Author(s): Valmari, Antti

Title: Adapting Formal Logic for Everyday Mathematics

Year: 2022

Version: Published version

Please cite the original version:

Valmari, A. (2022). Adapting Formal Logic for Everyday Mathematics. In M. Cukurova, N. Rummel, D. Gillet, B. McLaren, & J. Uhomoibhi (Eds.), CSEDU 2022 : Proceedings of the 14th International Conference on Computer Supported Education. Vol. 2 (pp. 515-524). SCITEPRESS - Science and Technology Publications. CSEDU. https://doi.org/10.5220/0011063300003182

Adapting Formal Logic for Everyday Mathematics

Antti Valmari[a]
University of Jyväskylä, Faculty of Information Technology, PO Box 35, FI-40014 University of Jyväskylä, FINLAND
antti.valmari@jyu.fi

Keywords: High school mathematics, Elementary university mathematics, Higher order thinking skills

Abstract: Although logic is considered central to mathematics and computer science, there is evidence that teaching logic has not been a great success. We identify three issues where what is typically taught conflicts with what is needed by those who are supposed to apply logic. First, what is taught about the notion of implication often disagrees with human intuition. We argue that in some cases human intuition is wrong, and in some others teaching is to blame. Second, the formal concepts of logical consequence, logical equivalence and tautology are not the similar concepts that everyday mathematicians and computer scientists need. The difference is small enough to go unnoticed but big enough to cause confusion. Third, how to deal with undefined operations such as division by zero is left informal and perhaps fuzzy. These problems also harm development of computer tools for education. We present suggestions about how to address them in teaching.

1 INTRODUCTION

Logical skills are widely believed to be important for mathematics and computer science. For instance, please see (Bronkhorst et al., 2020), (Hammack, 2018) and (Association for Computing Machinery (ACM) and IEEE Computer Society, 2013).

Despite this, logic has never been a great success in computer science education. Among the 75 topics whose importance were surveyed by (Lethbridge, 2000), it is perhaps not a surprise that Specific programming languages, Data structures, Software design and patterns, and Software architecture were considered the four most important. More surprisingly, typical engineering mathematics topics such as Differential / integral calculus were near the opposite end. Logic was in the middle ground. More recent surveys have made similar findings. Please see (Niemelä et al., 2018) for a discussion.

The situation of logic in computer science is presented vividly in the Preface to 2003 edition of a textbook that was originally published commercially as (Reeves and Clarke, 1990). The 2003 edition was made, because of repeated demands from around the world for more copies of the book. However, it was not officially published but instead made freely available on the web, because "One is that no company, today, thinks it worth publishing … The publishers look around at all the courses which teach short-term skills rather than lasting knowledge and see that logic has little place, and see that a book on logic for computer science does not represent an opportunity to make monetary profits."

(Mathieu-Soucy, 2016) wrote on the situation of logic in mathematics: "This paper shows once again that extensive knowledge of formal logic is not necessary to do mathematics. However, what this research brings is the whole idea of alertness to logical characteristics, which is an interesting asset for mathematics students. … This brings us to expand our reflection to the teaching of logic: what kind of knowledge should be taught and in what way to promote students' understanding and diminish logical mistakes, in order to make logic courses as efficient as possible?"

Many concepts in formal logic have been developed for theoretical studies on what can be expressed and proven, not for actually carrying out proofs. As a result, some of them do not serve applications well. For instance, while 1 and $<$ have a fixed meaning in everyday mathematics, the formal notion of logical consequence assumes that their meanings can vary within the limits set by the assumptions.

Therefore, "if $x < 1$ then $x \leq 1$" is not a logical consequence. It fails, for instance, if the meanings of $<$ and $\leq$ are swapped. Instead, $x \leq 1$ is a logical consequence of the two formulas $x < y \leftrightarrow x \leq y \land \neg(x = y)$ and $x < 1$. It is denoted by $\{x < y \leftrightarrow x \leq y \land \neg(x = y), x < 1\} \models x \leq 1$. Indeed, formal logic has notation for "I derive $x \leq 1$ from the laws of real numbers and

[a] https://orcid.org/0000-0002-5022-1624

$$
\begin{aligned}
&x \times 0 = x \times (0+0) && \text{(because } 0 = 0+0\text{)}\\
&\quad\Longrightarrow\ x \times 0 = x \times 0 + x \times 0 && \text{(by the distributivity)}\\
&\quad\Longrightarrow\ 0 + (x \times 0) = x \times 0 + x \times 0 && \text{(by the existence of zero)}\\
&\quad\Longrightarrow\ 0 = x \times 0 && \text{(by the cancellation)}
\end{aligned}
$$

Figure 1: A proof from (Stefanowicz, 2014) page 14

$x < 1$" but not for "I derive $x \leq 1$ from $x < 1$, taking the laws of real numbers for granted".

So formal logic lacks practical notation for expressing reasoning chains. This has made many to use $\Rightarrow$ for the purpose. Figure 1 shows an example. However, when $\Rightarrow$ is taught, this meaning is almost never taught. Instead, $\Rightarrow$ is often taught to mean so-called material implication and sometimes to mean logical consequence. This causes confusion. The situation is made worse by the fact that some correct aspects of implication seem counter-intuitive to many at first.

In Section 2 we introduce material implication, implication as an operator for expressing a reasoning step or reasoning rule, and logical consequence. It is a main claim of this paper that one of the reasons why students have difficulties with logic is failure to clarify the differences between these three concepts.

Some difficulties with material implication are discussed in Section 3. First, Wason's famous selection task is used to argue that original human intuition on implication is not necessarily correct. However, after explanation, there is no disagreement about the correct solution. This is different from the next topic: principle of explosion. It is so central to everyday mathematics that it cannot be avoided. It is also so counter-intuitive that there has been a long debate related to it among researchers. Finally an example is analysed where "if … then …" cannot be intepreted as material implication but can as a reasoning rule, demonstrating that the difference between these concepts is important to understand.

Section 4 presents $\Rightarrow$ as a reasoning operator. Its meaning is explained in a way which we believe easier for students than contrasting $\Rightarrow$ to logical consequence. Our $\Rightarrow$ has proven suitable for educational software that checks reasoning (Valmari, 2021).

A different problem for both humans and educational software is that standard systems of logic lack machinery for undefined operations, such as division by 0. As a consequence, teaching about how to deal with them in reasoning is scarce and mostly consists of informal rules. This is unsatisfactory, because undefined operations are ubiquitous in mathematics and computer science. Section 5 addresses this problem.

The paper ends with a brief conclusions section. Throughout the paper we make recommendations about how these diffculties could be addressed in teaching.

