Classical and constructive logic

Jeremy Avigad

September 19, 2000

In these notes and lectures we will discuss some of the differences between classical and constructive logic. In the first few sections we will try to place the issues in a broader philosophical, mathematical, and historical context. After that, we will discuss what one can concretely say about the relationship between the two kinds of logic.

1 The classical and constructive viewpoints

Logic can be described as the science of reasoning. Some of the central questions that logicians try to address are: "What constitutes a logical argument?" "What does it mean to say that a certain statement is a logical consequence of another?"

These questions are extremely broad, and to make any kind of progress we need to narrow our focus. In everyday life, we use different modes of reasoning in different contexts. We can reason about our experiences, and try to determine causal relations between different types of events; this forms the basis of scientific inquiry. We can reason probabilistically, and try to determine the "odds" that the Pirates will win the World Series; or we can employ subjunctive reasoning, and wonder what would have happened had Bill Clinton lost the election. We can reason about events occuring in time, or space; we can reason about knowledge, and belief; or we can reason about moral responsibility, and ethical behavior. We can even try to reason about properties that are vague and imprecise, or try to draw "reasonable" conclusions from vague or incomplete data.

For the most part, in this course we will be interested in one very specific kind of reasoning, namely, the kind that is appropriate for reasoning about abstract mathematical data like numbers, sets, functions, trees, sequences (lists), graphs, and so on. Consider the statement "every even number greater than two can be written as a sum of two primes," or "every binary tree of depth n has at most $2^n - 1$ nodes." Faced with assertions like these, we may wonder whether or not they are true. But typically we do not ask for the "odds" that they are true, whether it is morally proper for them to be true, how long they will be true, or whether they could have been false in some alien universe. In short, we will focus the "abstract" or "universal" aspect of logic, which aims to show that certain statements are necessarily true, independent of time or the particular state of our world.

Later in the course we will see that for computational purposes, it is often useful to consider more general types of logic. For example, when reasoning about computations, one may want to include *temporal* considerations: we may want to show that a process will *never* hang, every print request to an operating system will *eventually* be serviced, or an interrupt request to a processor will be handled at the *next* clock cycle. Or, for nondeterministic computations, we may wish to include *modal* considerations, to reason about the states our system might *possibly* be in at some point in time. For the time being, though, let us just focus on the "mathematical" core of logic described in the last paragraph.

Applied to mathematical reasoning, what, then, is *constructive* logic? The word "constructive" is usually used in contrast to "classical" or "non-constructive." The corresponding forms of logic reflect different understandings, or viewpoints, as to the nature of reasoning about abstract objects.

Roughly speaking, from a classical point of view, every meaningful statement is assumed to be either true or false independent of whether or not we know which is the case, and quantifiers like "all" or "some" are assumed to range over a well-defined domain. For example, the Goldbach conjecture asserts that every even number greater than two can be written as the sum of two primes. This statement has neither been proved nor disproved, but classically speaking, it is either true or false. After all, the set of natural numbers is assumed to be a well-defined collection, and so a statement of the form "every number has property X" is assumed to have a well-defined truth value.

From the constructive point of view, the emphasis is not on "truth in the abstract," but, rather, on the means of verification. From a constructive viewpoint, we are justified in asserting that a statement is true only when we have verified that it is true, and we can correctly assert that it is false only when we have verified that it is false. Similarly, we are allowed to assert "A or B" only once we have either verified A or verified B. Thus, from a constructive point of view, we are not now justified in asserting "either the Goldbach conjecture is true, or the Goldbach conjecture is false."

This divergence of viewpoints is not just a verbal one; it affects that type of arguments that are regarded as valid. For example, consider the following question. Remember that a real number r is called rational if it can be written as a quotient of two integers, $r = m/n$, and irrational if it can't be. For example, $-329/517$ is rational, but $\sqrt{2}$ is not. We know that it is possible to raise an irrational number to a rational power, and get a rational result: for example, $\sqrt{2}^2 = 2$. What is less clear is whether it is possible to raise an irrational number to an *irrational* power, and get a rational result. The following theorem answers this in the affirmative:

Theorem 1.1 *There are irrational numbers a and b such that a^b is rational.*

Proof. Consider $\sqrt{2}^{\sqrt{2}}$. If this is rational, we are done: we can let $a = b = \sqrt{2}$. Otherwise, it is irrational. Then we have

$$(\sqrt{2}^{\sqrt{2}})^{\sqrt{2}} = \sqrt{2}^{\sqrt{2}\times\sqrt{2}} = \sqrt{2}^2 = 2,$$

which is certainly rational. So, in this case, let a be $\sqrt{2}^{\sqrt{2}}$, and let b be $\sqrt{2}$. $\square$

Is this a valid proof? Most mathematicians agree that it is. But there is something unsatisfying here: we have proved the existence of a pair of real numbers with a certain property, without being able to say *which* pair of numbers it is. A constructivist would object, insisting that a proper existence proof should provide an explicit, unconditional description of the objects it asserts to exist. He or she would point out that we have proved the theorem only under the supposition "either $\sqrt{2}^{\sqrt{2}}$ is rational, or it is not," but would argue that this assertion requires further justification.

In fact, the classical and constructive viewpoints differ even as to what constitutes a legitimate *definition*. To give a simple example, let a be the real number that is $\sqrt{2}$, if $\sqrt{2}^{\sqrt{2}}$ is rational, and $\sqrt{2}^{\sqrt{2}}$ otherwise. From a classical point of view, this is a perfectly good description of a real number, while, from a constructive point of view, it is not.

