorous (by currently accepted standards). It requires, however, on the part of the reader a certain familiarity with standard mathematical reasoning (...). No formal logic is involved here; it is a typical reasoning that Aberdein has aptly characterized as the `informal logic` of mathematical proof.

He then goes on to describe what to do for obtaining a version of this proof that can be automatically checked, claiming that this needs, 'as must have been noted, a logician's know-how'. Here is the proof as given in the text, including some preliminaries about group theory (we omit Rav's bracketed observations on the proof):

Recall that a group is defined as a set G endowed with a binary operation (to be written as juxtaposition) having a distinguished identity element, denoted by 'e' and satisfying the following first-order axioms:

- (i) $(\forall x)(\forall y)(\forall z)[x(yz) = (xy)z]$; (associativity)
- (ii) $(\forall x)(ex = x)$; ('e' is a left-identity)
- (iii) $(\forall x)(\exists y)(yx = e)$. (every element has a left-inverse)

On the basis of these axioms, one proves that the postulated left-identity is also a right-identity - in symbols, $(\forall x)(xe = x)$ - from which it will follow that the identity element is unique. Here is the proof:

Let u be an arbitrary element of G. By axiom (iii), there exists t such that (1) $tu = e$; once more by axiom (iii), for the t just obtained, there exists $s \in G$ such that (2) $st = e$. Hence: (3) $ue = e(ue) = (eu)e = [(st)u]e = [s(tu)]e = (se)e = s(ee) = se = s(tu) = (st)u = eu = u$. (...) Since u is an arbitrary element of G and by (3) $ue = u$, we conclude that $(\forall x)(xe = x)$.

Now if there were a second element e' with the property that $(\forall x)(e'x = x)$, we would conclude from what has just been proven that $(\forall x)(xe' = x)$ also, and hence $e = ee' = e'$. Thus, the identity element is unique. Hence, we have proven the following:
Theorem: (a) $(\forall x)(xe = x)$ (b) $(\forall y)(\forall x)[yx = x \implies y = e]$.

The following Naproche version of this proof, due to Marcos Cramer, is accepted by the current version of the system:

Suppose that there is a function $*$ and an object e satisfying the following axioms:
Axiom 1. $\forall x \forall y \forall z (x * (y * z) = (x * y) * z)$.
Axiom 2. $\forall x (e * x = x)$.
Axiom 3. $\forall x \exists y (y * x = e)$.

Theorem A: For all x, we have $x * e = x$.
Proof: Let u be given. By Axiom 3, there is a t such that $t * u = e$. By Axiom 3, there is an s such that $s*t = e$. Then $u*e = e*(u*e) = (e*u)*e = ((s*t)*u)*e = (s*(t*u))*e = (s*e)*e$ and $(s*e)*e = s*(e*e) = s*e = s*(t*u) = (s*t)*u =$

$e * u = u$. Thus $\forall x(x * e = x)$. Qed.

Theorem B: If $\forall x \forall y(x * y = x)$, then $y = e$.
Proof: Assume that e_1 is such that $\forall x(e_1 * x = x)$. Then by Theorem A, $\forall x(x * e_1 = x)$. Hence $e = e * e_1 = e_1$. Thus $e_1 = e$. Qed.

Again, the transition from the original text to the Naproche text is quite elementary and needs neither a logician's expertise nor knowledge of group theory. It is particularly interesting to note that the equivalence of these texts can be observed even without a knowledge of the meaning of the corresponding terms. This demonstrates again that it is distinctiveness, not clarity, which is required for the transition as well as for judging the adequacy of such a translation