2 IMPLICATION AND RELATED CONCEPTS

Let us denote the truth values with F (false) and T (true). Implication as a connective is a propositional logic connective that inputs two truth values and outputs a truth value according to the following table:

	F	T
F	T	T
T	F	T

It is also called *conditional*, *material implication* and *material conditional*. It is usually denoted with $\rightarrow$ (especially in theoretical sources) or $\Rightarrow$. We use $\rightarrow$.

The only case when material implication yields F is when its first argument yields T and the second argument yields F. That is, $P \rightarrow Q$ is logically equivalent to $\neg P \vee Q$. (That they are logically equivalent, is sometimes called the *rule of material implication*.)

The connective $\leftrightarrow$ or $\Leftrightarrow$ is called *biconditional*, *material equivalence* or *material biconditional*. We have $P \leftrightarrow Q$ if and only if $(P \rightarrow Q) \wedge (Q \rightarrow P)$.

The symbols $\Rightarrow$ and $\Leftrightarrow$ are also often used informally in a somewhat different meaning, to express a reasoning step that may be chained to other reasoning steps, like in Figure 1. On page 1680 (Bronkhorst et al., 2020) write "If people smoke or inhale particulate matter, then it will affect their health and thus shorten their life", and then represent it as "(smoking $\vee$ inhaling particulate matter) $\Rightarrow$ unhealthy $\Rightarrow$ shorter life". When solving a pair of equations, it is handy to write $2x + 2y = 8 \wedge 3x - 2y = 7 \Rightarrow 5x = 15 \Leftrightarrow x = 3$.

These symbols are also often used to express rules that justify such steps, such as $x < y \Leftrightarrow x \leq y \wedge x \neq y$.

In mathematics, "if $x < 1$ then $x \leq 1$" and "if $x \leq 1$ then $x < 1$" may be interpreted as material implications $x < 1 \rightarrow x \leq 1$ and $x \leq 1 \rightarrow x < 1$. The former yields T for all values of x, and the latter yields F when $x = 1$ and T for all other values of x. However, it is often more natural to interpret them as a correct reasoning rule $x < 1 \Rightarrow x \leq 1$ and an incorrect rule $x \leq 1 \Rightarrow x < 1$. The former is correct, because $x \leq 1$ is a mathematical consequence of $x < 1$. The latter is incorrect, because $x < 1$ is not a mathematical consequence of $x \leq 1$, since $x = 1$ is a counter-example.

In everyday mathematics, the difference between material implication and implication as a reasoning operator is often ignored or confused. For instance, (Hammack, 2018) writes on page 44: "In mathematics, whenever we encounter the construction *"If P, then Q,"* it means exactly what the truth table for $\Rightarrow$ expresses." This clearly refers to material implication. On the other hand, on page 57 we have "*"If P, then Q,"* is a statement. This statement is true if it's impossible for P to be true while Q is false. It is false if there is at least one instance in which P is true but Q is false." This is definitely not material implication. It resembles much more a reasoning rule.

Hammack's idea becomes understandable on page 56: "In mathematics, whenever $P(x)$ and $Q(x)$ are open sentences concerning elements x in some set X (depending on context), an expression of form $P(x) \Rightarrow Q(x)$ is understood to be the *statement* $\forall x \in X, P(x) \Rightarrow Q(x)$". That is, he always treats $\Rightarrow$ in a reasoning operator -like fashion, but for closed formulas it is equivalent to material implication, and he uses this fact to explain his approach. Around page 44, he tried to restrict P and Q to closed formulas, but failed. (A formula is closed if and only if every occurrence of every variable in it is quantified by $\forall$ or $\exists$. Hammack's "statement" is a slightly different notion, but this is not essential for the present discussion.)

We recommend using $\rightarrow$ for material implication and $\Rightarrow$ as a reasoning operator. In this notation (and : instead of ,), Hammack's idea is that $P(x) \Rightarrow Q(x)$ is correct if and only if $\forall x \in X : P(x) \rightarrow Q(x)$ holds.

This idea works well in the basic case. Indeed, we will show in Section 3.4 that it resolves a much debated paradox. On the other hand, it runs into trouble in some other cases. For instance, it is true that if x is non-negative, then if $x^2 \geq 1$ then $x \geq 1$. It might seem natural to express this as $x \geq 0 \Rightarrow (x^2 \geq 1 \Rightarrow x \geq 1)$. However, transforming it as above yields $\forall x : x \geq 0 \rightarrow (\forall x : x^2 \geq 1 \rightarrow x \geq 1)$. Because $(-1)^2 \geq 1$ but not $-1 \geq 1$, the closed formula $\forall x : x^2 \geq 1 \rightarrow x \geq 1$ yields F, making the translated formula false as a whole.

Also another difficulty remains. We will address it in Section 5. It can be illustrated with the following reasoning. Its conclusion is unacceptable, although each step matches the interpretation of $\Rightarrow$ discussed above: On real numbers $x > y$ if and only if not $x \leq y$. Furthermore, if $\frac{1}{x} > 0$, then $x > 0$. By contraposition, if $x \leq 0$, then $\frac{1}{x} \leq 0$. We can similarly derive that if $x \geq 0$, then $\frac{1}{x} \geq 0$. Because $0 \leq 0$, letting $x = 0$ we get both $\frac{1}{0} \leq 0$ and $\frac{1}{0} \geq 0$. So $\frac{1}{0} = 0$.

Yet another potential problem is confusing $\Rightarrow$ with the notion of *logical consequence* in formal logic. Roughly speaking, the latter means a consequence that holds for every possible interpretation of all proposition, constant, function and relation symbols other than $=$, for all value combinations of variables. The precise definition is too complicated to be shown here, because it needs the notions of signature, domain of discourse, structure, assignment, interpretation and model. Two (sets of) formulas are *logically equivalent*, if and only if in both directions, one is a logical consequence of the other.

The essential difference is that in everyday mathematics, constant symbols (for instance, 35 and π), function symbols (for instance, $+$ and $\sqrt{\ }$) and relation symbols (for instance, $\leq$) are assumed to have fixed meanings, while in formal logic their meanings may vary (with the exception that many authors fix the meaning of $=$ to its familiar meaning in mathematics). As was explained in Section 1, $x \leq 1$ is not a logical consequence of $x < 1$, because $\leq$ and $<$ can be given meanings and x a value such that $x < 1$ holds but $x \leq 1$ does not (for instance, swap their standard meanings and choose $x = 1$).