The following scenario might help you better understand the constructive point of view. Suppose I came up to you one day and announced that I had determined a natural number x, with the property that if x is prime, the Goldbach conjecture is true, and if x is composite, the Goldbach conjecture is false. Great news! Whether the Goldbach conjecture is true or not is one of the big open questions of mathematics, and here I seem to have reduced the problem to one of calculation, that is, to the determination of whether a specific number is prime or not.

What is the magic value of x? Let me describe it as follows: x is the natural number that is equal to 7 if the Goldbach conjecture is true, and 9 otherwise.

Angrily, you cry foul. From a classical point of view, the description above really determines a unique value of x, and so constitutes a perfectly good "definition." But from a constructive point of view, any proper definition of a mathematical object should allow us to compute with it. For example, from the definition of a natural number we should be able to determine the first digit, or whether the number is even or odd, or whether it is greater than 100 or not, and so on.

At this point it will be useful to introduce a distinction between the notions of "logical argument" and "mathematical argument." Roughly, we will think of mathematics as consistig of logical reasoning about mathematical objects. In other words, we will think of a mathematical proof as being a logical proof from mathematical axioms and assumptions.

Intuitively, this makes sense. Certain features of a mathematical argument seem to lie on the "logic" side, notably uses of words like "and," "or," "if…then," "every," and so on. Other aspects seem to belong more to the realm of mathematics: notions like "number," "function," "point," "line," for example. However, there is no clear line between the two. For example, do notions like "set" and "property" count as logical concepts, or mathematical ones? In the 19th century Gottlob Frege claimed that all of mathematics could be explained in terms of logical constructions, making mathematics a part of logic. This "logicist" claim was pursued by Bertrand Russell in the early part of his career, culminating in the great three volume work, *Principia Mathematica*, written by Russell and Alfred North Whitehead and first published between 1911 and 1914. But most philosophers today consider the logicist program to have failed, since, in order to do mathematics, Russell and Whitehead were forced to introduce axioms that did not seem strictly logical.

Though the distinction between "logic" and "mathematics" is not a sharp one, it is nonetheless useful. We can now characterize constructive mathematics as constructive reasoning about constructively presented objects, and classical mathematics as classical reasoning about classically presented objects. In the next few sections we will discuss some of the differences between classical and constructive mathematics. After that, we will focus on the underlying logic.

2 Examples from number theory

It will be helpful to illustrate the differences between the classical and constructive viewpoints with some examples from number theory.

If a and b are natural numbers, a *divides* b if and only if b is a multiple of a, i.e. there is an natural number c such that $ac = b$. A natural number greater than 1 is *prime* if the only numbers dividing it are 1 and itself.

Theorem 2.1 *Every natural number greater than* 1 *can be written as a product of primes.*

Proof. Use induction. Suppose a greater than 1. If a is prime, we are done. Otherwise, a can be written as a product of smaller numbers, $a = b \times c$. Inductively write b and c as products of primes. $\square$

This argument is acceptable constructively, as well as classically. The next theorem was known to Euclid, and appears in his *Elements*.

Theorem 2.2 *There are infinitely many prime numbers. More precisely: given any number M, there is a prime number bigger than M.*

Proof. Suppose there were only finitely many prime numbers, that is, for some natural number M all the prime numbers are less than or equal to M. Then we could make an explicit list of all of them, say, $p_1, \ldots, p_k$.

Let $N = p_1 \cdot p_2 \cdot \ldots \cdot p_k + 1$. Notice that none of $p_1, \ldots, p_k$ divide N, since each leaves a remainder of 1. Now, either N is prime, or it can be factored into primes. So there is at least one prime number q dividing N. But q can't be any of the numbers $p_1, \ldots, p_k$, so q is a prime number that is not on the list. This contradicts our assumption.

Assuming that there were only finitely many prime numbers led to a contradiction. So, the assumption has to be false, and there are infinitely many primes. $\square$

This proof has a classical flavor, since it is a proof by contradiction: we have proved that there is a prime number bigger than M by showing that it can't be the case that there isn't. We expect a constructive proof to show us, explicitly, how to find a prime number bigger than M. But the proof above is more constructive than it appears at first glance. In fact, the following rewording shows that this constructive information is not hard to find:

Proof. Given any M, make a list of all the primes less than equal to M, say $p_1, \ldots, p_k$. Let $N = p_1 \cdot p_2 \cdot \cdots \ldots p_k + 1$. Factor N into primes, and let q be any prime factor. Then q is a prime that is not in the list $p_1, \ldots, p_k$, and so, q is a prime number bigger than M. □

Let us consider another example. If a and b are positive natural numbers greater than or equal to 1, the *greatest common divisor* of a and b is exactly what it sounds like: the biggest number that divides both a and b.

This definition makes sense, because we know that at least the number 1 divides them both. This provides an algorithm for finding the greatest common divisor of a and b: just test all the positive numbers less than the smaller of the two, and take the biggest one that divides them both. Our constructive proof of the following theorem will provide a better algorithm; but, once again, we will start with a nonconstructive proof.

Theorem 2.3 *Let a and b be positive natural numbers, and let d be the greatest common divisor of a and b.*

1. *There are integers x and y such that $d = ax + by$.*
2. *If e is any common divisor of a and b, then e divides d.*

Proof. The second statement follows from the first: if e divides a and b, then it divides $ax + by$, and hence d.

To prove the first statement, let A be the set of all integral linear combinations of a and b. In other words, A is the set of numbers that can be written in the form $au+bv$ for some pair of integers u and v. In set-theoretic notation,

$$A = \{au + bv \mid u \in \mathbb{Z} \wedge y \in \mathbb{Z}\}.$$

Let B be the set of all positive numbers in A. B is nonempty, since, for example, $a + b$ is in B. So B has a smallest element, d. Let x and y be the integers such that $d = ax + by$.