Let us briefly recall at this point how Naproche proceeds to check a text like the above: From the text, it builds a representation of the intended logical structure of the proof. These proof representation structures (PRSs) contain the occuring statements along with their intended relations, as A is to be assumed, B is to be deduced, C is to be used in the step from A to B etc. Other than a formal derivation which, by its definition, cannot be false, a PRS is completely neutral to the correctness or soundness of the represented argument: What it represents is the distinctively given intentional structure of a proof. Notably, presuppositions hidden in certain formulations (e.g. uniqueness in the use of definite articles as in 'Let n be the smallest natural number such that...') are made explicit. In the next step, an attempt is made to construct a fully formal proof from the PRS: In particular, when B is, according to the PRS, claimed to follow at a certain step (possibly with the help of an earlier derived statement A), then an automatic theorem prover is used to attempt to prove B from the available assumptions at this point, where indicators like explicit citations of earlier statements help to direct the proof search. In particular, the automatic prover will attempt to apply definitions of the occuring notions as well as results in which they appear: It will, when a notion T occurs, quite naturally try to use the 'inference package' associated with T for creating the actual deductions claimed to exist. It will, in other words, attempt to do something analogous to fulfilling the intentions given in the distinctive representation of the proof.

4.2 Distinctiveness is not sufficient for natural checking

The degree of understanding obtained by distinctly disclosing the structure of an argument is not sufficient for performing a proof check. In fact, we can have perfect distinctiveness and still be completely agnostic concerning correctness. This is already indicated by our examples above. One reason for this is that, in a natural argument, we do not have a fixed, manageable supply of inference rules justifying each step. When checking a step, we often use some mental representation ('image') of the objects under consideration. This representation is different from a formal definition and usually precedes it. [9] However, these images are imperfect in directing us towards the objects we mean and may carry false preconceptions concerning these objects. If, for example, concept B is a generalization of concept A, there is a certain tendency of assuming properties of A for B. There is a vast amount of frequent mistakes compatible with a distinct presentation. A strong source of mistakes is some kind of a closed world assumption that excludes objects we can't really imagine. This danger remains even after we come to know about counterexamples. Imagining e.g. a continuous function from $\mathbb{R}$ to $\mathbb{R}$ as a 'drawable line' is often very helpful, but it also misdirects us in many cases. For another example, in spite of strong and repeated efforts, some students in set theory courses never acknowledge the existence of infinite ordinals and keep subtracting 1 from arbitrary ordinals . The idea of a non-zero 'number' without a predecessor is apparently hard to accept.

Such preconceptions derived from a misinterpretation of mental images are a common source of mistakes even in actual mathematical research practice. [MO1] contains a long and occasionally amusing list of common misperceptions in mathematics, most of which are instances of such a misinterpretation. We will get back to this below when we consider classical examples of false proofs.

Of course, this kind of perception of mathematical objects is all but a dispensible source of mistakes: In fact, it is exactly this ability that steers the process of proving and creating mathematics, thereby making the human mathematician so vastly superiour to any existing automatic prover.

Let us now briefly reconstruct this example-driven account in the more general terminology set up above for this purpose. Given a distinct presentation, we only see what is claimed to happen, a 'proof intention'. We can at this point already spot some mistakes, e.g. circularity, or that the

[9] Historically, at least. The deductive style dominant in mathematical textbooks confronts the student with the inverse problem: Namely making sense of a seemingly unmotivated given formal definition.

statement actually proved is not the statement claimed; but we cannot check the proof: A ‘proof map’, as helpful as it is for finding one’s way through a more involved argument, does not require for its creation an understanding or checking of the proof. To check a proof, it is necessary to fulfill these inferential intentions; but for this, clarity is needed.

5 Distinctiveness, Clarity and Formalization

In this section, we consider the question what kind of understanding is necessary for carrying out a full formalization of a proof in a common system of first-order logic like, say, natural deduction. We argue that Rav is indeed right in claiming that such an understanding allows checking and that, in fact, the checking is almost inevitably carried out in the process of obtaining such an understanding.

To do this, let us reflect on the process of formalization. A formal proof is one in which the manipulation of symbols is justified without any reference to a meaning of these symbols. It is clear how a certain symbol may be treated without knowing what it means, without even taking into account that it might mean anything. This is achieved by replacing semantic reference by formal definitions. For instance, the meaning of the word ‘ball’, representing a certain geometric shape, will be replaced by rules that allow certain syntactical operations once a string of the form $ball(x)$ shows up. Still, the formal definition must capture the natural meaning if the formal proof is to be of any semantic relevance, not just a symbolic game. So the formal definitions have to be adequate in a way. How do we arrive at adequate formal definitions? Obviously by observing the role a certain object plays in proofs and then formulating precisely what about this object is used. The first step in formalizing a notion is hence to perform an eidetic reduction. Then, the notion of the object is replaced by the statements used about it. (See [Ti1] and [Ti2] for thorough discussions of phenomenological aspects involved in the forming and clarification of mathematical concepts.)