It seems that "logical consequence", "logically equivalent" and "tautology" are typically taught via truth tables. This matches their meaning in formal propositional logic, but leaves it open how to use them elsewhere. Because the idea of varying the meaning of 0, $+$ and $\leq$ is strange to students, and because no other phrase has been taught, this runs the risk of students thinking that, for instance, $x \leq 1$ is a logical consequence of $x < 1$. We recommend teaching that it is a *mathematical consequence*. When the meaning of 0, $+$, $\leq$ and so on may vary, we have "logical consequence", "logically equivalent" and "tautology"; and when their meaning is fixed to the standard meaning, we suggest using *mathematical consequence*, *mathematically equivalent* and *mathematical fact*.

Logical consequence is denoted with $\Gamma \models P$, where Γ is a set of formulas and P is a formula. For the time being, we define $P \Rightarrow Q$ as $\Gamma \cup \Upsilon \cup \{P\} \models Q$, where Γ specifies the properties of the mathematical system in question (such as real numbers) and additional symbols (such as $<$ in terms of $\leq$), and Υ lists the local assumptions (such as the specification of a case in a proof by cases). So $\Rightarrow$ and $\Leftrightarrow$ denote mathematical consequence and mathematical equivalence. "Reasoning step / rule" refer to applications of $\Rightarrow$ and $\Leftrightarrow$. Formally, step / rule are the same, but typically "step" is used about reasoning, and "rule" about what justifies it. We will fine-tune the definition and present it in a more student-friendly form in Section 4.

In formal logic, $\Gamma \vdash P$ denotes that P can be proven from Γ. *Modus ponens* is the rule that $\{P, P \rightarrow Q\} \vdash Q$. *Deduction theorem* says that if $\Gamma \cup \{P\} \vdash Q$, then $\Gamma \vdash P \rightarrow Q$. Both hold in standard systems of logic. Together they constitute a tight link between $\rightarrow$ and $\Rightarrow$, partially explaining why $\rightarrow$ and $\Rightarrow$ are often confused with each other. Modus ponens will remain appreciated throughout this paper, but we will find a problem with Deduction theorem in Section 5.

As evidence to the fact that $\Rightarrow$ may be confused with both $\rightarrow$ and $\models$, we cite the Wikipedia page on Logical equivalence: "The logical equivalence of p and q is sometimes expressed as $p \equiv q$," … "or $p \iff q$, depending on the notation being used. However, these symbols are also used for material equivalence, so proper interpretation would depend on the context. Logical equivalence is different from material equivalence, although the two concepts are intrinsically related."

3 PROBLEMS WITH IMPLICATION

3.1 Wason's selection task

Wason's selection task (Wason, 1968) is a famous psychological experiment on material implication. The literature discussing it is extensive. We used (Ragni et al., 2017) as our source.

The original experiment was of the following kind. There are four cards on the table. The participants are told that each card has a letter on one side and a number on the opposite side. The visible sides of the cards on the table show D, K, 3 and 7. The participants are asked to choose as few cards as possible so that by looking at their hidden sides one can check whether or not the four cards obey the following rule:

> If a card has D on one side, then it has 3 on the other side.

The only way to violate this rule is to have D on the letter side and some other number than 3 on the number side. So the correct answer is D and 7.

Please notice that in this task, implication occurs in a bare-boned form. The task does not depend on whether implication is thought of as a connective or as a reasoning operator. Those issues do not arise where material implication is an inappropriate formalization of implication in natural language.

In the original experiment, less than 10 % of the participants chose the right cards. In the meta-analysis by (Ragni et al., 2017), 19 % solved a similar task correctly; 36 % chose only the equivalent of D, 39 % chose the equivalent of D and 3, and 5 % chose the equivalent of D, 3 and 7.

However, the results improve dramatically, if the abstract letters and numbers are replaced by concrete familiar items. For instance, one side of each card could contain the name of a drink (beer or orange juice), the opposite side the age of a person (16 or 25), and the rule could be

> If a person is drinking beer, then the person must be over 19 years of age.

Now in the meta-analysis, 64 % chose the right cards, 13 % chose only the equivalent of beer, 19 % chose the equivalent of beer and 25, and 4 % chose the equivalent of beer, 16 and 25.

(Ragni et al., 2017) have recognized 15 distinct theories that aim at explaining this or related results. We believe that the result can be explained by the idea of two modes of human thinking: a fast instinctive mode and a slower more rational mode (Kahneman, 2011). Apparently the fast mode of many people lacks correct treatment of even the bare-boned material implication in its abstract form, while has rules for concrete applications in familiar situations such as age and alcohol.

On the other hand, the slower mode seems to master the bare-boned version. In the words of (van Benthem, 2008): "A psychologist, not very well-disposed toward logic, once confessed to me that despite all problems in short-term inferences like the Wason Card Task, there was also the undeniable fact that he had never met an experimental subject who did not understand the logical solution when it was explained to him, and then agreed that it was correct."

We suggest telling students about results on Wason's selection task or something similar, to make them realize that their original intuition may lead them astray. This may encourage them to learn, understand, trust and use the laws of propositional logic instead.

3.2 The debate on Carroll's paradox

In (Carroll, 1894), the following question was presented in this form and in the form of a story:

> There are two Propositions, *A* and *B*. It is given that
>
> 1. If *C* is true, then, if *A* is true, *B* is not true;
> 2. If *A* is true, *B* is true.
>
> The question is, can *C* be true?

In the story there was a barbershop with three barbers, Allen, Brown and Carr. At least one of them is not out, giving rise to (1), where *A* denotes that Allen is out, and similarly with *B* and *C*. Allen is very shy and does not go out without Brown, hence (2).

The barbershop paradox is not a problem for modern propositional logic. It is possible that *C* is true and *A* is false (and *B* may be either). So *C* can be true.

It is, however, worth noticing that it definitely was a problem at its time. In the story, Uncle Joe reasoned that if *C* is true, then we have simultaneously "if *A* is true, *B* is not true" and "if *A* is true, *B* is true", which is a contradiction, because the same premise *A* leads simultaneously to the conflicting conclusions *B* and not-*B*. Since this contradiction was obtained by assuming *C*, the principle of reductio ad absurdum yields that *C* cannot be true. (Carroll, 1894) mentioned that the dispute had already lasted more than a year, with conflicting opinions by several practised logicians. Indeed, John Cook Wilson, the Wykeham Professor in Logic of the University of Oxford, held Uncle Joe's view (Moktefi, 2007). The dispute continued for more than a decade (Jones, 1905).

Uncle Joe was wrong in claiming that "if A is true, B is not true" and "if A is true, B is true" cannot hold simultaneously. Instead, that happening means simply that A is not true. However, this forces us to accept that a false claim may simultaneously imply a claim and its negation, because A, B and not-B provide an example. We will return to this observation in the next subsection.