We need to show that (1) d divides a and b, and (2) if some other number e divides both a and b, then $e \leq d$.

Let us do (1) first. Suppose d does not divide a. Then dividing a by d yields an integer quotient, q, and a nonzero remainder, r. In other words, we have $a = qd + r$, where r is some number between 0 and d. But now we can write

$$r = a - qd = a - q(ax + by) = a(1 - qx) + b(qy),$$

which is a smaller positive linear combination of a and b. This is a contradiction, so d divides a. The same argument can be used to show that d divides b.

(2) is easy: we have already noted that if e divides both a and b, then e divides d. This implies that $e \leq d$. □

This proof also has a classical flavor. For example, (1) above was proved indirectly; we assumed that d did not divide a, and derived a contradiction. But most constructivists accept the "law of the excluded middle," p or not p, in this special case: since we can decide whether d divides a by simply tesing all the possible divisors, which are less than a, we are allowed to assume that either d divides a, or it doesn't.

A more dubious step occurs in the definition of B; our definition does not, on the surface, come with an explicit procedure that enables us to determine whether or not a given number is in B, let alone find the least element. But if we examine the proof more closely, we find that it *does* tell us that there is at least one element of B; for example, $a + b$. And the proof tells us that if we have a number in B that is not the greatest common divisor of A and B, we can find a smaller number in B.

These observations can be used to turn the proof above into a constructive one, which then yields the well known Euclidean algorithm for determining the GCD of a and b. Indeed, suppose a is greater than b. Write $a = bq + r$. If r is equal to 0, b is the greatest common divisor. Otherwise, any number dividing a and b also divides $r = a - bq$; so we are reduced to computing the greatest common divisor of b and r. If we iterate this process, the remainder gets smaller at each step. When the remainder is finally 0, we have the greatest common divisor, d. From the chain of equations, we can find values x and y satisfying $d = ax + by$.

3 Other mathematical examples

The examples above may convey the impression that every classical proof has a constructive version. But this is not the case: there are classical theorems that are constructively false; and there are classical theorems that may well have constructive proofs, but none are known. In this section we will discuss some examples.

An example of a classical theorem that is not provable constructively is the statement that every bounded increasing sequence of real numbers has a least upper bound. To explain this, let me sketch some of the background definitions. Some of you are used to thinking of a real number as an infinite decimal. This is not far from the standard, classical mathematical viewpoint. The constructive viewpoint is a slight variation: one thinks of a real number r as given by a function which, on input n, returns a rational approximation to r which is within, say, $1/10^n$. More formally, a real number is a function f from the natural numbers, $\mathbb{N}$, to the rationals, $\mathbb{Q}$, such that for every n and $m \geq n$, $|f(m) - f(n)|$ is less than $1/10^n$. Such a sequence $f(0), f(1), f(2), \ldots$ is called a *Cauchy sequence with a fixed rate of convergence*. The main difference between the classical and constructive viewpoints is that a classical mathematician allows f to be an "arbitrary" function, while the constructive mathematician expects f to be computable, in some sense. We will say more about this difference later.

Now let $a_0, a_1, a_2, \ldots$ be an increasing sequence of real numbers between 0 and 1. A real number b is an *upper bound* to this sequence if $a_i \leq b$ for every i. A real number b is a *least upper bound* if b is an upper bound and for every other upper bound c, $b \leq c$.

Draw a picture: first, put down a pair of x-y coordinate axes. On the x axis, mark of the natural numbers $0, 1, 2, \ldots$. On the y axis, mark off the interval $[0, 1]$. Then plot an increasing sequence of real numbers $a_0, a_1, a_2, \ldots$, all between 0 and 1.

Your intuition may tell you that such a sequence has a least upper bound. Classically, the argument goes roughly as follows: first divide the interval $[0, 1]$ into tenths, marking off $1/10, 2/10, 3/10, \ldots$. Let n be the least number (less than or equal to 10) such that all the a_i are less than $n/10$. If $n = 0$, 0 is the least upper bound. Otherwise, $(n - 1)$ is the first decimal digit of the least upper bound. Then, divide the interval $[(n - 1)/10, n/10]$ into ten pieces, and proceed as before, to determine the second digit. Keep going. The resulting sequence of digits determines the least upper bound, b.

Constructively, this proof is not valid. The problem is this: given a description of the sequence $a_0, a_1, a_2, \ldots$, in general there isn't an algorithmic procedure to determine whether or not all the a_i are less than a particular fraction (like $1/10$).

In fact, constructively, the theorem is false. The argument that this is the case uses the fact that the halting problem is unsolvable; in other words, given a reasonable numbering of the Turing machines $M_0, M_1, M_2, \ldots$, there is no algorithm which, on input i, determining whether or not Turing machine M_i halts when started on an empty tape. We will describe a sequence of real numbers (in fact, of rational numbers) such that a least upper bound yields a solution to the halting problem, and so, is not computable.

The sequence of rationals $a_0, a_1, a_2, \ldots$ is constructed as follows. At stage i, simulate machines $M_0, \ldots, M_i$ for i steps. For each number j between 0 and i, let $p_{i,j}$ be 0 if Turing machine M_j has not halted in i steps, and 1 if it has. Let

$$a_i = \sum_{j=0}^{i} \frac{p_{i,j}}{2^{j+1}}.$$

You can think of a_i as a binary fraction, in which the jth digit after the "binary point" gets switched on if the jth machine has halted by the ith stage.

It is not hard to see that the sequence $a_0, a_1, a_2, \ldots$ is increasing, since as i grows, more machines halt. With some work, one can justify the claim that if f is a least upper bound, then f provides a solution to the halting problem.

