Logic and Proof

Supervision 1 – Solutions

1. Introduction

1. Determine whether the truth value of the following sentences is true, false, unknown, or if the sentence is not a proper logical statement. Justify all of your answers, and give satisfying and falsifying valuations of the variables where appropriate.

   a) Broccoli is delicious.

   b) x likes broccoli.

   c) The area of a circle of radius r is πr^2.

   d) Larry Paulson was born in x.

   e) Larry Paulson est né aux États-Unis.

   f) This sentence is false.

   g) Every even number greater than 2 can be written as the sum of two primes.

   h) There exists a unique prime n that satisfies $a^n + b^n = c^n$ for some positive integers a, b and c.

   i) There are two people in Peterborough with the exact same number of hairs on their head.

> This exercise aims to demonstrate some interesting considerations in the study of logic and proof theory, in particular the blurry and ambiguous nature of human language and the nuanced relationship between statements and their proofs. Even considering whether the sentences above are valid logical statements assumes that we are working within the syntax and semantics of the English language – they would be meaningless (or, more precisely, not syntactically correct) if our syntax was propositional logic, for example! Even then, almost none of these can be unequivocally labelled as true, false, unknown or invalid, though objections can get quite weird and pedantic. Any answer with a suitable justification can be accepted. Some ideas:
>
> a) *Broccoli is delicious.* Invalid, because a subjective opinion, though devout broccoli fans/haters may argue that this is an universal fact/lie.
>
> b) *x likes broccoli.* True or false, if x is interpreted as a person that likes/doesn't like broccoli. Then again, "liking" things is not always a clear, binary choice so perhaps the statement cannot be given a truth value.
>
> c) *The area of a circle of radius r is πr^2.* True, as it is the well-known formula for the area of the circle. Interpreting the formula requires switching to the syntax of mathematics, with its own rules and conventions; one may say that π is a free variable and the statement is unknown until we give the interpretation $\pi \mapsto 3.14159265\ldots$ (which would of course only satisfy the statement if the value assigned is *exactly* π, not some finite-decimal-expansion approximation of it). Finally – and this is not even an unnecessarily pedantic point – the statement can only hold on the Euclidean plane;

the formula is not valid for the area of a circle on a sphere or hyperbolic plane.

d) *Larry Paulson was born in x.* Satisfiable with interpretation $x \mapsto$ United States, but also with $x \mapsto 1955$; we do not have notions of "types" that constrain the nature of domains in which the variables can be instantiated. Of course, there is nothing stopping us from studying *many-sorted logic* where the variable can be constrained to a particular sort/type, and quantifying over elements of a specific set. For example, "$\exists x \in \mathbb{N}$. Larry Paulson was born in x" will only admit 1955 as a witness. It would also be fair to say that Larry Paulson may not necessarily refer to the course lecturer and without a valuation telling us what to do with the variable "Larry Paulson" we cannot assign a truth value to the statement.

e) *Larry Paulson est n'e aux 'Etats-Unis.* The sentence means "Larry Paulson was born in the United States", which is true; but it uses the syntax of French, rather than English. Thus, if we have made the assumption that we take English as the base syntax, we have to conclude that the sentence is syntactically not well-formed, despite semantically being true. Again, this may seem pedantic, but appreciating the distinction between the two concepts is very important for getting a grasp of the formal study of logic.

f) *This sentence is false.* A well-known *liar paradox* which is true if and only if it is false. Philosophers have been debating the truth, falsity, or validity of this sentence for a long time and will probably continue doing so for the foreseeable future. On the surface it is clearly nonsensical, but it is difficult to properly explain why (or why not!): it's not an opinion, question or command but a "perfectly normal" declarative assertion. The only fishy thing about it is the self-reference, but disallowing that would also disallow perfectly fine self-referential statements as well, and one may always try to cheat the system by elaborate indirectly self-referential statements. Even Alfred Tarski's trick of introducing an infinite hierarchy of "meta-levels" of language that can only assert the truth or falsity of statements in lower "object-levels" fails, since one may assert valid and meaningful statements about the infinite levels themselves that cannot be placed in the hierarchy, making the system incomplete. We could throw our hands up and say that the liar paradox is neither true nor false, but then consider "This sentence is not true": saying that it is neither true nor false (viz. not true and not false) must in particular imply that it is not true, from which the paradox still follows. For very detailed discussions of the liar paradox see the Wikipedia and Stanford Encyclopaedia of Philosophy pages on the subject (as well as a worrying number of Reddit posts claiming to have solved it).

g) *Every even number greater than 2 can be written as the sum of two primes.* A well-known unsolved mathematical statement called *Goldbach's Conjecture*. It is not known to be true or false since no proof or counterexample is known; it has been verified for integers below 4×10^{18} but of course one cannot extrapolate from that to all natural numbers. The conjecture is certainly true or not true (in the "law of excluded middle"

sense), but we don't have a constructive evidence of for the truth of either disjunct.