If we replace an informal by a purely formal proof, we have to make all implicit references to the content explicit to eliminate them in the formalization. This means that the role of the object in the argument must be clear. Consequently, to obtain a fully formal derivation from an informal proof, we must have distinctiveness **and** clarity. But when all hidden information is made explicit as part of a complete understanding of the argument, i.e. if all intentions are fulfilled, mistakes will inevitably become apparent. The only questionable part remaining is then the connection between the original semantic references and the formal definitions [10]. But in established areas of mathematics, these definitions have stood the test of time, and even though, particularly in new areas, there are debates about the adequacy of definitions and though the focus occasionally shifts from one definition to another providing a deeper understanding of the subject (often indicated by amending the original notion with expressions like 'normal', 'acceptable', 'good' etc.), this issue virtually doesn't come up in mathematical practice. Even if it does, it is usually considered to affect the degree to which the result is interesting, not the correctness of the proof. [11]

Considering a distinct proof presentation, a good automatic prover will be able to draw from formal definitions what we draw from correct intuition. Note that we make no claim on the question whether formal definitions can exhaust semantic content, nor do we need such a claim. The process of replacing steps refering to understanding and perception of abstract objects by derivations from formal definitions is what corresponds to the activity of fulfilling intentions involved in the course of a clarification. Conceptually and mentally, this may well be a very different operation. However, as explained above, Gödel's completeness theorem ensures that, whenever an argument can be brought to clarity, there will be a derivation from the definitions. This is the reason why an automatic proof checker, using an enhanced formalism as described above, can give us information on the possibility of clarification. For a more concrete picture of what information this kind of automatic checking can give us, let us suppose that we have given a certain proof B to our proof-checker (e.g. Naproche) and got a negative feedback, i.e. that our assumptions were found to be contradictory or that some supposed consequence could not be reproduced by the system. We are thereby made aware of a particular proof step that needs further explanation, and it is here that we may become aware of an actual mistake in our intended proof. This can already be seen in the most basic examples above, e.g. the 'proof' that $2 = 1$: Here, the demand to divide by $(a - b)$ will trigger the presupposition that $(a - b) \neq 0$ as an intermediate proof goal which will then be given to the

[10] An excellent example of the delicate dialectics involved in forming definitions of intuitive concepts is the notion of polyhedron in Euler's polyhedron formula as discussed in [La]. Sometimes, this is the really hard part in creating new mathematics. Another prominent example is the way how the intuitive notion of computability was formalized by the concept of the Turing machine.

[11] Suppose, for example, that someone came up with a non-recursive function that one can evaluate without investing original thought so that one is inclined to accept the evaluation of this function as an instance of 'calculation', thus disproving the Church-Turing thesis. As a consequence, recursiveness would lose its status as an exact formulation of the intuitive concept of calculation. But this would not affect the correctness of recursion theory.

automatic theorem prover. It is then a matter of seconds that we will be informed that this proof goal was found to contradict the assumptions which will make it very obvious what goes wrong in the intended proof. Here, we thus have a clear example how automatic proof checking can lead to an epistemic gain.

It is not so obvious what information there is to be drawn from the opposite scenario, e.g. a positive feedback from the system. Certainly, we are informed that the proof goal is actually formally provable as it should be, but that doesn't imply that our proof was correct. Getting information about our proof from a positive feedback would need a close connection between our natural way to think about missing proof steps and the automatic theorem prover, which is certainly a fascinating subject for further study, but currently far remote from reality. This point will be discussed in section 7.

6 An Historical Example

In this section, we apply the notions obtained above to a famous historical example of a false proof. Our goal is to demonstrate that this proof shows a sufficient degree of distinctiveness for a formalization in a Naproche-like system and hence that automatic checking could indeed have contributed in this case to the development of mathematics. This example further demonstrates that even incomplete distinctivication can be sufficient for automatic checking and that actual mistakes may occur already in the margin between the degree of distinctiveness necessary for formalization and complete distinctiveness.