We have seen examples of classical theorems that are also constructive theorems, and examples of classical theorems that are not constructive theorems. There are als examples of classical theorems for which the constructive proofs are much harder, or for which it is not known whether or not they can be proved constructively.

A proof in the first section used the fact that "either $\sqrt{2}^{\sqrt{2}}$ is rational, or it isn't." Classically, this has a trivial proof; it is just a logical axiom, the law of the excluded middle. In fact, one can prove constructively that $\sqrt{2}^{\sqrt{2}}$ is irrational, though both the classical and constructive proofs are hard.

A more interesting example is Fermat's Last Theorem. This is now almost universally accepted has having been proved. But the proof is very long, and uses very advanced methods from algebraic and analytic number theory, that have, for the most part, been developed classically. So, at present, it is unknown whether or not this can be proved by constructive methods.

4 Historical and philosophical issues

Before the 19th century there was not a sharp distinction between classical and constructive proofs in mathematics, for the simple reason that most proofs were more or less constructive. But the latter half of the 19th century witnessed what many consider to have been a revolution in mathematics, which involved the introduction of new and nonconstructive methods of reasoning about the infinitary structures. Such reasoning involved "arbitrary" real numbers, "arbitrary" functions, and "arbitrary" sets of objects, defined using abstract methods that did not correspond to explicit algorithmic procedures.

These new methods were controversial. Some mathematicians, like Leopold Kronecker, objected, insisting that all mathematical objects be described in terms of symbolic representations and explicit algorithms. In his 1883 *Grundlagen einer allgemeinen Mannigfaltigkeitslehre* (Foundation of a general theory of manifolds), Georg Cantor responded to this criticism with a passage that is now famous:

> It is not necessary, I believe, to fear, as many do, that these principles contain any danger to science. For in the first place the designated conditions, under which alone the freedom to form numbers can be practised, are of such a kind as to allow only the narrowest scope for discretion. Moreover, every mathematical concept carries within itself the necessary corrective: if it is fruitless or unsuited to its purpose, then that appears very soon through its uselessness, and it will be abandoned for lack of success. But every superfluous constraint on the urge to mathematical investigation seems to me to bring with it a much greater danger, all the more serious because in fact absolutely no justification for such constraints can be advanced from the essence of the science — for the *essence* of *mathematics* lies precisely in its *freedom*.

By the turn of the century, the stakes were raised. Problems were discovered with some uses of Cantor's notion of set (indeed, Cantor was aware of some of these dangers), and in 1903 Russell discovered a related paradox in Frege's logical foundations for mathematical reasoning. So now the issue was not only whether or not the new methods in mathematics were correct; but, indeed, whether or not they were even consistent.

In the 1910's, L. E. J. Brouwer mounted a campaign to "refound" mathematics on a logic which rejected the law of the excluded middle, based on his "intuitionist" philosophy of mathematics. (The words "constructive" and "intuitionistic" are used today almost interchangeably.) David Hilbert, one of the most influential mathematicians of all time, responded angrily that "denying a mathematician use of the principle of excluded middle is like denying an astronomer his telescope or a boxer the use of his fists." Just as colorfully, he declared that "no one will drive us from the paradise that Cantor has created for us."

Hilbert proposed a means of rescuing classical mathematics from the threat of inconsistency and intuitionist challenges. The idea was to represent classical mathematics with formal axiomatic systems, and to prove the consistency of these systems with secure, "finitary" methods that were acceptable to even the strictest of constructivists. Gödel's incompleteness theorems of 1931 showed that, unfortunately, Hilbert's program could not succeed: any reasonable system of classical mathematics would have to include Hilbert's finitary methods, and Gödel showed that no such system can prove its own consistency, unless it, itself, is inconsistent.

Since then, these issues have played a central role in the philosophy of mathematics. From the historical considerations above, we can extract a number of arguments in favor of constructive methods. One might argue that they are

- more secure, i.e. there is less likelihood of falling into error or inconsistency;
- more meaningful, i.e. more in tune with the nature of mathematical objects and knowledge; or
- more useful, in terms of applications to the sciences.

The usual classical response is that constructive methods place too much of a burden on mathematics, invalidating central theorems and detracting from its elegance. Indeed, many justifications of classical mathematics also point to its usefulness, and applications in the development of science and technology.

On an abstract level, philosophers would like to give a general account of mathematical existence and mathematical knowledge that might help adjudicate the dispute. On a less abstract level, one wants to characterize the general goals and methods of mathematics, and try to determine which methods are best suited to the various goals.

How do things stand today? Philosophical debates as to the metaphysical basis for mathematics continue, but there is no single account that is generally accepted as unproblematic. Lo and behold, mathematics still goes on. Classical methods of reasoning, with reasonable restrictions the general notion of a set, have not revealed any inconsistencies, and today most mathematicians reason classically. In that sense, the 19th century revolution is complete; Cantorian methods are now commonplace. On the other hand, with the rise of the computer, there is also a good deal of interest in algorithmic aspects of mathematics; in that sense, Kronecker's insistence on symbolic representation and algorithms was ahead of its time.

With regard to the question as to whether mathematical reasoning should be classical or constructive, one need not adopt an "either/or" attitude. As noted above, aspects of both are found in everyday mathematics; and so a better question is to determine what kinds of reasoning are appropriate in which contexts, or towards which goals. The attempt to characterize the general methods used by contemporary mathematicians and understand them better is still a central focus in the field of mathematical logic. Meanwhile, the advent of computer science has brought new issues to the fore, as well a fresh perspective on issues related to constructivism.