h) *There exists a unique prime n that satisfies $a^n + b^n = c^n$ for some positive integers a, b and c.* The statement itself is true for $n = 2$ by our beloved *Pythagorean Theorem*: $a^2 + b^2 = c^2$ has infinitely many positive integer solutions (and the theorem probably has infinitely many proofs, but at least 112). The fact that $n = 2$ is the unique such prime (or indeed, natural number not less than 2) follows from Pierre de Fermat's infamous *Last Theorem*, which was an unsolved problem for hundreds of years until it was finally proved by Andrew Wiles in 1994. Fermat claimed to have discovered a "truly marvellous proof of this" which he could not fit in the margin of the book he was annotating; all other similar comments he made were subsequently verified, but his Last Theorem (which was technically a conjecture until 1994) resisted most direct and indirect approaches. Mathematicians found proofs of special cases (for all composite numbers and certain classes of primes), but most attempts at the general statement resulted in failure. Andrew Wiles built on rather advanced results of algebraic geometry and number theory to establish a stronger result called the *Taniyama–Shimura–Weil conjecture* (now known as the *modularity theorem*) that shows a connection between concepts called elliptic curves and modular forms, and – as shown by Gerhard Frey – implies FLT as a corollary. In addition to being an inspiring underdog story (not that Wiles wasn't already an established mathematician, but anyone crazy enough to tackle FLT after 358 years of failed attempts would face an almost insurmountable challenge), it also highlights an interesting phenomenon in logic and mathematics: the simplicity of the statement of a theorem often has no relation to the complexity of the required proof. Finding solutions even to propositional logic formulas is NP-complete (intractable) in the general case, though proof systems and other methods you learn about in this course can handle non-pathological formulas with ease. Fermat's Last Theorem is very easy to state as a first-order logic formula on natural numbers, but proving it seems to require deriving it from a far more involved and general result because elementary number theory (which is certainly what Fermat had access to) is simply not powerful enough. While we can't know if Fermat really had a valid, simple and marvellous proof of the proposition, he almost certainly didn't make use of 20th century algebraic geometry which seems to be the only good avenue towards the result. Having said that, there isn't that much ongoing research or interest trying to tackle FLT by alternative means; Wiles' primary contribution was proving a crucial step of the very influential modularity conjecture, and the fact that it implies Fermat's Last Theorem is a simple (and mathematically relatively uninteresting!) corollary that ticked a big box in the history of mathematics.

i) *There are two people in Peterborough with the exact same number of hairs on their head.* Bald people exist, so this is obviously true for $n = 0$. Surprisingly, even when restricted to non-bald people, this is almost certainly a true statement: there are "only" 100-150 thousand hairs on a person's head on average, but the population of

> Peterborough is above 150,000 so by the Pigeonhole Principle it is (almost) impossible for everyone to have different numbers of hairs. This is in the category of "unexpected consequences of seemingly obvious theorems": results that are very easy to state and intuitively accept, but may often require some work and ingenuity to prove and have surprising implications. Another well-known example is the fact that there must be two opposite (antipodal) points on the surface of Earth with *exactly* the same temperature and pressure – this follows from the intermediate value theorem, which essentially states that if a continuous curve starts below a line and ends above the line, it must cross the line at some point. A similarly atmospheric example is the fact that there must exist a point on Earth without any blowing wind, as a consequence of the hairy ball theorem which can be summarised as "you can't comb a coconut". Note that all of these are *pure existence* theorems that do not provide a concrete witness for the existence: they don't explicitly name the two Peterborough residents or geographic coordinates where the statements hold, only that they must exist. One may reasonably say that existence should really be established by a *constructive* proof that either states the witness or gives an algorithm for computing it; *constructive mathematics* takes such algorithmic processes as the basis of proof theory and has surprising connections to computing (see the Part II Types course).

2. a) Write two closed statements (without variables): one true and one false.

   b) Write three statements with at least two variables: one valid, one satisfiable and one unsatisfiable. Give satisfying and falsifying interpretations where appropriate.

   c) Write a set S of at least three statements containing at least one (common) variable such that every statement is satisfiable, but the set is inconsistent.

2. Propositional logic

1. Verify de Morgan's laws and Peirce's Law using truth tables.
2. Each of the following formulas is satisfiable but not valid. Exhibit an interpretation that makes the formula true and another that makes the formula false.

$$P \to Q \qquad \neg(P \vee Q \vee R) \qquad P \vee Q \to P \wedge Q \qquad \neg(P \wedge Q) \wedge \neg(Q \vee R) \wedge (P \vee R)$$

3. Associate each of the following terms with one of the propositional equivalences on Slide 205. Give a number-theoretic example for each property.

   unit, distributivity, idempotence, commutativity, inverse, associativity, annihilation

4. Convert each of the following propositional formulas into Conjunctive Normal Form and also into Disjunctive Normal Form. For each formula, state whether it is valid, satisfiable, or unsatisfiable; justify each answer.

$$(P \to Q) \wedge (Q \to P) \qquad ((P \wedge Q) \vee R) \wedge \neg(P \vee R) \qquad \neg(P \vee Q \vee R) \vee ((P \wedge Q) \vee R)$$

3. Proof systems for propositional logic

1. Briefly compare and contrast the following three formal proof systems:

   - Hilbert-style deductive system

     > A formal proof system for propositional logic centred around the modus ponens logical deduction rule. Hilbert systems are very minimal, with implication as the only connective and a small number of axiom schemas (such as K, S, and DN) that can be instantiated with propositions and then combined using MP. Despite its simplicity, the Hilbert-style deduction is a sound and complete proof system for propositional logic: all the theorems it derives are (semantically) true, and it can derive any true proposition. The flip side is that the system is really difficult to use by hand and proving even the simplest propositions (such as $A \to A$) takes several nontrivial steps. The main issue is that we have no notion of hypothetical assumptions that can be used in the required places of the proof, so every step of a Hilbert proof is a tautology (i.e. a proposition that is true without any assumptions). Nevertheless, one may prove of any classical proposition *in principle*, and often that's all we need!