Because material implication was not easy for professional logicians, we should not expect it to be easy for students either.

The barbershop paradox can also be used to illustrate the power of formal manipulation. The premises (1) and (2) can be represented as $(C \rightarrow (A \rightarrow \neg B)) \wedge (A \rightarrow B)$. Replacing each $P \rightarrow Q$ by $\neg P \vee Q$ results in $(\neg C \vee \neg A \vee \neg B) \wedge (\neg A \vee B)$, which simplifies to $\neg A \vee (B \wedge \neg C)$. So either Allen is in (and Brown and Carr may be anywhere), or Brown is out and Carr is in (and Allen may be anywhere).

3.3 Principle of explosion

Material implication is *truth-functional*, that is, its result is a truth value which depends on nothing else than the incoming truth values. This sometimes clashes with the feeling by many people that for "if P then Q" to be true, Q must somehow depend on P. For instance, "if it is not sunny tomorrow, I will stay at home" sounds sensible, while "if I do not stay at home tomorrow, it will be sunny" sounds odd, because it seems to suggest that I could cause sunshine by leaving home. However, as material implications, they are logically equivalent.

This has led to a vast body of research; see (Egré and Rott, 2021) for an up-to-date survey. Many approaches have rejected truth-functionality at the cost of making the logic more complicated. However, following that path would take us far from everyday mathematical practice.

Instead, we keep material implication, and suggest to emphasize students that it does not capture every aspect that our intuitive notion of implication may cover. In particular, material implication does not pay attention to what is the cause and what is the effect. It only deals with what combinations of truth values are possible and what are impossible. If $P \rightarrow Q$, then P true and Q false is impossible, and the remaining three combinations are possible (unless something else makes them impossible). That is all.

It is also worth emphasizing students that within its scope, material implication is reliable, while intuition tends to give wrong results every now and then. The students may also be told that there have been attempts to find notions of implication that match intuition better, but the results have not been good enough to replace material implication.

A related problem with intuition arises if P cannot be true or Q cannot be false, because then $P \rightarrow Q$ yields T, although there is not necessarily any sensible connection between P and Q. For instance, the following are true:

- If Earth is flat, then nobody likes coffee.
- If Earth is flat, then every natural number can be expressed as a sum of three squares of natural numbers.

They are true because the only way to make $P \rightarrow Q$ yield F is to make P true and Q false, but "Earth is flat" cannot be made true. On the other hand, "nobody likes coffee" and "... sum of three squares ..." do not depend on "Earth is flat". This makes the above examples seem counter-intuitive.

Indeed, in standard logic, any false claim implies just anything. This is called the *principle of explosion*. It underlies proof by contradiction and proof by contrapositive, which are central in everyday mathematics. In the experience of the present author, this principle is difficult for many students. Apparently it was difficult also for John Cook Wilson. The barbershop paradox illustrates that it is necessary to accept that a false claim may simultaneously imply a claim and its negation. Although accepting it does not necessarily mean accepting the principle of explosion in its full generality, it is at least a long step in that direction.

Now assume that a person who is held in a secure prison says "if it rains tomorrow, I will stay in prison". The sentence is true, but sounds ironic, because the prisoner cannot leave the prison, no matter what the weather is. This is an example of material implication that is true not because of a sensible connection between P and Q, but because Q cannot be made false. Indeed, in standard logic, any true claim is implied by just anything. This principle can be thought of as a dual to the principle of explosion.

We recommend teaching students to rely on the idea that a claim is true if and only if it has no counter-examples. Similarly, a reasoning rule is incorrect if and only if it has a counter-example. A counter-example consist of a combination of values of variables that is allowed by the assumptions made in the context where the claim or reasoning rule is stated, and makes the claim false or rule incorrect. For instance, in the case of real numbers, $x = \frac{1}{2}$ and $y = 0$

is a counter-example to $x > y \rightarrow x \geq y+1$. However, in the case of integers, x cannot be $\frac{1}{2}$. Indeed, $x > y \rightarrow x \geq y+1$ is true on integers.

If P cannot be true, then we cannot have P true and Q false. So $P \rightarrow Q$ has no counter-examples. That is, the principle of explosion agrees with the idea that a claim is true if and only if it has no counter-examples.

3.4 Confusion with Reasoning Operator

This example has been modified from the Wikipedia page on Paradoxes of material implication. The following is assumed to hold:

3. If John is in London then he is in England.

If "if ... then ..." is interpreted as material implication, then the following can be proven, although it seems plain wrong:

4. If John is in London then he is in France, or if he is in Paris then he is in England.

To prove it, let (5) and (6) denote "if John is in London then he is in France" and "if he is in Paris then he is in England", respectively. Now (4) can be rewritten as ((5) or (6)). If "John is in London" is true, then by (3) also "he is in England" is true. Then by the dual to the principle of explosion, also (6) is true. If "John is in London" is not true, then by the principle of explosion, (5) is true. So no matter where John is, at least one of (5) and (6) is true. Therefore, ((5) or (6)) is true. Q.E.D.

A perhaps even more striking example is "if John is in London then he is in Paris, or if he is in Paris then he is in Brussels." Namely, if John is in Paris then the first implication holds, and if he is not in Paris then the second implication holds. Indeed, no matter what P, Q and R are, $(P \rightarrow Q) \vee (Q \rightarrow R)$ yields $\top$.

Hammack's idea in Section 2 chases this paradox away. That is, each material implication $P \rightarrow Q$ gives rise to the reasoning rule $P \Rightarrow Q$. The rule $P \Rightarrow Q$ is correct if and only if $P \rightarrow Q$ yields $\top$ for all interpretations of P and Q that are possible in the context. In other words, $P \Rightarrow Q$ is incorrect if and only if the context allows at least one interpretation that makes P true and Q false. This is similar to mathematics, where "if $x \geq 1$ then $x > 1$" is doomed incorrect by the fact that it fails when $x = 1$, although it works okay for all other values of x.

When understood as a reasoning rule, (3) is correct by assumption. However, (5) and (6) are incorrect, because the real Europe is a counter-example that is allowed by (3). Furthermore, (4) must be interpreted as saying that (5) is a correct reasoning rule or (6) is a correct reasoning rule. Under this interpretation, (4) is indeed wrong, matching intuition.

More formally, let $x \in L$ denote that John is in London, $x \in E$ that he is in England, and so on. The reasoning rule interpretation of (4) yields $(\forall x : (x \in L \rightarrow x \in F)) \vee (\forall x : (x \in P \rightarrow x \in E))$, while the material implication interpretation and the proof of (4) only provide $\forall x : ((x \in L \rightarrow x \in F) \vee (x \in P \rightarrow x \in E))$.