5 The role of symbolic logic

In trying to characterize the differences between classical and constructive logic, we have made a number of rather vague claims:

1. Constructive logic depends on a certain understanding of the meaning of logical terms, which is different from the classical viewpoint.

2. A constructive proof should enable one to "construct" the objects asserted to exist.

3. Constructive proofs have algorithmic content.

4. Constructive logic, by its nature, has applications where computation is concerned.

These claims might help you understand what we mean by the term "constructive logic," but they are fairly imprecise. We would like to explore the issues in a more down-to-earth, concrete way.

The methods of *symbolic logic* provide a very useful approach. The idea is to provide formal mathematical descriptions of the kinds of logic we are interested in, and use these descriptions to flesh out the claims above. For example, in this course we chose to use symbols $\wedge$, $\vee$, $\supset$, and $\perp$ to represent basic logical terms like "and," "or," "implies," and "false." We then presented precise rules that govern the use of these symbols. In a sense, these rules provide a clear explanation of the constructive understanding of the logical terms.

We have also begun to explore the relationship between constructive logic and computation. Our formal presentation of constructive logic makes it clear that the rules for forming propositions and proving that they are true bear a striking similarity to rules for forming data types and constructing elements of those types. The "Curry-Howard isomorphism" provides one way of making this correspondence precise, as do formal presentations of type theory. On a practical level, this correspondence will help us understand ways in which we can extract computer programs from constructive proofs, or even view proofs as programs themselves.

In short, symbolic logic helps us clarify our intuitions, and puts us in a better position to address the philosophical issues. In the next section, we will use the formal analysis to explore the relationship between classical logic and constructive logic. And in the section after that, we will turn to a discussion of the relationship between classical and constructive mathematics.

6 The relationship between classical and constructive logic

The rules of inference you have seen so far reflect the constructive understanding of the logical connectives. What do we need to do to get classical logic? The classical viewpoint is more liberal than the constructive one, in the sense that any constructive proof is classically valid. So we can obtain classical logic by adding some additional rules to the constructive ones.

One option is to add the *law of the excluded middle*: for every proposition A, the judgement

$$A \vee \neg A \text{ true}.$$

We will denote this axiom EM. Another option is to add the axiom

$$\neg\neg A \supset A \text{ true}$$

for every proposition A. We will call this DN, for *double negation elimination.* The most natural option, however, takes the form of a rule representing the classically valid practice of proving a statement by contradiction. Classically, to prove a proposition A, it is valid to assume $\neg A$ and obtain a contradiction. In rule form:

$$\begin{array}{c} \overline{\neg A}^{\,u} \\ \vdots \\ \dfrac{\perp}{A}\, RAA^u \end{array}$$

The letters RAA stand for "reductio ad absurdum." Do not confuse this rule with the following:

$$\dfrac{\begin{array}{c}\dfrac{}{A}\,u \\ \vdots \\ \bot\end{array}}{\neg A}\,u$$

Since we have defined $\neg A$ to be $A \supset \bot$, this is just an instance of the introduction rule for $\supset$. So this rule is constructively valid, in contrast to RAA.

The following theorem tells us that it doesn't really matter which option we choose.

Theorem 6.1 *With any one of the three additions above (EM, DN, or RAA), one can derive any other.*

Proof. We will show that RAA and DN are equivalent, and that with RAA one can prove the law of the excluded middle. For homework, you will be asked to show that, using the law of the excluded middle, one can derive RAA.

The following is a proof of $\neg\neg A \supset A$, using RAA:

$$\dfrac{\dfrac{\dfrac{\dfrac{}{\neg\neg A}\,u \qquad \dfrac{}{\neg A}\,v}{\bot}}{A}\,RAA^v}{\neg\neg A \supset A}\,u$$

For brevity, we have left off the annotation $\supset I$ at the inference labelled u. The following shows that using $\neg\neg A \supset A$, if one can derive $\bot$ from $\neg A$, then A follows:

$$\dfrac{\dfrac{}{\neg\neg A \supset A} \qquad \dfrac{\begin{array}{c}\dfrac{}{\neg A}\,u \\ \vdots \\ \bot\end{array}}{\neg\neg A}\,u}{A}$$

The following is a proof of $A \lor \neg A$, using RAA:

$$\dfrac{\dfrac{\dfrac{}{\neg(A \lor \neg A)}\,u \qquad \dfrac{\dfrac{\dfrac{\dfrac{}{\neg(A \lor \neg A)}\,u \qquad \dfrac{\dfrac{}{A}\,v}{A \lor \neg A}}{\bot}}{\neg A}\,v}{A \lor \neg A}}{\bot}}{A \lor \neg A}\,\mathrm{RAA}^u$$

Modulo the homework assignment, this completes the proof. $\square$

Let us explore some of the ways in which classical logic differs from constructive logic. Right off the bat, we know that $A \vee \neg A$ is derivable classically. But it is not derivable constructively, in general. To see this, remember that anything derivable constructively in fact has a *normal* derivation. Remember that in a normal derivation, every formula appearing is either a subformula of the conclusion or a subformula of one of the hypotheses. Working backwards, such a derivation would have to either be of the form

$$\dfrac{\begin{array}{c}\vdots\\ A\uparrow\end{array}}{A\vee\neg A\uparrow}$$

or

$$\dfrac{\dfrac{\begin{array}{c}\overline{\ A\downarrow\ }^{\,u}\\ \vdots\\ \bot\uparrow\end{array}}{\neg A\uparrow}\,u}{A\vee\neg A\uparrow}$$

In the first case, with no additional knowledge about A, we will not be able to derive it. In the second case, the only rule applicable in is $\supset$ elimination

$$\dfrac{\begin{array}{c}\overline{\ A\downarrow\ }^{\,u}\\ \vdots\\ B\downarrow\end{array}\qquad\begin{array}{c}\vdots\\ \neg B\uparrow\end{array}}{\bot\uparrow}$$

where B is a subformula of A. But the only subformula of A is A itself, and replacing B by A represents no progress at all.