   - Gentzen-style natural deduction

     > Natural deduction is an improvement over Hilbert systems in that it is *natural*: logical connectives are treated on their own merit, and the logical rules correspond to strategies we would naturally use to use or prove a given connective. The emphasis is therefore on deduction rules, divided into *introduction* and *elimination* rules which construct or consume a proposition involving a logical connective. The only axiom is **t**, the introduction rule for truth. Writing a proof in natural deduction involves constructing a derivation tree rooted at the intended proposition, with all branches corresponding to a particular inference rule. For example, if we have a derivation of A and a derivation of B, we can combine the two into a derivation of $A \wedge B$.
     >
     > An added flexibility and complication is the use of hypothetical assumptions, which allow us to use a proposition as an "axiom" (a leaf node) as long as the derived formula is hypothesised on this assumption: the introduction rule for implication derives $A \to B$ from a derivation of B given a hypothetical assumption $[A]$. Similarly, the elimination form for disjunction expresses case analysis: if we can show C both by assuming $[A]$ and independently by assuming $[B]$, we can deduce C from $A \vee B$.
     >
     > Dealing with lots of hypothetical assumptions can get quite unwieldy since we need to keep track of when they are "in scope" and when the assumptions have already been consumed. An alternative presentation of natural deduction uses *hypothetical judgments* of the form $\Gamma \vdash A$, where A is the proposition to be proved and Γ is the set of assumptions we have "in scope" at a particular point of the proof. As we progress through the derivation, this context of assumptions varies as we add and discharge hypotheses; for example, if we derive $\Gamma, A \vdash B$ (assuming Γ and A, we can derive B),

we can further derive $\Gamma \vdash A \to B$ by the introduction implication rule. Propositions proved in the empty context of assumptions $\vdash A$ are called *theorems*; the soundness and completeness of ND states that theorems and tautologies (true propositions for every interpretation of propositional letters) "coincide".

While natural deduction is indeed far more natural than Hilbert systems, using it to construct derivation trees is still difficult: the elimination rules seen "bottom-up" become introduction rules, and quite often involve inventing one or more propositions that get consumed at the application of the rule and therefore do not appear in the hypothesis. For example, modus ponens (implication introduction) derives B from $A \to B$ and A, so if we encounter a B in a bottom-up derivation, we may not have any way of knowing what hypothesis A to introduce in order to continue the subderivations. Even worse, there is nothing telling us that the last rule that must have been applied is modus ponens – it could well have been disjunction or conjunction elimination, all of which end with the derivation of an arbitrary formula. For this reason, natural deduction in the general case is not suitable for hand-written proofs or automatic proof search – there is an unbounded number of cases that could be considered.

- Gentzen-style sequent calculus

Gentzen's sequent calculus avoids the problem above by making every rule into an introduction rule – thus, sequent proofs have a clear bottom-up reading, each step eliminating one connective (thereby also bounding the size of the proof by the syntactic depth of the formula). Introducing connectives via *right* rules resembles natural deduction, but using assumptions is done via *left* rules rather than elimination. This necessitates generalising the hypothetical judgments $\Gamma \vdash A$ to *sequents* $\Gamma \Rightarrow \Delta$, where both Γ and Δ are sets of propositions. The interpretation of $\Gamma \Rightarrow \Delta$ is that assuming *all* the formulas in Γ implies *at least one* formula in Δ: $\bigwedge \Gamma \vDash \bigvee \Delta$. This duality in the sequents themselves leads to a very elegant and compact proof system that embraces the duality of propositional connectives: proving conjunction is similar to assuming a disjunction (a branching rule), assuming a conjunction is similar to proving a disjunction (turning the connective into a comma), and negation is self-dual. Implication may seem less intuitive: the right rule adds the hypothesis to the set of assumptions as expected, but the left rule splits it up in a somewhat backwards-looking way:

$$\frac{\Gamma \Rightarrow \Delta, A \qquad B, \Gamma \Rightarrow \Delta}{A \to B, \Gamma \Rightarrow \Delta}\ (\to l)$$

To see why it nevertheless makes sense (a.k.a. why the rule is *sound*), we need need to think about how an assumption $A \to B$ could be used in a proof. As a hypothetical statement, it is not immediately useful: "If I had a million pounds, I would move to Hawaii" doesn't give me a million pounds, and neither does it let me move anywhere. However, if the hypothesis is discharged, we can make use of the consequent in the rest of the proof – this is precisely what modus ponens says. Thus, if in addition to

$A \to B$ we also had an assumption A, we could apply MP "behind the scenes" and make use of B. The only place we could get this extra A assumption from is from Γ, hence the first hypothesis of the rule; then, the B conclusion can be further assumed in the proof of Δ after the implicit application of MP. Of course, we should still allow the option of proving Δ from Γ directly, and use the existing Γ assumptions to establish Δ, which justifies the principle of carrying Γ and Δ around in all the proofs.

Even though sequent calculus doesn't explicitly make use of elimination rules, they still lurk in the background of the right-side inferences. A similar analysis can be performed with the $\neg l$ rule: the negated assumption $\neg A$ is not necessarily useful until it can be confronted with an additional assumption A derived from Γ and the resulting contradiction is enough to deduce Δ. Sequent calculus rules are satisfyingly symmetric/dual, but they are interpreted as very different reasoning principles.

Remark. Note that the lecture slides use the notation $S \vdash A$ as the metatheoretic assertion that A is deducible from elements of S, and the course does not dwell on natural deduction long enough to introduce the $\Gamma \vdash A$ notation as the alternative for the bracketing and crossing out of hypotheses in natural deduction (and misleadingly suggests that an assumption environment is characteristic of the multiple-conclusion sequent calculus only). While it could be seen as a clash of notation, the two concepts really mean the same thing: A being syntactically derivable/provable from a set of propositions Γ is the same thing as the natural deduction judgment $\Gamma \vdash A$ with a finite derivation, which, in turn, is the same as a singleton-conclusion sequent $\Gamma \Rightarrow \{A\}$ also with a finite derivation. Thus, $\Gamma \vdash A$ can be seen as the "syntactic truth" assertion in contrast to the "semantic truth" $\Gamma \vDash A$, with soundness and completeness of propositional logic expressible as $\vdash A \leftrightarrow \vDash A$ for every formula A.