Example (Cauchy 1821)[12]

Claim: Let $(f_i|i \in \mathbb{N})$ be a convergent sequence of continuous functions from $\mathbb{R}$ to $\mathbb{R}$, and let s be its limit. Then s is continuous.

Proof: Define $s_n(x) := \Sigma_{i=1}^{n} f_i(x)$, $r_n(x) := \Sigma_{i=n+1}^{\infty} f_n(x)$. Also, let $\varepsilon > 0$. Then, as each f_i is continuous and finite sums of continuous functions are continuous, we have $\exists\delta\forall a(|a| < \delta \implies |s_n(x+a) - s_n(x)| < \varepsilon)$.

As the series $(f_i|i \in \mathbb{N})$ converges at x, there is $N \in \mathbb{N}$ such that, for all $n > N$, we have $|r_n(x)| < \varepsilon$.

Also, the series converges at $x + a$, so there is N such that, for all $n > N$, we have

[12] This formulation is sometimes disputed as not correctly capturing the argument Cauchy had in mind. Some claim that Cauchy meant the variables implicit in his text to not only range over what is now known as the set of reals, but also over infinitesimals. However, the formulation we offer captures the way the proof was and still is understood and at first sight considered correct by many readers, so we will not pursue this historical question further.

$|r_n(x+a)| < \varepsilon$.
So we get: $|s(x+a) - s(x)| = |s_n(x+a) + r_n(x+a) - s_n(x) - r_n(x)| \leq |s_n(x+a) - s_n(x)| + |r_n(x)| + |r_n(x+a)| \leq 3\varepsilon$.
Hence s is continuous.

This example is taken from [Ri] and closer analyzed in the appendix of [La]. The mistake becomes obvious when one focuses on the dependencies between the occuring quantities: The δ shown to exist in line 3 of the proof depends on ε, x and n. The N from line 4, on the other hand, only depends on ε and x. However, the N used in line 6 obviously also depends on a. Hence N is in a subtle way used in two different meanings. The dependence on a can only be eliminated if there is some M bigger than $N(\varepsilon, x+a)$ for all $|a| < \delta(\varepsilon, x, n)$. This property means that $(f_i | i \in \mathbb{N})$ is uniformly convergent. which is much stronger than mere convergence.
Simple as this mistake may seem, it has a long success story (see again [La]): The (wrong) statement it supposedly proves was considered trivially true for quite a while by eminent mathematicians, and when the first counterexamples occured, they were considered either as pathologies that shouldn't be taken seriously as functions or violently re-interpreted as examples. It was no other then Cauchy who first felt the urge to give a proof and published the above argument in his monograph [Cau]. It took several decades before the mistake was spotted and the statement was corrected by strengthening the assumption to uniform convergence.
Reproducing the understanding of this argument shows what is going on: In the arguing for the existence of N, one gets the imagination of a 'sufficiently large number' and then reuses the object in a new context in an inappropriate way because hidden properties of the object - its dependencies on others in its construction - are ignored. That is, while the train of thought described here gives distinct intentions to certain objects N and N' which are then identified, a fulfillment of these intentions is not possible.
Now, suppressing the arguments on which an object depends is quite common in mathematical writings. A formalizer, of course, must reconstruct this information. The way a Naproche-like system models a text can easily allow for such a convention. Apart from that, the text is certainly not lacking distinctiveness. It also uses only very little natural language and not in any complicated way. It would be quite feasible to enrich the vocabulary of e.g. Naproche to process it in the precise form given here. But when the formalization is carried out, the proof breaks down. It will be very interesting to actually carry this out on concrete systems once they are sufficiently developed.