For another example, consider "DeMorgan's law," which is classically valid:

$$\neg(A \wedge B) \leftrightarrow \neg A \vee \neg B.$$

Both constructively and classically, the reverse direction is valid. But the forwards direction,

$$\neg(A \wedge B) \supset \neg A \vee \neg B.$$

holds classically but not constructively. The following is a classical proof:

$$
\begin{prooftree}
\AxiomC{}
\RightLabel{u}
\UnaryInfC{$\neg(A \wedge B)$}
\AxiomC{}
\RightLabel{w}
\UnaryInfC{$\neg(\neg A \vee \neg B)$}
\AxiomC{}
\RightLabel{x}
\UnaryInfC{$\neg A$}
\UnaryInfC{$\neg A \vee \neg B$}
\BinaryInfC{$\bot$}
\RightLabel{RAA^x}
\UnaryInfC{A}
\AxiomC{}
\RightLabel{w}
\UnaryInfC{$\neg(\neg A \vee \neg B)$}
\AxiomC{}
\RightLabel{y}
\UnaryInfC{$\neg B$}
\UnaryInfC{$\neg A \vee \neg B$}
\BinaryInfC{$\bot$}
\RightLabel{RAA^y}
\UnaryInfC{B}
\BinaryInfC{$A \wedge B$}
\BinaryInfC{$\bot$}
\RightLabel{RAA^w}
\UnaryInfC{$\neg A \vee \neg B$}
\RightLabel{u}
\UnaryInfC{$\neg(A \wedge B) \supset \neg A \vee \neg B$}
\end{prooftree}
$$

Let us show that $\neg(A \wedge B) \supset \neg A \vee \neg B$ is not constructively valid, by showing that it has no normal proof. Working backwards, a normal proof of an implication looks as follows:

$$
\begin{prooftree}
\AxiomC{}
\RightLabel{u}
\UnaryInfC{$\neg(A \wedge B)\downarrow$}
\noLine
\UnaryInfC{$\vdots$}
\noLine
\UnaryInfC{$\neg A \vee \neg B \uparrow$}
\RightLabel{u}
\UnaryInfC{$\neg(A \wedge B) \supset \neg A \vee \neg B \uparrow$}
\end{prooftree}
$$

Working backwards again, the second-to-last inference can only be $\vee$ introduction, from $\neg A$ or $\neg B$, or $\bot$ elimination. In the last case, we have to complete the following proof:

$$
\begin{prooftree}
\AxiomC{}
\RightLabel{u}
\UnaryInfC{$\neg(A \wedge B)\downarrow$}
\noLine
\UnaryInfC{$\vdots$}
\noLine
\UnaryInfC{$\bot \uparrow$}
\UnaryInfC{$\neg A \vee \neg B \uparrow$}
\RightLabel{u}
\UnaryInfC{$\neg(A \wedge B) \supset \neg A \vee \neg B \uparrow$}
\end{prooftree}
$$

A short argument shows that it is not possible to get, in general, from $\neg(A \wedge B)$ to $\bot$. We will omit this argument, since a similar one occurs below. We will therefore consider only the first case, since the second is symmetric:

$$
\begin{prooftree}
\AxiomC{}
\RightLabel{u}
\UnaryInfC{$\neg(A \wedge B)\downarrow$}
\noLine
\UnaryInfC{$\vdots$}
\noLine
\UnaryInfC{$\neg A \uparrow$}
\UnaryInfC{$\neg A \vee \neg B \uparrow$}
\RightLabel{u}
\UnaryInfC{$\neg(A \wedge B) \supset \neg A \vee \neg B \uparrow$}
\end{prooftree}
$$

Again, the preceeding rule is determined:

$$\begin{array}{c} \overline{\neg(A \wedge B)\downarrow}^{\,u} \qquad \overline{A\downarrow}^{\,v} \\ \vdots \qquad\qquad \vdots \\ \dfrac{\dfrac{\dfrac{\perp\uparrow}{\neg A\uparrow}}{\neg A \vee \neg B\uparrow}}{\neg(A \wedge B) \supset \neg A \vee \neg B\uparrow}\,u \end{array}$$

Now the preceeding rule can only be $\supset$ elimination, applied to a subformula of one of the hypotheses. There is only one candidate:

$$\dfrac{\dfrac{\dfrac{\dfrac{\overline{\neg(A \wedge B)\downarrow}^{\,u} \qquad \begin{array}{c} \overline{\neg(A \wedge B)\downarrow}^{\,u} \qquad \overline{A\downarrow}^{\,v} \\ \vdots \qquad\qquad \vdots \\ (A \wedge B)\uparrow \end{array}}{\perp\uparrow}}{\neg A\uparrow}\,v}{\neg A \vee \neg B\uparrow}}{\neg(A \wedge B) \supset \neg A \vee \neg B\uparrow}\,u$$

The preceeding rule must be $\wedge$ introduction:

$$\dfrac{\dfrac{\dfrac{\dfrac{\overline{\neg(A \wedge B)\downarrow}^{\,u} \qquad \dfrac{\dfrac{\overline{A\downarrow}^{\,v}}{A\uparrow} \qquad \begin{array}{c} \overline{\neg(A \wedge B)\downarrow}^{\,u} \qquad \overline{A\downarrow}^{\,v} \\ \vdots \qquad\qquad \vdots \\ B\uparrow \end{array}}{(A \wedge B)\uparrow}}{\perp\uparrow}\,\mathrm{v}}{\neg A\uparrow}}{\neg A \vee \neg B\uparrow}}{\neg(A \wedge B) \supset \neg A \vee \neg B\uparrow}\,u$$