2. a) Proof systems employ many "if X then Y"-style connectives and assertions on different reasoning levels. Explain, with examples, the differences and similarities between

$$A \wedge B \to C \qquad A, B \vdash C \qquad A, B \vDash C \qquad A, B \Rightarrow C \qquad \frac{A \quad B}{C} \qquad \begin{array}{c}[A \wedge B] \\ \vdots \\ C\end{array}$$

Formal logic is a zoo of implication-like symbols and constructions that are quite easy to get confused about – and that is more-or-less the intention of this question!

$$A \wedge B \to C$$

This is a formula of propositional logic, and $\to$ is one of the *propositional connectives* declared in the syntax. Treated purely symbolically it does not denote or correspond to implication in the semantic (if $A \wedge B$ then C) sense, simply because symbols of the logic do not have an intrinsic meaning until such a meaning is assigned to them via truth tables. Indeed, there are several other symbols used for implication, such as

$A \wedge B \Longrightarrow C$ and (especially in older texts) $A \wedge B \supset C$ – they're all different symbols that traditionally get interpreted as implication. In implementations the connective would correspond to a constructor of the datatype of propositional formulas.

$$A, B \Rightarrow C$$

The notation for *sequent*s used in the course: a conditional assertion with some number of hypotheses, and some number of consequents. For the general sequent calculus both sides of the sequent are sets/sequences of any number of formulas. The $\Rightarrow$ symbol is therefore a metasyntactic connective between sets of formulas, acting as a separator between the assumptions and the conclusions. Once again, the distinction between assumptions and conclusions is merely our intuitive understanding of sequents; syntactically, all we care about is that $\Gamma, A \Rightarrow \Delta, A$ is an axiom (schema), and the sequent rules decompose and move symbols around the $\Rightarrow$ separator – "assuming" and "proving" is our human narrative to explain how and why the rules are sound, but they are nothing more than syntax tree transformations. Another way to appreciate this is to imagine implementing sequent calculus in a functional language: the sequent $\Rightarrow$ can simply be implemented as a pair of lists of formulas, without having to make any implementation distinction between the "hypothesis" list and the "consequent" list.

$$A, B \vdash C$$

A so-called *simple conditional assertion*, a special case of a sequent where the set of conclusions must have exactly one element: $\Gamma \vdash A$. While this particular notation is not used in the course (and, conversely, most other literature uses $\vdash$ for the general sequent rather than $\Rightarrow$), you've almost certainly encountered it in the IB Semantics course as a separator between the type environment Γ and type judgment $e : T$ where it serves a similar purpose: with the assumptions (variable typing environment) in Γ we have a finite derivation of A (a term e of type T). Such single-conclusion sequents usually feature in natural deduction systems where the formula on the right is operated on by introduction and elimination rules, while the formulas in Γ are left untouched (other than the set Γ itself evolving throughout the proof). The $\Gamma \vdash A$ notation is also used as a general statement of deducibility/derivability/provability/syntactic truth of a proposition A from a set of assumptions Γ but this essentially coincides with the analogous assertions in natural deduction and the sequent calculus (since provability is a statement of derivability in a particular proof system).

$$\begin{array}{c} [A \wedge B] \\ \vdots \\ C \end{array}$$

The original notation for *hypothetical assumptions* in Gentzen's natural deduction, analogous to the $\vdash$ presentation above. It suggests temporarily introducing an "axiom" $A \wedge B$ (i.e. a formula that needs no proof), and if from this we can derive C, we have a

derivation of the implication $A \wedge B \to C$. The "temporariness" of the axiom is exhibited by having to cross it out as soon as the rule requiring it (e.g. $\to I$ or $\vee E$) is applied. In the conditional assertion presentation $\Gamma \vdash A \to B$, the introduction of a new "axiom" corresponds to extending Γ with a new assumption $\Gamma, A \vdash B$. Thus, whenever a proof of A is required in a further derivation, we may use the hypothesis $\Gamma, A \vdash A$. Of course, a formula can only be considered a theorem if it has no assumptions, i.e. no uncrossed assumptions $[P]$ in the proof tree or, analogously, an empty assumption context $\emptyset \vdash A$.

$$\frac{A \quad B}{C}$$

An *inference rule* that is not specific to propositional logic, but used for inductive definitions of sets via rules and axioms. As such, it is also not about logical implication per se, but simply acts as a branch node for constructing derivation trees: if we have all the subtrees rooted the hypotheses of the rule (A and B, for example), we can create a new tree rooted at the conclusion of the rule C. Rules with no assumptions are simply the leaf nodes of the derivation tree – the axioms. Most formal systems and relations you encounter can be represented as inductively defined sets: syntax definitions (of programming languages, logics, regular expressions), typing rules, logical deduction rules, reduction rules, etc.

$$A, B \vDash C$$

The *semantic entailment* assertion which states that any interpretation that satisfies A and B (or a set Γ in general) also satisfies C. This is the only judgment that concerns the semantics of propositional logic by talking about interpretations of propositional letters and thus corresponds to the "actual" meaning of truth as defined by truth tables. A semantic entailment with no assumptions is known as a tautology: all interpretations satisfy the formula. A sound and complete proof system guarantees syntactic theorems $\vdash A$ exactly correspond to semantic tautologies $\vDash A$.

b) Separate the statements above based on whether they belong to the *syntax* or the *semantics* of propositional logic. Which proof systems are the constructs associated with?

Semantic entailment $\Gamma \vDash A$ is the only semantic assertion. The syntactic entailment $\Gamma \vdash A$ states that A is derivable from Γ in the particular proof system in question and is commonly used in the Hilbert-style (where Γ is always empty) and the Gentzen styles too. This course specifically uses $\Rightarrow$ for the multiple-conclusion sequent calculus. The implication $\to$ is defined in the syntax of propositional logic and is not tied to a particular proof system. Finally, the inference rule $\frac{A \quad B}{C}$ is a general tool for inductively defining proof systems and other formal relations.