7 Discussion

We have explained above how automatic checking with Naproche-like systems can lead to an epistemic gain concerning a proof. We will now take a close look at the assumptions on which our account relies, thereby sharpening the picture when such gains can be expected and when not.
Under what circumstances, then, do we get new information from Naproche and what is that information? Certainly, it is informative if Naproche spots a non-intended contradiction - in this case, the proof contains a mistake. The false proofs of $1 = 2$ and $-1 = 1$ above are examples where a tacit assumption contradicts the information given, which an automatic proof checker can spot and report. But what does it mean in general when Naproche fails to confirm a proof? And what does it mean for the correctness of the proof if it succeeds? An informal proof, as found in a mathematical journal, a math exam or a math olympiad is roughly seen as correct when it convinces the critical expert: I.e. when it provides a person with the right background with the information needed to construct a complete, detailed argument. This notion of 'right background' is highly context-dependent: An original research paper at the frontier of some area of core mathematics may leave proof steps to the reader as 'clear' that would require from the average beginner student several years of study and concentrated work to complete; on the other hand, proof exercises for beginner students of mathematics often require details even - occasionally especially - for steps that are supposedly immediately clear. This is only one respect in which the correctness of proof texts is a delicate notion, involving sociological aspects like what background knowledge and what heuristic power is to be expected by the audience addressed. The criterion realized in Naproche in its current implementation is comparably weak, namely whether each claim can be formally deduced from the information given at that point with limited ressources (typically within 3 seconds processing time). This is a very rough model of the inferences regarded as admissible for bridging proof steps by human readers to whom formally extremely complex inferences may be clear based on spatial intuition, analogy with previous arguments etc. Consequently, a failure of Naproche to confirm a proof does not necessarily mean that this proof is not a rational and sensible way of convincing the reader of the correctness of the conclusion.Conversely, at least in principle, we could have false positive results. If the checking succeeds, then a formal proof of the conjecture in question was generated, so a correct proof is confirmed to exist; but this does not necessarily mean that the original informal proof was correct by the standards for informal proofs. Suppose e.g. that, in a proof of a statement A, the end-result A is deduced as an intermediate step in the process

of bridging a gap in that proof; in this case, the formalization should certainly not be seen as a confirmation that the original argument was sound. Similarly, if the work needed to confirm a certain inference in a proof is much more complex than the whole proof itself, one has a reason to doubt the proof. This becomes particularly obvious when one thinks of tutorial contexts: When asked to prove something, some students try to trick the corrector by making some deductions from the information given forwards, some steps backwards from the goal and then simply writing the results next to each other, claiming that the latter follows from the former. Even though this claim may be right, these texts fail to show that the author knew how to do it, and usually, the step left out is as complicated as the original problem itself. It is an interesting field of further study to see what kind of inferences may be 'left to the reader' and build formal models of those. More generally, the notion of correctness for informal proofs, in contrast to that for informal proofs, certainly deserves further attention.

Still, granting these objections and as decent as our currently existing software may be, the experience with Naproche and our analysis thereof above should suffice to demonstrate that there is more to get from automatic proof checking than Rav (and many others) might expect. However, one should not forget that this kind of checking relies on various convenient circumstances; for example, it assumes a stable formal framework which in particular allows to replace the understanding of a term by a formal definition for all purposes of deduction. Where such a framework is missing, the emulation of fulfillment of mathematical intentions by automatic proving will not work. Such a framework is, of course, not always present; rather, it typically occurs as a rather late stage in the development of a mathematical theory. The clarification and development of notions is an important part of mathematics, and proofs play a role as a part of this process (as e.g. Lakatos has impressively demonstrated in [Lak]). A similar caveat holds with respect to the use of axiomatic systems: Not always is the content of a theory canonically codified in an axiomatic system; furthermore, axiomatic systems codifying a theory can be complemented when defects become apparent in its usage. (A well-known examples is Zermelo's proof of the well-ordering principle introducing the axiom of choice.)