Working forwards, we can only apply an elimination rule to $\neg(A \wedge B)$:

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\neg(A\land B)\downarrow}u \qquad \dfrac{\dfrac{\dfrac{}{A\downarrow}v}{A\uparrow} \qquad \dfrac{\dfrac{\dfrac{\dfrac{}{\neg(A\land B)\downarrow}u}{\neg(A\land B)\uparrow} \qquad \dfrac{\dfrac{\dfrac{}{A\downarrow}v}{A\uparrow} \qquad \begin{array}{cc}\dfrac{}{\neg(A\land B)\downarrow}u & \dfrac{}{A\downarrow}v \\ \vdots & \vdots \\ B\uparrow & \end{array}}{(A\land B)\uparrow}}{\bot\uparrow}}{B\uparrow}}{(A\land B)\uparrow}}{\bot\uparrow}\text{v}}{\neg A\uparrow}}{\neg A\lor\neg B\uparrow}}{\neg(A\land B)\supset\neg A\lor\neg B\uparrow}u
$$

But now we are back where we were a few inferences ago. So, in general, it is impossible to derive B from $\neg(A \land B)$ and A.

7 The double-negation translation

We have characterized classical logic as constructive logic together with an additional rule. So, in a sense, constructive logic is *weaker* than classical logic. We will now consider an interpretation of classical logic in constructive logic that paints an entirely different picture; we will see that, with this interpretation, one can view classical logic as a subset of constructive logic that is, in fact, less expressive.

The interpretation is usually referred to as the Gödel/Gentzen double-negation translation, and was presented independently by these two logicians in the early 1930's. It works not only for propositional logic, but also for predicate logic, which extends propositional logic with the quantifiers "for all" and "for some." In fact, Gödel and Genzten originally presented the translation for first-order arithmetic, which is predicate logic together with a certain set of axioms for the natural numbers. An earlier version of the double-negation translation was presented in 1925 by Kolmogorov.

Suppose A is a proposition, expressed in a language with $\land$, $\lor$, $\supset$, and $\bot$, and some basic, indecomposable propositions. The double-negation translations maps A to another proposition, A^N, defined recursively as follows:

- $A^N = \neg\neg A$, if A is a basic proposition
- $\bot^N = \bot$

- $(A \wedge B)^N = A^N \wedge B^N$
- $(A \vee B)^N = \neg(\neg A^N \wedge \neg B^N)$
- $(A \supset B)^N = A^N \supset B^N$

Theorem 7.1 *The double-negation translation has the following two properties, for every proposition A:*

1. *Classically, A^N is equivalent to A.*
2. *If A is provable classically, then A^N is provable constructively.*

This implies that for every proposition A, A is provable classically if and only if A^N is provable constructively. One can strengthen the second statement: if A is provable from a set of propositions, S, using classical logic, then A^N is provable from the double-negation translations of the formulae in S, using constructive logic.

This provides a nice description of the relationship between classical and constructive logic. Suppose a classical logician claims to have proved a statement of the form $A \vee B$. The constructive logician examines the proof, and insists that in fact the classical logician has only proved $\neg(\neg A^N \wedge \neg B^N)$. The classical logician won't argue with this; as far as he or she is concerned, the two statements are equivalent. In short, the constructive logician makes finer distinctions than the classical one; but the constructive logician can translation the classical viewpoint into his or her own terms.

Notice that the central clause in the translation has to do with the reinterpretation of $\vee$. (For predicate logic, $\exists$ has to be reinterpreted in a similar way.) This supports the intuition that the constructive "or" is *stronger* than the classical "or"; to a constructive logician, $\neg(\neg A \wedge \neg B)$ is weaker than $A \vee B$. In fact, $\neg(\neg A \wedge \neg B)$ is constructively equivalent to $\neg\neg(A \vee B)$, from whence the translation gets its name.

8 The relationship between classical and constructive mathematics

Remember that we are thinking of constructive mathematics as consisting of constructive reasoning about constructively presented mathematical objects. From the constructive point of view, any presentation of a collection of mathematical objects should explain how these objects are constructed; these constructions then dictate the rules for reasoning about them. You have already seen some examples, with the cross product and disjoint sum data types. You will see many more examples as the semester progresses, including the natural numbers, pairs, lists, trees, and so on.

From the classical point of view it is also desirable to explain how mathematical objects are constructed from more basic ones; but the classical mathematician claims more latitude in these constructions, and the general structural properties of the set of objects constructed are more important. For example, classically one can construct the set of real numbers using Cauchy sequences, as alluded to above. But here one relies on the notion of an "arbitrary" function, and the end goal is to have an ordered field satisfying the least upper bound principle.

What can one say about the relationship between classical mathematics and constructive mathematics? The story is less clear cut than in the previous section. It turns out that a good deal of modern mathematics has been developed constructively. Erret Bishop's book, *Foundations of Constructive Analysis*, published in 1967, was a landmark in that regard. (The book *Constructive Analysis*, cited below, is a revised and expanded version, cowritten with Bishop's student, Douglas Bridges.) As noted above, many classical theorems fall by the wayside. For example, some of you may recall the intermediate value theorem from calculus: given any continuous function f defined on the interval $[0, 1]$, if $f(0) = -1$ and $f(1) = 1$, then there is some real number x between 0 and 1 such that $f(x) = 0$. This does not hold constructively, but one can prove a weaker statement that has the same flavor. This characterizes constructive mathematics to a large extent: it requires more work, and the theorems are not as easy to state; but the proofs typically have a clear algorithmic content, so even the classical mathematician usually admits that the proofs yield additional information.