3. Prove the following sequents:

$$\neg\neg A \Rightarrow A \qquad A \wedge B \Rightarrow B \wedge A \qquad (A \vee B) \wedge (A \vee C) \Rightarrow A \vee (B \wedge C)$$
$$\neg(A \vee B) \Rightarrow \neg A \wedge \neg B \qquad \Rightarrow (A \wedge \neg A) \to B \qquad \Rightarrow ((A \to B) \to A) \to A$$

(You can write the proof trees upside-down if you prefer, so you don't have to reserve space.)

> There is some flexibility in the order of rule applications, but the general shape of the trees will be similar.
>
> $$\dfrac{\dfrac{\overline{A \Rightarrow A}}{\Rightarrow A, \neg A}\,(\neg r)}{\neg\neg A \Rightarrow A}\,(\neg l) \qquad \dfrac{\overline{A, B \Rightarrow B} \quad \overline{A, B \Rightarrow A}}{A, B \Rightarrow B \wedge A}\,(\wedge r)}{A \wedge B \Rightarrow B \wedge A}\,(\wedge l) \qquad \dfrac{\dfrac{\dfrac{\overline{A \Rightarrow B, A}}{A, \neg A \Rightarrow B}\,(\neg l)}{A \wedge \neg A \Rightarrow B}\,(\wedge l)}{\Rightarrow (A \wedge \neg A) \to B}\,(\to r)$$
>
> $$\dfrac{\dfrac{\dfrac{\overline{A \Rightarrow A, B}}{\Rightarrow \neg A, A, B}\,(\neg r) \quad \dfrac{\overline{B \Rightarrow A, B}}{\Rightarrow \neg B, A, B}\,(\neg r)}{\Rightarrow \neg A \wedge \neg B, A, B}\,(\wedge r)}{\Rightarrow \neg A \wedge \neg B, A \vee B}\,(\vee r)}{\neg(A \vee B) \Rightarrow \neg A \wedge \neg B}\,(\neg l) \qquad \dfrac{\dfrac{\overline{A \Rightarrow A, B}}{\Rightarrow A, A \to B}\,(\to r) \quad \overline{A \Rightarrow A}}{(A \to B) \to A \Rightarrow A}\,(\to l)}{\Rightarrow ((A \to B) \to A) \to A}\,(\to r)$$
>
> $$\dfrac{\dfrac{\overline{A, A \vee C \Rightarrow A, B} \quad \overline{B, A \vee C \Rightarrow A, B}}{A \vee B, A \vee C \Rightarrow A, B}\,(\vee l) \qquad \dfrac{\overline{A \vee B, A \Rightarrow A, C} \quad \overline{A \vee B, C \Rightarrow A, C}}{A \vee B, A \vee C \Rightarrow A, C}\,(\vee l)}{A \vee B, A \vee C \Rightarrow A, B \wedge C}\,(\wedge r)}{A \vee B, A \vee C \Rightarrow A \vee (B \wedge C)}\,(\vee r)}{(A \vee B) \wedge (A \vee C) \Rightarrow A \vee (B \wedge C)}\,(\wedge l)$$

4. Derive the sequent calculus rules for the connectives $\leftrightarrow$ (iff) and $\oplus$ (exclusive or). Note that other connectives must not appear in these rules.

> The sequent calculus for the standard propositional connectives is sound and complete, which means that if a formula is derivable, so is any formula logically (that is, semantically) equivalent to it. We of course has the logical equivalence $A \leftrightarrow B \simeq (A \to B) \wedge (B \to A)$, and a derivation tree of the latter (with arbitrary sets of formulas Γ and Δ) on each side of a sequent would have the following form:
>
> $$\dfrac{\dfrac{A, \Gamma \Rightarrow \Delta, B}{\Gamma \Rightarrow \Delta, A \to B}\,(\to r) \qquad \dfrac{B, \Gamma \Rightarrow \Delta, A}{\Gamma \Rightarrow \Delta, B \to A}\,(\to r)}{\Gamma \Rightarrow \Delta, (A \to B) \wedge (B \to A)}\,(\wedge r)$$
>
> $$\dfrac{\dfrac{\Gamma \Rightarrow \Delta, A, B \quad \overline{A, \Gamma \Rightarrow \Delta, A}}{B \to A, \Gamma \Rightarrow \Delta, A}\,(\to l) \qquad \dfrac{\overline{B, \Gamma \Rightarrow \Delta, B} \quad B, A, \Gamma \Rightarrow \Delta}{B, B \to A, \Gamma \Rightarrow \Delta}\,(\to l)}{A \to B, B \to A, \Gamma \Rightarrow \Delta}\,(\to l)}{(A \to B) \wedge (B \to A), \Gamma \Rightarrow \Delta}\,(\wedge l)$$
>
> Thus, if bi-implication were in the sequent calculus, any application of $(\leftrightarrow l)$ or $(\leftrightarrow r)$ should be replaceable with these derived rules, since the two formulas are equivalent. But