This resolution of the paradox suggests that people do not think of each "if John is in x then he is in y" as material implication but as a reasoning rule. It resembles the *strict implication* in Section 3.1 of (Egré and Rott, 2021), but aims at less deviation from standard logic.

Unfortunately, we saw in Section 2 that this idea does not always work. Therefore, we recommend to teach that "if ... then ..." may translate to material implication or to a reasoning rule, and the student has to choose the right one. If neither one matches intuition, then the student should perhaps ask for clarification on the intended meaning of the sentence. If it translates to a reasoning rule but is part of a bigger formula like ((5) or (6)) above, then $\forall$ may have to be added as was illustrated above. It is a task of the student to choose how many $\forall$ are added and where they are added to capture the intended meaning. No simple general rule can be given, because different choices are needed by (4) and by the "if x is non-negative, then if $x^2 \geq 1$ then $x \geq 1$" example in Section 2.

4 REASONING OPERATORS

In this section we describe the meaning of $\Rightarrow$, $\Leftrightarrow$ and $\Leftarrow$ as reasoning operators. The ideas have been developed from (Valmari and Hella, 2017).

A *reasoning chain* is of the form $P_0 \sim_1 P_1 \sim_2 \ldots \sim_n P_n$, where $n \geq 1$, each P_i is a formula and each $\sim_i$ is from the set $\{\Rightarrow, \Leftrightarrow, \Leftarrow\}$. The same reasoning chain may not contain both $\Rightarrow$ and $\Leftarrow$. Each $P_{i-1} \sim_i P_i$ is a *reasoning step*.

Each reasoning chain occurs in a *context*. It specifies those properties of proposition, constant, variable, function and relation symbols that are allowed to be used in writing formulas and reasoning. To write formulas one needs to know such thing as $+$ denotes a function from a pair of real numbers to a real number; variable x contains a real number; and variable $\vec{v}$ contains a 3-dimensional vector. To reason one needs to know that if Allen is out then Brown is out as well; and for every x we have $x+0 = x$. The specification need not be exhaustive. For instance, the barbershop paradox allows 5 and rules out only 3 combinations of locations of persons. As another example, the group axioms have infinitely many different models.

The context is not a concrete piece of text but an abstract entity. Properties that the reader may be expected to already know, need not be mentioned explicitly. For instance, if it is said that the domain of discourse is real numbers, then the familiar properties of 0, π, $+$, cos and so on are automatically available.

Each proposition, constant, function and relation symbol has the same meaning in every formula to which the same context applies, and $\forall$ and $\exists$ cannot be applied to them. The values of variables are not specified by the context. They may vary between formulas, and within a formula as determined by $\forall$ or $\exists$. Information may be temporarily added to the context by such phrases as: "to derive a contradiction assume that p is not a prime number", "consider first the case that p is even" and "so there is an integer $k \geq 2$ such that $p = 2k$".

The difference between reasoning steps and material implication is easier to keep in mind, if we do not think of the former returning a truth value, but being correct or incorrect. A reasoning step is *correct* if and only if it has no counter-examples. A *counter-example* to $P \Rightarrow Q$ is any combination of variable values that is allowed by the context such that P yields T but Q does not yield T. The step $P \Leftarrow Q$ is correct if and only if $Q \Rightarrow P$ is correct. The step $P \Leftrightarrow Q$ is correct if and only if both $P \Rightarrow Q$ and $P \Leftarrow Q$ are correct. That is, for each value combination of variables that is allowed by the context, either both P and Q yield T, or neither of them yields T. A reasoning chain is correct if and only if every step in it is correct.

A *reasoning rule* is a reasoning step. Typically "rule" is used for steps whose instantiations (for example, $2n+1$ in place of x) are intended for later use.

By the above definition, $P \Rightarrow Q \Rightarrow R$ is correct if and only if both $P \Rightarrow Q$ and $Q \Rightarrow R$ are correct, and similarly with $P \Rightarrow Q \Leftrightarrow R$, and so on. Figure 1 illustrates that this is how $\Rightarrow$ is used by many. Because reasoning operators occur *between*, not in, formulas, $P \Rightarrow (Q \Rightarrow R)$ is a syntax error and thus means nothing. This is analogous to common practice with relation symbols in mathematics, where, for instance, $0 \leq x < 1$ means $0 \leq x \wedge x < 1$ and $x \in A \subseteq B$ means $x \in A \wedge A \subseteq B$; and $0 \leq (x < 1)$ means nothing. It saves us from having to decide whether $\neg(x \geq 0 \Rightarrow x \geq 1)$ means $0 \leq x < 1$ or that $x \geq 0 \Rightarrow x \geq 1$ is an incorrect reasoning step.

It would be analogous to interpret $P \rightarrow Q \rightarrow R$ as $(P \rightarrow Q) \wedge (Q \rightarrow R)$, but few, if any, do so. Instead, many interpret it as $P \rightarrow (Q \rightarrow R)$, some interpret it as $(P \rightarrow Q) \rightarrow R$, and many reject it as ambiguous because of lacking (and). All these interpretations are different. For instance, if P is $x \geq 0$, Q is $x^2 \geq 1$ and R is $x \geq 1$, then

formula	yields T
$(P \rightarrow Q) \wedge (Q \rightarrow R)$	when $-1 < x < 0 \vee x \geq 1$
$P \rightarrow (Q \rightarrow R)$	for every x
$(P \rightarrow Q) \rightarrow R$	when $x \geq 0$

That is, $\rightarrow$ is treated like $+$ and $\cup$ in mathematics. It is also how other binary logical connectives are treated. Indeed, $P \vee Q \vee R$ definitely does not mean the same as $(P \vee Q) \wedge (Q \vee R)$!

We observe that in literature, when used as a logical symbol, $\rightarrow$ almost always denotes material implication, while the use of $\Rightarrow$ is somewhat fuzzy but has aspects of a reasoning operator. This is why we recommend teaching $\rightarrow$ and not $\Rightarrow$ as the material implication, making it clear what reasoning operators are, and teaching $\Rightarrow$ as a reasoning operator.

We recall from Section 2 that $\Rightarrow$ as a reasoning operator is different from logical consequence, because it uses the standard meaning of 0, $+$, $\leq$ and so on, while logical consequence uses every meaning allowed by Γ. Therefore, to students whose interest is in applying logic, $\Rightarrow$ is both more useful and much easier to teach than logical consequence.

