Formal and Natural Proof - A phenomenological approach

Merlin Carl

February 19, 2013

Abstract

It is frequently claimed (see e.g. [Rav]) that the formalization of a mathematical proof requires a quality of understanding that subsumes all necessary acts 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 writeup, but does not subsume checking.

1 Introduction

It is a striking consequence of Gödel's completeness theorem that, whenever there is a correct mathematical proof of a certain sentence ϕ from a fixed set of axioms, then there is also a formal derivation of ϕ in the sense of a system of formal deduction. 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.

On the other hand, there is the program of formal mathematics, where mathematics is actually carried out in a strictly formal framework. This is 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 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 be 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 praxis', namely the fact that mathematics, considered as a social praxis, is remarkably stable when compared to other social practices such as art, religion, politics, philosophy and even the natural sciences. The notion of 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 praxis 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 the 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 ones 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 receipees 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 receipee 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 so 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 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 argument will not result in any epistemic gain concerning the argument. In particular, it does not strengthen the position that the 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.

This argument has certainly a good degree of persuasive power. We want to evaluate this criticism closer. On the surface, we claim that the underlying image of an 'algorithmic system' for this 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 obviously don't 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. 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 writeup, 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 the consciousness has to be taken into account.

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 present in the mind, applying the approach of Husserl's analysis of judgements (see [Hu1]), in particular his notions of 'distinctiveness' and 'clarity', to proofs. We want to 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 the Naproche system, a system for the automatic checking of natural mathematical arguments, which is currently under development. Its aim is exactly, as in Rav's words, 'to do the work of automatic checking even an informal proof'. 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 as natural deduction or the sequence 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 several historical examples of false proofs in these terms, considering whether or not and how a Naproche-like system might have helped to spot the mistake. In section 6*b*, we suggest further examples for future consideration. In section 7, we give 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 to the human reader. If one was presented 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 a modified 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 Clarity and Distinctiveness

In this section, we introduce Husserl's notions of the distinctiveness and clarity of a judgement and indicate its transfer to proofs.

3.1 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 turns 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. [1] [2]
A judgement is given in a `distinct` manner when its parts and their references to each other are explicitly given to us. The intentions indicated by its parts

[1]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.
[2]See [Lo] for a further discussion of this point.

may remain unfulfilled, but the compositional structure of the partial intentions is apparent. This explication proves certain intentions such as those of the form $P \wedge \neg P$ principally unexecutable.
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 unexecutability 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 unfullfilled. The everyday experience with the process of understanding mathematical arguments suggests that something corresponds to these notions in 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]

3.2 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 to 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,

[3] In our transfer, we ignore the level corresponding to what Husserl calls 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'. It is hard to imagine anything being considered 'proof' by anyone that is not subject to this condition.

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 converse.

But now for the two notions in question: One may, in a first go, think of 'distinctiveness' as a 'syntactical' category. Distinctiveness about an argument 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. A distinct argument does not need to be correct, not even its logical structure must be sound; for example, a distinct argument can well be circular. However, from a distincive perception of the argument, it will be apparent that it is. Clarity includes consciousness of a sequential structure of the argument. To some extent, it is necessary whenever we even attempt to formulate it in a language. [4]
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 exactly where each assumption enters the argument, where each auxiliar lemma is used etc. Often, in understanding natural proofs, we encounter some mixture of distinct deductive steps and imaginative thought experiments.

As for 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 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

[4]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.

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 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. 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, one might consider one of the many 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' the countability of $\mathbb{R}$ etc.

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 at least to a certain degree, 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 mathematical argument written up for human readers (see [Pr]) to its counterparts in the Naproche language.

Example: There are infinitely many primes.
Euclid considers any finite set S of primes. The key idea is to consider the product of all these numbers plus one:

$$N = 1 + \prod_{p \in S} p$$

Like any other natural number, N is divisible by at least one prime number (it is possible that N itself is prime).

None of the primes by which N is divisible can be members of the finite set S of primes with which we started, because dividing N by any of these leaves a remainder of 1. Therefore the primes by which N is divisible are additional primes beyond the ones we started with. Thus any finite set of primes can be extended to a larger finite set of primes. QED.

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

Let S be a finite set of prime numbers. Then there is a function P and a number r such that, for every n in S, there is a k such that $k \leq r$ and $P(k) = n$. Obviously, $\Pi_1^n P \neq 0$. Define N to be $\Pi_1^n P + 1$. N is nontrivial. So there is a prime number q such that q divides N.
Assume that q is in S. Then there is an i such that $1 \leq i \leq r$ and $q = P(i)$. $P(i)$ divides $\Pi_1^n P$. Then q divides 1 by lemma DivMin. Contradiction.
Thus S is not the set of prime numbers. QED.

We see some extra complications because Naproche at the moment allows finite products only for functions, not for sets, and hence the set S has to be turned into a function P first. Also, the claim of the first proof that N leaves a remainder of 1 by a member of S is now supported by an argument. (DivMin refers to a lemma earlier in the text stating that if $l|m$ and $l|m+n$, then $l|n$ - see section 2 above.)
The first complication can be easily overcome by providing Naproche with the notion in question. 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 'Define N to be $\Pi S + 1$' 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, leaving out heuristical and historical indications like 'Like any other number', 'Euclid considers' or 'The key idea is'. 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 will suffice.