> then we can just take these as the derived inference rules for $\leftrightarrow$ to begin with!
>
> $$\frac{A,\Gamma \Rightarrow \Delta,B \qquad B,\Gamma \Rightarrow \Delta,A}{\Gamma \Rightarrow \Delta,A \leftrightarrow B}\ (\leftrightarrow r) \qquad \frac{\Gamma \Rightarrow \Delta,A,B \qquad B,A,\Gamma \Rightarrow \Delta}{A \leftrightarrow B,\Gamma \Rightarrow \Delta}\ (\leftrightarrow l)$$
>
> These are sound by construction, since any derivation involving these rules should have a derivation involving $\rightarrow$ and $\wedge$ whose associated rules are sound.
>
> We can do something similar to derive exclusive or: we have the logical duality $A \oplus B \simeq \neg(A \leftrightarrow B)$, so just like with $\wedge$ and $\vee$, we expect the right rule of $\oplus$ to resemble the left rule of $\leftrightarrow$ and vice versa. This can of course be verified by a derivation of $(A \wedge \neg B) \vee (\neg A \wedge B)$.
>
> $$\frac{\Gamma \Rightarrow \Delta,A,B \qquad B,A,\Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta,A \oplus B}\ (\oplus r) \qquad \frac{A,\Gamma \Rightarrow \Delta,B \qquad B,\Gamma \Rightarrow \Delta,A}{A \oplus B,\Gamma \Rightarrow \Delta}\ (\oplus l)$$

4. First-order logic

1. To test your understanding of quantifiers, consider the following formulas: *everybody loves somebody* vs. *there is somebody that everybody loves*:

   $$\forall x.\, \exists y.\, \mathsf{loves}(x,y) \qquad \exists y.\, \forall x.\, \mathsf{loves}(x,y)$$

   Does the first imply the second? Does the second imply the first? Consider both the informal meaning and the formal semantics defined in the course.

2. Let $\approx$ be a 2-place predicate symbol, which we write using infix notation as $x \approx y$ instead of $\approx(x,y)$.

   a) Give three formulas in FOL describing the axioms that make $\approx$ an *equivalence relation*.

   > Reflexivity, symmetry and transitivity:
   >
   > Ⓡ $\forall x.\, x \approx x$ Ⓢ $\forall x,y.\, x \approx y \rightarrow y \approx x$ Ⓣ $\forall x,y,z.\, x \approx y \wedge y \approx z \rightarrow x \approx z$

   b) Let the universe be the set $\mathbb{N}$ of natural numbers. Which axioms hold if $I[\approx]$ is:

   (i) the empty relation, $\emptyset$?

   > Ⓢ and Ⓣ. Since the relation never holds, the hypotheses of symmetry and transitivity are never satisfied so they vacuously hold. Reflexivity requires at least the identity loops for all elements of the domain, but since the domain is nonempty, the empty relation is not reflexive.

   (ii) the universal relation, $\{(x,y) \mid x,y \in \mathbb{N}\}$?

   > Ⓡ, Ⓢ and Ⓣ. This is not to say the universal relation trivially satisfies everything – it violates negative properties like irreflexivity and asymmetry.

   (iii) the equality relation, $\{(x,x) \mid x \in \mathbb{N}\}$?

   > Ⓡ, Ⓢ and Ⓣ. Equivalence wouldn't really be a good generalisation of equality if equality wasn't an equivalence!

(iv) the relation $\{(x, y) \mid x, y \in \mathbb{N} \wedge x + y \text{ is even}\}$?

> Ⓡ, Ⓢ and Ⓣ. Two numbers are related if both are even or both are odd – in other words, $x \approx y$ iff $x \equiv y \pmod 2$ and congruence is an equivalence relation.

(v) the relation $\{(x, y) \mid x, y \in \mathbb{N} \wedge x + y = 100\}$?

> Ⓢ. Not reflexive, since the only number that is related to itself is 50. Not transitive either: $30 \approx 70$ and $70 \approx 30$, but $30 \not\approx 30$.

(vi) the relation $\{(x, y) \mid x, y \in \mathbb{N} \wedge x \leq y\}$?

> Ⓡ and Ⓣ. $\leq$ is a partial order so it is reflexive and transitive; instead of symmetry it satisfies *antisymmetry*.

3. Taking R as 2-place relation symbol and $=$ denoting equality, consider the following axioms:

$$\forall x.\ \neg R(x, x) \tag{1}$$
$$\forall x, y.\ \neg(R(x, y) \wedge R(y, x)) \tag{2}$$
$$\forall x, y, z.\ R(x, y) \wedge R(y, z) \to R(x, z) \tag{3}$$
$$\forall x, y.\ R(x, y) \vee (x = y) \vee R(y, x) \tag{4}$$
$$\forall x, y.\ R(x, z) \to \exists y.\ R(x, y) \wedge R(y, z) \tag{5}$$

> These properties correspond to: irreflexivity (different from R just not being reflexive – we explicitly forbid any loops); asymmetry (again, different from both the negation of symmetry and antisymmetry); transitivity; linearity (every pair of elements is related in some way); density (for any two related elements there exists one between them).

a) Exhibit two interpretations that satisfy axioms 1–5.

> Interpret R as the ordering $<$ on rational or real numbers (or an interval thereof). The less-than ordering on numbers is irreflexive, asymmetric, transitive and linear in general, and density is ensured by considering a domain in which every nonempty interval has e.g. a halfway point.

b) Exhibit two interpretations that satisfy axioms 1–4 and falsify axiom 5.

> Without the need for density we can consider $<$ on domains such as naturals or integers, or other examples of strict total orders such as the strict lexicographical order on a finite Cartesian product of totally ordered sets, e.g. $\mathbb{N} \times \mathbb{N}$.

c) Exhibit two interpretations that satisfy axioms 1–3 and falsify axioms 4 and 5.

> Now we need a strict relation that can be partial, i.e. we can have two distinct elements which are not related to itself in any direction. The usual example is the strict subset order $\subset$, since for example $\{1\} \not\subset \{2\}$ and $\{2\} \not\subset \{1\}$. Another is strict divisibility on integers: $d \mid n \leftrightarrow \exists k \in \mathbb{Z}.\ k \neq 1 \wedge n = k \cdot d$.

Hint: Consider simple examples such as $R(x, y) = x < y$ on some appropriate domain.