A vast majority of contemporary mathematical work, however, does indeed work in such stable environments where we have formal definitions and a universally accepted axiomatic background. Still, there are parts of mathematicians proof practice (arguably the philosophically most interesting ones) that evade automatic checking in the sense explained above. For those, there may indeed be fundamental reasons to expect that automatic checking will not be helpful. Let us look at one example: Namely Turing's work ([Tu]) on the Entscheidungsproblem, which was widely accepted as settling this problem. It depends on a technical part (the Entscheidungsproblem is not solvable by a Turing machine) as well as a concept analysis (Turing machine computability captures the informal notion of computability, as intended in the formulation of the Entscheidungsproblem). The first part is, technical difficulties aside, open to automatic checking. But the second part seems to be of an entirely different nature: Can a machine even in principle help us to decide whether some formalization of a certain intuitive notion is adequate? Apparently, we can't have told the machine what such a notion means without knowing it ourselves. If, by some other means (like automatic learning) we had a machine answering 'yes' or 'no' to such questions, how could we know it is right? The machine would have to take part in the discussion arising, providing experiences, thought experiments exploring the borders of the informal notion etc. In some cases, the underlying informal understanding of such notions will be rooted far outside of mathematics in what may be called the life-world. Is the set of reals an adequate model for a 'line' or for time? Does Turing computability indeed capture computability in the intuitive sense? Does ZFC provide a reasonable understanding of what it means to be a set and are the set-theoretical formalizations of mathematical notions adequate to them? These questions play a crucial role in the acceptance of various formal proofs as answers to mathematical questions, and they are only non-mathematical when one arbitrarily limits the scope of mathematics to formalism. In all of these cases, one can easily imagine how the discussion about them will touch on aspects of human thinking and experience far outside of mathematics. These notions are human notions, made by humans for humans. Human understanding is the ultimate criterion for their adequacy, so that no outer authority like a computer can tell us what they mean. It is hence quite plausible that we cannot write a program of which we are justified to believe that it gives correct answers to questions concerning the adequacy of a concept analysis, but there is nothing mysterious about this 'non-mechanical nature of the mind': It is a mere consequence of the specific evidence type of (human) concept clarification. [13]

[13] Even more, taking the speculation a bit further, being able to meaningfully and convincingly participate in this kind of discourse is a plausible criterion for not calling something a 'machine' any more. (Moreover, the definiteness of a machine's response, which is a main motivation for striving for automatization in the first place, is lost when a machine becomes merely another participant of a discourse.) Of course, the rules of a discourse are made by its participants; so in the end, the possibility of computers becoming influential even in the conceptual part of mathematics might boil down to the question whether we are willing to accept a computer as a participant in such a debate on equal terms. However,

To briefly summarize our discussion: There are indeed important aspects and parts of mathematician's proof practice that are likely beyond automatic checking in principle, and definitely in the current state or any state to be expected in the foreseeable future. The great consensus on the Church-Turing-Thesis and the general acceptance of Turing's work as a solution to the Entscheidungsproblem (or Matiyasevich's work on diophantine equations as a solution to the 10th problem, see [Ma]) is something that Azzouni's derivation-indicator-view can hardly explain. But many great mathematician's have done their work in stable frameworks and a great deal of mathematics takes part in those: Here, automatic proof checking can lead to epistemic gains. And here, Naproche-like systems work, demonstrating that the derivation-indicator view is not as easily discarded by arguments like those of Rav.

8 Conclusions and Further Work

We hope to have made it plausible that phenomenological considerations and the corresponding shift of focus can be fruitfully applied to questions concerning the philosophy of mathematical practice with a relevance to mathematical research itself. Namely, we have argued that, in spite of the claims against it, automatic proof checking can lead to an epistemic gain about an argument in providing evidence that the indicated intentional acts can be carried out in a distinct and clear manner. The reason for this was that human proof checking needs clarity about a proof, while automatic checking can be performed once a certain degree of distinctiveness is obtained. For this argument, we crucially used the phenomenological turn from proofs in the way they are usually considered to the ways in which they occur.

A phenomenological theory of proof perception has, to our knowledge, not yet been given. It would certainly be interesting in its own right. As one consequence, it would contain a thorough study of proof mistakes, which, on the one hand, might become relevant in pedagogical considerations, but would also sharpen our understanding of what automatic proof checkers can add to our trust in a proof and how they can do this.