To demonstrate this further, consider the common misjudgement that Euclid's proof shows that $\Pi P + 1$ is prime if P is an initial segment of the sequence of primes. Whoever believes this can easily alter the Naproche proof above by supplying the extra condition on S and replacing the final clause. Thereby, he obtains a presentation of his (wrong) train of thoughts sufficient for automatic checking without noticing the mistake, which would be spotted in the process of automatic checking.

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 tu 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.

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 use some mental representation ('image') of the objects under consideration. This representation is different from a formal definition and usually precedes it. [5] 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 in the exercises. The idea of a non-zero 'number' without a predecessor is apparently

[5] 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.

hard to accept.
Such preconceptions derived from a misinterpretation of mental images are a common source for 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. Meanwhile, we want to remark that this kind of perception of mathematical objects is all but a dispensible source of mistakes: In fact, it is eactly this ability that steers the process of proving and creating mathematics, thereby making the human mathematician so vastly superiour to any existinc automatic prover.

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 stating 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 implicite 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 [6]. But in established areas of mathematics, these definitions have

[6] 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.

stood the test of time, and even though, particularly in new areas, there are debates about the adequacy of definitions and while 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. [7]
Considering a distinct proof presentation, a good automatic prover will be able to draw from formal definitions what we draw from correct intuition. We make no claim on the question whether formal definitions can exhaust semantic content, and we don't need such a claim. This 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, this may well be a very different operation. However, the 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.

6 A Historical Examples

In this section, we apply the notions obtained above to a famous historical examples of false proofs. 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 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 occur in the margin between the degree of distinctiveness necessary for formalization and complete distinctiveness.

Example (Cauchy 1821)[8]
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 $s_n(x) := \Sigma_{i=1}^{n} f_i(x)$, $r_n(x) := \Sigma_{i=n+1}^{\infty} r_n(x)$. Also, let $\varepsilon > 0$. Then, as each f_i is continuous and finite sums of contiuous 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 $|r_n(x+a)| < \varepsilon$.

[7]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.

[8]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.

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 subtle way used in two different meanings. The dependence on a can only be eliminated if there is some M bigger then $N(\varepsilon, x+a)$ 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 'large enough number', then re-uses 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 Naproche-like system models a text can easily allow for such a convention. Apart from that, the text is certainly not lacking distinctiveness. It uses only very little natural language and not in any complicated way. It would be quite feasible to enrich the vocabularly 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.

6.1 Further Examples

We give some more advanced historical cases of false proofs that possibly fall under our analysis, without a thorough discussion. Namely, there is a (very natural and therefore frequently arising) way of considering them in which an unclarified intuition brings us to the belief that the argument is correct, while it really isn't. Generally, such examples are hard to find as the presentation of mathematics is usually a historically artificial development leading to the current state of affairs, ignoring mistakes and sideways. [MO2] contains a list of some more examples that might be considered. We would welcome further suggestions.

- A notorious source for failed proof attempts is Fermat's last theorem. However, most proofs discussed in [FF] are given by layman and either contain very elementary mistakes (like deducing $x = x'$ and $y = y'$ from $x + y = x' + y'$) or demonstrate a complete lack of understanding what a proof is altogether. There are a few interesting examples, however. Most noteworthy are two attempts by Lame and Cauchy, both of whom apparently worked with Gaussian integers and used unique prime factorization, failing to notice that the latter property fails over the complex integers. (See [Si] and [Ku].) This is a typical instance of a misperception of the kind mentioned above: Gaussian integers behave a lot like integers, including the fact that the notion of a prime and a prime decomposition makes sense for them. This easily misdirects one to take the property of uniqueness, which is closely tied to the existence in the case of real integers, for granted, giving rise to a completely distinct proof presentation, which is nevertheless wrong.

- Another example is apparently the original 'proof' of the four colour conjecture, which was considered to be correct for quite some time. Unfortunately, we have not yet been able to take a look at this.

- More recent examples are easily obtained, though - due to their deeper technical involvement - not as easily analyzed. In 2011, there were e.g. announcements of proofs for $P \neq NP$, the Collatz conjecture and even the inconsistency of Peano Arithmetic, all by distinguished mathematicians, all of which turned out to be flawed. We do not know these arguments well enough to tell whether they fit in our framework, but the responses pointing out the mistake indicate that they were indeed based on plausible misperceptions.

7 Conclusions and Further Work

We hope to have made it plausible that phenomenological consideration 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 argument' and the number of lines in a corresponding derivation. [9] A next step is hence to consider common elementary operations that are performed in supplying such proofs steps and give formal background theories to replace them. The goal of this is to make the automatic prover's activity more resemblant to a real 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 it would of course be much nicer not just to check the theorem, but also the argument. For this, the checker has to become 'pragmatically closer' to a 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.

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

[9] 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.

[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.

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

[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)

[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 Mathematicuqes pures et appliquees 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)

[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*

[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)