4. Some textbook and paper authors like to use nonstandard notation, or even define their own syntax for some particular formal system. While often difficult to read and generally not recommended, there is nothing inherently wrong with this practice – as long as the meaning (semantics) is defined appropriately, the syntax can be whatever the author chooses to use.

   Suppose that a particularly creative logician decides to reinvent the syntax of first-order logic. Below are three logically valid formulas written in the syntax which should hold for any formula A and predicate P and Q containing a free variable x.

   $$\overbrace{\left[P(x), \boxed{P(x)}\right]}^{x} \qquad \boxed{A} \triangleright A \qquad \overbrace{\left[P(x), \boxed{Q(x)}\right]}^{x} \bowtie \underbrace{\left\langle \boxed{P(x)}, Q(x) \right\rangle}_{x}$$

   > The aim of this question is to demonstrate the decoupling of syntax and semantics in formal logics by asking you to give meaning to a notation that you have never seen before and therefore should not have any preconceptions about. A problem with studying logic at such a foundational level is that we humans are already used to standard notational conventions and it becomes difficult to see $\wedge$, $\rightarrow$ and $=$ as mere syntax and $A \wedge B \rightarrow C$ as a sequence or tree of symbols. As such, things like the truth definition of first-order logic may seem needlessly verbose and complicated while also "stating the obvious": $A \wedge B$ is true if A is true and B is true. Well, let us throw off the shackles of common sense notation and consider a syntax where things are not as obvious!

   a) Describe the formal syntax of this logic as a grammar, giving the different syntactic forms of a formula A (see e.g. Slide 201). You may assume that the syntax already includes atomic formulas made up of a relation symbol $P, Q, R \ldots$ applied to some number of terms.

      > We fix a set $\mathcal{L} = \{x, y, \ldots\}$ of variables. The syntax of the language can be read off without having to understand what the formulas mean. If A, B are formulas, so are
      >
      > $$\boxed{A} \qquad \langle A, B \rangle \qquad [A, B] \qquad A \triangleright B \qquad A \bowtie B \qquad \overbrace{A[x]}^{x} \qquad \underbrace{A[x]}_{x}$$

   b) Consider a formula A in the syntax presented above. Let $\mathcal{I}$ be an interpretation of the symbols and V a valuation for the formula. Give a *truth definition* for the formula A by defining a predicate $\vDash_{\mathcal{I},V} A$ which holds when A is true in $\mathcal{I}$ under V. Make sure that the three formulas above are true with your definition.

      > Here we need to perform some detective work to figure out what the connectives are actually supposed to mean. We know that all formulas are valid, i.e. satisfied by any interpretation. The most telling may be the second one, which suggests that two boxes around a formula are related to no boxes around a formula – it resembles

double negation elimination, if box means negation and $\triangleright$ means implication or bi-implication. With this interpretation of box, we see the first formula suggesting that boxed and unboxed versions of the same formula are somehow related; again, this resembles the law of excluded middle $A \vee \neg A$, assuming the square brackets denote disjunction. The overbrace surrounds the disjunction with a variable x so it is likely a binding construct. We can't in general say that there exists an x for which either $P(x)$ or not $P(x)$ hold since that wouldn't hold in the empty domain (as it at least requires the existence of an element); but we can say that if x is universally quantified. Finally, the last formula relates $\neg(\forall x.\, P(x) \vee \neg Q(x))$, and some combination of $\neg P(x)$ and $Q(x)$. It is not difficult to see that an application of the generalised and propositional de Morgan laws to the LHS gives $\exists x.\, \neg P(x) \wedge Q(x)$, so we can guess that the angle brackets mean conjunction and the underbrace is existential quantification. The $\bowtie$ connective may again mean implication or bi-implication – the formulas are valid with both interpretations of $\triangleright$ and $\bowtie$. We actually get to make a choice on how to interpret them, which is not something we usually get to do when the interpretations are "obvious"! Purely based on the symmetry of the $\bowtie$ symbol we can use it for bi-implication and $\triangleright$ for implication – perhaps this was not the intended interpretation of the logician, but it is a legitimate option based purely on the three valid formulas.

Now we actually get to give the truth definition for the language. It is entirely analogous to Tarski's definition, but with the modified syntax. The interpretation $\mathcal{I}$ and valuation V are required to interpret atomic formulae $P(a, f(x)), Q(g(y, b)), \ldots$ where we need interpretations for constants a, b, function symbols f, g, relation symbols P, Q, and variables x, y. The rest proceeds by recursing on the formula syntax, with the variable cases extending the valuation V with the newly bound variable.

For an interpretation $\mathcal{I} = (D, I)$ on domain D and valuation V, we define a formula A to be *true* in $\mathcal{I}$ under V, denoted $\vDash_{\mathcal{I},V}$, by case analysis on the syntax of formulas from part (a):

- $\vDash_{\mathcal{I},V} P(t_1, \ldots, t_n)$ if $I[P](\mathcal{I}_V[t_1], \mathcal{I}_V[t_n]) = 1$.
- $\vDash_{\mathcal{I},V} \boxed{A}$ if $\vDash_{\mathcal{I},V} A$ does not hold
- $\vDash_{\mathcal{I},V} \langle A, B \rangle$ if $\vDash_{\mathcal{I},V} A$ holds and $\vDash_{\mathcal{I},V} B$ holds
- $\vDash_{\mathcal{I},V} [A, B]$ if $\vDash_{\mathcal{I},V} A$ holds or $\vDash_{\mathcal{I},V} B$ holds
- $\vDash_{\mathcal{I},V} A \triangleright B$ if $\vDash_{\mathcal{I},V} A$ does not hold or $\vDash_{\mathcal{I},V} B$ holds
- $\vDash_{\mathcal{I},V} A \bowtie B$ if $\vDash_{\mathcal{I},V} A$ and $\vDash_{\mathcal{I},V} B$ both hold or neither hold
- $\vDash_{\mathcal{I},V} \overbrace{A[x]}^{x}$ if $\vDash_{\mathcal{I},V\{m/x\}} A$ holds for all $m \in D$
- $\vDash_{\mathcal{I},V} \underbrace{A[x]}_{x}$ if there exists some $m \in D$ such that $\vDash_{\mathcal{I},V\{m/x\}} A$ holds

c) The formula below is a statement about arithmetic on natural numbers. Give an interpretation $\mathcal{I} = (D, I)$ of the constant, function, and relation symbols, and a valuation V of the

free variables, which satisfy this formula.