A concrete application of such considerations would be the development of proving tools suitable for Naproche-like systems. Such a prover is supposed to bridge steps in natural proofs which are assumed to be supplied by the reader. In a sense, these proofs are hence 'easy' and 'short'. Of course, such steps often take place in e.g. spatial or temporal intuition rather than formal reasoning. There is therefore no obvious relation between a 'simple, short

we now have reached a level of speculation at which it is better to stop.

argument' and the number of lines in a corresponding derivation. [14] A next step is hence to consider common elementary operations that are performed in supplying such proof steps and give formal background theories to replace them. The goal of this would be to make the automatic prover's activity more resemblant to an actual human reader. (Suppose e.g. that the automatic prover proves an auxiliary lemma in a proof in a very complicated way, obtaining the final theorem as an intermediate step. We would certainly not call this a valid reconstruction of the argument.) This could help to considerably increase the contribution of natural-language oriented automatic proof-checkers: In areas like elementary number theory, where crucial appeal to intuition is rare and proofs can be translated rather naturally, a Naproche reconstruction of an informal proof will usually correspond well to the proof intended. Even if it doesn't, we gain trust in the theorem from a positive checking, as we obtain a formal proof, whether it adequatly captures the original proof or not. But of course, the goal of a proof checker is not just to check whether the theorem claimed to be proved is provable, but whether the purported proof actually is one. For succeeding at this task, the checker would have to become 'pragmatically closer' to the intended human reader. Once sufficient background theories are build up, one should actually carry out the examples given above and others to see what Naproche does with them. Will it find the 'right' mistake? This asks for a systematic study of wrong proofs in e.g. flawed research papers, wrong student's solutions etc. Such a reconsideration of well-known mistakes can serve both as a source of inspiration for the development of natural proof-checkers and as powerful demonstration of what has been achieved.

References

[Az1] J. Azzouni. Tracking Reason. Proof, Consequence, and Truth. Oxford University Press (2005)

[Az2] J. Azzouni. Why do informal proofs confirm to formal norms? Foundations of Science 14: 9-26.

[14] To appreciate this difference, one might consider the equation $((a+b)+c)+d = ((d+c)+b)+a$ over the reals, which is obviously true for a human reader who thinks of addition as taking the union of two quantities. In our experience with number-theoretical texts in Naproche ([Ca]), the automatic prover, having to derive this from commutativity and associativity of addition, often got lost in the countless alternative possibilities which rule to apply. This is a striking example for the pragmatical difference between formal definitions and intuitive concepts.

[Az3] J. Azzouni. Is there still a Sense in which Mathematics can have Foundations? In G. Sica, ed., Essays on the Foundations of Mathematics and Logic. Polimetrica, 9-47.

[BB] C.E. Benzmüller, C.E. Brown. The Curious Inference of Boolos in Mizar and OMEGA*. STUDIES IN LOGIC, GRAMMAR AND RHETORIC 10 (23) 2007

[Bo] G. Boolos. A Curious Inference. Journal of Philosophical Logic, Vol. 16, No. 1, Feb., 1987

[Ca] M. Carl. An Introduction to elementary number theory for humans and machines. Work in progress.

[Cau] A. Cauchy. Cours d'Analyse, p. 120. (1821)

[CCK] M. Carl, M. Cramer, D. Kühlwein. Chapter 1 of Landau in Naproche, the first chapter of our Landau translation. Available online: `http://www.naproche.net/inc/downloads.php`

[CK] M. Carl, P. Koepke. Interpreting Naproche - An algorithmic approach to the derivation-indicator view, paper for the International Symposium on Mathematical Practice and Cognition at the AISB 2010.

[CKKS] M. Cramer, P. Koepke, D. Kühlwein, and B. Schröder: The Naproche System, paper for the Calculemus 2009.

[Cr] M. Cramer. Naproche version of 'There are infinitely many primes'. Unpublished notes.

[FM1] P. Koepke. J. Schlöder. The Gödel Completeness Theorem for Uncountable Languages. J. Formalized Math. 30 (3) (2012)

[FM2] K. Pak. Brouwer Fixed Point Theorem in the General Case. J. Formalized Math. 19 (3) (2011)

[FM3] J. Avigad, K. Donnelly, D. Gray, P. Raff. A Formally Verified Proof of the Prime Number Theorem. ACM Transactions on Computational Logic (2006)

[FF] A. Fleck, Ph. Maennschen, O. Perron. Vermeintliche Beweise des Fermatschen Satzes. Archiv der Mathematik und Physik. Vol. 14. (1909)