In this framework, solving an equation, inequation or a system of them consists of deriving a formula that is mathematically equivalent to the original system and shows explicitly the value combinations of variables that make the formula yield T. We have already started an example by deriving $2x+2y = 8 \wedge 3x-2y = 7 \Rightarrow x = 3$. We may continue $3x-2y = 7 \wedge x = 3 \Rightarrow 9-2y = 7 \Leftrightarrow y = 1$ and $2x+2y = 8 \wedge 3x-2y = 7 \Leftrightarrow x = 3 \wedge y = 1$. That there are no roots is expressed by F, and that every value combination is a root by T. For instance, with real numbers, $x^2+1 = 0 \Leftrightarrow \mathsf{F}$.

We believe that many students learn solving (systems of) (in)equations first as a mechanical procedure that they follow without asking why it yields the correct roots. Seeing that it is actually application of more general logical reasoning may widen their understanding and promote high-level thinking skills.

5 DEALING WITH UNDEFINED EXPRESSIONS

In Section 2 we presented a fake proof that $\frac{1}{0} = 0$. It was based on ignoring the fact that $\frac{1}{0}$ is undefined. Standard systems of logic assume that every application of each function symbol is defined, and are thus unable to deal with undefined operations. Textbooks on mathematics may warn about them, but it is hard to find a systematic discussion. Compared to the amount of effort devoted to teaching the zero product property or the truth table of $\rightarrow$, little is done to give students

$$
\begin{array}{lcl}
 & 3\sqrt{|x|-1} = x+1 & \\
\Leftrightarrow & (x<0 \wedge 3\sqrt{-x-1} = x+1) \;\vee\; (x \geq 0 \wedge 3\sqrt{x-1} = x+1) & x<0 \text{ and } |x| = -x \text{ or } x \geq 0 \text{ and } |x| = x \\
\Leftrightarrow & (x<0 \wedge x=-1) \;\vee\; (x \geq 0 \wedge (x=2 \vee x=5)) & \text{solve each equation elsewhere} \\
\Leftrightarrow & x=-1 \vee x=2 \vee x=5 & \text{combine roots}
\end{array}
$$

Figure 2: A solution via analysis by cases presented in logical notation

tools for avoiding such errors as in our proof of $\frac{1}{0}=0$.

Some mathematicians, logicians and computer scientists have taken this problem seriously. As a matter of fact, there is a debate on how undefined expressions should be treated. Unfortunately, none of the earlier solutions seems to match everyday mathematical thinking. Therefore, (Valmari and Hella, 2017) presented an initial version of a system of our own. Since then it has been given a sound and Gödel-complete proof system and compared to many other solutions, please see (Valmari and Hella, 2021).

Perhaps the first (insufficient) idea is that when using $\frac{1}{x}$, the domain of discourse is not $\mathbb{R}$ but $\mathbb{R}\setminus\{0\}$; and with $\sqrt{x}$, it is $\{x \in \mathbb{R} \mid x \geq 0\}$. It runs into trouble with the reasoning in Figure 2. It shows the splitting of an equation to two cases and the combination of the results of the cases, with the cases solved via other reasoning chains that are not shown. The idea makes the domain of discourse of the second line be the empty set.

Despite this, the idea works well and is widely used with arithmetic comparison chains such as $\frac{x^2-9}{x^2+x-6} = \frac{(x+3)(x-3)}{(x+3)(x-2)} = \frac{x-3}{x-2}$. In the absence of a remark to the effect that $x \neq -3$, the second $=$ is widely considered incorrect, but the first $=$ is often accepted. The unconscious rule seems to be that when both sides of a comparison are undefined for precisely the same combinations of values of variables, then the remark need not be made. This does not mean, however, treating undefined as equal to itself, since mathematicians do not accept 0 as a root of $\frac{1}{2x}+x=\frac{1}{x}$.

In logic, a *term* is an expression whose value is in the domain of discourse (as opposed to a truth value). Many computer scientists entertain the idea that every undefined term yields a value from the domain of discourse, but we do not necessarily know what value (Gries and Schneider, 1995). This approach is not Gödel-complete. It makes it unknown whether 0 is a root of $\frac{1}{x}=x$, while in everyday mathematics it is definitely not a root. Furthermore, it makes $\frac{1}{0} \in \mathbb{R}$ and thus $\frac{1}{0}+0=\frac{1}{0}$, so 0 becomes a root of $\frac{1}{2x}+x=\frac{1}{x}$.

In our fake proof that $\frac{1}{0}=0$, the error took place in the step from "if $\frac{1}{x}>0$, then $x>0$" to "if $x \leq 0$, then $\frac{1}{x} \leq 0$". There $\frac{1}{0} \leq 0$ was derived from $\neg(\frac{1}{0}>0)$. However, mathematicians consider both $\frac{1}{0}>0$ and $\frac{1}{0} \leq 0$ as not true. This suggest that an undefined formula is neither false nor true. Indeed, many people share this idea. For instance, when over 200 software developers were asked how various undefined situations should be interpreted, between 74 % and 91 % chose "error/exception" instead of "true", "false", and "other (provide details)" (Chalin, 2005).

Therefore, we introduce a third truth value U (undefined) and have that $\neg\mathsf{U}$ yields U. At least three options for $\wedge$ and $\vee$ have been suggested in the literature. The one that matches the intention in Figure 2 is obtained by thinking of F less true than U which is less true than T, and letting $\wedge$ and $\forall$ pick the minimum and $\vee$ and $\exists$ the maximum of the truth values of their arguments. A relation yields U if and only if at least one of its arguments is undefined.

For $3\sqrt{|x|-1} = x+1 \Leftrightarrow x=-1 \vee x=2 \vee x=5$ to hold when $x=0$, we have to accept that $\mathsf{U} \Leftrightarrow \mathsf{F}$. To not lose the symmetry of $\Leftrightarrow$, we also have $\mathsf{F} \Leftrightarrow \mathsf{U}$. To avoid reasoning $\mathsf{U} \Leftrightarrow \neg\mathsf{U} \Leftrightarrow \neg\mathsf{F} \Leftrightarrow \mathsf{T}$, we no longer let $P \Leftrightarrow Q$ give permission to replace P by Q in a bigger formula $R(P)$ of which P is a part.

Instead, we introduce a new reasoning operator $\equiv$ that does not treat U as equivalent to F. That is, $P \equiv Q$ if and only if for every value combination of variables that is allowed by the context, P and Q yield the same truth value. If $P \equiv Q$, then $P \Leftrightarrow Q$ and $R(P) \equiv R(Q)$, and thus $R(P) \Leftrightarrow R(Q)$.