$$\left\langle \overbrace{\left[n \sim \bullet, \underbrace{n \sim k^{+}}_{k} \right]}^{n}, \left\langle v \sim \bullet^{+++}, \boxed{\underbrace{\bullet \sim \ell^{+}}_{\ell}} \right\rangle \right\rangle$$

Above we gave a syntax and semantics for the formula language of this logic. The other variable aspect of first-order syntax is the term language, which makes up the syntactic entities that we can actually reason about. This follows naturally from the generalisation of propositional letters $P, Q, \ldots$ to predicates $P(a, f(x)), Q(g(y, b)), \ldots$ that now relate terms built from variables, constants and function symbols: if our logic involves variables, we need something that the variables can be instantiated with. Again, a difficulty with getting an intuition for the term syntax is that it is either entirely abstract, or too concrete and familiar: it's hard to see the point of saying "let 0 denote zero and $+$ denote addition and $=$ denote equality". Hence this question asks you to give an interpretation to symbols that you are not familiar with based entirely on a formula that the interpretation must satisfy.

The first step is rewriting the formula in the familiar syntax of FOL:

$$\left(\forall n.\, (n \sim \bullet) \vee \exists k.\, n \sim k^{+}\right) \;\wedge\; (v \sim \bullet^{+++}) \;\wedge\; \left(\neg \exists \ell.\, \bullet \sim \ell^{+}\right)$$

Next, we analyse the arities of the various symbols featuring in the formula, based on how they are notated: $\bullet$ is a constant symbol, $(-)^{+}$ is a unary function symbol written in superscript, $\sim$ is a binary relation symbol written infix (it can't be a function symbol since $(v \sim \bullet^{+++})$ is used as a formula). Now, we know that the formula expresses a property of arithmetic on natural numbers, and with a bit of trial and error it seems likely that $\bullet$ denotes 0, $(-)^{+}$ denotes successor and $\sim$ is equality. The first conjunct states that any natural number n is either zero or is the successor of another natural number, and the third conjunct states that there does not exist a number ℓ whose successor is zero. These are axioms of Peano arithmetic. The middle conjunct involves a free variable v and simply states that it is equal to 3 (the third successor of 0) – the only valuation that will satisfy the formula is one that maps v to 3. Thus, we define the interpretation and valuation as follows:

$$D = \mathbb{N} \qquad V = v \mapsto 3$$

$$I[\bullet] = 0 \in \mathbb{N} \qquad I[(-)^{+}] = n \mapsto n + 1 \in \mathbb{N} \to \mathbb{N} \qquad I[\sim] = \{(n, m) \mid n = m\} \subseteq \mathbb{N}^2$$

This exercise hopefully gave you a better grasp of symbols, first-order formulas, interpretations and truth definitions, and how syntax of formulas and terms really isn't constrained by anything other than notational conventions. This of course doesn't mean that you should question the semantics of symbols every time you encounter

them ("Excuse me, what is the intended interpretation of the symbols '£3.50' on that price list?") but when studying formal systems one must remember not to make assumptions about the syntax based on perceived semantics and accidentally muddle the distinction between the two. For example, it may be tempting to apply a logical equivalence in the middle of a sequent proof e.g. to reparenthesise a nested conjunction from $\Gamma \Rightarrow \Delta, (A \wedge B) \wedge C$ to $\Gamma \Rightarrow \Delta, A \wedge (B \wedge C)$ in order to get access to A right away. The step may seem innocuous (after all, conjunction *is* associative) but crucially it is an application of a semantic property in a purely syntactic proof and assumes that the proof system is sound and complete, and $\wedge$ actually stands for an associative operator. From the point of view of syntax, $\wedge$ is merely a branch node in the abstract syntax tree of propositional logic and the tree of $(A \wedge B) \wedge C$ is not interchangeable with the tree of $A \wedge (B \wedge C)$. If we did something similar in the syntax above to rewrite $(A \triangleright B) \triangleright C$ as $A \triangleright (B \triangleright C)$, we would clearly be making a mistake! You may of course say "well I obviously won't do that because I know that implication is not associative" but again, that comes from a semantic understanding of the symbol $\triangleright$; an automated theorem prover searching for a sequent calculus proof of the formula will have no idea what it's doing other than manipulating syntax trees according to strictly predetermined rules.

To take this to its logical conclusion, there is nothing stopping us from completely butchering the syntax of first-order logic by mixing up all the symbols:

$$(\rightarrow n.\,(n\,(+1)\,=) \exists \wedge k.\,n\,(+1)\,0k)\ \neg\ (\nu\,(+1)\,000 =)\ \neg\ (\forall \wedge \ell.\,=\,(+1)\,0\ell)$$

Silly as this may look, it has a perfectly consistent formula syntax and interpretation, e.g. with $=$ interpreted as zero and $\neg$ as disjunction, but of course it completely violates all notational conventions and could very much be described as *cursed*. Deciphering this was actually the original iteration of the question, but it seemed less cryptic to use more cryptic symbols instead!

Optional exercise

1. Using OCaml, define datatypes for representing propositions and interpretations. Write a function to test whether or not a proposition holds under an interpretation (both supplied as arguments). Write a function to convert a proposition to Negation Normal Form.