[FS] G. Bauer, T. Nipkow. Flyspeck I: Tame Graphs. Archive of Formal Proofs. Available online at http://afp.sourceforge.net/devel-entries/Flyspeck-Tame.shtml

[GP] Monty Hall Problem. Wikipedia-article
available at `http://en.wikipedia.org/wiki/Monty_Hall_problem`

[Hu1] E. Husserl. Formale und transzendentale Logik. Versuch einer Kritik der logischen Vernunft. Niemeyer, Tübingen (1900)

[Hu2] E. Husserl. Erfahrung und Urteil. Untersuchungen zur Genealogie der Logik. Felix Meiner Hamburg. (1999)

[Hu3] E. Husserl. Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. Meiner Hamburg (2009)

[Hu4] E. Husserl. Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. Drittes Buch. Haag Matrinus Nijhoff (1952)

[AK] M. van Atten, J. Kennedy. On the Philosophical Development of Kurt Gödel. Bull. Symb. Logic, Vol. 9, No. 4 (2003), pp. 425-476

[Kr] G. Kreisel. Informal Rigor and Completeness Proofs. In: I. Lakatos (Edt.) Problems in the Philosophy of Mathematics. Proceedings of the International Collquium in the Philosophy of Science, vol. 1 (1965)

[KR] H. Kamp, U. Reyle. From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory. Springer (2008)

[Ku] E. Kummer. Extrait d'une lettre de M. Kummer a M. Liouville. Journal de Mathematiques pures et appliquées 12 (1847)

[La] E. Landau. Grundlagen der Analysis. Heldermann, N (2004)

[Lak] I. Lakatos. Proof and Refutation. Cambridge University Press (1976)

[Lo] D. Lohmar. Phänomenologie der Mathematik: Elemente einer phänomenologischen Aufklärung der mathematischen Erkenntnis nach Husserl. Kluwer Academic Publishers. (1989)

[Ma] Y. Matiyasevich. Hilbert's 10th Problem. MIT Press Series in the Foundations of Computing. Foreword by Martin Davis and Hilary Putnam. Cambridge, MA: MIT Press.

[ML] P. Martin-Löf. Truth of a Proposition, Evidence of a Judgement, Validity of a Proof. Synthese 73 (1987), 407-420

[ML2] P. Martin-Löf. Intuitionistic Type Theory. (1980). Notes of Giovanni Sambin on a series of lectures given in Padova. Available online: http://www.cip.ifi.lmu.de/~langeh/test/1984\%20-\%20Loef\%20-\%20Intuitionistic\%20Type\%20Theory.pdf

[MO1] MathOverflow-Discussion: Examples of common false beliefs in mathematics. http://mathoverflow.net/questions/23478/examples-of-common-false-beliefs-in-mathematics-closed

[MO2] MathOverflow-Discussion: Widely accepted mathematical results that were later shown wrong. http://mathoverflow.net/questions/35468/widely-accepted-mathematical-results-that-were-later-shown-wrong

[NWI] Naproche Web Interface. Available at http://www.naproche.net/inc/webinterface.php

[Pr] Prime number. Wikipedia entry. Available at http://en.wikipedia.org/wiki/Prime_number

[Rau] W. Rautenberg. A Concise Introduction to Mathematical Logic. Springer (2006)

[Rav] Y. Rav. A Critique of a Formalist-Mechanist Version of the Justification of Arguments in Mathematicians' Proof Practices. Philosophia Mathematica (III) 15 (2007) pp. 291-320

[Ri] V.F. Rickey. Cauchy's Famous Wrong Proof. Available online: http://www.math.usma.edu/people/rickey/hm/CalcNotes/CauchyWrgPr.pdf

[Si] S. Singh. Fermats letzter Satz. Die abenteuerliche Geschichte eines mathematischen Rätsels. dtv (2000)

[Ti1] R. Tieszen. Phenomenology, Logic, and the Philosophy of Mathematics. Cambridge University Press. (2005)

[Ti2] R. Tieszen. After Gödel. Platonism and Rationalism in Mathematics and Logic. Oxford University Press. (2011)

[Tu] A. Turing. On computable numbers, with an application to the Entscheidungsproblem. Proceedings of the London Mathematical Society, (Ser. 2, Vol. 42, 1937)