Many parts of classical mathematics, however, have not been developed constructively. As noted above, Fermat's last theorem provides an example; we do not know whether or not this theorem has a constructive proof.

Can we, as we did in the last section, carry out a formal analysis of the relationship between classical and constructive mathematics? We can, to the extent that we can find appropriate formal systems that characterize the two. A collection of axioms known as Zermelo-Fraenkel set theory is often accepted as a foundational basis for classical mathematics, but it is awfully difficult to understand these axioms in constructive terms. It turns out, however, that a good deal of mathematics can be carried out in weaker theories than set theory, and many of these have been analyzed in classical terms. Though the classical and constructive viewpoints are very different, in some respects they can be reconciled.

9 Constructive logic and computer science

As noted in the introduction, one of the goals of this course is to clarify the nature of constructive logic, and explore its relationship to computation. Constructive reasoning is quite general, however, and can be applied in other domains as well. Later in the course we will see that it is even easy to incorporate classical reasoning into constructive frameworks, essentially adding the law of the excluded middle where it is appropriate to the situation. We will explore numerous applications in computer science, involving the design of programming languages and the use of logic as a means of specification and verification.

As this section should make clear, the issues we have been addressing lie at the intersection of a number of fields, including computer science, philosophy, and mathematics. Our pursuits join together a philosophical analysis of constructive logic, a mathematical analysis thereof, and computational applications, bringing together a number of different perspectives in an illuminating way.

10 Exercises

1. Show that the proof that there are infinitely many primes, discussed in Section 2, provides an algorithm for finding a prime number bigger than N. Is it a good algorithm?

2. Show that given any number N, there is a prime number between N and $N!+1$. (In fact, it is known that there is always a prime number between N and $2N$, but the proof is more difficult.)

3. Let $a = 7590$ and let $b = 1155$. Use the Euclidean algorithm to find the GCD of a and b, and call it d. Find values of x and y such that $ax + by = d$.

4. If you can, write a computer program that solves GCD problems like the preceeding one.

5. Fill in the details of the proof that the statement "every bounded sequence of real numbers has a least upper bound" implies "there is a function f that solves the halting problem," sketched in Section 3. (This is hard!)

6. Show that using the law of the excluded middle, one can derive a proposition A by deriving $\perp$ hypothetically from $\neg A$. In other words,

from arbitrary instances of $B \vee \neg B$ and a proof of $\perp$ from $\neg A$, derive A.

7. Consider the dual form of the DeMorgan equivalence considered in Section 6:

$$\neg(A \vee B) \leftrightarrow (\neg A \wedge \neg B).$$

Which directions are constructively valid?

8. Consider the classically valid equivalence

$$(A \supset B) \leftrightarrow (\neg A \vee B)$$

   (a) Show that the reverse direction is constructively valid; i.e. give a constructive proof of $(\neg A \vee B) \supset (A \supset B)$

   (b) Show that from $\neg(\neg A \vee B)$ and $A \supset B$ one can derive $\neg A$ constructively

   (c) Show that from $\neg(\neg A \vee B)$ and $\neg A$ one can derive $\perp$ constructively

   (d) Give a classical proof (using RAA) of the forwards direction of the equivalence above, i.e. $(A \supset B) \supset (\neg A \vee B)$. (Hint: using RAA, it suffices to assume $A \supset B$ and $\neg(\neg A \vee B)$ and derive a contradiction.)

   (e) Using the fact that every constructively valid proposition has a normal proof, show that $(A \supset B) \supset (\neg A \vee B)$ is not constructively valid.

9. Consider the classically valid equivalence:

$$((A \supset B) \vee (A \supset C)) \leftrightarrow (A \supset B \vee C).$$

   (a) Give a constructive proof of the forwards direction.

   (b) Give a classical proof of the reverse direction.

   (c) Show that the reverse direction is not constructively valid.

10. Show that $\neg\neg(A \vee B)$ is constructively equivalent to $\neg(\neg A \wedge \neg B)$.

11. If you were the ruler of the world, would you command that mathematics be done constructively? Classically? Some other way? Or would you let mathematicians choose individually?

Keep in mind that you are paying every mathematician's salary. (But, you fear, that banishing mathematics outright might result in a revolution.) As ruler of the world, you can enlist advisors to do research and gather information to help you make your decision. What kinds of information would you take into consideration?

Further reading

Further reading on the history and philosophy of mathematical logic, and the foundations of mathematics:

1. Benacerraf, Paul and and Hillary Putnam, *Philosophy of Mathematics: selected readings*, Cambridge University Press, second edition, 1983.

2. van Heijenoort, Jean, *From Frege to Gödel: A sourcebook in mathematical logic, 1879–1931*, Harvard University Press, 1967.

3. Ewald, William, *From Kant to Hilbert: A sourcebook in the foundations of mathematics*, Oxford Science Publications, 1996.

Further reading on intuitionism and constructive mathematics:

1. Troelstra, A.S. and Dirk van Dalen, *Constructivity in Mathematics: An Introduction* (two volumes), North-Holland, 1988.

2. Beeson, Michael, *Foundations of Constructive Mathematics*. Springer-Verlag, 1985.

3. Bishop, Erret and Douglas Bridges, *Foundations of Constructive Mathematics*, McGraw Hill, 1985.

See also the entries on Intuitionistic Logic and Constructive Mathematics in the online *Stanford Encyclopedia of Philosophy*, at http://plato.stanford.edu/entries/.