We let every function symbol f have a corresponding formula $\lfloor f \rceil$ that specifies when f is defined. For instance, with real numbers, $\lfloor \sqrt{x} \rceil$ is $x \geq 0$, $\lfloor \frac{x}{y} \rceil$ is $y \neq 0$ and $\lfloor x+y \rceil$ is T, denoting that $x+y$ is always defined. To ensure that $\lfloor f \rceil$ itself is always defined, we require that if it mentions any function symbol g, then $\lfloor g \rceil$ is T.

This idea extends naturally to terms and formulas. For instance, $\frac{f}{g}$ is defined if and only if both f and g are defined and g does not yield 0. That is, $\lfloor \frac{f}{g} \rceil$ is $\lfloor f \rceil \wedge \lfloor g \rceil \wedge g \neq 0$. Furthermore, $\lfloor f \leq g \rceil$ is $\lfloor f \rceil \wedge \lfloor g \rceil$, $\lfloor \neg P \rceil$ is $\lfloor P \rceil$, and $\lfloor P \wedge Q \rceil$ is $(\lfloor P \rceil \wedge \lfloor Q \rceil) \vee (\lfloor P \rceil \wedge \neg P) \vee (\lfloor Q \rceil \wedge \neg Q)$. For any formula P and combination of values of variables, precisely one of P, $\neg P$ and $\neg\lfloor P \rceil$ yields T.

Assume that $P(Q)$ is a formula that contains no other logical connectives or quantifiers than $\neg$, $\wedge$, $\vee$,

$\forall$ and $\exists$, and has Q as a sub-formula. If Q is within the scope of an even number of $\neg$, then $P(Q) \Leftrightarrow P(\lfloor Q \rceil \wedge Q)$, and otherwise $P(Q) \Leftrightarrow P(\neg \lfloor Q \rceil \vee Q)$. This facilitates reduction of problems to a form where nothing is undefined, and then solving them using classical two-valued logic. For instance, $\frac{x^2-9}{x^2+x-6} = 2$

$\Leftrightarrow\ x^2+x-6 \neq 0 \wedge x^2-9 = 2(x^2+x-6)$
$\Leftrightarrow\ x^2+x-6 \neq 0 \wedge (x = -3 \vee x = 1)$
$\Leftrightarrow\ x = 1.$

It is also correct to reason $\frac{x^2-9}{x^2+x-6} = 2 \Rightarrow x^2-9 = 2(x^2+x-6) \Leftrightarrow x = -3 \vee x = 1$ and then check the roots, rejecting -3 and accepting 1.

This idea has been used in computer science, and also many mathematicians seem to use it. Its correctness is not trivial to prove. Furthermore, it fails if a connective $*$ is allowed such that $*\mathsf{F} \equiv *\mathsf{T} \equiv \mathsf{T}$ and $*\mathsf{U} \equiv \mathsf{F}$. Please notice that $\lfloor\ \rceil$ is not such a connective, because it is not a connective, because it acts on formulas instead of truth values.

Familiar laws on the domain of discourse remain correct, such as $x - x = 0$ and $0x = 0$; but an issue arises regarding their use: clearly $\frac{1}{x} - \frac{1}{x} = 0$ and $0 \cdot \frac{1}{x} = 0$ are not correct when $x = 0$. Therefore, we have to teach students to distinguish between values and terms. Constant and variable symbols (such as 1 and x) denote values and are always defined. However, when a more complicated term (such as $\frac{1}{x}$) is put in the place of x (this is called instantiation of x with $\frac{1}{x}$), it is not necessarily always defined.

Fortunately, there are special cases where everything works like in classical two-valued logic, and there are theorems that tell how to deal with many of the remaining cases. We cannot cover here the topic in full, but we try to give a feeling.

Let f and g be terms and $P(x)$ a formula. If for each variable value combination that is allowed by the context either $f = g$ or both f and g are undefined, then $P(f) \equiv P(g)$. Therefore, if $f = g$ has been promised under the convention mentioned earlier without any remark about the domain, we need not worry about the domain. For instance, we may replace $\frac{(x+3)(x-3)}{(x+3)(x-2)}$ for $\frac{x^2-9}{x^2+x-6}$ even if x may be -3.

In particular, the condition holds automatically with instantiations of $=$-laws whose both sides contain precisely the same variables. So the law may be instantiated without worrying whether what goes in the place of the variables is defined. For instance, $x + y = y + x$ is such, but $x - x = 0$ is not.

Unfortunately, the same does not apply to $\Leftrightarrow$- and $\Rightarrow$-laws. For instance, the rule that a product is 0 if and only if at least one of its factors is 0, is vulnerable to mis-reasoning $x \cdot \frac{1}{x} = 0 \Leftrightarrow x = 0 \vee \frac{1}{x} = 0 \Leftrightarrow x = 0$. We recommend teaching that a product is 0 if and only if at least one of its factors is 0 *and all factors are defined*. This is different from the law $x > y \Leftrightarrow \neg(x \leq y)$, which does not need a similar remark. Even $x > y \equiv \neg(x \leq y)$ holds unconditionally.

To deal with such cases as $\frac{x}{x} = 1$ or $\frac{1}{x} - \frac{1}{x} = 0$, there is a theorem that assumes that the logic has a property known as *regularity*. A sufficient but not necessary condition for regularity is that there are no other connectives and quantifiers than $\neg$, $\wedge$, $\vee$, $\forall$ and $\exists$. Then $\lfloor f \rceil \Rightarrow f = g$ implies $P(f) \Rightarrow P(g)$.

In addition to the above, the presence of U affects propositional reasoning. For instance, the law of contraposition now has many forms, including "if $P \Rightarrow Q$, then $\neg Q \vee \neg \lfloor Q \rceil \Rightarrow \neg P \vee \neg \lfloor P \rceil$" and "if $P \Rightarrow \neg Q \vee \neg \lfloor Q \rceil$, then $Q \Rightarrow \neg P \vee \neg \lfloor P \rceil$". Of course, $P \vee \neg P \Leftrightarrow \mathsf{T}$ must be replaced by $P \vee \neg P \vee \neg \lfloor P \rceil \equiv \mathsf{T}$. The law $P \wedge \neg P \Leftrightarrow \mathsf{F}$ still holds, but only because $\mathsf{U} \Leftrightarrow \mathsf{F}$; we do not have $P \wedge \neg P \equiv \mathsf{F}$. We have $P \wedge \neg P \wedge \lfloor P \rceil \equiv \mathsf{F}$. These effects are mostly easy to take into account in reasoning. Furthermore, almost all widely mentioned propositional laws that do not use $\rightarrow$ or $\leftrightarrow$ are valid also in the presence of U.

An interesting exception is that $P \wedge (\neg P \vee Q)$ and $P \vee (\neg P \wedge Q)$ are no longer equivalent to $P \wedge Q$ and $P \vee Q$. If P yields F, U and T, then $P \wedge (\neg P \vee Q)$ yields F, U and the same as Q yields, respectively. This is exactly how the "and" operator of many progamming languages behaves, if U is interpreted as program crash. Three-valued logic thus naturally represents a phenomenon that is important in progamming but ignored by classical two-valued logic.

There are two main conventions of what $\rightarrow$ should mean in three-valued logic: one by Jan Łukasiewicz and another by Stephen Cole Kleene. Modus ponens holds for both, but Deduction theorem for neither. Neither of them matches $\Rightarrow$, because $\mathsf{U} \Leftrightarrow \mathsf{F}$. Łukasiewicz's convention is not regular but Kleene's is, since it makes $P \rightarrow Q$ equivalent to $\neg P \vee Q$.

Let f be free for x in $P(x)$. The law $\forall x : P(x) \Rightarrow P(f)$ must be replaced by $\lfloor f \rceil \wedge \forall x : P(x) \Rightarrow P(f)$. If the logic is not regular, then $P(f) \Rightarrow \exists x : P(x)$ must be replaced by $\lfloor f \rceil \wedge P(f) \Rightarrow \exists x : P(x)$.

There is strong evidence that the phenomena we discussed above, cover all deviations from familiar reasoning rules. Namely, so-called first-order logic covers much of the use of predicate logic. Unlike higher-order logics, it is *complete* in the sense that all logical consequences can be proven. (Valmari and Hella, 2021) proved that also our three-valued first-order logic is complete. Furthermore, only five proof rules of our complete proof system differ from the corresponding system for classical two-valued logic: the above-mentioned quantifier rules, $\emptyset \vdash P \vee \neg P \vee \neg \lfloor P \rceil$, $\{P\} \vdash \lfloor P \rceil$, and $\{\lfloor f \rceil\} \vdash f = f$.

6 CONCLUSIONS

The success of teaching of formal logic as a practical tool has been mediocre. We pointed out issues that may be problematic for students.

We suggested showing students examples where intuition has led people astray, such as Wason's selection task or the debate on Carroll's paradox. It may help them reject incorrect intuitive rules of reasoning in favour of rules of formal logic.

We suggested teaching students the following: Trust on the principle that "if ... then ..." holds if and only if it has no counter-examples. The principle of explosion follows from this, so accept it, although it may seem counter-intuitive at first. "If ... then ..." may mean material implication or a reasoning rule. Use $\forall$ appropriately to capture the intended meaning. Denote material implication with $\rightarrow$, and the similar reasoning rule with $\Rightarrow$. Use the phrases "mathematical consequence" and so on instead of "logical consequence" and so on, because the latter do not assume that 0, $+$, $\leq$ and so on have their standard meaning.

We presented conventions for reasoning operators and the treatment of undefined operations. They have proven well-defined and rigorous enough to be used in educational software written by us (Valmari, 2021).

REFERENCES

Association for Computing Machinery (ACM) and IEEE Computer Society (2013). *Computer Science Curricula 2013: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science*. ACM, New York, NY, USA.

Bronkhorst, H., Roorda, G., Suhre, C., and Goedhart, M. (2020). Logical reasoning in formal and everyday reasoning tasks. *International Journal of Science and Mathematics Education*, 18:1673–1694.

Carroll, L. (1894). A logical paradox. *Mind*, 3(11):436–438.

Chalin, P. (2005). Logical foundations of program assertions: What do practitioners want? In Aichernig, B. K. and Beckert, B., editors, *Third IEEE International Conference on Software Engineering and Formal Methods (SEFM 2005), 7-9 September 2005, Koblenz, Germany*, pages 383–393. IEEE Computer Society.

Egré, P. and Rott, H. (2021). The Logic of Conditionals. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2021 edition.

Gries, D. and Schneider, F. B. (1995). Avoiding the undefined by underspecification. In *Computer science today*, volume 1000 of *Lecture Notes in Comput. Sci.*, pages 366–373. Springer, Berlin.

Hammack, R. (2018). *Book Of Proof*. Self-published, 3rd edition. Approved by the American Institute of Mathematics' Open Textbook Initiative.

Jones, E. E. C. (1905). Lewis Carroll's logical paradox. *Mind*, 14(56):576–578.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux, New York.

Lethbridge, T. (2000). What knowledge is important to a software professional? *IEEE Computer*, 33(5):44–50.

Mathieu-Soucy, S. (2016). Should university students know about formal logic: an example of nonroutine problem. In *INDRUM 2016: First conference of the International Network for Didactic Research in University Mathematics*. hal-01337943.

Moktefi, A. (2007). Lewis Carroll and the British nineteenth-century logicians on the barber shop problem. In Cupillari, A., editor, *Proceedings of The Canadian Society for the History and Philosophy of Mathematics' Annual Meeting*, pages 189–199.

Niemelä, P., Valmari, A., and Ali-Löytty, S. (2018). Algorithms and logic as programming primers. In McLaren, B. M., Reilly, R., Zvacek, S., and Uhomoibhi, J., editors, *Computer Supported Education - 10th International Conference, CSEDU 2018, Funchal, Madeira, Portugal, March 15-17, 2018, Revised Selected Papers*, volume 1022 of *Communications in Computer and Information Science*, pages 357–383. Springer.

Ragni, M., Kola, I., and Johnson-Laird, P. (2017). The Wason selection task: A meta-analysis. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, pages 980–985.

Reeves, S. and Clarke, M. (1990). *Logic for Computer Science*. Addison-Wesley.

Stefanowicz, A. (2014). *Proofs and Mathematical Reasoning*. University of Birmingham, Mathematics Support Centre.

Valmari, A. (2021). Automated checking of flexible mathematical reasoning in the case of systems of (in)equations and the absolute value operator. In Csapó, B. and Uhomoibhi, J., editors, *Proceedings of the 13th International Conference on Computer Supported Education, CSEDU 2021, Online Streaming, April 23-25, 2021, Volume 2*, pages 324–331. SCITEPRESS.

Valmari, A. and Hella, L. (2017). The logics taught and used at high schools are not the same. In Karhumäki, J., Matiyasevich, Y., and Saarela, A., editors, *Proc. of the Fourth Russian Finnish Symposium on Discrete Mathematics*, number 26 in TUCS Lecture Notes, pages 172–186. Turku Centre for Computer Science.

Valmari, A. and Hella, L. (2021). A completeness proof for a regular predicate logic with undefined truth value. arXiv:2112.04436.

van Benthem, J. (2008). Logic and reasoning: Do the facts matter? *Studia Logica*, 88(1):67–84.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, 20(3):273–281.